

WIOLETTA GRZENDA

ANALIZA PŁODNOŚCI KOBIEC W POLSCE Z WYKORZYSTANIEM BAYESOWSKIEGO MODELU REGRESJI POISSONA¹

1. WSTĘP

Zagadnienia dotyczące płodności są obecnie bardzo szeroko analizowane ze względu na postępujący proces starzenia się ludności oraz niski wskaźnik dzietności, który uniemożliwia zastępowanie pokoleń. Dla większości krajów europejskich współczynnik dzietności² waha się od 1.2 do 1.9, dla Polski wynosi 1.3 ([2]). Utrzymanie się w długim okresie obserwowanych zmian w liczbie i strukturze ludności może istotnie wpłynąć na rozwój gospodarczy oraz stabilność istniejących rozwiązań instytucjonalnych systemu ubezpieczeń społecznych i emerytalnych. Analizom ekonomicznym skutków zachodzących procesów demograficznych, powinno towarzyszyć badanie wpływu czynników społeczno-ekonomicznych na te właśnie procesy ([1]). Jednocześnie, aby dotrzeć do istoty badanych przemian, trzeba je zgłębiać na poziomie uwarunkowań zachowań jednostkowych, które następnie przekładają się na procesy ludnościowe na poziomie makro ([43]).

Zachowania prokreacyjne mogą być warunkowane wieloma różnymi czynnikami społecznymi, kulturowymi, ekonomicznymi, demograficznymi oraz innymi, związanymi m.in. z wykorzystaniem środków kontroli płodności. Zatem w analizie mechanizmów przyczynowo-skutkowych zjawisk demograficznych konieczne jest uwzględnienie różnych aspektów życia człowieka. Celem niniejszej pracy jest zbadanie zachowań prokreacyjnych Polek poprzez identyfikację czynników je determinujących z wykorzystaniem metod bayesowskich. Analiza płodności kobiet została przeprowadzona na podstawie danych indywidualnych pochodzących z polskiego badania retrospektywnego „Przemiany rodziny i wzorce dzietności w Polsce”³.

¹ Badanie zostało zrealizowane w ramach badań własnych „Bayesowska estymacja modelu regresji Poissona” nr 03/E/0028/09, Szkoła Główna Handlowa.

² Współczynnik dzietności (TFR) określa średnią liczbę dzieci, którą urodziłaby kobieta w ciągu całego okresu rozrodczego, zakładając, że intensywność urodzeń pozostałaby niezmienna na poziomie badanego roku ([2]).

³ Badanie zrealizowane przez Instytut Statystyki i Demografii SGH oraz Główny Urząd Statystyczny w 1991 roku w ramach europejskiej rundy badań „Family and Fertility Survey” koordynowanej przez Population Activity Unit Commission for Europe, UN Geneva.

Badanie determinant płodności wymaga nie tylko odpowiednich danych, ale również metod analizy. Opisowa charakterystyka determinant płodności oraz trendów jej zmian dla Polski w latach 1950-2005 ze szczególnym uwzględnieniem okresu transformacji ustrojowej zawarta jest m.in. w [2] i [44]. Inne podejście polega na użyciu odpowiedniego modelu statystycznego, który umożliwi badanie wpływu różnych czynników na dzietność. W badaniach procesu tworzenia się rodziny wykorzystywane są również metody i modele analizy historii zdarzeń (np. [20], [21]), modele te na ogólnie wymagają znajomości rozkładu zależnej zmiennej losowej⁴. Ponadto w literaturze można znaleźć wiele opracowań dotyczących modelowania decyzji prokreacyjnych z wykorzystaniem odpowiednich modeli regresji ([1], [6], [68])⁵.

Pierwsze próby modelowania płodności przy użyciu modelu regresji można znaleźć w pracy I. Adelman ([1]). Celem tej publikacji było zbadanie wpływu determinant społeczno-ekonomicznych na płodność i umieralność. W badaniu współczynnika urodzeń żywych, dla wybranych grup wiekowych kobiet uwzględniono następujące czynniki społeczno-ekonomiczne: wielkość dochodu narodowego na mieszkańca, wskaźnik procentowy zatrudnienia poza rolnictwem, poziom wykształcenia matki oraz gęstość zaludnienia.

