



Katarzyna Pawlukiewicz

Estymacja liczby cudzoziemców nielegalnie
przebywających w Polsce z wykorzystaniem
hierarchicznego modelu Gamma-Poissona

Estimation of the number of irregular
immigrants in Poland using hierarchical
Gamma-Poisson model

Praca magisterska

Promotor: dr Maciej Beręsewicz

Pracę przyjęto dnia:

Podpis Promotora

Kierunek: Informatyka i ekonometria

Specjalność: Informatyka w gospodarce i administracji

Poznań 2019

Spis treści

Wstęp	3
1 Cudzoziemcy w Polsce	4
1.1 Terminologia i aspekt prawny	4
1.2 Źródła danych o cudzoziemcach w Polsce	11
1.3 Cudzoziemcy w Polsce w świetle dostępnych źródeł danych	15
1.4 Podsumowanie	20
2 Metody estymacji wielkości populacji trudnych do zbadania	21
2.1 Populacja trudna do zbadania	21
2.2 Metody estymacji wielkości populacji	24
2.3 Hierarchiczny model Gamma-Poissona (Zhang, 2008)	28
2.3.1 Podstawowe założenia modelu	28
2.3.2 Estymacja parametrów metodą największej wiarygodności	32
2.3.3 Implementacja modelu w pakiecie R	35
2.4 Podsumowanie	38
3 Estymacja liczby cudzoziemców nielegalnie przebywających w Polsce 2018 roku	39
3.1 Źródła danych wykorzystane na potrzeby badania	39
3.2 Weryfikacja założeń modelu	40
3.3 Estymacja parametrów modelu oraz jego ocena	43
3.3.1 Wyniki estymacji parametrów modelu zlinearyzowanego i Gamma-Poissona	43
3.3.2 Ocena modelu Gamma-Poissona	44
3.4 Szacunek wielkości populacji cudzoziemców nielegalnie przebywającej w Polsce	46

Podsumowanie	51
Bibliografia	52
Spis tabel	54
Spis rysunków	55
Spis skryptów oprogramowania	56
A Dodatki	57
A.1 Dane wykorzystane w modelu	57
A.2 Zestawienie definicji wykorzystanych w pracy	59

Wstęp

W ostatnich latach Polska stała się jednym z nielicznych krajów Organizacji Współpracy Gospodarczej i Rozwoju (OECD), które przyjmują znaczącą liczbę cudzoziemców¹. Przykładowo, na przestrzeni 2016-2018 liczba cudzoziemców legalnie zatrudnionych w Polsce zwiększyła się z 200 tys. do ponad 600 tys, a liczba cudzoziemców zameldowanych na pobyt czasowy lub stały z 95 tys. do ponad 200 tys. Większość cudzoziemców, którzy przybywają do Polski pochodzą z Ukrainy, Białorusi oraz Rosji i są imigrantami zarobkowymi. Taki napływ jest stosunkowo nowym zjawiskiem w aktualnej historii Polski, który jest widoczny zarówno w rejestrach administracyjnych, jak i statystykach publikowanych przez Główny Urząd Statystyczny.

Aktualnie, w Polsce brak jest jednego zunifikowanego źródła danych o cudzoziemcach. Istniejące rejestry administracyjne nie są ze sobą zintegrowane i pokrywają pewne subpopulacje cudzoziemców. Na przykład, bazy danych Ministerstwa Spraw Zagranicznych w zakresie wiz nie są połączone z rejestrami Ministerstwa Cyfryzacji (PESEL), Zakładem Ubezpieczeń Społecznych (ZUS) czy bazami danych Urzędu do Spraw Cudzoziemców co utrudnia weryfikację pobytu czy legalności pracy cudzoziemców, którzy otrzymali wizę. Rozdrobnienie źródeł informacji oraz fakt pokrywania pewnych subpopulacji jest również problematyczny dla Głównego Urzędu Statystycznego (GUS). Aby nadążać za zmieniającą się rzeczywistością, GUS rozszerza Program Badań Statystycznych Statystyki Publicznej (PBSSP) o wcześniej niewykorzystywane rejestry na poziomie jednostkowym (np. wize) lub poszerzając zakres informacyjny już wykorzystywanych źródeł danych (np. rejestry ZUS).

Należy podkreślić, że statystyka publiczna (w Polsce jak i na świecie) zainteresowana jest populacją *de facto* czyli zarówno zarejestrowaną, jak i niezarejestrowaną ludnością podczas gdy administracyjne źródła danych dostarczają informacji o populacji *de iure* (zarejestrowanej) zdefiniowaną przez ich gestorów. Dlatego, dla statystyki publicznej, istotne jest określenie zarówno całkowitej populacji cudzoziemców, jak i subpopulacji przebywającej nielegalnie na terenie Pol-

¹W polskim prawie przez *cudzoziemce* rozumie się osobę, która nie posiada polskiego obywatelstwa.

ski. Ta ostatnia jest w centrum zainteresowania niniejszej pracy magisterskiej.

Na potrzeby pracy, za słownikiem Urzędu do Spraw Cudzoziemców przyjęto, że migracja nieregularna (nielegalna) to migracja z naruszeniem obowiązującego prawa (bez wiz, bez dokumentów podróży lub jakichkolwiek dokumentów, poza przejściami granicznymi, w wyniku przestępstw przemytu ludzi lub handlu ludźmi). Natomiast zgodnie z polskimi przepisami nielegalny pobyt definiowany jest jako pobyt niezgodny z przepisami dotyczącymi warunków wjazdu cudzoziemców na terytorium Rzeczypospolitej Polski i ich pobytu na tym terytorium. Wybór takich definicji podyktowany jest danymi udostępnionymi przez Straż Graniczną. Dodatkowo, w pracy naprzemiennie używane będzie pojęcie nielegalnego lub nieregularnego imigranta / cudzoziemca.

Głównym celem badawczym niniejszej pracy magisterskiej było oszacowanie liczby cudzoziemców, którzy czasowo lub stale przebywali nielegalnie na terenie Polski na koniec III kw. 2018 roku. Przyjęcie definicji czasowego lub stałego nielegalnego pobytu związane jest z zastosowaniem modelu Zhang (2008), który jako populację referencyjną (ang. proxy) dla populacji nieregularnej wykorzystuje populację regularną (zarejestrowaną na pobyt czasowy lub stały zgodnie z definicją meldunku wykorzystywaną w rejestrze PESEL). Dla ułatwienia dalszych rozważań przyjęto, skróconą wersję tj. nielegalny pobyt, który odnosi się zarówno do pobytu czasowego, jak i stałego.

Jak wspomniano powyżej, w pracy zastosowano podejście zaproponowane przez Zhang (2008), które opiera się na hierarchicznym modelu Gamma-Poissona i jest nowatorskim podejściem do estymacji nieregularnej imigracji. Model ten w przeciwieństwie do istniejących w literaturze rozwiązań nie wymaga dostępu do danych jednostkowych, a jedynie danych zagregowanych. Wymaga również dostępu wyłącznie do trzech źródeł danych – populacji zarejestrowanej, danych policyjnych oraz danych o ujawnionych nielegalnie przebywających cudzoziemcach. Dzięki temu możliwy jest do implementacji w statystyce publicznej bez dodatkowych ograniczeń związanych z dostępem do wrażliwych, identyfikowalnych danych.

Praca ma charakter teoretyczno-empiryczny i składa się z trzech rozdziałów. W pierwszym przytoczono zasady prawne dotyczące cudzoziemców przekraczających granice Polski i przebywających na terenie kraju. Przedstawiono jak w polskim prawie definiuje się *cudzoziemca* i przytoczono sytuacje, w których pobyt na terenie Rzeczypospolitej staje się nielegalny. Określono również źródła danych, z których można czerpać informację o różnych grupach cudzoziemców. Znajduje się tutaj również podrozdział, w którym przedstawiono zaagregowane dane o cudzo-

ziemcach na podstawie różnych rejestrów.

Rozdział drugi przedstawia metody estymacji wielkości populacji, które są trudne do zbadania. Określono jak definiuje się populację trudną do zbadania, do której należy populacja cudzoziemców nielegalnie przebywających na terenie danego kraju. Przedstawiono tutaj również hierarchiczny model Gamma-Poissona Zhang (2008) wraz z założeniami modelu oraz estymacją parametrów z wykorzystaniem metody największej wiarygodności. Rozdział kończy implementacja modelu w pakiecie statystycznym R.

Rozdział trzeci jest rozdziałem empirycznym, w którym dokonano próby estymacji populacji cudzoziemców za pomocą pakietu statycznego R, zgodnie z teoretycznymi założeniami przedstawionymi w rozdziale drugim. Głównie wykorzystywanym pakietem był pakiet *maxLik*, omówiony w rozdziale 2. Rozdział trzeci rozpoczyna się od przedstawienia wykorzystanych źródeł danych i zadeklarowania zmiennych. W badaniu wykorzystano dane z rejestru PESEL (wg stanu na 30.09.2018r.), dane ze Straży Granicznej o osobach zatrzymanych na terenie kraju, którzy nie mieli pozwolenia na pobyt (dla II półrocza 2018r.) i dane policyjne z systemu „Krajowy System Informacyjny Policji” (KSIP) o zameldowanych cudzoziemcach, którzy popełnili przestępstwo lub wykroczenie (dla II półrocza 2018r.). Dane pozyskano na podstawie zapytania o informację publiczną. W pracy ograniczono się do nieregularnej populacji w wieku 18 i więcej. Na końcu rozdziału trzeciego zamieszczono wyniki estymacji dla dwóch modeli – uwzględniającego oraz nieuwzględniającego Ukrainy.

Do pracy dołączono również dodatki w postaci podsumowania definicji populacji cudzoziemców jak i tabeli z danymi wejściowymi do oszacowania wielkości populacji.

Rozdział 1

Cudzoziemcy w Polsce

1.1 Terminologia i aspekt prawny

W polskim prawie głównymi aktami prawnymi, które regulują zasady i warunki pobytu, przejazdu przez terytorium państwa, właściwe organy i tryb postępowania w sprawach dotyczących cudzoziemców jak i pobyt członków UE na terytorium Polski są:

- Ustawa z dnia 12 grudnia 2013 r. o cudzoziemcach. (Dz.U. 2013 poz. 1650 z późn. zmianami);
- Ustawa z dnia 14 lipca 2006 r. o wjeździe na terytorium Rzeczypospolitej Polskiej, pobycie oraz wyjeździe z tego terytorium obywateli państw członkowskich Unii Europejskiej i członków ich rodzin (Dz. U. Nr 144, poz. 1043 z późn. zmianami);
- Akty wykonawcze.

Z racji członkostwa Polski w Unii Europejskiej nadrzędne są prawa ustanowione w ramach UE. To znaczy, że ustawa o cudzoziemcach nie dotyczy obywateli państw UE jak i państw członkowskich Europejskiego Stowarzyszenia Wolnego Handlu (EFTA) i stron umowy, czyli obywateli Europejskiego Obszaru Gospodarczego (EOG) i Szwajcarii oraz członków ich rodzin („Dz.U. 2013 poz. 1650 z późn. zmianami”, [2013](#)).

Mimo, iż powyższe ustawy określają obowiązki i sytuacje, w których osoby, których dotyczą mogą być wydalone z państwa, nigdzie nie zostało zdefiniowane pojęcie nielegalnego pobytu w odniesieniu do cudzoziemców posiadających obywatelstwo jednego z państw członkowskich UE. Dlatego w późniejszych rozważaniach z grona osób, których pobyt na terenie Polski może być nielegalny wykluczono takie osoby.

Z prawnego punktu widzenia cudzoziemiec jest to każda osoba, która nie posiada obywatelstwa polskiego („Dz.U. 2013 poz. 1650 z późn. zmianami”, 2013).

Pobyt cudzoziemca na terenie Polski jest nielegalny w momencie, gdy jest niezgodny z przepisami wjazdu cudzoziemców i ich pobytu na terenie Rzeczypospolitej Polskiej zawartymi w dwóch wyżej wymienionych ustawach. Każdy cudzoziemiec jest zobowiązany do posiadania ważnych dokumentów podróży oraz dokumentów uprawniających go do wjazdu i pobytu na terenie Polski o ile są one wymagane. Organami uprawnionymi do kontroli legalności pobytu są Straż Graniczna i Policja w przypadku ustalenia stanu faktycznego przestrzegania prawa dotyczącego cudzoziemców, Szef Urzędu i wojewoda w zakresie działań dotyczących cudzoziemców, naczelnik urzędu celno-skarbowego na zasadach zawartych w następujących ustawach: Prawo o ruchu drogowym, o transporcie drogowym, o udziale Rzeczypospolitej Polskiej w Systemie Informacyjnym Schengen oraz Wizowym Systemie Informacyjnym, o Krajowej Administracji Skarbowej, o grach hazardowych, o gromadzeniu, przetwarzaniu i przekazywaniu informacji kryminalnych. Podczas kontroli cudzoziemiec może zostać zobowiązany do okazania dokumentów takich jak dokument uprawniający go do pobytu na terytorium Polski, do wykonywania pracy zarobkowej, prowadzenia własnej działalności gospodarczej lub powierzenia wykonywanej pracy, bądź dokument określający cel i warunki jego pobytu. Cudzoziemiec może również zostać poproszony o udowodnienie posiadania wystarczających środków finansowych na pokrycie kosztów związanych z pobytem na terenie Rzeczypospolitej, podróży powrotnej do miejsca pochodzenia, bądź zamieszkania lub kosztów tranzytu osoby przez obszar Polski do kraju państw trzecich.

Dokumentami umożliwiającymi cudzoziemcowi wjazd i pobyt na terytorium Polski są wizy. Istnieją dwa rodzaje wiz, mianowicie wiza krajowa i wiza Schengen. Wiza krajowa będąca wizą długoterminową, uprawnia cudzoziemca do wjazdu i pobytu lub kilku następujących po sobie pobyków, które łącznie trwają więcej niż 90 dni. Okres ważności wizy ustala się na podstawie celu pobytu zadeklarowanego przez cudzoziemca i rozpoczyna się najpóźniej 3 miesiące od wydania wizy i nie może przekroczyć roku. Drugim rodzajem wizy jest wiza Schengen, która jest wizą krótkoterminową, która uprawnia do pobytu na terenie państw członkowskich strefy Schengen (26 państw) nie przekraczającym 90 na każdy 180-dniowy okres. Wizy wydawane są przez polskiego konsula we właściwej polskiej jednostce dyplomatycznej lub konsulacie.

Istotną kwestią jest, którzy obywatele innych państw potrzebują wizy według polskiego prawa. Według umów międzynarodowych obywatele, których dotyczy ruch bezwizowy przy

wjeździe na terytorium państwa strefy Schengen na krótki okres, nie dłuższy niż 90 dni w ciągu 180-dniowego okresu, nie są zobowiązani do posiadania wizy. Są to obywatele państw:

Tabela 1.1. Państwa, których obywatele mogą podróżować do Polski bez wiz (Do 90 dni)

Państwa	Dodatkowe informacje
Albania, Bośnia i Hercegowina, Była Jugosłowiańska Republika Macedonii, Czarnogóra, Gruzja, Ukraina	Zwolnienie z wymogu posiadania wiz odnosi się jedynie do posiadaczy paszportów biometrycznych
Mołdawia	Zwolnienie z wymogu posiadania wiz odnosi się jedynie do posiadaczy paszportów biometrycznych wydanych zgodnie z normami ICAO.
Serbia	Oprócz posiadaczy paszportów serbskich wydanych przez serbską Dyрекcję ds. Koordynacji (w jęz. serbskim: Koordinaciona uprava) Zwolnienie z wymogu posiadania wiz odnosi się jedynie do posiadaczy paszportów biometrycznych
Andora, Antigua i Barbuda, Argentyna*, Australia, Barbados, Brazylia, Brunei Darussalam, Chile*, Gwatemala, Honduras*, Izrael*,	

<p> Japonia*, Kanada, Korea Południowa*, Kostaryka*, Malezja*, Mauritius, Meksyk*, Monako, Nikaragua*, Nowa Zelandia, Panama*, Paragwaj, Saint Kitts i Nevis, Salwador, San Marino, Seszele, Singapur*, Stany Zjednoczone Ameryki Północnej*, Urugwaj*, Watykan, Wenezuela, Wyspy Bahama </p>	
<i>Obywatele specjalnych regionów administracyjnych Chińskiej Republiki Ludowej</i>	
Hongkong SAR*	Zwolnienie z wymogu uzyskania wizy stosuje się tylko do posiadaczy paszportu „Hong Kong Special Administrative Region”
Makao SAR*	Zwolnienie z wymogu uzyskania wizy stosuje się tylko do posiadaczy paszportu „Região Administrativa Especial de Macau

<i>Obywatele brytyjscy, którzy nie są obywatelami Zjednoczonego Królestwa Wielkiej Brytanii i Irlandii Północnej do celów prawa Unii</i>	
1. zamorscy Brytyjczycy (ang. British nationals Overseas)) 2. obywatele brytyjskich terytoriów zamorskich (ang. British Overseas Territories Citizens) 3. zamorscy obywatele brytyjscy (ang. British Overseas Citizens) 4. osoby podlegające zwierzchnictwu brytyjskiemu (ang. British Protected Persons) bezpieczeństwa poddani Korony Brytyjskiej (ang. British Subjects)	
<i>Obywatele jednostek oraz władz terytorialnych, które nie są uznawane jako państwa przez przynajmniej jedno państwo członkowskie</i>	
Tajwan	Zwolnienie z obowiązku wizowego dotyczy wyłącznie posiadaczy paszportów wydanych przez władze Tajwanu, w których figuruje numer dowodu tożsamości

***Obywatele tych państw lub specjalnych regionów administracyjnych mogą wjeżdżać do Polski ponownie w ramach ruchu bezwizowego bez konieczności zachowania 180 - dniowego limitu, na podstawie umów bilateralnych regulujących ruch bezwizowy, zawartych przez Polskę z tymi państwami przed przystąpieniem Polski do UE.*

Źródło: Opracowanie własne na podstawie (Urząd ds. Cudzoziemców, [2019](#))

Cudzoziemiec wjeżdżający na terytorium Polski mimo braku wymogu posiadania wizy musi spełnić następujące warunki:

- musi posiadać ważne dokumenty podróży,
- musi potrafić uzasadnić cel pobytu i jego warunki,

- musi posiadać wystarczającą ilość środków finansowych na pokrycie pobytu, lub udowodnić, że jest w stanie takie środki zgodnie z prawem zgromadzić,
- nie może być to osoba wpisana na listę osób, których pobyt na terenie państw strefy Schengen jest niepożądany (rejestr SIS),
- nie może być to osoba, która stanowi zagrożenie dla bezpieczeństwa państwa, porządku publicznego, zdrowia publicznego lub stosunków międzynarodowych.