Modelowanie decyzji prokreacyjnych gospodarstw domowych z wykorzystaniem uogólnionego modelu regresji Poissona w oparciu o klasyczną metodę największej wiarygodności przedstawiono natomiast w pracy [68]. W. Wang i F. Famoye pokazali, kiedy wybrany model jest lepszy w porównaniu ze zwykłym modelem regresji Poissona i modelem ujemnym dwumianowym, jednak ostatecznie otrzymano, że wartości wyznaczonych parametrów są podobne w obu modelach Poissona. Wykorzystany model dał możliwość weryfikacji elementów ekonomicznej teorii płodności Beckera-Lewisa poprzez uwzględnienie w badaniu wybranych zmiennych społeczno-ekonomicznych.

W pracy Ch. Hilgemana i C. T. Buttsa [35] przeprowadzono analizę związków pomiędzy płodnością kobiet, a ich udziałem w rynku pracy przy użyciu bayesowskiego modelu hierarchicznego z wykorzystaniem niejednorodnego procesu Poissona z funkcją kawałkami stałą, w estymacji wykorzystano metody Monte Carlo oparte na łańcuchach Markowa. Badaniu poddano 20 krajów bez uwzględnienia Polski. Zmienną zależną była całkowita liczba dzieci kiedykolwiek urodzonych (total number of children ever born) do momentu przeprowadzenia badania. Zastosowany model umożliwił badanie wybranych czynników indywidualnych oraz socjalno-ekonomicznych dla poszczególnych państw. W przeciwieństwie do innych badań ([9], [61]), których wyniki sugerują, że kraje o wyższej dzietności charakteryzowała również wyższa aktywność zawodowa kobiet, w pracy [35] otrzymano, że wzrostowi aktywności zawodowej kobiet w okresie lat 80-tych towarzyszył spadek płodności.

⁴ Do modeli analizy historii zdarzeń zaliczamy również modele parametryczne, wówczas znana jest analityczna postać gęstości rozkładu prawdopodobieństwa.

⁵ W pracach tych nie wykorzystywano podejścia bayesowskiego; w niniejszej pracy po raz pierwszy wykorzystano bayesowski model regresji w analizie płodności na danych polskich.

W niniejszej pracy podjęto próbę zbadania wpływu wybranych czynników na ogólną liczbę dzieci posiadanych przez kobiety, które okres prokreacyjny miały już za sobą. Uwzględniając charakter zmiennej objaśnianej można przyjąć, że odpowiednim narzędziem analizy jest model regresji Poissona. Model ten umożliwia określenie kierunku i skali wpływu wybranych czynników na liczbę dzieci posiadanych przez kobiety.

Dotychczas w powszechnie prowadzonych badaniach dotyczących modelowania płodności kobiet w Polsce wykorzystywano tzw. podejście klasyczne, które opiera się wyłącznie na informacji zawartej w próbie losowej. Zaproponowane w niniejszej pracy podejście bayesowskie umożliwia włączenia do wnioskowania statystycznego wiedzy a priori, pochodzącej z wyników innych dotychczasowych badań. W pracy oszacowano modele, w których wykorzystano nieinformacyjne i informacyjne rozkłady a priori. Wybierając rozkłady nieinformacyjne uzyskano wyniki analogiczne do tych, które można by otrzymać wykorzystując podejście klasyczne, ale należy podkreślić, że w podejściu bayesowskim nieznanne parametry modelu są zmiennymi losowymi, zatem interpretacja otrzymanych wyników w tych dwóch podejściach jest różna. Inne zalety stosowania metod bayesowskich przedstawione są m.in. w publikacjach [8], [26] oraz [19]⁶.

2. ZAKRES BADAŃ

Istnieje wiele teorii płodności, w kontekście których można rozważać obserwowane w Polsce zachowania prokreacyjne przed okresem transformacji. Właściwą podstawę dla tych rozważań mogą stanowić ekonomiczne teorie płodności oraz koncepcja drugiego przejścia demograficznego.

Podstawowe założenia ekonomicznej teorii płodności zostały sformułowane przez H. Leibensteina ([46]). Oparł on swoją teorię na dwóch podstawowych aksjomatach. Pierwszy z nich to przekonanie, że racjonalny wybór jest czynnikiem regulującym wszelkie sfery ludzkiej aktywności, zatem rodzice decydując o liczbie dzieci zachowują się w sposób racjonalny. W myśl drugiego aksjomatu dziecko traktowane jest jako dobro dostarczające rodzicom trzech form użyteczności: konsumpcyjnej, produkcyjnej i zabezpieczającej.