Karta pobytu jest dokumentem, który uprawnia cudzoziemca wraz z innym dokumentem potwierdzającym tożsamość na wielokrotne przekraczanie granic i pobyt tzn. na zamieszkanie czasowe lub osiedlenie się na terenie Rzeczypospolitej bez konieczności posiadania wizy.

Cudzoziemiec w różnych sytuacjach może ubiegać się o kartę pobytu, która wydawana jest, gdy udzielone zostało pozwolenie na pobyt czasowy lub stały, lub pozwolenie na pobyt rezydenta długoterminowego UE. Mogą też zaistnieć względy humanitarne lub uzyskanie przez obcokrajowca statusu uchodźcy lub ochrony uzupełniającej. W zależności od przyczyny okres ważności wynosi 10 lat dla zezwolenia na osiedlenie się, 5 lat dla rezydenta długoterminowego. W sytuacjach, gdy cudzoziemcowi udzielona jest ochrona karta pobytu jest ważna 3 lata dla osoby, której przyznano status uchodźcy, 2 lata w przypadku ochrony uzupełniającej, 1 rok gdy udzielona została zgoda na pobyt tolerowany.

Sytuacje wyjątkowe, gdy cudzoziemiec może ubiegać się o status uchodźcy i ochronę międzynarodową reguluje ustawa z dnia 13 czerwca 2003 r. o udzielaniu cudzoziemcom ochrony na terytorium Rzeczypospolitej Polskiej. W Polsce cudzoziemcom można udzielić ochrony przez nadanie statusu uchodźcy, udzielenie azylu, udzielenie zgody na pobyt tolerowany bądź udzielenie ochrony czasowej. Okoliczności, które muszą zaistnieć, aby cudzoziemiec mógł być uznany za uchodźcę określone są w Konwencji Genewskiej i Protokole Nowojorskim. Status uchodźcy nadaje się również współmałżonkowi i dziecku uwzględnionym we wniosku o nadanie statusu uchodźcy.

Urząd do Spraw Cudzoziemców podaje taką definicję statusu uchodźcy: *„międzynarodowa forma ochrony udzielana osobom, które na skutek uzasadnionej obawy przed prześladowaniem w kraju pochodzenia z powodu rasy, religii, narodowości, przekonań politycznych lub przynależności do określonej grupy społecznej nie mogą lub nie chcą korzystać z ochrony tego kraju. Status uchodźcy nadawany jest także małoletniemu dziecku cudzoziemca, który uzyskał status uchodźcy w Rzeczypospolitej Polskiej, urodzonemu na tym terytorium”* (Urząd ds. Cudzoziem-

ców, 2019).

Z kolei azyl jest to „forma ochrony udzielana cudzoziemcom w sytuacji, gdy jest to niezbędne do zapewnienia im ochrony oraz gdy przemawia za tym ważny interes Rzeczypospolitej Polskiej. Decyzje w sprawach udzielania i pozbawiania azylu wydaje Szef Urzędu do Spraw Cudzoziemców, po uzyskaniu zgody ministra właściwego do spraw zagranicznych. Cudzoziemcowi, któremu udzielono azylu, Szef Urzędu do Spraw Cudzoziemców wydaje zezwolenie na osiedlenie się oraz kartę pobytu” (Urząd ds. Cudzoziemców, 2019).

Osoba, której wniosek o nadanie statusu cudzoziemca został odrzucony, może starać się o ochronę uzupełniającą, gdy istnieje ryzyko doznania krzywdy w przypadku powrotu do kraju pochodzenia takiego jak kara śmierci lub egzekucja, tortury, poniżające traktowanie lub przymus stosowana wobec ludności cywilnej ze względu konfliktu zbrojnego.

Według prawa obowiązującego w Polsce można wyszczególnić następujące sytuacje, gdy pobyt cudzoziemca jest nielegalny, które jednocześnie mogą być podstawą do wydania decyzji zobowiązującej cudzoziemca do powrotu do kraju pochodzenia.

1. Cudzoziemiec znajduje się na terytorium państwa poprzez nielegalne przekroczenie granicy poza miejscem do tego przeznaczonym. Przekracza granicę nie posiadając odpowiednich dokumentów lub za pomocą sfałszowanych dokumentów.
2. Osoba posiadająca zezwolenie na przekraczanie granic w ramach małego ruchu przygranicznego znajduje się poza strefą przygraniczną.
3. Obcokrajowiec przybywa lub przebywał na terytorium Polski bez dokumentu uprawniającego go do wjazdu i pobytu w przypadku, gdy istnieje taki wymóg np. posiadania wizy.
4. Cudzoziemiec przebywa na terenie Rzeczypospolitej mimo, iż minął okres dopuszczalnego pobytu w państwach strefy Schengen w ramach ruchu bezwizowego w okresie 180 dni.
5. Cudzoziemiec jest w Polsce mimo, iż upłynął okres ważności wizy Schengen w okresie 180 dni lub wizy krajowej.
6. Pozwolenie na pobyt cudzoziemca w RP zostało cofnięte przez uprawniony do tego organ.
7. Minął czas, w którym cudzoziemiec może przebywać na terenie Polski w ramach małego ruchu granicznego, chyba że umowy międzynarodowe stanowią inaczej.

8. Cel i warunki pobytu cudzoziemca zadeklarowane nie są zgodne z rzeczywistymi.
9. Cudzoziemiec przebywa na terenie kraju mimo, iż wniosek o nadanie statusu uchodźcy lub udzielenie ochrony międzynarodowej lub ochronie uzupełniającej został odrzucony, postępowanie zostało umorzone, lub została wydana decyzja o pozbawieniu cudzoziemca specjalnych statusów i nakaz opuszczenia kraju w wyznaczonym terminie.
10. Cudzoziemiec podjął się nielegalnego zatrudnienia lub nielegalnie prowadzi działalność gospodarczą.
11. Cudzoziemiec nie posiada wystarczających środków finansowych do pokrycia pobytu w Polsce i jego powrotu do kraju pochodzenia.
12. Cudzoziemiec widnieje w Systemie Informacyjnym Schengen do celów odmowy wjazdów jako osoba, której pobyt na terenie Rzeczypospolitej Polskiej jest niepożądany.
13. Cudzoziemiec zagraża wewnętrznemu bezpieczeństwu państwa, porządku publicznemu, zdrowiu publicznemu lub ogólnie interesowi RP.
14. Cudzoziemiec zawarł pozorne małżeństwo wyłącznie w celu legalizacji swojego pobytu.

Według polskiego prawa sam nielegalny pobyt nie stanowi przestępstwa, dopóki nie wiąże się z popełnieniem innego przestępstwa, może być jednak podstawą do zobowiązania cudzoziemca do powrotu w wyznaczonym terminie lub do przymusowego wydalenia w przypadku przekroczenia danego terminu. Wspólne normy dotyczące wydalenia cudzoziemców, którzy przebywają nielegalnie na terenie któregoś państwa członkowskiego UE zostały opisane w Dyrektywie Parlamentu Europejskiego i Rady 2008/115/WE z dnia 16 grudnia 2008 r. w sprawie wspólnych norm i procedur stosowanych przez państwa członkowskie w odniesieniu do powrotów nielegalnie przebywających obywateli państw trzecich.

1.2 Źródła danych o cudzoziemcach w Polsce

W celu oszacowania liczby cudzoziemców, którzy przebywają na terenie Polski niezgodnie z prawem należy zgromadzić dane statystyczne, które mogą pochodzić z różnych źródeł. Źródła mogą być różnie sklasyfikowane. Jedną z propozycji podziału jest podział na źródła statystyczne i pozastatystyczne. Rozróżnia się je w tym przypadku na administracyjne i statystyczne ze

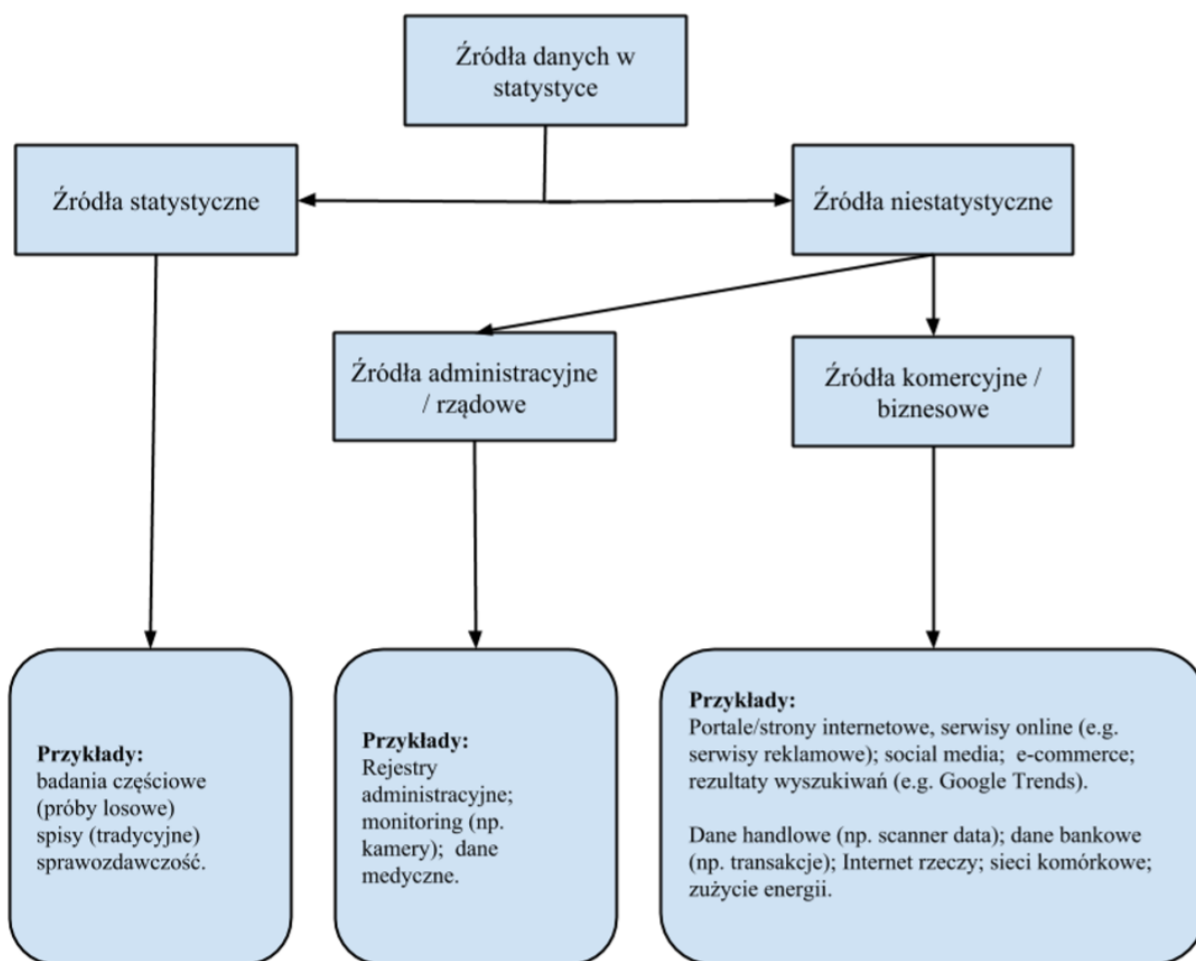
względu na cel, dla którego zostały zgromadzone. Dane administracyjne są gromadzone do celów niestatystycznych jednak są wykorzystywane w statystyce, a dane statystyczne są gromadzone bezpośrednio w celu tworzenia statystyk opisujących stan i rozwój określonych populacji, gdy żadne inne odpowiednie dane nie są dostępne (Nordbotten, 2010).

Klasycznymi źródłami danych w statystyce publicznej są spisy powszechne, które mogą być przeprowadzone na całej zbiorowości, bądź na wylosowanej próbie jako badania reprezentacyjne. Problem jest okres, który dzieli kolejne badania (np. 10 lat), który w obecnych czasach i gospodarkach opartych na informacji jest zbyt długim ostępem czasowym. Innym źródłem są badania reprezentacyjne prowadzone w określonym celu badawczym np. BAEL – Badanie Aktywności Ekonomicznej Ludności. Takie badania są prowadzone z dużo większą częstotliwością niż spisy powszechne np. miesięczną (Beręsewicz & Szymkowiak, 2015).

Oprócz klasycznych źródeł danych istnieją też nowe, nieklasyczne źródła danych takie jak np. rejestry administracyjne i źródła internetowe (portale społecznościowe, serwisy aukcyjne, fotoradary, kamery, GPS, zdjęcia satelitarne itd.), które pojawiły się wraz z rozwojem nowych technologii (Beręsewicz & Szymkowiak, 2015). Szczególnie popularnym terminem jest Big Data, czyli zbiory danych, które można opisać za pomocą trzech charakterystyk tzw. 3V 'objętość' (ang. *volume*) nawiązująca do dużego wolumenu danych, różnorodność (ang. *variety*) dane nie są ustrukturyzowane, występują ich różne typy, prędkość (ang. *velocity*) dane aktualizowane w czasie rzeczywistym, duża prędkość przetwarzania danych (Beyer & Laney, 2012). Na rysunku 1.1 przedstawiona została klasyfikacja źródeł danych w formie grafu.

Urządzenia, które połączone są ze sobą za pomocą sieci generują bardzo duże ilości danych. W tych czasach każde przedsiębiorstwo w zależności od skali działalności gromadzi dane dotyczące niemalże każdego aspektu jego działalności. Dlatego wyzwaniem staje się przetwarzanie tych danych (Fawcett & Provost, 2015). W tych czasach problem, który ogranicza wykorzystanie Big Data w statystyce publicznej jest jakość danych, od której zależy wartość poznawcza wyników przeprowadzonych analiz. Dane muszą charakteryzować się dokładnością, czyli powinny być obarczone jak najmniejszym błędem, aktualnością przedstawiający jak najbardziej aktualny stan rzeczywistości i przydatnością, czyli można je wykorzystać do badań naukowych (Cieślak, 2005).

Dane z rejestrów administracyjnych są dostępne natychmiast jednak są one różnorodne, różnią się jakością i właściwościami. Ważna jest staranna ocena, przetwarzanie i integracja (Beręsewicz & Gołata, 2019). Problemy, z którymi spotykają się naukowcy w przypadku ko-



Rysunek 1.1. Klasyfikacja źródeł danych w statystyce z uwzględnieniem ‘nowych’ źródeł oraz ich gestorów

Źródło: Beręsewicz (2016, Rozdział 1)

rzystania z tego źródła danych jest ich integracja, estymacja pośrednia, imputacja, kalibracja, stosowanie GIS (Geograficznych Systemów Informacyjnych), badania internetowe, upowszechnianie danych, ochrona ich poufności, konstruowanie niestandardowych operatów losowania (Baffour, King & Valente, 2013).

W przypadku szacowania liczby cudzoziemców przebywających w Polsce natykamy się na problem braku źródła danych, która przedstawia rzeczywistość ich liczbę, gdyż rejestry administracyjne dostarczają informacje o cudzoziemcach *de iure* czyli zarejestrowanych a statystykę interesuje sytuacja *de facto* tzn. zarejestrowani i niezarejestrowani (Beręsewicz, Gudaszewski & Szymkowiak, 2019).

W Polsce głównymi źródłami danych o cudzoziemcach są:

- System „Pobyt v.2” czyli główne narzędzie wykorzystywane w Urzędzie do Spraw Cudzo-

ziemców służące prowadzeniu krajowego zbioru rejestrów, ewidencji i wykazu w sprawach cudzoziemców.

- Rejestr PESEL czyli Powszechny Elektroniczny System Ewidencji Ludności, który jest rejestrem zawierającym informacje między innymi o obywatelach polskich oraz o cudzoziemcach lub członkach ich rodzin, którzy zostali zameldowani na pobyt stały lub czasowy i otrzymali:
 - prawo stałego pobytu,
 - zezwolenie na pobyt stały,
 - zezwolenie na pobyt rezydenta, długoterminowego Unii Europejskiej,
 - status uchodźcy,
 - ochronę uzupełniającą,
 - azyl,
 - zgodę na pobyt tolerowany,
 - ochronę czasową,
 - zgodę na pobyt ze względów humanitarnych <https://obywatel.gov.pl/dokumenty-i-dane-osobowe/czym-jest-rejestr-pesel>.
- Centralny Rejestr Ubezpieczonych, czyli zbiór wszystkich osób fizycznych zgłoszonych do ubezpieczeń prowadzony przez Zakład Ubezpieczeń Społecznych.
- Ministerstwo Rodziny, Pracy i Polityki Społecznej statystyki dotyczące zezwoleń na prace, rejestracji oświadczeń pracodawców o zamiarze powierzenia pracy cudzoziemcowi, informacji o zatrudnieniu cudzoziemców w Polsce.
- Rejestr prowadzony przez Policję w sprawie osób, które popełniły przestępstwo lub wykroczenie w zakresie danych osobowych włączając kraj obywatelstwa.
- Statystyki dostarczane przez Straż Graniczną dotyczące ruchu granicznego i cudzoziemców nielegalnie przebywających na terenie Polski.
- Dane gromadzone przez Ministerstwo Spraw Zagranicznych dotyczące liczby przyjętych wniosków wizowych, wydanych wiz oraz decyzji negatywnych z podziałem na placówki. Dane są dostępne w Internecie pod adresem: <https://dane.gov.pl/dataset/1576,209e8d50-b703-4b71-a5c2-e1bff4ce6189>.

1.3 Cudzoziemcy w Polsce w świetle dostępnych źródeł danych

Różne rejestry w Polsce w zależności od ich przeznaczenia zawierają dane o odmiennych grupach cudzoziemców. Danych pochodzących z tych rejestrów nie można jednak sumować ze względu na duże prawdopodobieństwo tego, że dana osoba może widnieć w więcej niż jednym rejestrze i może to prowadzić do wyciągnięcia błędnych wniosków o faktycznej liczbie cudzoziemców (Szałtys i in., 2018). Oczywiście rejestry nie zawierają wszystkich danych, które mogłyby świadczyć o faktycznej liczbie osób posiadających obywatelstwo inne niż Polskie, gdyż poza rejestrami są jeszcze osoby, których pobyt nie jest uregulowany prawnie.

Zmiany w ustawach dotyczących cudzoziemców często są zauważalne w zmianach wielkości populacji notowanych w rejestrach. np. w 1 maja 2014 w życie weszła nowa ustawa o cudzoziemcach, której celem było uproszczenie procedury ubiegania się o zezwolenie na pobyt cudzoziemców w Polsce. Wprowadzono także zmiany w nazewnictwie mianowicie "zamieszkanie na czas oznaczony" zastąpiono nazwą "pobyt czasowy", a "osiedlenie się" zastąpiono nazwą "pobyt stały". Wprowadzono jedno zezwolenie na pobyt i pracę, ułatwienia dotyczyły również osób studiujących w polskich uczelniach i ich absolwentów. Osoby posiadający "Kartę Polaka" mogą od wejścia tych zmian ubiegać się o prawo do pobytu stałego. Wydłużono okres pobytu czasowego z 2 lat na 3 i czas, w którym można składać wnioski o przedłużenie wizy i pobyt. Według danych i raportów urzędu do spraw cudzoziemców miało to znaczący wpływ na wzrost zainteresowania różnymi formami legalizacji pobytu. W porównaniu do roku obowiązywania starej ustawy i rok po wprowadzeniu nowej ustawy od 1 maja 2014 do 1 maja 2015 roku nastąpił wzrost liczby wszystkich złożonych wniosków o 76% (Urząd ds. Cudzoziemców, 2015).

Z końcem roku 2017 zaczęła obowiązywać ustawa z dnia 24 listopada 2017 r. o zmianie ustawy o repatriacji, ustawy o Karcie Polaka oraz ustawy o cudzoziemcach, w której wdrożono dyrektywę Parlamentu Europejskiego i Rady Unii Europejskiej, w której określone zostały warunki wjazdu i pobytu cudzoziemców spoza UE, którzy zostają przeniesieni w ramach jednego przedsiębiorstwa na terytorium Rzeczypospolitej. Ustawa ta miała również udoskonalić ramy prawne w zakresie zarządzania migracjami w Polsce. Nowelizacje te dotyczyły również zwiększenia pomocy finansowej dla repatriantów, a także objęcie statusem repatrianta małżonków repatriantów. Można przypuszczać, że udogodnienia, które zaproponowano w tej ustawie mogły skłonić niektóre osoby do migracji i osiedlenia się w Polsce.

Z początkiem 2018 roku w życie weszła Ustawa z dnia 27 października 2017 r. o zmianie ustawy o ewidencji ludności. W tejże ustawie zmienione zostały zasady nadawania numeru

PESEL cudzoziemcom rejestrującym swój pobyt na terenie Polski. Tzn. że od wprowadzania ustawy każdy cudzoziemiec, który otrzymuje pozwolenie na pobyt czasowy lub pobyt stały automatycznie otrzymuje numer PESEL bez konieczności składania osobnego wniosku. Wynikiem tej nowelizacji jest m.in. obowiązek ewidencjonowania w rejestrze PESEL każdego cudzoziemca, który przebywa w Polsce dłużej niż 30 dni i otrzymał pozwolenie na pobyt czasowy lub stały. W związku z tą ustawą już w drugiej połowie 2018 odnotowano ponad dwukrotny wzrost liczby cudzoziemców z aktualnym pozwoleniem na pobyt czasowy posiadających numer PESEL.

Poniżej zostały przedstawione dane z wybranych rejestrów. Szczególną uwagę zwrócono na rok 2018, który był podstawą późniejszych szacunków populacji cudzoziemców nielegalnie przebywających w Polsce.

Tabela 1.2 przedstawia podsumowanie liczby cudzoziemców posiadających numer PESEL w podziale na aktualny status zameldowania i wiek. Jest to stan na koniec półrocza i roku w latach 2014-2018. Liczba cudzoziemców posiadających pozwolenie na pobyt stały od 2015 roku do 2018 wzrastała z roku na rok. W 2015 roku liczba osób zarejestrowanych w rejestrze PESEL posiadających pozwolenie na pobyt stały wyniosła 40 916 a w 2018 roku 50 150 jest to wzrost o 22.6%.

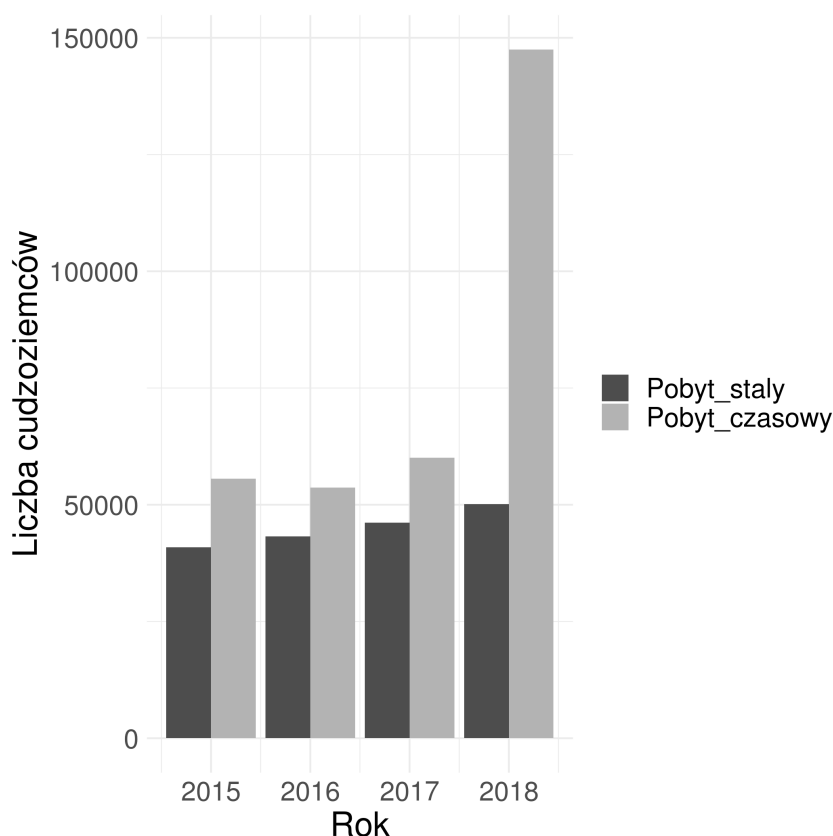
Tabela 1.2. Liczba cudzoziemców posiadających numer PESEL zgodnie ze stanem na koniec 2015-2018

		2015		2016		2017		2018	
Wiek	Status	I poł.	II poł.	I poł.	II poł.	I poł.	II poł.	I poł.	31.09
Do 17	Pobyt czasowy	6 930	6 280	6 071	6 245	7 032	7 523	14 998	20 898
	Pobyt stały	2 407	2 566	2 720	2 823	3 001	3 222	3 544	3 714
18+	Pobyt czasowy	61 312	55 540	52 592	53 645	60 757	60 056	137 290	147 463
	Pobyt stały	39 756	40 916	42 112	43 247	44 531	46 163	48 818	50 150

Źródło: Opracowanie własne na podstawie rejestru PESEL.

Biorąc pod uwagę aktualny pobyt czasowy i cudzoziemców, którzy ukończyli 18 rok życia posiadających numer PESEL zauważalna jest tendencja spadkowa w latach 2015-2016 z 61 312 osób zarejestrowanych na koniec pierwszej połowy 2015 roku do 53 645 na koniec 2016 roku. Jednak z końcem 2017 roku odnotowano wzrost liczby cudzoziemców w porównaniu z rokiem 2016 o 12% a w 2018 o 145% w porównaniu z rokiem poprzednim. Na podstawie danych z rejestru PESEL można zauważyć, iż liczba osób posiadających numer PESEL była stosunkowo nie-

wielka do 2018 roku. Rok 2018 szczególnie wyróżnia się nagłym wzrostem liczby cudzoziemców, którzy posiadają pozwolenie na pobyt czasowy i jednocześnie zostali wpisani do rejestru PESEL, a jest to spowodowane zmianą ustawy o ewidencji ludności, która zaczęła obowiązywać z początku 2018 roku. Na wykresie 1.2 jest to najbardziej odstająca obserwacja w 2018 roku. Większość cudzoziemców było pochodzenia Ukraińskiego z czego 110 335 posiadało pozwolenie na pobyt czasowy a 13 305 na pobyt stały, jest to stan na 30.09.2018.



Rysunek 1.2. Liczba cudzoziemców posiadających aktualny status pobytu zarejestrowanych w rejestrze PESEL

Źródło: Opracowanie własne na podstawie rejestru PESEL

Dane o osobach zgłoszonych do ubezpieczeń społecznych do ZUS w latach 2016-2017 zawiera tabela 1.3. Wielkość populacji cudzoziemców w tym rejestrze jest znacznie liczniejsza niż w rejestrze PESEL. Zauważalny jest wzrost liczby osób pochodzących z zagranicy pracujących w Polsce niezależnie od rozpatrywanej grupy z roku na rok. Dla grupy "pracownicy" wyniósł on około 48,5% w 2017 dla każdego typu ubezpieczenia. W 2017 roku cudzoziemcy, którzy odprowadzali składki na ubezpieczenia stanowili grupę około 439 565 osób w tym 17 644 to osoby prowadzący własną działalność gospodarczą i 251 115 to pracownicy, z czego większość osób fizycznych, bo 315 986 to cudzoziemcy z obywatelstwem Ukrainy.

Tabela 1.3. Liczba cudzoziemców odprowadzających składki na ubezpieczenia społeczne w ZUS zgodnie ze stanem na koniec 2016-2017

Typ ubezpieczenia	Grupa	2016	2017
CHOROBOWE	Działalność gospodarcza	9 602	10 309
	Osoby fizyczne	206 831	298 576
	Pracownicy	169 045	251 115
EMERYTALNE I RENTOWE	Działalność gospodarcza	16 003	17 644
	Osoby fizyczne	292 686	439 565
	Pracownicy	169 045	251 115
ZDROWOTNE	Działalność gospodarcza	19 406	21 350
	Osoby fizyczne	315 749	464 834
	Pracownicy	168 418	250 578

Źródło: Opracowanie własne na podstawie rejestru ZUS.

Kolejna tabela 1.4 zawiera dane o osobach nielegalnie przebywających w Polsce zatrzymanych na terenie kraju przez Straż Graniczną w latach 2015-2018 w poszczególnych półroczach, w podziale na płeć i grupy wiekowe. Komenda Główna Straży Granicznej gromadzi dane i publikuje je na stronie Komendy Głównej Straży Granicznej w formie statystyk na koniec każdego kwartału. W 2015 roku zatrzymano 3 180 cudzoziemców, w 2016r. – 4 458, a w 2017r. – 5 958, w pierwszej połowie 2018 roku – 3 103. Z roku na rok wzrastała liczba cudzoziemców zatrzymanych przez służby straży granicznej nie posiadających zezwolenia na pobyt. W 2017 roku liczba ta wzrosła o 33,6% w stosunku do roku poprzedniego. Najwięcej osób zatrzymanych było w wieku produkcyjnym czyli dla kobiet jest to przedział 18-59 lat, a dla mężczyzn 18-64 lata z czego największą grupę stanowili mężczyźni.

Tabela 1.4. Liczba cudzoziemców nielegalnie przebywających na terenie Polski zatrzymanych przez Straż Graniczną według półroczy w latach 2015-2018

		Kobiety			Mężczyźni		
Rok	Razem	do 17 lat	18-59 lat	60+	do 17 lat	18-64 lata	65+
2015	1 467	49	338	20	60	999	1
	1 713	55	393	26	55	1 176	8
2016	2 033	73	508	28	62	1 353	9
	2 425	70	599	28	124	1 600	4
2017	3 028	193	822	46	186	1 769	12
	2 930	141	682	33	159	1 909	6
2018	3 103	123	736	50	111	2 068	15

Źródło: Opracowanie własne na podstawie rejestru Straży Granicznej

Ostatnie źródło informacji o cudzoziemcach w Polsce brane pod uwagę w tej pracy to rejestr policyjny osób pochodzących z zagranicy, które złamały polskie prawo. Policja gromadzi dane w Krajowym Systemie Informacyjnym Policji (KSIP). Tabela 1.5 zawiera podsumowanie danych

o cudzoziemcach zarejestrowanych przez Policję w latach 2016-2018 w poszczególnych półroczach, którzy byli podejrzani o popełnienie przestępstwa i posiadali numer PESEL. Dane odnoszą się do rejestracji procesowych, notowań kryminalnych i wykroczeń w ruchu drogowym. W 2016 roku odnotowano 9 244 osób, zaś w 2017 było tych osób o 37% (12 692) więcej w porównaniu z rokiem poprzednim. W 2018 roku liczba odnotowań utrzymała się na podobnym poziomie jak w 2017 roku i wyniosła 12 576 cudzoziemców. Największą liczbę osób stanowili cudzoziemcy pochodzenia Ukraińskiego w 2016 roku było to 3 233 osób, w 2017 – 4 745 osób, a w 2018 – 5 620. Drugą i trzecią najbardziej liczną grupę stanowili Armeńczycy i Białorusini.

Tabela 1.5. Liczba cudzoziemców posiadających numer PESEL zatrzymanych przez Policję, podejrzanych o pełnienie przestępstwa, notowania kryminalne lub wykroczenie drogowe latach 2014-2018

Rok	I półrocze	II półrocze	Suma
2016	4456	4788	9244
2017	6201	6491	12692
2018	7294	5282	12576

Źródło: Opracowanie własne na podstawie Krajowego Systemu Informacyjnego Policji.

W Polsce coraz popularniejsze staje się badanie kwestii dotyczących cudzoziemców, szczególnie cudzoziemców na polskim rynku pracy. Największym ośrodkiem zajmującym się tym tematem jest Ośrodek Badań nad Migracjami przy Uniwersytecie Warszawskim, specjalizujący się on w interdyscyplinarnych badaniach procesów migracyjnych w Polsce i w Europie. Głównymi obszarami badawczymi są m.in.:

- Napływ i funkcjonowanie cudzoziemców na polskim rynku pracy.
- Teorie i metody w badaniach nad migracjami.
- Powstawanie, funkcjonowanie i integracja społeczności imigranckich w Polsce.
- Polityka migracyjna w Polsce i w UE.
- Funkcjonowanie granicy i migracje nielegalne.
- Mobilność społeczna imigrantów.

Również Narodowy Bank Polski we współpracy z Fundacją Ośrodka Badań nad Migracjami prowadzi badania dotyczące obywateli Ukrainy pracujących w Polsce.

1.4 Podsumowanie

W rozdziale pierwszym wyjaśniono zjawisko nielegalnego pobytu cudzoziemca na terenie RP w oparciu o ustawy obecnie obowiązujące w Polsce i przegląd literatury. Przytoczone zostały sytuacje, w których pobyt cudzoziemca staje się nielegalny. Przedstawiono również źródła danych, z których można czerpać informacje na temat liczebności różnych grup cudzoziemców przebywających na terenie Polski zarówno legalnie jak i nielegalnie w oparciu o oficjalne dane pochodzące z różnych rejestrów. Celem tego rozdziału było przytoczenie informacji na temat populacji cudzoziemców w Polsce, włączając nielegalną imigrację, której próbę oszacowania podjęto w dalszej części tej pracy.

Rozdział 2

Metody estymacji wielkości populacji trudnych do zbadania

2.1 Populacja trudna do zbadania

Może wydawać się, że wszystkie istniejące populacje są trudne do zbadania jednak niektóre z nich ze względu na różne czynniki stanowią większe wyzwanie w badaniach. Populacje te mogą być rzadkie, ukryte, trudne do odnalezienia bądź kontaktu, lub trudne w kooperacji z ankieterem (Tourangeau, Edwards & Johnson, 2014). Na podstawie badania ankietowego populacji na każdym jego etapie od losowania próby, przez identyfikację członków, wywiady, próby kontaktu z respondentami i ich odpytywania mogą pojawić się problemy, które są charakterystyczne dla populacji trudnej do zbadania. Cechy, którymi wyróżniają się populacje trudne do zbadania na tle innych populacji (Tourangeau i in., 2014):

1. Trudność w wylosowaniu próby (ang. *hard-to-sample*).

Mając do czynienia z populacjami trudnymi do zbadania właściwy i wyczerpujący operat losowania zazwyczaj nie istnieje. Z tego powodu stosuje się nielosowe metody poboru jednostek, aby utworzyć próbę np. kuli śnieżnej czy metoda doboru sterowaną przez respondenta (Beręsewicz i in., 2019).

W przypadku rzadko występujących populacji oprócz tego, iż stanowią niewielki odsetek innej populacji generalnej, to częstość występowania w poszczególnych podgrupach próbkowania też może być bardzo zróżnicowany. Czasami możliwe jest zidentyfikowanie grup przed pobraniem próby, w których częstość występowania rzadkiej podgrupy

jest stosunkowo wysoka przez dużą koncentrację populacji. Kolejną przeszkodą jest koszt przeprowadzania wywiadów gdy populacja jest rzadka. Przeszkodą w próbkowaniu jest również mobilność i nieuchwytność populacji. Dotyczy to osób, które nie są związane z żadnym miejscem na stałe jak np. osoby bezdomne lub osoby migrujące za pracą. W takim przypadku rozwiązaniem może być pobranie próby po zlokalizowaniu miejsca, w którym takie osoby mogą przebywać np. ze schronisk dla bezdomnych (Tourangeau i in., 2014). Do oszacowania wielkości populacji w takiej sytuacji stosuje się metody *capture-recapture* (CR) czyli metodę tzw. wielokrotnych złowień. Metoda ta jest szeroko stosowana przez biologów i ekologów do badania populacji biologicznych szczególnie dla populacji zamkniętych przy szacowaniu ich wielkości (Chao, 1987). Metody CR mimo, iż początkowo opracowywano do szacowania wielkości populacji zwierząt obecnie są stosowane do szacowania populacji ludzkich (Tourangeau i in., 2014).

2. Trudność w identyfikacji (ang. *hard-to-identify*).

Czasami ujawnienie cech danej populacji przez osobę do niej należącej wiąże się z obawą przed stygmatyzacją. Informacje wrażliwe, których respondenci mogą nie chcieć ujawnić mogą dotyczyć np. orientacji seksualnej lub nielegalnego pobytu. W takich przypadkach utrudnione jest zidentyfikowanie osób należących do takiej populacji. Przykłady takich populacji, to mniejszości seksualne, osoby uzależnione od narkotyków, alkoholicy, nielegalni cudzoziemcy itd. (Beręsewicz i in., 2019).