W ekonomicznej teorii płodności wyróżnia się dwa przeciwstawne sobie nurty: „model szkoły chicagowskiej” oraz „model szkoły pensylwańskiej”. Głównym przedstawicielem pierwszego nurtu jest G. S. Becker, który wyrażał przekonanie o możliwości analizy decyzji prokreacyjnych w ramach ekonomicznej teorii zachowań konsumencjnych ([4]). Uważał on, że dzieci stanowią trwałe dobro konsumpcyjne oraz produkcyjne i traktował je na równi z innymi dobrami konsumpcyjnymi takimi jak domy i samochody. Jego przekonania podzielali min. R. J. Willis, M. R. Rosenzweig, P. Schulz oraz

⁶ Publikacja ta zawiera porównanie wersji klasycznej i bayesowskiej modelu regresji Poissona, otrzymano m.in., że w przypadku testowania hipotez o istotności parametrów modelu, rezultaty uzyskane tymi dwoma metodami są równoważne, natomiast w przypadku estymacji parametrów lepsze wyniki daje podejście bayesowskie.

Y. Ben-Porath. Modyfikacją koncepcji Beckera wraz z odwołaniem się do stwierdzeń Leibensteina była, określana mianem modelu „szkoły pensylwańskiej”, teoria płodności R. A. Easterlina ([18]). Proponował on model zachowań prokreacyjnych wykraczający poza ramy teorii mikroekonomicznej i uwzględniający również zmienne socjoekonomiczne, a także proces zmiany historycznej. Uważał, że podejmowane decyzje prokreacyjne wynikają z kalkulacji korzyści i kosztów związanych z posiadaniem dzieci, a nie z wysokości dochodu. Szersze rozważania na temat teorii ekonomicznych można znaleźć w opracowaniu [63].

Koncepcja drugiego przejścia demograficznego dotyczy przemian demograficznych zachodzących od lat 60 w wysoko rozwiniętych krajach Europy Zachodniej, a od początku lat 90 w krajach Europy środkowo-Wschodniej. Przemiany strukturalne związane z uprzemysłowieniem, urbanizacją, ogólną poprawą poziomu życia, wzrostem poziomu wykształcenia i większą rolą kobiet w życiu ekonomicznym i społecznym przyczyniły się do nasilania dążeń samorealizacyjnych jednostki i odrzucania wartości tradycyjnych na rzecz liberalnych. W odniesieniu do płodności obserwowane wówczas procesy przejawiały się głównie spadkiem dzietności poniżej poziomu prostej zastępowalności pokoleń, opóźnianiem decyzji o urodzeniu pierwszego dziecka, wzrostem liczby urodzeń pozamałżeńskich oraz spadkiem skłonności do zawierania małżeństw. Za twórców tej teorii uznaje się przede wszystkim D. Van de Kaa ([66], [67]) i R. Lesthaeghe'a ([47], [48], [49]). Analiza zmian demograficznych w Polsce w kontekście drugiego przejścia demograficznego jest tematem wielu opracowań ([23], [44], [45]) i dotyczy głównie okresu transformacji systemowej – wtedy to przedstawione procesy uległy znacznej intensyfikacji.

W nawiązaniu do powyższych teorii i innych opracowań przedstawimy teraz potencjalne zmienne objaśniające i związane z nimi problemy badawcze.

Większość opracowań prezentuje różne mierniki rodności, płodności i ich trendy ogólnie dla Polski oraz z podziałem na miasto i wieś ([2], [36], [44]). Uwzględniając w badaniu miejsce zamieszkania, należy jednak pamiętać, że lata 1950-1990 pozostawały pod silnym wpływem migracji wewnętrznych, zwłaszcza przemieszczeń ze wsi do miast. Duże migracje między obszarami wiejskimi a miejskimi charakteryzowały zwłaszcza pierwszą dekadę podanego okresu. Do roku 1980 obserwowano stale rosnące saldo dla miast i dopiero kryzys lat osiemdziesiątych doprowadził do istotnego zmniejszenia mobilności przestrzennej ludności ([43]). Migranci przynosili do miast model rodziny wielodzietnej, co miało wpływ na stosunkowo wysoki poziom dzietności. Jednak w całym okresie obserwowano wyższą płodność na wsi, niż w miastach, co można m.in. tłumaczyć na gruncie ekonomicznej teorii płodności, według Beckera ([4]) dzieci niekiedy mogą także dostarczać dochodu pieniężnego i wtedy stanowią dobro produkcyjne. Na wsi dzieci pracowały, zatem przyczyniały się do wzrostu dochodu rodziny, w mieście posiadanie dzieci okazało się zbyt kosztowne ([45]). Biorąc pod uwagę powyższe rozważania zbadamy, w jakim stopniu miejsce zamieszkania wpływało na liczbę posiadanych dzieci w badanym okresie.