3. Trudność dotarcia do osób (ang. *hard-to-reach*)

Członkowie populacji mogą być trudni do zlokalizowania (ang. *hard-to-locate*) lub trudno się z nimi skontaktować. Grupy, które są trudne do zlokalizowania to członkowie koczowniczych ludów np. plemię Beduinów z południowo-zachodniej Azji i Tuareg z Afryki Północnej; mniejszości wędrownie np. Romowie tworzący diasporę czyli rozproszoną grupę etniczną, bądź naród¹; osoby tymczasowo mobilne lub przesiedlone (imigranci, uchodźcy, bezdomni); osoby które są mobilne na pewnym etapie życia jak np. studenci. Mobilność może utrudniać zlokalizowanie członków danej populacji.

Niekiedy występują bariery w dostępie do osób (ang. *barriers-to-access*). Wiąże się to z celowym izolowaniem się np. strzeżone osiedla w krajach rozwiniętych. Rozwój technologii i urządzeń mobilnych, które dają możliwość filtrowania niechcianych połączeń

¹[https://pl.wikipedia.org/wiki/Diaspora\(naród\)](https://pl.wikipedia.org/wiki/Diaspora(naród)).

i zapytań.

4. trudność w przekonaniu osoby do wzięcia udziału w badaniu (ang. *hard-to-persuade*)

Nawet gdy osoba zostanie wybrana do próby nadal może wystąpić problem z uzyskaniem zgody na przeprowadzenie badania. Potencjalnymi przeszkodami mogą być brak czasu, niskie zaangażowanie obywatelskie lub niechęć do prowadzenia rozmowy na dany temat. Przykłady takich populacji: osoby aktywne zawodowo czy cudzoziemcy (Beręsewicz i in., 2019).

5. trudność w przeprowadzaniu wywiadu (ang. *hard-to-interview*)

Ostatnim etapem, na którym pojawiają się trudności jest przeprowadzenie badania. Niektóre populacje wymagają specjalnego traktowania bądź zgody opiekuna prawnego, rodzica bądź przełożonego na przeprowadzenie wywiadu np. dzieci, więźniowie, kobiety w ciąży. Mogą wystąpić bariery w postaci niepełnosprawności fizycznej lub umysłowej, które uniemożliwiają przeprowadzenie rozmowy w standardowym schemacie. Również osoba badana może nie rozumieć języka, w którym kwestionariusz został sporządzony. Wszystkie te problemy utrudniają zebranie informacji, co nie oznacza, że uniemożliwiają zdobycie informacji w ogóle.

Powyższe cechy można przypisać do populacji cudzoziemców nielegalnie przebywających w Polsce. Trudno jest wylosować jednostki do próby statystycznej, gdyż nie istnieje pełen operat losowania i brakuje całościowych źródeł statystycznych, które umożliwiłyby dostęp do informacji o cudzoziemcach w Polsce. Występuje również problem z identyfikacją takich osób, gdyż przez to, iż mogą narazić się na konsekwencje prawne związane z brakiem legalnego pozwolenia na pobyt, nie są one skłonne do ujawniania takich informacji. Wyzwaniem staje się również dotarcie do takich osób i skontaktowanie się z nimi, gdyż miejsce ich zamieszkania nie jest stałe, lub gdy mamy do czynienia z celowym izolowaniem się od społeczeństwa w obawie przed ujawnieniem łamanego przez nich prawa. Z tego również powodu trudnością może być przekonanie tych osób do wzięcia udziału w badaniu. Istnieje także duże prawdopodobieństwo wystąpienia barier językowych. Wszystkie wymienione przeszkody prowadzą do konkluzji, iż oszacowanie populacji cudzoziemców przebywających nielegalnie na terenie Polski może być zadaniem o dużej złożoności. Badacze są zainteresowani różnymi charakterystykami populacji jednak w tej pracy nałożono nacisk na jej wielkość. Dlatego podjęto próbę oszacowania liczby cudzoziemców przebywających nielegalnie w Polsce wykorzystując do tego celu metodę estymacji opartą

na technikach CR, zaproponowaną przez Li-Chung Zhang'a wykorzystującą hierarchiczny model gamma Poisson'a. Źródłami danych, na których model zastosowano były dostępne rejestry administracyjne w Polsce przedstawiające różne grupy cudzoziemców.

2.2 Metody estymacji wielkości populacji

W literaturze istnieje kilka podejść do zagadnienia szacowania liczby nielegalnych imigrantów. Jandl (2004) w swoim artykule wymienia techniki szacowania, których zastosowana metodologia jest jasno określona:

1. Metody rezydualne (ang. *residual methods*)

Podstawowa idea podejścia rezydualnego szacowania populacji polega na porównaniu analitycznego bądź demograficznego oszacowania populacji legalnie zamieszkującej państwo z wielkością populacji na jaką wskazuje spis powszechny lub badania ankietowe, aby otrzymać szacunkową wartość populacji cudzoziemców nielegalnie rezydujących w kraju. Powstałe różnice podlegają dalszym korektom w celu określenia stopnia niedoszacowania legalnych i nielegalnych imigrantów w ankiecie (Passel, 2007).

Warianty tej metody różnią się w następujących obszarach (Passel, 2007):

- źródeł danych użytych do badania,
- definicji populacji branej pod uwagę,
- geograficznych i demograficznych cech populacji,
- metod stosowanych do szacowania składników zmian demograficznych,
- założeń dotyczących zasięgu spisu / ankiety,
- definicji legalnej populacji i pomiar składników legalnej populacji.

2. Metody mnożnikowe

Przy stosowaniu tych metod oczekuje się lepszego oszacowania liczby cudzoziemców nielegalnie przebywających w kraju. Jednak należy w przypadku takiego podejścia założyć, iż zmienna, która będzie szacowana w tym przypadku wielkość populacji cudzoziemców nielegalnie przebywających na terenie danego kraju charakteryzuje się stałym stosunkiem do zmiennych obserwowalnych np. wielkość populacji legalnych imigrantów. Tak więc istotą jest znalezienie odpowiedniego mnożnika (Jandl, 2004).

Bardziej zaawansowane podejścia metod mnożnikowych wiążą się z bardziej zaawansowanymi obliczeniami obejmującymi metody demograficzne, którymi można posłużyć się do oszacowania wielkości populacji np.:

- porównywanie struktur wiekowych populacji,
- wskaźnik urodzeń i zgonów wśród nielegalnych rezydentów w porównaniu z odpowiadającymi im wskaźnikami pośród legalnych imigrantów,

oraz inne metody:

- szacowanie zużycia energii, elektrycznej,
- szacowania konsumpcji chleba.

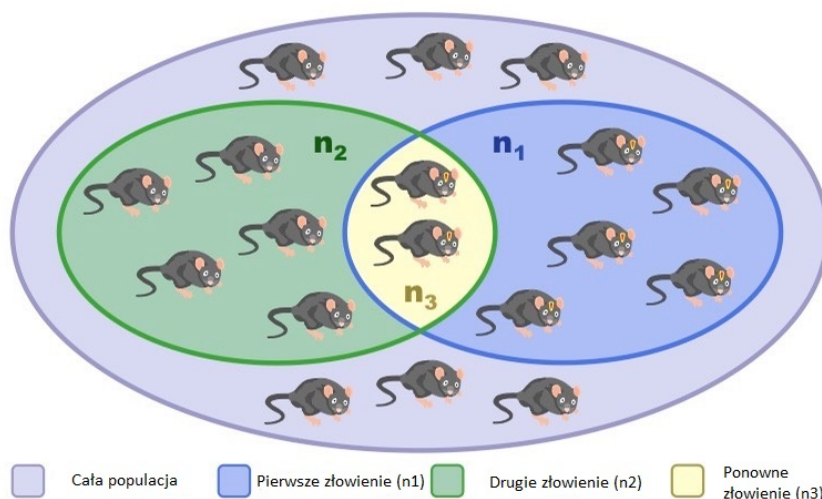
3. Badania oparte na źródłach ankietowych

Szczególny wariant metody mnożnikowej polega na badaniu "eksperckim" tematu i opracowywaniu własnych estymacji do odpowiadających mnożników (Jandl, 2004).

Przykładem wykorzystania tej metody było badanie pracodawców w różnych sektorach biznesowych przeprowadzone w 2001 roku w Szwajcarii przez Piguet i innych (Piguet & Losa, 2002). Badacze przeprowadzili ankietę wśród 800 pracodawców pytając o ich opinię na temat odsetka nielegalnych pracowników z zagranicy zatrudnionych w ich branży. Na podstawie wyników szacowana była populacja nielegalnie zatrudnionych obcokrajowców. Jednak metoda ta w dużej mierze opiera się na zgadywaniu danej wartości. W zależności od wybranej przez autora metody czy będzie to średnia ważona czy mediana, wartość w tej pracy według autorów szacuje się na zakres 70,000 – 180,000 osób (Piguet & Losa, 2002).

4. Metoda tzw. wielokrotnych złowień (ang. *Capture-Recapture method*)

Metody te wywodzą się z nauk przyrodniczych, jednak zaczęto stosować je również do szacowania populacji ludzkich. W przypadku zwierząt typowym sposobem jest umieszczenie siatek, które służą do wielokrotnego wyłapywania jednostek danej populacji. Każdy osobnik jest znakowany i następnie wypuszczany. W każdej kolejnej próbie znajduje się osobniki złapane po raz pierwszy i notuje się jednostki, które wcześniej były już złapane. Na koniec analizuje się historię złowień. Ten system nazywany jest *mark-recapture*, *tag-recapture*, lub *multiple-record system* (Beręsewicz i in., 2019).



Rysunek 2.1. Metoda capture-recapture

Źródło: <https://ib.bioninja.com.au/options/option-c-ecology-and-conser/c5-population-ecology/population-sampling.html>

Rysunek 2.1 przedstawia ideę stojącą za metodą CR. Do zbadania liczebności badanej populacji należy wyznaczyć obszar, na którym występuje. Następnie na podstawie cech populacji dostosowuje się harmonogram 'złowień'. Pierwszego dnia dokonuje się pierwszego połowu części populacji. Złapane jednostki zostają trwale oznaczone a następnie uwolnione. Pierwszą złowioną grupę oznaczono na rysunku 2.1 symbolem n_1 . Po upływie określonego w harmonogramie czasu, który ma umożliwić ponowną integrację schwytanych osobników z resztą populacji dokonuje się drugiego połowu i grupę tę oznaczamy n_2 . Podczas drugiego połowu zarówno nieoznaczone jednostki są liczone jak i oznaczone w pierwszym złowieniu i oznacza się je symbolem n_3 . Na podstawie trzech wartości n_1, n_2, n_3 przy użyciu estymatora Lincolna-Petersena (LP) szacuje się wielkość całej populacji N . Estymator LP można wyliczyć za pomocą formuły:

$$\hat{N} = \frac{n_1 n_2}{n_3}. \quad (2.1)$$

Dla dwóch pobranych próbek lub źródeł szacowanie populacji odzwierciedla proporcję jednostek wykrytych w obu próbkach. Podstawowym ryzykiem tej metody jest założenie, że złapanie (ang. *capture*) i ponowne złapanie (ang. *recapture*) danej jednostki są niezależne. W populacjach rzadkich "nieuchwytność" jednostek jest zmienna lub jest różna w ramach warstw próbkowania co powoduje błędy w mierzeniu korelacji (Tourangeau i in., 2014). Uproszczona interpretacja wygląda następująco: w przypadku niewielkiej

liczby kolejnych złowień osobników w kolejnych próbach można założyć, iż wielkość całej populacji jest większa niż unikatowych wystąpień jakie złowiono. W odwrotnej sytuacji gdy liczba ta jest stosunkowo wysoka można wnioskować, że złowiono większość jednostek danej populacji (Beręsewicz i in., 2019).

W ramach metod *capture-recapture* występuje wiele różnych modeli statystycznych np. modele log-liniowe, modele klas ukrytych czy uogólnione modele liniowe. Wybór metody uzależniony jest od dostępności źródeł i spełnienia założeń odpowiednich metod (Beręsewicz i in., 2019).

Założenia, które muszą być spełnione w przypadku, gdy dostępne jest tylko jedno źródło danych szacując wielkość populacji z wykorzystaniem metod *capture-recapture* (Wolter, 1986):

- (a) każda jednostka populacji charakteryzuje się dodatnim prawdopodobieństwem wystąpienia w danych źródłach,
- (b) populacja jest zamknięta,
- (c) źródła danych są niezależne,
- (d) nie występują błędy pokrycia i duplikaty,
- (e) nie występują błędy łączenia źródeł,
- (f) jednorodne prawdopodobieństwa wystąpienia w co najmniej jednym źródle danych.

Niespełnienie tych założeń będzie skutkowało obciążonymi szacunkami wielkości populacji.

W literaturze można odnaleźć wiele przykładów wykorzystania metody *capture-recapture* np. Bakker i in. (2014) podjął się próby oszacowania liczby cudzoziemców nielegalnie przebywających w Holandii. Dla Norwegii Zhang (2008) oszacował liczbę nieautoryzowanych cudzoziemców w pracy pod tytułem "Developing methods for determining the number of unauthorized foreigners in Norway". Praca badawcza pod tytułem "Cudzoziemcy na krajowym rynku pracy w ujęciu regionalnym", zawiera badania i szacunki populacji cudzoziemców, którzy czasowo przebywali na terenie Polski ze szczególnym uwzględnieniem osób pracujących w Polsce (Szałtys i in., 2018). Celem pracy pt. "Estimating the number of HIV-infected injection drug users in Bangkok: a capture–recapture method" było oszacowanie liczby osób używających narkotyków iniekcyjnych zakażonych wirusem HIV w Bangkoku, aby umożliwić planowanie usług zdrowotnych dla tej populacji obciążonej chorobą AIDS (Mastro i in., 1994).

2.3 Hierarchiczny model Gamma-Poissona (Zhang, 2008)

2.3.1 Podstawowe założenia modelu

Populacja cudzoziemców nielegalnie przebywających w Polsce jak już wspomniano wcześniej jest populacją trudną do zbadania. Cudzoziemców czyli osoby nie posiadające obywatelstwa polskiego można klasyfikować według następujących kryteriów:

- Status wjazdu: legalny, nielegalny,
- Status pobytu: stały, czasowy, tolerowany, legalny, nielegalny
- Status zatrudnienia: legalny, nielegalny, brak

Celem tej pracy jest oszacowanie liczby cudzoziemców nielegalnie przebywających na terenie Polski więc skupiono się na statusie pobytu jako głównej charakterystyce. W rozdziale pierwszym w pierwszym podrozdziale zostały przybliżone sytuacje, w których wjazd i pobyt na terenie kraju jest legalny bądź nielegalny.

Przyjęcie założenia, że cudzoziemiec przebywający nielegalnie na terenie Polski to osoba, która fizycznie przebywa w danym momencie na terenie kraju niezgodnie z Polskim prawem byłoby poprawne jedynie w momencie zbierania danych od drzwi do drzwi w spisie ludności, co z oczywistych względów jest rzeczą niemożliwą do wykonania. Nie można zatem przyjąć takiego założenia, gdyż dane dotyczące obcokrajowców wymagają, aby były akumulowane w czasie i niemożliwa jest weryfikacja w określonym punkcie czasowym.

Może wystąpić również problem z identyfikacją osób, co może prowadzić do wielokrotnego rejestrowania tych samych jednostek. Ponadto status legalności pobytu w czasie gromadzenia danych może zmieniać się kilkakrotnie.

Należałoby również rozróżnić nielegalny tymczasowy pobyt cudzoziemca od nielegalnego rezydowania w Polsce. Biorąc pod uwagę to, że są sytuacje gdy wnioski o udzielenie prawa do pobytu są składane przez osoby, które już są na jego terenie i starają się np. o status uchodźcy. Wobec tego nie każdy pobyt nielegalny powinien być w taki sposób identyfikowany. Kolejnym problemem jest przypadkowa zmienność, która naturalnie może wystąpić nawet w jednodniowym odstępie czasowym podczas badania. Dlatego oczekuje się teoretycznego stabilnego rozmiaru, aby uniknąć przypadkowej zmienności losowej (Zhang, 2008).

Metoda rezydualna jest obecnie najlepiej zbadaną i udokumentowaną metodą estymacji (Zhang, 2008). Metoda, która stosowana jest od ponad 50 lat przez Primary Health Care (PHC) i Urban Institute w Stanach Zjednoczonych przedstawia się następująco:

Łączna liczba nielegalnych cudzoziemców mieszkających na terenie kraju równa się całkowitej liczbie wszystkich cudzoziemców pomniejszonej o liczbę osób legalnie zamieszkujących kraj (Passel, 2007).

$$U = A - L, \quad (2.2)$$

gdzie: U to całkowita liczba cudzoziemców przebywających nielegalnie w kraju (wykrytych i niewykrytych), A to całkowita liczba cudzoziemców (legalnie i nielegalnie przebywających w kraju) oraz L to całkowita liczba cudzoziemców legalnie przebywających w kraju.

Ze względu na różne zdarzenia jak np. śmierć czy emigracja, zmienne po prawej stronie wymagają regulacji. Biorąc pod uwagę wcześniejsze rozważanie możliwe jest jedynie wyznaczenie wartości teoretycznych lub oczekiwanych nie zaś wartości rzeczywistych dla danego okresu. To znaczy, że idąc za tym rozumowaniem w tej metodzie otrzymujemy wartość U jako wartość przybliżoną, teoretyczną i stabilną.

Na podstawie powyższego modelu Zhang (2008) zdefiniował następujące równanie (2.5)

$$\xi = E(M|N) \quad (2.3)$$

gdzie: M to wielkość populacji cudzoziemców nielegalnie przebywających w danym dniu na terenie określonego kraju, N to wielkość populacji cudzoziemców przebywających legalnie w danym dniu na terenie określonego kraju oraz $f(M|N)$ to rozkład warunkowy zmiennej M względem N .

M jest zmienną losową, podczas gdy N jest wartością znaną (na przykład stanem rejestru populacji na określony dzień). Zatem parametr docelowy ξ jako wartość teoretyczna określająca liczbę nielegalnych cudzoziemców stanowi wartość oczekiwaną zmiennej losowej M zależnej od danej wartości N . Wartość ta jest pozbawiona fałszywej zmienności dzięki warunkowej wartości oczekiwanej, pozbawiona jest sezonowości a populacja jest stała. Wprowadzenie N jako warunku funkcji miało na celu objaśnianie rozmiaru zmiennej M i jej interpretacji. Zmienność M jest silnie związana ze zmiennością N . Zmienna N nie podlega dużym wahaniom sezonowym za czym idą niewielkie sezonowe wahania M . Wartość teoretyczna nie przedstawia

wartości tylko dla danego dnia, ale również dla okresu wokół tej daty.

Idąc dalej za rozważaniami Zhang (2008) należy przyjąć nowe zmienne, mianowicie m , która reprezentuje liczbę osób nie posiadających pozwolenia na pobyt w kraju, zaobserwowanych w określonym przedziale czasowym (np. w ciągu roku, kwartału).

Wprowadzono również zmienną n , która przedstawia liczbę osób pochodzących z zagranicy, którzy złamali prawo i zostali odnotowani w rejestrze Policji. W tym przypadku również wybiera się przedział czasowy.

Wybór okresu zamiast jednego dnia tożsamego z aktualnością danego rejestru jest podyktowany liczbą obserwacji oraz przypadkowymi wahaniami, które w takim dniu mogłyby wystąpić. Dodatkowo, dostęp do tak szczegółowych danych może być utrudniony. Również należy wziąć pod uwagę, to iż policja czy Straż Graniczna zwykle publikują lub przekazują dane w ujęciu kwartalnym czy rocznym.

Model, który zaproponował Zhang (2008) opiera się na zmiennych n , m , i N . Zmiennej n nie można jednak bezpośrednio przekładać na zmienną m lub N , gdyż na podstawie dostępnych danych nie można jednoznacznie stwierdzić czy cudzoziemiec, który łamie prawo, przebywa w kraju legalnie czy też nie. Jednak można założyć, że część z tych osób jest nielegalnymi imigrantami.

Wskaźniki n/N i m/M nie mogą być również rozważane jako wskaźniki złapań (ang. *catch rate*) dla populacji referencyjnej i celowej. Dlatego uproszczony model taki jak w równaniu (2.4):

$$E(m/M|n, N) = n/N, \quad (2.4)$$

mógłby prowadzić do błędnych wniosków dotyczących zależności między m czyli zidentyfikowaną liczbą osób przebywających nielegalnie, a M czyli hipotetyczną wielkością populacji cudzoziemców, którzy nie posiadają dokumentów uprawniających do pobytu. Dlatego należy przyjąć model, który pozwoli na oszacowanie wartości zmiennej M biorąc pod uwagę n , m i N . Zhang (2008) w swojej pracy zaproponował modelowanie uwzględniające efekty stałe i mieszane.

Niech $i = 1, \dots, t$ stanowi kolejne subpopulacje zarówno dla populacji celu i referencyjnej pogrupowane na podstawie kraju pochodzenia i obywatelstwa. W modelu tym założono, że obserwowana liczba nielegalnych imigrantów ma rozkład Poissona z parametrem λ_i oznacza-

nym przez:

$$m_i \sim \text{Poisson}(\lambda_i). \quad (2.5)$$

Zhang (2008) przyjął, że parametr λ_i powinien być zależny od dwóch innych wielkości, mianowicie od:

1. całkowitej liczby cudzoziemców nielegalnie rezydujących w kraju pochodzących z kraju i oznaczonej przez M_i ,
2. prawdopodobieństwa obserwacji tzn. że osoba, bez pozwolenia na pobyt będzie figurowała w rejestrze straży granicznej lub innej instytucji oznaczone przez p_i . Wynika z tego, że $\lambda_i = M_i p_i$.

Dodatkowo, niech:

$$u_i = M_i p_i / E(M_i p_i | n_i, N_i) \quad (2.6)$$

gdzie $E(M_i p_i | n_i, M_i)$ oznacza warunkową wartość oczekiwaną $M_i p_i$ dla danego n_i i N_i a u_i to efekt losowy, który odpowiada za heterogeniczność wariancji dla badanych krajów.

Biorąc pod uwagę powyższe relacje otrzymujemy:

$$\lambda_i = \mu_i u_i, \quad (2.7)$$

gdzie

$$\mu_i = E(M_i p_i | n_i, N_i) = E(M_i | N_i) \cdot E(p_i | M_i, n_i, N_i). \quad (2.8)$$

W tym modelu Zhang (2008) przyjął następujące założenia

$$\xi_i = E(M_i | N_i) = N_i^\alpha, \quad (2.9)$$

$$E(p_i | M_i, n_i, N_i) = E(p_i | n_i, N_i) = \left(\frac{n_i}{N_i}\right)^\beta, \quad (2.10)$$

$$u_i \sim \text{Gamma}(1, \phi), \quad (2.11)$$

gdzie $\text{Gamma}(1, \phi)$ to rozkład Gamma z wartością oczekiwaną $E(u_i) = 1$ i wariancją $V(u_i) = 1/\phi$.

Równania od (2.9) do (2.11) tworzą razem hierarchiczny model gamma Poissona. Zauważono, że z modelu wynika, że:

$$E(m_i / M_i | M_i, n_i, N_i) = (n_i / N_i)^\beta. \quad (2.12)$$

Termin hierarchiczność wprowadzono z powodu zmienności losowej, która występuje na dwóch różnych poziomach. Na poziomie populacji parametr Poissona λ_i zależy od u_i która jest losową zmienną z rozkładu Gamma dla poszczególnych narodowości. Z kolei μ_i zawiera efekty α i β , które są stałe we wszystkich krajach. Na drugim poziomie biorąc pod uwagę u_i czyli również λ_i obserwacja m_i podlega przypadkowemu błędowi poboru próby zgodnie z rozkładem Poissona.

Równania (2.9) i (2.10), których iloczyn daje oczekiwaną wartość parametru Poissona i pozwolił na wprowadzenie do modelu wartości M_i , gdyż nie jesteśmy zainteresowani modelem, który wyjaśnia jedynie obserwowalne wartości m_i dla kolejnych $i = 1, \dots, t$.

Parametry α i β są identyfikowalne i model może być testowanych empirycznie pod kątem dopasowania ich do obserwowanych danych. Gdy mamy już oszacowanie α możemy wykorzystać je do oszacowania $\xi_i = E(M_i | N_i)$. Wiarygodność otrzymanych wyników zależy od tego czy spełnione zostały warunki modelu.

2.3.2 Estymacja parametrów metodą największej wiarygodności

Parametr ξ wraz z estymatorem badanej zmiennej ξ opisują odpowiednio wzory:

$$\xi = \sum_{i=1}^t E(M_i | N_i) = \sum_i N_i^\alpha, \quad (2.13)$$

$$\hat{\xi} = \sum_i N_i^{\hat{\alpha}}, \quad (2.14)$$

gdzie $\hat{\alpha}$ jest estymatorem parametru α . Dalej użyjemy metody największej wiarygodności (MNW). Oznaczmy przez $L(\eta, \mathbf{m})$ prawdopodobieństwo dla $\eta = (\alpha, \beta, \phi)$ dla m_i dla kolejnych $i = 1, \dots, t$. Zgodnie z modelem Gamma-Poisson'a określonym w poprzednim podrozdziale

mamy:

$$f(m_i, u_i; \eta) = \frac{e^{-\mu_i u_i} (\mu_i u_i)^{m_i}}{m_i!} \cdot \frac{\phi^\phi u_i^{\phi-1} e^{-\phi u_i}}{\Gamma(\phi)} = \frac{\mu_i^{m_i} \phi^\phi}{m_i! \Gamma(\phi)} e^{-u_i (\mu_i + \phi)} u_i^{m_i + \phi - 1}, \quad (2.15)$$

gdzie $\Gamma()$ jest funkcją gamma. Zatem,

$$\begin{aligned} f(m_i; \eta) &= \int_0^\infty f(m_i, u_i; \eta) d(u_i) \\ &= \frac{\mu_i^{m_i} \phi^\phi}{m_i! \Gamma(\phi)} \int_0^\infty e^{-(\sqrt{u_i})^2 (\mu_i + \phi)} (\sqrt{u_i})^{2(m_i + \phi - 1)} 2\sqrt{u_i} d(\sqrt{u_i}) \\ &= \frac{\mu_i^{m_i} \phi^\phi}{m_i! \Gamma(\phi)} (\mu_i + \phi)^{-(m_i + \phi)} \Gamma(m_i + \phi), \end{aligned} \quad (2.16)$$

na podstawie tożsamości $\int_0^\infty e^{-\gamma z^2} z^k dz = \frac{1}{2} \gamma^{-\frac{k+1}{2}} \Gamma(\frac{k+1}{2})$, dla $z = \sqrt{u_i}$ i $k = 2(m_i + \phi) - 1$. Zależnie od m_i, u_i ma rozkład gamma z średnią $(m_i + \phi)/(\mu_i + \phi)$ i wariancją $(m_i + \phi)/(\mu_i + \phi)^2$. Prawdopodobieństwo wyznaczone jest przez funkcję wiarygodności

$$L(\eta; \mathbf{m}) = \prod_{i=1}^t f(m_i; \eta). \quad (2.17)$$

Dla ułatwienia estymacji stosuje się logarytm funkcji wiarygodności, który bez uwzględnienia elementów stałych dany jest wzorem (2.18)

$$l(\eta; \mathbf{m}) = \sum_{i=1}^t l_i(\eta), \quad (2.18)$$

gdzie po odpowiednich przekształceniach (2.16) otrzymujemy

$$\begin{aligned} l_i(\eta) &= m_i \log \mu_i - (m_i + \phi) \log(\mu_i + \phi) + \log \Gamma(m_i + \phi) + \phi \log \phi - \log \Gamma(\phi) \\ &\doteq m_i \log \mu_i - (m_i + \phi) \log(\mu_i + \phi) + \phi \log \phi \\ &\quad + (m_i + \phi - 0.5) \log(m_i + \phi) - (m_i + \phi) - (\phi - 0.5) \log(\phi) + \phi \\ &= m_i \log \mu_i - (m_i + \phi) \log(\mu_i + \phi) + (m_i + \phi - 0.5) \log(m_i + \phi) + 0.5 \log \phi, \end{aligned} \quad (2.19)$$

wykorzystując przybliżenie Stirling'a $\log \Gamma(z) \doteq (z - 0.5) \log(z) + 0.5 \log(2\pi) - z$.

Średnia μ_i jest funkcją liniową w skali logarytmicznej oznaczoną przez $\log \mu_i = x_i^T \gamma$

z wektorem zmiennych x_i i parametrem γ .

Pierwsze pochodne funkcji $l_i(\eta)$ względem γ i μ_i mają następującą postać

$$\frac{\partial l_i(\eta)}{\partial \gamma} = \frac{\partial l_i(\eta)}{\partial \mu_i} \frac{\partial \mu_i}{\partial \log \mu_i} \frac{\partial \log \mu_i}{\partial \gamma} = \frac{\partial l_i(\eta)}{\partial \mu_i} \mu_i x_i = \frac{m_i - \mu_i}{\mu_i + \phi} \phi x_i, \quad (2.20)$$

gdzie $\partial l_i(\eta) / \partial \mu_i = m_i / \mu_i - (m_i + \phi) / (\mu_i + \phi)$,

$$\frac{\partial l_i(\eta)}{\partial \phi} = -\log(\mu_i + \phi) - \frac{m_i + \phi}{\mu_i + \phi} + \log(m_i + \phi) + \frac{m_i + \phi - 0.5}{m_i + \phi} + \frac{1}{2\phi}. \quad (2.21)$$

Dodatkowo, aby wyznaczyć błędy standardowe należy określić hessian czyli pochodne drugiego stopnia względem badanych parametrów, które dane są poniżej

$$\frac{\partial^2 l_i(\eta)}{\partial \gamma \partial \gamma^T} = \frac{\partial^2 l_i(\eta)}{\partial \mu_i^2} \mu_i x_i \frac{\partial \mu_i}{\partial \gamma^T} + \frac{\partial l_i(\eta)}{\partial \mu_i} x_i \frac{\partial \mu_i}{\partial \gamma^T} = - \left(\frac{m_i + \phi}{\mu_i + \phi} \mu_i \phi \right) x_i x_i^T, \quad (2.22)$$

$$\frac{\partial^2 l_i(\eta)}{\partial \phi^2} = - \frac{2\mu_i + \phi - m_i}{(\mu_i + \phi)^2} + \frac{m_i + \phi + 0.5}{(m_i + \phi)^2} - \frac{1}{2\phi^2}, \quad (2.23)$$

oraz

$$\frac{\partial^2 l_i(\eta)}{\partial \gamma \partial \phi} = \left(\partial \left(\frac{\partial l_i(\eta)}{\partial \mu_i} \right) / \partial \phi \right) \mu_i x_i = - \frac{\mu_i - m_i}{(\mu_i + \phi)^2} \mu_i x_i = \left(\frac{\partial^2 l_i(\eta)}{\partial \phi \partial \gamma^T} \right)^T. \quad (2.24)$$

W celu wyznaczenia błędów standardowych oszacowań po wyznaczeniu pochodnych drugiego rzędu otrzymujemy macierz kwadratową Hessa zawierającą pochodne cząstkowe optymalizowanej funkcji względem szacowanych parametrów. Macierz odwrotna stanowi oszacowanie macierzy kowariancji parametrów funkcji i na głównej przekątnej tej macierzy znajdują się elementy do oszacowania ich wariancji a pierwiastki kwadratowe tych pierwiastków stanowią wartości błędów standardowych.

Aby wyznaczyć wektor parametrów η należy rozwiązać równanie dane wzorem (2.25)

$$\frac{\partial l(\eta; \mathbf{m})}{\partial \eta} = \sum_{i=1}^t \frac{\partial l_i(\eta)}{\partial \eta} = 0. \quad (2.25)$$

Do rozwiązywania układu równań wiarygodności można posłużyć się metodą Newton'a-

Raphson'a, która jest metodą iteracyjną. W każdym kroku wyznacza się lokalne oszacowanie parametrów funkcją kwadratową logarytmu funkcji wiarygodności, czego celem jest wyznaczenie punktu maksymalizującego funkcję wiarygodności (Biecek, 2013).

Równanie (2.10) można przedstawić w postaci zlogarytmowanej.

$$\log(E(p_i | n_i, N_i)) = \beta \log\left(\frac{n_i}{N_i}\right). \quad (2.26)$$

Wartości dla n_i / N_i i m_i / M_i są w przybliżeniu do wartości binarnych w związku z tym dokonano transformacji logistycznej gdzie:

$$\text{logit}(p_i) = \log\{p_i / (1 - p_i)\} = \log(p_i) - \log(1 - p_i), \quad (2.27)$$

Przekształcenie logarytmiczne pozwala na połączenie dwóch równań modelu (2.9) i (2.10) na tej samej skali w celu uzyskania:

$$\mu_i = N_i^\alpha \left(\frac{n_i}{N_i}\right)^\beta, \quad (2.28)$$

Z równania (2.28) wynika, że:

$$\log(\mu_i / N_i) = (\alpha - 1) \log N_i + \beta \log(n_i / N_i), \quad (2.29)$$

czyli:

$$\log(m_i / N_i) = (\alpha - 1) \log N_i + \beta \log(n_i / N_i) + \varepsilon_i, \quad (2.30)$$

Ta postać zlinearyzowana modelu posłuży jako oszacowanie startowe. Używamy oszacowań α i β jako wartości początkowe dla parametrów modelu gamma Poisson'a i odwrotność oszacowanego $V(\varepsilon_i)$ jako wartość początkową ϕ . Do oszacowania wartości modelu można wykorzystać pakiet `maxLik` (Henningsen & Toomet, 2011) zaimplementowanym w R (R Core Team, 2018).

2.3.3 Implementacja modelu w pakiecie R

W tym podrozdziale zaimplementowano hierarchiczny model gamma Poissona w języku programowania R. Oszacowanie parametrów $\gamma(\alpha, \beta)$ i ϕ dla funkcji wiarygodności i maksymalizację tej funkcji z użyciem metody Newtona-Raphsona wygenerowano za pomocą kodu, który

zamieszczono poniżej. Do maksymalizacji funkcji wiarygodności wykorzystano pakiet `maxLik`. Pakiet `maxLik` jest to zestaw funkcji i narzędzi do estymacji modeli metodą największej wiarygodności. W pakiecie zaimplementowano różne algorytmy do wyznaczania estymatorów funkcji największej wiarygodności takie jak np. metoda Newton’a-Raphson’a, której wynikiem jest obiekt typu *maxLik* (Henningsen & Toomet, 2011). Najważniejszą funkcją pakietu jest funkcja `maxLik`, która przyjmuje następujące argumenty:

```
maxLik(logLik, grad = NULL, hess = NULL, start, method,
constraints=NULL, ...)
```

1
2

- `logLik` – Funkcja wiarygodności, której pierwszym argumentem musi być wektor parametrów, które mają zostać oszacowane. Funkcja musi zwracać pojedynczą wartość logarytmu, albo numeryczny wektor, w której każdy komponent jest wartością logarytmu prawdopodobieństwa odpowiadającej mu obserwacji.
- `grad` – Gradient funkcji wiarygodności. Funkcja musi posiadać jako pierwszy argument wektor parametrów do oszacowania. Musi zwrócić pojedynczy wektor gradientu o długości równej liczbie parametrów lub macierz, w której każdy rząd jest wektorem gradientu odpowiadającej mu indywidualnej obserwacji. Dla `grad=NULL`, zostanie użyty gradient numeryczny.
- `hess` – Hessian funkcji wiarygodności. Musi zawierać wektor parametrów jako pierwszy argument. Musi zwrócić macierz kwadratową. Dla `hess=NULL`, zostanie użyty numeryczny Hesjan.
- `start` – Wektor numeryczny, początkowa wartość parametrów. Jeśli ma nazwy, będą one również użyte do nazwania wyników.
- `method` – Metoda maksymalizacji funkcji wiarygodności. metoda maksymalizacji „NR” dla Newtona-Raphsona, „BFGS” dla Broyden-Fletcher-Goldfarb-Shanno, „BFGSR” dla algorytmu BFGS zaimplementowanego w R, „BHHH” dla Berndta-Halla -Hall-Hausman, „SANN” dla Simulated annealing, „CG” dla metody gradientu sprzężonego lub „NM” dla Nelder-Meada. Dozwolone są małe litery (takie jak „nr” dla Newtona-Raphsona). W razie nie zadeklarowania metody odpowiednia metoda zostanie wybrana automatycznie.

- constraints - Wartość NULL dla nieograniczonej maksymalizacji lub lista, określająca ograniczenia.
- ... – parametry przekazywane funkcjom podanym w argumentach loglik, grad i hess.

Wynikiem tej funkcji jest obiekt typu *maxLik*, na którym można zastosować funkcje *summary* w celu przedstawienia wyników, *coef* do wyodrębnienia szacunku parametrów, *vcov* do obliczania macierzy kowariancji wariancji szacowanych parametrów, *stdEr* do obliczenia błędów standardowych.

```
ll <- function(param, data) {
  alpha <- param[1]
  beta <- param[2]
  phi <- param[3]
  m <- as.matrix(data$m)
  n <- as.matrix(data$n)
  N <- as.matrix(data$N)
  mu <- N^alpha*(n/N)^beta
  res <- m * log(mu) - (m + phi) * log(mu + phi)
  + (m + phi - 0.5) * log(m + phi) + 0.5 * log(phi)
  res
}
```

Skrypt 2.1. Implementacja funkcji największej wiarygodności danej we wzorze (2.19)

```
## gradient
grad <- function(param, data) {
  alpha <- param[1]
  beta <- param[2]
  phi <- param[3]
  m <- as.matrix(data$m)
  n <- as.matrix(data$n)
  N <- as.matrix(data$N)
  mu <- N^alpha*(n/N)^beta
  x <- cbind(log(N), log(n/N))
  d_gamma <- as.numeric((m - mu) / (mu + phi)) * (phi * x)
  d_phi <- -log(mu + phi) - (m + phi)/(mu + phi) +
    log(m + phi) + (m + phi - 0.5) / (m + phi) + 1/(2*phi)
  cbind(d_gamma, d_phi)}

## hessian
hess <- function(param, data) {
  alpha <- param[1]
  beta <- param[2]
  phi <- param[3]
  m <- as.matrix(data$m)
  n <- as.matrix(data$n)
  N <- as.matrix(data$N)
  mu <- N^alpha*(n/N)^beta
  x <- cbind(log(N), log(n/N))

  d2_gamma <- t(x) \% * \% (x * as.numeric((m+phi)/(mu+phi)*phi))
  d2_phi <- sum(-(2*mu+phi-m)/(mu+phi)^2
    + (m+phi + 0.5)/(m+phi)^2 - 1/(2*phi^2))
  d2_gamma_phi <- colSums(-as.numeric((mu - m) / (mu+phi)^2 * mu) * x)
```

```

h <- as.matrix(Matrix::bdiag(d2_gamma, d2_phi)) 34
colnames(h) <- rownames(h) <- c("alpha", "beta", "phi") 35
h[3,1:2] <- h[1:2,3] <- d2_gamma_phi 36
h 37
} 38

```

Skrypt 2.2. Implementacja gradientu oraz hessiana danych wzorami (2.20)–(2.24)

```
m0 <- lm(log(m/N) ~ -1 + log(N) + log(n/N), data = dt) 1
```

Skrypt 2.3. Estymacja modelu zlinearyzowanego danego wzorem (2.30) w funkcji lm

```

## model 1
wynik <- maxLik(logLik = ll, 2
                grad = grad, 3
                hess = hess, 4
                data = dt, 5
                method = "NR", 6
                start = c(alpha = coef(m0)[1]+1, beta = coef(m0)[2], phi = 7
                        1/sigma(m0)))
summary(wynik) 8
## estymacja 9
### 80% ufnosci 10
confint(wynik, level = 0.8) 11
### 95 % ufnosci 12
confint(wynik, level = 0.95) 13
14

```

Skrypt 2.4. Estymacja parametrów metodą największej wiarygodności z pakietem maxLik oraz przedziałów ufności dla parametrów

2.4 Podsumowanie

W rozdziale drugim zdefiniowano populację trudną do zbadania i w tym kontekście przedstawiono populację cudzoziemców nielegalnie przebywającą w Polsce. Następnie przedstawiono metody estymacji wielkości populacji. Bardziej szczegółowo przedstawiono idee metody *capture-recapture*, która była podstawą metody opracowanej przez Zhang (2008). Ten model, hierarchiczny model gamma Poissona zostanie w następnym rozdziale zastosowany do oszacowania populacji cudzoziemców nielegalnie przebywających w Polsce.

Rozdział 3

Estymacja liczby cudzoziemców nielegalnie przebywających w Polsce 2018 roku

3.1 Źródła danych wykorzystane na potrzeby badania

Dane, które wykorzystano do oszacowania wielkości populacji cudzoziemców nielegalnie rezydujących w Polsce pochodzą z trzech źródeł. Dane zostały pozyskane na podstawie zapytania o informację publiczną i nie pochodzą z systemu statystyki publicznej. Wybór dat tj. stanu na oraz przedziału czasowego wynikał z udostępnionych danych. W szczególności, w pracy wykorzystano następujące rejestry oraz dane w nich zawarte:

1. Rejestr PESEL – dane o cudzoziemcach zameldowanych na pobyt czasowy lub stały terytorium Rzeczypospolitej Polskiej stan na koniec III kwartału 2018 roku (tj. 30.09.2018) – zmienna N w modelu Zhang (2008).
2. System KSIP – rejestr osób, które popełniły przestępstwo lub wykroczenie (rejestracje procesowe, notowania kryminalne, wykroczenia w ruchu drogowym). Dane dotyczą wszystkich cudzoziemców posiadających numer PESEL wpisanych do rejestru KSIP w trakcie II półrocza 2018 – zmienna n w modelu Zhang (2008)
3. Rejestr Komendy Głównej Straży Granicznej – dane o cudzoziemcach nielegalnie przebywających na terenie Polski zidentyfikowanych przez funkcjonariuszy Straży Granicznej II półrocze 2018 roku. Dane uwzględniają wyłącznie osoby, które zweryfikowano tylko raz (brak duplikatów) – zmienna m w modelu Zhang (2008).

Dane ograniczono do cudzoziemców, którzy ukończyli 18 rok życia i pochodzą z tzw. z państw trzecich (nie będących członkami Unii Europejskiej). Pod uwagę brane były agregowane dane dla kraju pochodzenia:

Afganistan, Albania, Algieria, Arabia Saudyjska, Armenia, Australia, Azerbejdżan, Bangladesz, Białoruś, Bośnia i Hercegowina, Brazylia, Chiny, Egipt, Ekwador, Etiopia, Filipiny, Ghana, Gruzja, Indie, Indonezja, Irak, Iran, Izrael, Jordania, Kamerun, Kanada, Kazachstan, Kenia, Kirgistan, Kolumbia, Kongo, Korea Północna, Libia, Malezja, Maroko, Mołdawia, Mongolia, Nepal, Nigeria, Pakistan, Peru, Republika Południowej Afryki, Rosja, Rwanda, Salwador, Serbia, Sri Lanka, Stany Zjednoczone, Syria, Tadżykistan, Tajlandia, Tunezja, Turcja, Ukraina, Uzbekistan, Wenezuela, Wietnam oraz Pozostałe.

Jako 'Pozostałe' zakwalifikowano obywateli państw, które nie wystąpiły, w którymś z rejestrów przez co nie zostały spełnione założenia: $m > 0$, $n > 0$ i $n/N < 1$ oraz osoby, które w rejestrach zostały zaklasyfikowane jako bezpaństwowcy lub miały nieustalony kraj pochodzenia.

Do oszacowania wielkości posłużono się hierarchicznym modelem Gamma-Poissona, który należy do rodziny modeli mieszanych z efektami losowymi. Dodatkowo, na potrzeby analizy wrażliwości policzono dwa modele: 1) uwzględniający Ukrainę oraz 2) nieuwzględniający Ukrainy.

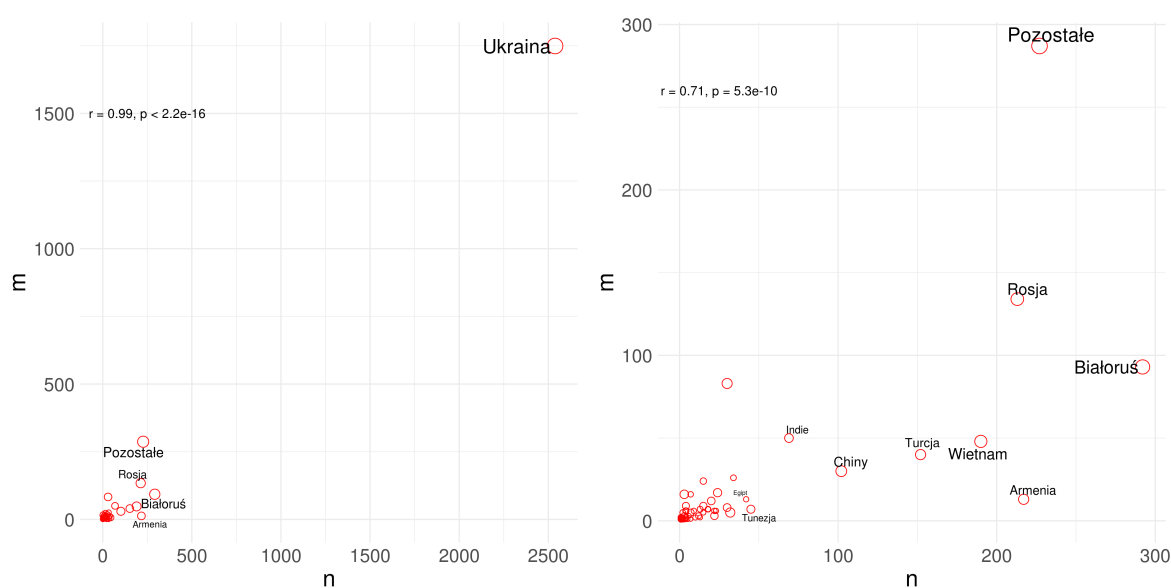
3.2 Weryfikacja założeń modelu

W celu weryfikacji założeń modelu skorzystano z postaci zlinearyzowanej danej wzorem (2.30). Stworzono wykresy rozrzutu między: m_i i n_i , $\log N_i$ i $\log(m_i/N_i)$ oraz $\log(n_i/N_i)$ i $\log(m_i/N_i)$, gdzie wielkość punktów odzwierciedla pierwiastek wielkości populacji w rejestrze PESEL (N_i) oraz uwzględniono nazwy krajów, aby zwiększyć ich interpretację. Dodatkowo, wykresy zawierają współczynniki korelacji liniowej Pearsona oraz prostą będącą oszacowaniem prostej regresji liniowej między osią X oraz Y. Wykresy przedstawiają zależności dla dwóch wersji modelu w pierwszym w zbiorze danych uwzględniono 58 państw wraz z Ukrainą, zaś w drugim nie uwzględniono Ukrainy.

Na rysunku 3.1 przedstawiona została relacja między cudzoziemcami obserwowanymi w rejestrze Policji (n_i) oraz nielegalnie przebywającymi na terenie Polski, którzy zostali zaobserwowani przez straż graniczną (m_i). Istnieje silna dodatnia korelacja ($R=0,99$ dla górnego wykresu i $R=0,71$ dla dolnego) między zmiennymi. Zgodnie z założeniami, M rośnie wraz z N a tempo

wzrostu maleje wraz ze wzrostem N .

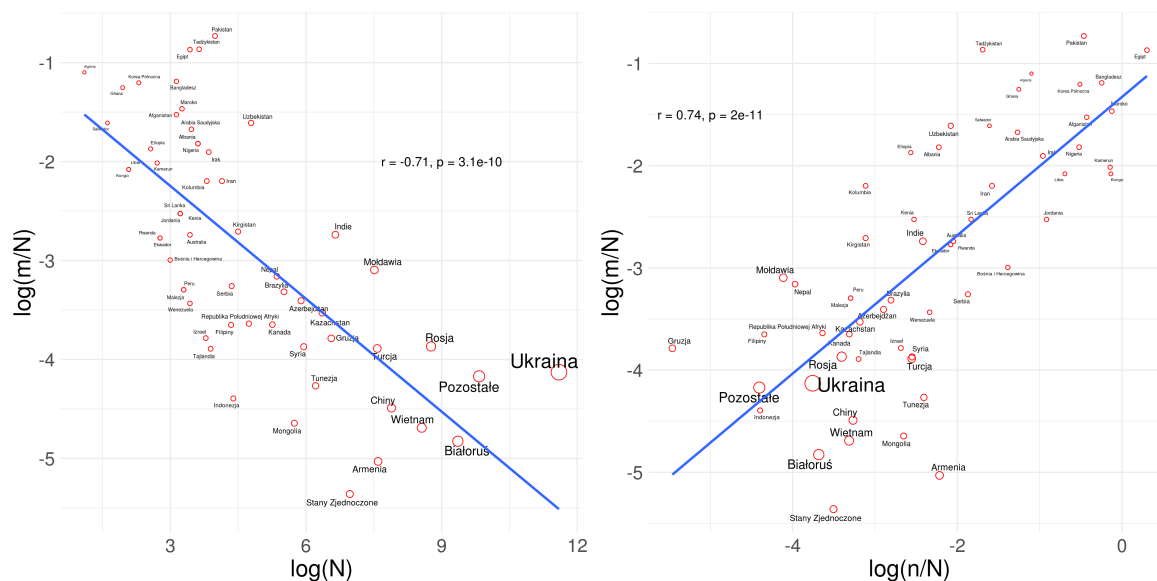
Dla postaci zlogarytmowanej modelu stworzono wykresy 3.2 i 3.3 przedstawiające zależności $\log(m_i/N_i)$ odpowiednio dla $\log(N_i)$ (lewy) i $\log(n_i/N_i)$ (prawy). W obu przypadkach zauważalna jest silna liniowa zależność oznaczona prostą na wykresach, która wynika z założenia modelu, które stanowi, że wraz ze wzrostem n/N rośnie m/M , a szybkość wzrostu maleje wraz ze wzrostem n/N . Dla $\log(m_i/N_i)$ i $\log(N_i)$ istnieje silna ujemna zależność $r=-0,71$ (wykres bez Ukrainy $r=-0,73$). Dla $\log(m_i/N_i)$ i $\log(n_i/N_i)$ współczynnik Pearsona wyniósł $r=0,74$ zarówno uwzględniając Ukrainę i nie uwzględniając.



Rysunek 3.1. Wykres zależności między danymi policji (n_i) a straży granicznej (m_i) w zakresie cudzoziemców według kraju obywatelstwa w 2 półroczu 2018 z Ukrainą (lewy) i bez Ukrainy (prawy)

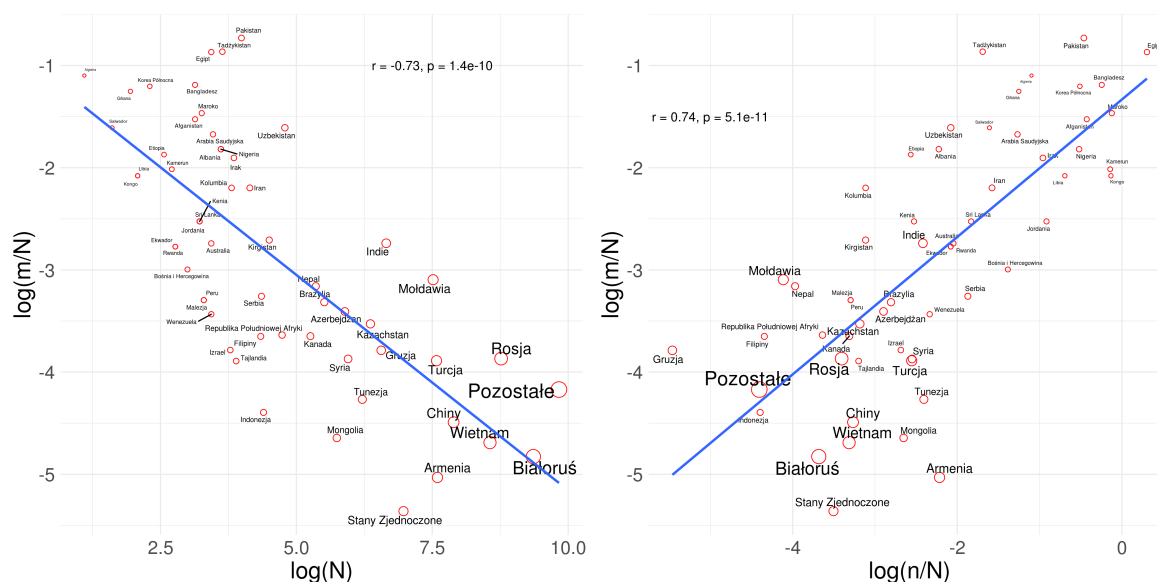
Źródło: Opracowanie własne.

Wykresy przedstawione powyżej sugerują słuszność wyboru modelu zlinearyzowanego, a tym samym mogą sugerować spełnienie założeń modelu Zhang (2008).



Rysunek 3.2. Wykresy przedstawiające zależności między logarytmem relacji nieregularnymi i regularnymi cudzoziemcami ($\log(n_i/N_i)$), a logarytmem liczby regularnych cudzoziemców ($\log(N_i)$; lewy wykres) i relacją regularnych cudzoziemców obserwowanych w danych policji i regularnymi cudzoziemcami obserwowanymi w rejestrze PESEL ($\log(n_i/N_i)$) prawy wykres)

Źródło: Opracowanie własne.



Rysunek 3.3. Wykresy przedstawiające zależności między logarytmem relacji nieregularnymi i regularnymi cudzoziemcami ($\log(n_i/N_i)$), a logarytmem liczby regularnych cudzoziemców ($\log(N_i)$; górny wykres) i relacją regularnych cudzoziemców obserwowanych w danych policji i regularnymi cudzoziemcami obserwowanymi w rejestrze PESEL ($\log(n_i/N_i)$; dolny wykres) bez uwzględnienia Ukrainy

Źródło: Opracowanie własne.

3.3 Estymacja parametrów modelu oraz jego ocena

3.3.1 Wyniki estymacji parametrów modelu zlinearyzowanego i Gamma-Poissona

Za pomocą skryptu 2.3 otrzymaliśmy następujące wartości oszacowań parametrów funkcji modelu zlinearyzowanego przedstawione w tabeli 3.1, która zawiera wyniki dla modelu z i bez Ukrainy. Wartości współczynników dla poziomu istotności 0.05 są istotnie różne od zera w obu przypadkach.

Wartość współczynnika determinacji dla modelu pierwszego R^2 wyniosła 0.9381 czyli wyjaśnia on 93,8% zmienności zmiennej objaśnianej. Sugeruje to, że model zlinearyzowany jest bardzo dobrze dopasowany do danych. Dla modelu drugiego wartość R^2 wyniosła 0.9418, oznacza to, że w tym przypadku model wykazuje lepsze dopasowanie do danych niż model z Ukrainą. W przypadku modelu pierwszego pierwiastek błędu średniokwadratowego wyniósł 0.7886 przy 56 stopniach swobody co świadczy o bardzo wysokiej jakości modelu gdyż wartość oszacowania może różnić się od wartości rzeczywistej średnio o 2,2, zaś dla modelu drugiego 0.7598. Model drugi, z którego wykluczono państwo Ukraina osiągnęło nieco lepsze wyniki oceny dopasowania modelu do danych dla modelu zlinearyzowanego.

Tabela 3.1. Podsumowanie wyników oszacowania wartości parametrów modelu zlinearyzowanego z i bez Ukrainy

	Oszacowana wartość parametru	Błąd standardowy	Wartość statystyki t	p-value
Model pierwszy uwzględniający Ukrainę				
α	-0.32557	0.05056	-6.439	2.91e-08
β	0.52878	0.09908	5.337	1.77e-06
Model drugi nieuwzględniający Ukrainy				
α	-0.36293	0.05134	-7.069	2.91e-09
β	0.47903	0.09788	4.894	9.02e-06

Źródło: Opracowanie własne.

Wartość funkcji wiarygodności dla 58 krajów wyniosła 17305.4 b biorąc pod uwagę pierwszy model. W przypadku drugiego modelu oszacowana wartość funkcji wiarygodności dla 57 państw wyniosła 4251.47. Bezpośrednio nie można porównać pod tym względem dwóch modeli z racji różnej liczby obserwacji. W szczególności, że większość cudzoziemców pochodzi z Ukrainy.

Szczegółowe dane dotyczące oszacowań wartości parametrów funkcji znajdują się w tabeli 3.2. Tabele zawierają również informacje o błędach standardowych, wartościach statystyki t

oraz wynikach testu istotności dla współczynników, z których wynika, że wartości współczynników dla poziomu istotności 0.05 są istotnie różne od zera dla obu modeli.

Tabela 3.2. Podsumowanie wyników oszacowań wartości parametrów funkcji wiarygodności dla modelu z i bez Ukrainy

	Oszacowana wartość parametru	Błąd standardowy	Wartość statystyki t	p-value
Model pierwszy uwzględniający Ukrainę				
α	0.71870	0.04338	16.569	2e-16
β	0.49642	0.08646	5.742	9.38e-09
ϕ	2.09024	0.50328	4.153	3.28e-05
Model drugi nieuwzględniający Ukrainy				
α	0.66676	0.04582	14.552	2e-16
β	0.41962	0.08732	4.805	1.54e-06
ϕ	2.29921	0.57836	3.975	7.03e-05

Źródło: Opracowanie własne.

3.3.2 Ocena modelu Gamma-Poissona

W celu znalezienia najlepiej dopasowanego modelu zweryfikowano dwie wersje modelu:

1. Dla 58 państw w tym dla państwa oznaczonego jako 'Pozostałe' i Ukrainy.
2. Dla 57 państw z wykluczeniem Ukrainy.

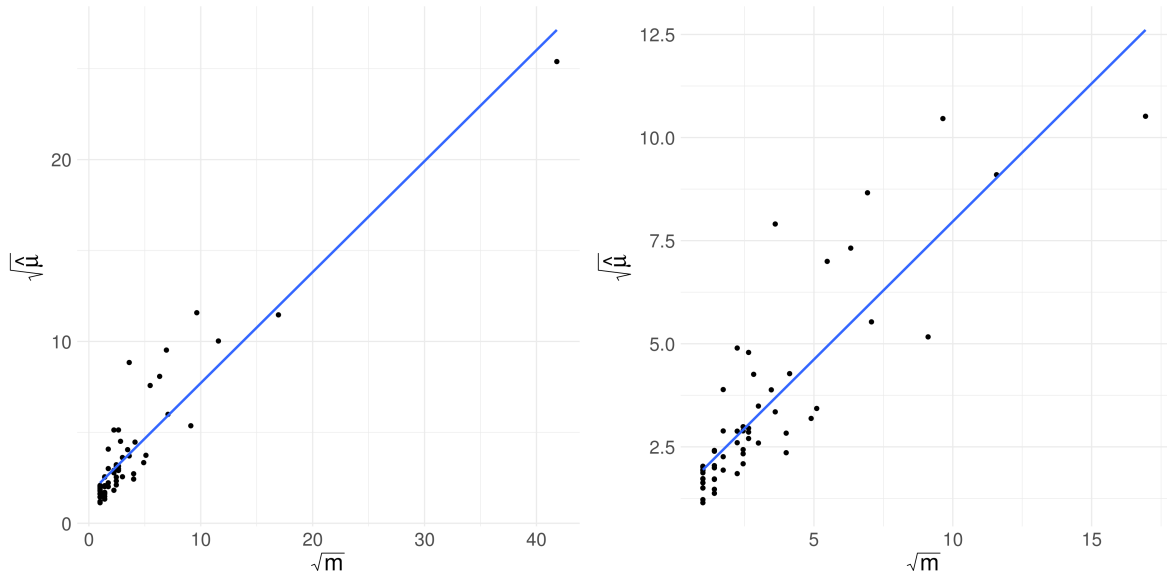
co spowodowane było znacząco odstającą obserwacją jaką była Ukraina. W celu zbadania dopasowania modelu hierarchicznego Gamma Poissona do danych wygenerowano wykresy od 3.4 do 3.5 dla modelu uwzględniającego Ukrainę.

Wykres 3.4 przedstawia wykres rozrzutu wartości pierwiastka oszacowań $\sqrt{\mu_i}$ względem $\sqrt{m_i}$. μ zostało obliczone zgodnie ze wzorem (2.28). Nic nie wskazuje na błąd w specyfikacji modelu dla oszacowania μ zarówno dla modelu z Ukrainą jak i bez. Wykres 3.6 przedstawia wartości reszt, które dla rozkładu Poissona definiuje się następująco:

$$r_i = \frac{3m_i^{\frac{2}{3}} - 3\mu_i^{\frac{2}{3}}}{2\mu_i^{\frac{1}{6}}} \quad (3.1)$$

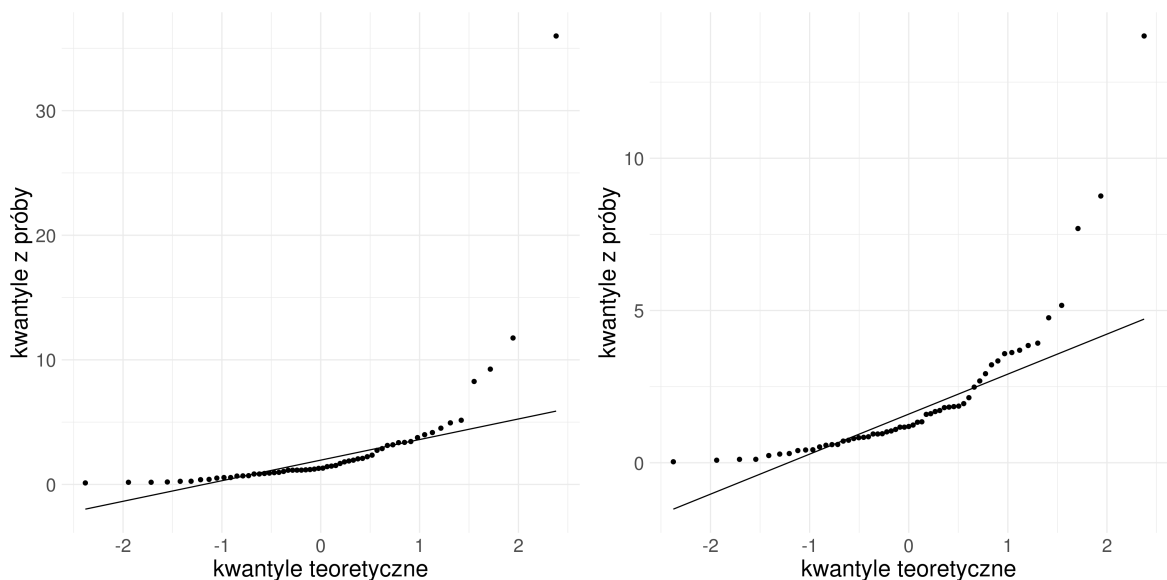
względem pierwiastka wartości oszacowań $\sqrt{\mu}$. Chodzi o przedstawienie zmiennej losowej w skali, w której przybliżenie zgodnie z rozkładem normalnym będzie najlepsze.

Wykresy kwantyl-kwantyl 3.5 pokazują, że rozkład reszt odstaje od rozkładu normalnego dla modelu nie uwzględniającego Ukrainy bardziej niż dla modelu uwzględniającego Ukrainę. W obu przypadkach występuje heteroskedastyczność reszt oznacza to, że szacunek wariancji estymatora może być obciążony.



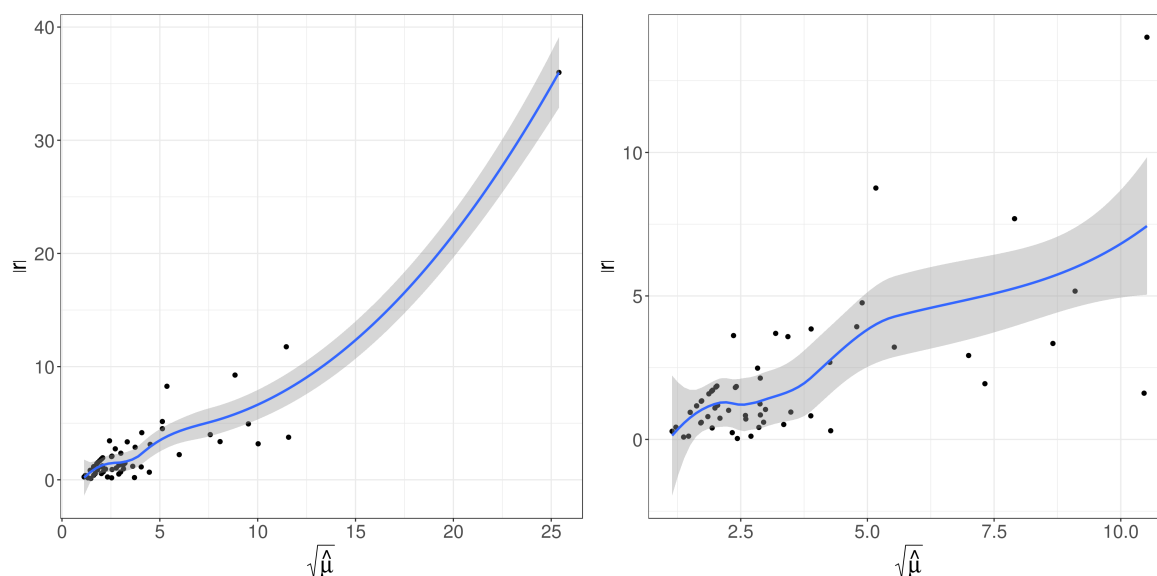
Rysunek 3.4. Wykres punktowy $\sqrt{m_i}$ względem $\sqrt{\mu_i}$ dla modelu 1 (lewy) oraz modelu 2 (prawy)

Źródło: Opracowanie własne.



Rysunek 3.5. Wykres kwantyl-kwantyl dla reszt Pearsona dla modelu 1 (lewy) oraz modelu 2 (prawy)

Źródło: Opracowanie własne.



Rysunek 3.6. Wykres dla wartości bezwzględnej reszt r względem $\sqrt{\mu_i}$ dla modelu 1 (lewy) oraz modelu 2 (prawy)

Źródło: Opracowanie własne.

Posługując się oszacowaną wartością wariancji efektu losowego można ocenić jak dobrze model objaśnia zmienną niezależną.

$$\hat{V}(u_i) = 1/\hat{\phi} \quad (3.2)$$

Wartość ta dla modelu z Ukrainą wyniosła $\hat{V}(u_i = 0.4784)$ (dla $\hat{\phi} = 2.09024$) a dla modelu bez Ukrainy $\hat{V}(u_i = 0,4349)$ (dla $\hat{\phi} = 2.29921$) oznacza to, że dla modelu uwzględniającego Ukrainę, gdzie oszacowanie ϕ jest nieco mniejsze, większa jest wariancja efektów losowych czyli mniejsza zmienność danych jest wyjaśniona przez ten model czyli model teoretycznie jest mniej skuteczny.

3.4 Szacunek wielkości populacji cudzoziemców nielegalnie przebywającej w Polsce

W tej sekcji przedstawiono wyniki oszacowań teoretycznej wielkości populacji cudzoziemców nielegalnie przebywających w Polsce na koniec 3 kw. 2019 roku. Zgodnie ze wcześniejszymi informacjami oszacowano dwa modele, które dostarczyły dwóch zestawów szacunków i dla tego rozważane są dwa zestawy szacunków ξ oznaczone odpowiednio ξ_1 i ξ_2 dla modelu 1

i modelu 2. Estymatory $\hat{\xi}$ dla tych modeli dane są wzorami

$$\hat{\xi}_1 = \sum_{i=1}^{68} N_i^{\hat{\alpha}_1}, \quad (3.3)$$

oraz

$$\hat{\xi}_2 = \sum_{i=1}^{67} N_i^{\hat{\alpha}_2} + N_{\text{Ukraina}}^{\hat{\alpha}_2}, \quad (3.4)$$

gdzie $\hat{\xi}_1$ obliczane jest na podstawie wszystkich krajów 68 krajów z wykorzystaniem $\hat{\alpha}_1$ oszacowanego na podstawie modelu 1 oraz $\hat{\xi}_2$ jest wyliczone jako suma 67 krajów oraz oddzielnie dla Ukrainy. To rozdzielenie zostało zaprezentowane tylko i wyłącznie dlatego aby podkreślić, że model 2 był szacowany na zbiorze pomniejszonym o ten kraj.

W przypadku estymacji przedziałowej zastosowano następujące wzory (niezależne od modelu)

$$\left(\hat{\xi}^{\text{Dolna}} = \sum_{i=1}^{68} N_i^{\hat{\alpha}^{\text{Dolna}}}, \hat{\xi}^{\text{Górna}} = \sum_{i=1}^{68} N_i^{\hat{\alpha}^{\text{Górna}}} \right), \quad (3.5)$$

gdzie $\hat{\xi}^{\text{Dolna}}$ i $\hat{\xi}^{\text{Górna}}$ oznaczają odpowiednio dolną i górną granicę przedziału ufności wyliczoną na podstawie $\hat{\alpha}^{\text{Dolna}}$ i $\hat{\alpha}^{\text{Górna}}$ przedziałów ufności dla parametru α .

Tabela 3.3 zawiera wyniki oszacowań punktowych oraz 80% i 95% przedziałów ufności dla oszacowań wielkości populacji cudzoziemców nielegalnie przebywających na terenie Polski w III kw. 2019 roku dla modelu 1 (z Ukrainą) oraz 2 (bez Ukrainy). Należy zauważyć, że model 2 charakteryzuje się blisko dwukrotnie mniejszym szacunkiem niż model 1. Według modelu z Ukrainą szacunek wynosi ponad 9,5 tys. cudzoziemców, podczas gdy drugi model sugeruje blisko 6 tys. Taki wynik sugerowałby, że Straż Graniczna lokalizuje blisko połowę (2 796/5 856) cudzoziemców nielegalnie przebywających w Polsce co w praktyce jest mało realne m.in. przez skończone nakłady finansowe, liczbę strażników czy nasilenie migracji. Pierwszy model jest pod tym względem bardziej realistyczny i sugeruje skuteczność na poziomie 30% co nadal wydaje się dość wysokim wynikiem.

Dodatkowo, patrząc na 95% estymację przedziałową model drugi sugeruje, że badana populacja jest mniejsza niż liczba cudzoziemców zweryfikowanych przez Straż Graniczną. W związku z tym ostatecznie zrezygnowano z opisu wyników na podstawie modelu drugiego i skupiono się na oszacowaniach na podstawie modelu pierwszego. Oznacza to, że pomimo iż Ukraina jest

Tabela 3.3. Wyniki oszacowań wielkości populacji cudzoziemców nielegalnie przebywających na terenie Polski w III kw. 2019 roku według modelu 1 i 2

Oszacowania	Model 1	Model 2
$\hat{\mu}$	9 555,63	5 856,68
80% CI $\hat{\mu}$	(5 661,68; 16 381,95)	(3 430,30; 10 196,59)
95% CI $\hat{\mu}$	(4 321,92; 21 917,25)	(2 607,02; 13 773,20)

Źródło: Opracowanie własne. Uwaga: CI oznacza *Confidence Interval* czyli przedział ufności.

krajem odstającym od reszty nie należało jej wyłączać z analizy.

Na podstawie modelu pierwszego wskazujemy, że 80% przedział ufności wynosi od 5,6 tys. do ponad 16 tys., a 95% ufności sugeruje, że wielkość populacji powinna być w przedziale od 4,4 tys. do blisko 22 tysięcy. Przedziały ufności są szerokie co wynika głównie z niewielkiej liczby obserwacji oraz wariancji w badanej zmiennej. Niemniej, porównując wyniki do populacji zarejestrowanej (197 613 cudzoziemców pełnoletnich) oznacza to, że w ostatnim przypadku udział cudzoziemców nielegalnie przebywających w Polsce jest na poziomie około 2.6% do 13%.

W 2018 zgłoszono do ubezpieczeń emerytalnego i rentowego 533 535 osób z obywatelstwem innymi niż polskie i innym niż państw UE. W tym przypadku oszacowanie stanowi 1,79% tej populacji. Suma blisko 9 tys. cudzoziemców stanowi bardzo mały odsetek całej populacji Polski. Niestety nie istnieją inne źródła ani szacunki przedstawiające ilu cudzoziemców przebywa nielegalnie w Polsce, z którymi można byłoby porównać oszacowanie uzyskane w tej pracy. Niemniej, na podstawie oceny statystycznej, jak i merytorycznej model na podstawie pełnych danych jest bliższy (nieznanej) prawdzie.

Podsumowanie

Głównym celem badawczym niniejszej pracy magisterskiej było oszacowanie liczby cudzoziemców, którzy przebywali nielegalnie na terenie Polski na koniec III kw. 2018 roku. W tym celu zastosowano nowatorski, hierarchiczny model Gamma-Poissona zaproponowany przez Zhang (2008). Głównym założeniem tego modelu jest uzależnienie wielkości populacji nielegalnie przebywającej w Polsce (oznaczonej jako M) od populacji zarejestrowanej (oznaczonej jako N). Estymowanym parametrem jest teoretyczna wielkość nieregularnej populacji cudzoziemców oznaczona przez $\xi = E(M|N) = N^\alpha$. Innymi słowy, model ten zakłada, że jest nieliniowy związek między teoretyczną wielkością nielegalnej ξ , a zarejestrowanej populacji N . Główną przesłanką za uwzględnieniem takiej zależności jest istnienie powiązań między tymi populacjami m.in. dlatego osoby, które nielegalnie przebywają na terenie danego kraju muszą mieć kontakty z legalną populacją aby mieć dostęp do miejsca pobytu czy pracy.

Aby ten model oszacować wymagane są dane pochodzące z trzech źródeł danych. Za publikacją Zhang (2008) wykorzystano następujące zmienne: 1) liczbę cudzoziemców zameldowanych na pobyt czasowy lub stały w rejestrze PESEL zgodnie ze stanem na koniec III kw. 2018r., 2) liczbę zarejestrowanych (posiadających numer PESEL) cudzoziemców, którzy popełnili przestępstwo lub wykroczenie drogowe zgodnie z danymi zawartymi w Krajowym Systemie Informacji Policji w II półroczu 2018r. oraz 3) liczbę cudzoziemców wobec, których stwierdzono nielegalny pobyt publikowaną przez Straż Graniczną w II półroczu 2018r. Dane zostały pozyskane od gestorów w ramach informacji publicznej.

W badaniu brane pod uwagę były jedynie osoby, które ukończyły 18 rok życia. Zbiór krajów ograniczono ze względu na założenia hierarchicznego modelu Gamma-Poissona do 58 z czego ostatni stanowi sztucznie stworzony kraj "Pozostałe" złożony z państw, które nie spełniły założeń modelu tzn. że nie wystąpiły w którymś z wykorzystanych rejestrów. O założeniach modelu można znaleźć więcej w rozdziale 2.

W części empirycznej postanowiono oszacować dwa modele: 1) z uwzględnieniem Ukrainy

oraz 2) bez uwzględnienia Ukrainy. Ostatecznie szacunek był przeprowadzony dla wszystkich danych, różnica polegała wyłącznie na ograniczeniu liczby obserwacji wyłącznie na potrzeby szacowania modelu. Podyktowane to było analizą wrażliwości modelu na istnienie wpływowej jaką niewątpliwie jest Ukraina. Obydwa modele, w postaci zlinearyzowanej, charakteryzowały się bardzo dobrym dopasowaniem na poziomie 95%, a oszacowane parametry charakteryzowały się niskimi błędami oszacowania.

Wyniki szacunku liczby cudzoziemców nielegalnie przebywających w Polsce w III kw. 2018 przedstawiały się następująco:

- Model 1 (model z Ukrainą): 9 555,63 osób; 95% przedział ufności (4 321,92; 21 917,25),
- Model 2 (model bez Ukrainy): 5 856,68 osób; 95% przedział ufności (2 607,02; 13 773,20).

Należy zauważyć, że model 2 charakteryzuje się blisko dwukrotnie mniejszym szacunkiem niż model 1. Według modelu z Ukrainą szacunek wynosi ponad 9,5 tys. cudzoziemców, podczas gdy drugi model sugeruje blisko 6 tys. Taki wynik sugerowałby, że Straż Graniczna lokalizuje blisko połowę (2 796/5 856) cudzoziemców nielegalnie przebywających w Polsce co w praktyce jest mało realne m.in. przez skończone nakłady finansowe, liczbę strażników czy nasilenie migracji. Pierwszy model jest pod tym względem bardziej realistyczny i sugeruje skuteczność na poziomie 30% co nadal wydaje się dość wysokim wynikiem.

Dodatkowo, patrząc na 95% estymację przedziałową model drugi sugeruje, że badana populacja jest mniejsza niż liczba cudzoziemców zweryfikowanych przez Straż Graniczną. W związku z tym ostatecznie zrezygnowano z opisu wyników na podstawie modelu drugiego i skupiono się na oszacowaniach na podstawie modelu pierwszego. Oznacza to, że pomimo iż Ukraina jest krajem odstającym od reszty nie należało jej wyłączać z analizy.

Na podstawie modelu pierwszego wskazujemy, że 80% przedział ufności wynosi od 5,6 tys. do ponad 16 tys., a 95% ufności sugeruje, że wielkość populacji powinna być w przedziale od 4,4 tys. do blisko 22 tysięcy. Przedziały ufności są szerokie co wynika głównie z niewielkiej liczby obserwacji oraz wariancji w badanej zmiennej. Niemniej, porównując wyniki do populacji zarejestrowanej (197 613 cudzoziemców pełnoletnich) oznacza to, że w ostatnim przypadku udział cudzoziemców nielegalnie przebywających w Polsce jest na poziomie około 2.6% do 13%.

Otrzymana wielkość stanowi niewielki odsetek całej populacji obywateli Rzeczypospolitej. Otrzymanej wielkości nie można skonfrontować z innymi badaniami gdyż takie nie istnieją, dlatego też trudna jest ocena tej wielkości jako dużej lub małej. Niemniej, na podstawie oceny

statystycznej, jak i merytorycznej model na podstawie pełnych danych jest bliższy (nieznanej) prawdzie.

Należy zaznaczyć, że ocena modelu wskazuje na heteroskedastyczność reszt co z jednej strony wskazuje, że błędy standardowe mogą być zaniżone, a z drugiej możliwe pominięcie dodatkowych zmiennych. Rozkład reszt wskazuje również na istnienie wartości odstających. Należy jednak zaznaczyć, że wstępna weryfikacja modelu sugeruje spełnienie założeń prezentowanego modelu tj. bardzo silna zależność między relacją cudzoziemców obserwowanych w danych policji, a zarejestrowaną populacją oraz relacją nielegalnie przebywających w Polsce, a zarejestrowaną populacją.

Kolejnym ograniczeniem modelu jest dostępność danych zarówno w kontekście ich definicji jak i jakości. Na potrzeby pracy wykorzystano dane pochodzące z rejestru PESEL pozyskane w ramach informacji publicznej co z jednej strony nie pozwala zweryfikować jaki jest faktyczny błąd pokrycia tego źródła (ilu faktycznie cudzoziemców przebywa w Polsce¹) a z drugiej jest czy jest to dobra populacja odniesienia dla populacji celu. Ewentualnie można rozważyć również dane pochodzące z Urzędu ds. Cudzoziemców w zakresie kart pobytu czy Ministerstwa Spraw Zagranicznych w zakresie wiz. Niemniej, rejestr PESEL zawiera informacje o osobach, które są aktualizowane o urodzenia i zgony podczas gdy pozostałe dotyczą dokumentów i ich ważności. Ponadto weryfikacja założeń modelu przeprowadzona w rozdziale 3 wskazała na bardzo silną zależność w przypadku wykorzystania rejestru PESEL.

Potencjalne dalsze kroki badań mogą dotyczyć uwzględnienia estymacji odpornej (tzw. estymatorów M czy MM) czy rozszerzenia modelu o dane z pozostałych okresów. Uwzględnienie efektów płci, wieku czy kraju pochodzenia może również pomóc w lepszy sposób oszacować teoretyczną wielkość badanej populacji ξ . W zależności od dostępnych danych możliwe jest również próba oszacowania populacji cudzoziemców nielegalnie przebywających na poziomie województw czy miast. Model ten również nie jest ograniczony wyłącznie do populacji nielegalnie przebywającej ale również nielegalnie pracującej.

¹Bez dostępu do danych jednostkowych oraz łączenia rejestru PESEL z innymi rejestrami nie będzie to możliwe.

Bibliografia

- Baffour, B., King, T. & Valente, P. (2013). The modern census: evolution, examples and evaluation. *International Statistical Review*.
- Bakker, B., Gerritse, S., van der Heijden, P. G., van der Laan, J., van der Vliet, R. & Cruyff, M. (2014). Estimation of non-registered usual residents in the netherlands, ultimo september 2010. W *Conference of European Statistics Stakeholders, Rome, 24-25 November 2014*.
- Beręsewicz, M. (2016). *Internet data sources for real estate market statistics*. Poznan University of Economics i Business.
- Beręsewicz, M. & Gołata, E. (2019). Przyszłość badań statystycznych wobec potrzeb społecznych i gospodarczych na przykładzie spisu ludności.
- Beręsewicz, M., Gudaszewski, G. & Szymkowiak, M. (2019). Estymacja liczby cudzoziemców w Polsce w latach 2015 i 2016 z wykorzystaniem rejestrów administracyjnych i metody capture-recapture. *Wiadomości Statystyczne*.
- Beręsewicz, M. & Szymkowiak, M. (2015). Big data w statystyce publicznej–nadzieje, osiągnięcia, wyzwania i zagrożenia. *Ekonometria*, 2, 9–22.
- Beyer, M. & Laney, D. (2012). *The Importance of 'Big Data': A Definition*.
- Biecek, P. (2013). *Analiza danych z programem R: modele liniowe z efektami stałymi, losowymi i mieszanymi*. Wydawnictwo naukowe PWN.
- Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, 783–791.
- Cieślak, M. (2005). Prognozowanie analogowe [w:] Prognozowanie gospodarcze. *Metody i zastosowania*, (2005) red. M. Cieślak, Wydawnictwo Naukowe PWN, Warszawa.
- Fawcett, T. & Provost, F. (2015). Data Science for Business. *Fawcett T., Provost F.–1 st ed–O'Reilly Media-2013*.

- Henningsen, A. & Toomet, O. (2011). maxLik: A package for maximum likelihood estimation in R. *Computational Statistics*, 26(3), 443–458. doi:10.1007/s00180-010-0217-1
- Jandl, M. (2004). The estimation of illegal migration in Europe. *Studi Emigrazione*, 141–156.
- Mastro, T. D., Kitayaporn, D., Weniger, B. G., Vanichseni, S., Laosunthorn, V., Uneklabh, T., ... Limpakarnjanarat, K. (1994). Estimating the number of HIV-infected injection drug users in Bangkok: a capture–recapture method. *American Journal of Public Health*.
- Nordbotten, S. (2010). The use of administrative data in official statistics–past, present and future: with special reference to the Nordic countries.
- Passel, J. (2007). *Unauthorized migrants in the United States: Estimates, methods. and characteristics*. Working Paper.
- Piguet, E. & Losa, S. (2002). *Travailleurs de l’ombre?: demande de main-d’oeuvre du domaine de l’asile et ampleur de l’emploi d’étrangers non déclarés en Suisse*. Seismo.
- R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.
- Szałtys, D., Adamski, M., Beręsewicz, M., Chmielewski, M., Filip, P., Gudaszewski, G., ... Wysocka, A. (2018). Cudzoziemcy na krajowym rynku pracy w ujęciu regionalnym.
- Tourangeau, R., Edwards, B. & Johnson, T. P. (2014). *Hard-to-survey populations*. Cambridge University Press.
- Urząd ds. Cudzoziemców. (2015). *Podsumowanie roku obowiązywania ustawy o cudzoziemcach*.
- Urząd ds. Cudzoziemców. (2019). Słownik.
- Ustawa z dnia 12 grudnia 2013 r. o cudzoziemcach. (2013). z późniejszymi zmianami.
- Wolter, K. M. (1986). Some coverage error models for census data. *Journal of the American Statistical Association*, 81(394), 337–346.
- Zhang, L.-C. (2008). Developing methods for determining the number of unauthorized foreigners in Norway. *Statistics Norway (SSB), Division for Statistical Methods and Standards*. www.ssb.no.(accessed July 28, 2008).

Spis tabel

1.1	Państwa, których obywatele mogą podróżować do Polski bez wiz (Do 90 dni) . .	6
1.2	Liczba cudzoziemców posiadających numer PESEL zgodnie ze stanem na koniec 2015-2018	16
1.3	Liczba cudzoziemców odprowadzających składki na ubezpieczenia społeczne w ZUS zgodnie ze stanem na koniec 2016-2017	18
1.4	Liczba cudzoziemców nielegalnie przebywających na terenie Polski zatrzymanych przez Straż Graniczną według półroczy w latach 2015-2018	18
1.5	Liczba cudzoziemców posiadających numer PESEL zatrzymanych przez Policję, podejrzanych o pełnienie przestępstwa, notowania kryminalne lub wykroczenie drogowe latach 2014-2018	19
3.1	Podsumowanie wyników oszacowania wartości parametrów modelu zlinearyzowanego z i bez Ukrainy	43
3.2	Podsumowanie wyników oszacowań wartości parametrów funkcji wiarygodności dla modelu z i bez Ukrainy	44
3.3	Wyniki oszacowań wielkości populacji cudzoziemców nielegalnie przebywających na terenie Polski w III kw. 2019 roku według modelu 1 i 2	48
A.1	Dane wykorzystane w zaproponowanym modelu	57

Spis rysunków

1.1	Klasyfikacja źródeł danych w statystyce z uwzględnieniem ‘nowych’ źródeł oraz ich gestorów	13
1.2	Liczba cudzoziemców posiadających aktualny status pobytu zarejestrowanych w rejestrze PESEL	17
2.1	Metoda capture-recapture	26
3.1	Wykres zależności między danymi policji (n_i) a straży granicznej (m_i) w zakresie cudzoziemców według kraju obywatelstwa w 2 półroczu 2018 z Ukrainą (lewy) i bez Ukrainy (prawy)	41
3.2	Wykresy przedstawiające zależności między logarytmem relacji nieregularnymi i regularnymi cudzoziemcami ($\log(n_i/N_i)$), a logarytmem liczby regularnych cudzoziemców ($\log(N_i)$; lewy wykres) i relacją regularnych cudzoziemców obserwowanych w danych policji i regularnymi cudzoziemcami obserwowanymi w rejestrze PESEL ($\log(n_i/N_i)$ prawy wykres)	42
3.3	Wykresy przedstawiające zależności między logarytmem relacji nieregularnymi i regularnymi cudzoziemcami ($\log(n_i/N_i)$), a logarytmem liczby regularnych cudzoziemców ($\log(N_i)$; górny wykres) i relacją regularnych cudzoziemców obserwowanych w danych policji i regularnymi cudzoziemcami obserwowanymi w rejestrze PESEL ($\log(n_i/N_i)$; dolny wykres) bez uwzględnienia Ukrainy	42
3.4	Wykres punktowy $\sqrt{m_i}$ względem $\sqrt{\mu_i}$ dla modelu 1 (lewy) oraz modelu 2 (prawy)	45
3.5	Wykres kwantyl-quantyl dla reszt Pearsona dla modelu 1 (lewy) oraz modelu 2 (prawy)	45
3.6	Wykres dla wartości bezwzględnej reszt r względem $\sqrt{\mu_i}$ dla modelu 1 (lewy) oraz modelu 2 (prawy)	46

Spis skryptów w języku R

2.1	Implementacja funkcji największej wiarygodności danej we wzorze (2.19)	37
2.2	Implementacja gradientu oraz hessiana danych wzorami (2.20)–(2.24)	37
2.3	Estymacja modelu zlinearyzowanego danego wzorem (2.30) w funkcji lm	38
2.4	Estymacja parametrów metodą największej wiarygodności z pakietem maxLik oraz przedziałów ufności dla parametrów	38

Dodatek A

Dodatki

A.1 Dane wykorzystane w modelu

Tabela A.1. Dane wykorzystane w zaproponowanym modelu

Rok	Numer ISO	n	m	N	Nazwa kraju
2018	004	15	5	23	Afganistan
2018	008	4	6	37	Albania
2018	012	1	1	3	Algieria
2018	031	20	12	362	Azerbejdżan
2018	036	4	2	31	Australia
2018	050	18	7	23	Bangladesz
2018	051	217	13	1989	Armenia
2018	070	5	1	20	Bośnia i Hercegowina
2018	076	15	9	248	Brazylia
2018	112	292	93	11606	Białoruś
2018	120	13	2	15	Kamerun
2018	124	7	5	192	Kanada
2018	144	4	2	25	Sri Lanka
2018	156	102	30	2675	Chiny
2018	170	2	5	45	Kolumbia
2018	178	7	1	8	Kongo
2018	218	2	1	16	Ekwador

2018	222	1	1	5	Salwador
2018	231	1	2	13	Etiopia
2018	268	3	16	706	Gruzja
2018	288	2	2	7	Ghana
2018	356	69	50	774	Indie
2018	360	1	1	81	Indonezja
2018	364	13	7	63	Iran
2018	368	18	7	47	Irak
2018	376	3	1	44	Izrael
2018	398	24	17	579	Kazachstan
2018	400	10	2	25	Jordania
2018	404	2	2	25	Kenia
2018	408	6	3	10	Korea Północna
2018	417	4	6	90	Kirgistan
2018	434	4	1	8	Libia
2018	458	1	1	27	Malezja
2018	496	22	3	312	Mongolia
2018	498	30	83	1835	Mołdawia
2018	504	23	6	26	Maroko
2018	524	4	9	212	Nepal
2018	566	22	6	37	Nigeria
2018	586	34	26	54	Pakistan
2018	604	1	1	27	Peru
2018	608	1	2	77	Filipiny
2018	643	213	134	6413	Rosja
2018	646	2	1	16	Rwanda
2018	682	9	6	32	Arabia Saudyjska
2018	688	12	3	78	Serbia
2018	704	190	48	5225	Wietnam
2018	710	3	3	114	Republika Południowej Afryki
2018	760	30	8	384	Syria

2018	762	7	16	38	Tadżykistan
2018	764	2	1	49	Tajlandia
2018	788	45	7	499	Tunezja
2018	792	152	40	1954	Turcja
2018	804	2538	1749	108621	Ukraina
2018	818	42	13	31	Egipt
2018	840	32	5	1064	Stany Zjednoczone
2018	860	15	24	120	Uzbekistan
2018	862	3	1	31	Wenezuela
2018	999	227	287	18583	Pozostałe

Źródło: Opracowanie własne.

A.2 Zestawienie definicji wykorzystanych w pracy

Populacja – Cudzoziemcy nielegalnie rezydujący na terenie Polski w wieku 18 lat lub więcej nie będący obywatelami państw członkowskich Unii Europejskiej i państw nienależących do UE, ale będących stroną układów i porozumień z UE tj. Islandii, Liechtensteinu, Norwegii, Szwajcarii, Monako, San Marino, Watykanu.

Cudzoziemiec – osoba, która nie posiada obywatelstwa polskiego („Dz.U. 2013 poz. 1650 z późn. zmianami”, 2013).

Nielegalny pobyt – Pobyt cudzoziemca na terenie Polski jest nielegalny w momencie, gdy jest niezgodny z przepisami wjazdu cudzoziemców i ich pobytu na terenie Rzeczypospolitej Polskiej zawartymi w Ustawie z dnia 12 grudnia 2013 r. o cudzoziemcach i Ustawie z dnia 14 lipca 2006 r. o wjeździe na terytorium Rzeczypospolitej Polskiej, pobycie oraz wyjeździe z tego terytorium obywateli państw członkowskich Unii Europejskiej i członków ich rodzin.

Karta pobytu - dokument, który wydawany jest cudzoziemcowi, który otrzymał pozwolenie na rezydowanie w Polsce z tytułu:

- Zezwolenia na pobyt czasowy;
- Zezwolenia na pobyt stały;
- Zezwolenia na pobyt rezydenta długoterminowego UE;

- Zgody na pobyt ze względów humanitarnych
- Ochrony międzynarodowej (nadany został w Polsce status uchodźcy lub udzielona została ochrona uzupełniająca).