Zgodnie z teorią inwestowania w kapitał ludzki Beckera ([5]) inwestowanie w wykształcenie daje możliwość uzyskania w przyszłości większych dochodów, ale jednocześnie powoduje spadek płodności. Posiadanie większej wiedzy daje kobietom inne możliwości samorealizacji głównie w postaci pracy zawodowej ([2]). Wpływ ten może mieć charakter bezpośredni poprzez m.in. poszerzenie wiedzy na temat środków kontroli urodzeń. Natomiast pośredni wpływ wykształcenia może być związany z podwyższaniem wieku zawierania małżeństw oraz ograniczeniem umieralności noworodków. W badaniach często wykształcenie traktowane jest jako determinanta aktywności zawodowej i dochodu, jednak warto zaznaczyć, iż takie znaczenie wykształcenia w gospodarkach centralnie planowanych było znacznie mniejsze ze względu na system regulacji wynagrodzeń. Według Leibensteina ([46]) wzrost dochodu na osobę zmniejsza potrzebę korzystania z dzieci jako źródła dochodów, ponadto rozwój systemów opieki społecznej zmniejsza znaczenie dziecka w tej roli. Można zatem przypuszczać, że kobiety bardziej wykształcone będą posiadały mniej dzieci.

W okresie powojennym przyspieszone procesy industrializacji spowodowały duże zapotrzebowanie na siłę roboczą, ponadto ze względu na popularyzację emancypacji zawodowej kobiet, były one zatrudniane na dużą skalę ([20]). Chęć poprawy jakości życia rodziny również skłaniała kobiety do podejmowania dodatkowych prac. Takie zachowania kobiet mogły z jednej strony negatywnie wpływać na płodność (np. [35]), z drugiej strony wzrost dochodu powinien prowadzić do wzrostu zarówno liczby, jak i jakości potomstwa (Becker [4]). W krajach zachodnich w badanym okresie kierunek korelacji pomiędzy dzietnością a aktywnością zawodową był różny. Polskę w okresie socjalizmu charakteryzował relatywnie wysoki poziom dzietności i aktywności zawodowej. Na poziom urodzeń wpływa również skuteczna polityka ludnościowa. Jednym z jej elementów jest prawo pracy i zakres uprawnień przysługujący pracującym kobietom w okresie ciąży, porodu i wychowania małego dziecka. Dopiero w 1974 r. skodyfikowano prawo i uchwalono kodeks pracy, który zapewniał kobietom m.in. trwałość stosunku pracy w okresie ciąży i płatne urlopy macierzyńskie, natomiast od 1981 r. kobiety mogą korzystać z 3 letniego urlopu wychowawczego oraz z zasiłku wychowawczego ([25]), zachowując przy tym ciągłość pracy. Mimo tego wkrótce rozpoczął się proces spadku liczby urodzeń, zatem należy zbadać, czy sytuacja na rynku pracy faktycznie determinowała decyzje prokreacyjne kobiet.

Pomimo zaobserwowanego malejącego współczynnika zawierania związków małżeńskich i rosnącego współczynnika rozwodów w latach 1960-1990, małżeństwo było i jest ważną instytucją społeczną w Polsce ([43]). Różnice w płodności kobiet zamężnych i niezamężnych zależą w dużym stopniu od charakterystyk społeczno-gospodarczych. Warto również podkreślić, że dla niektórych kobiet małżeństwo uprawnia je do posiadania dzieci, dla innych małżeństwo jest zawierane po urodzeniu dziecka w celu zalegalizowania tego urodzenia ([2]). Niewątpliwie istnieje związek pomiędzy liczbą posiadanych dzieci, a faktem pozostawania w związku małżeńskim, zatem warto zbadać jego skalę.

Polska jest krajem z mocno osadzonymi wartościami religijnymi i wpływem Kościoła. Można przypuszczać, że kobiety bardziej religijne przywiązane były do tradycyjnego modelu rodziny, zatem mniej chętnie uczestniczyły w pracy zawodowej, co mogło sprzyjać większej liczbie posiadanych przez nie dzieci. Uzasadnia to uwzględnienie w badaniu wyznania jako czynnika potencjalnie wpływającego na płodność.

Podsumowując, w badaniu płodności wskazane jest uwzględnienie wielu różnych zmiennych. Szczegółową charakterystykę poszczególnych zmiennych objaśniających zawiera tabela 1.

Tabela 1.

Zmienna	Opis zmiennej	Poziom	Struktura
m_zamieszkania	miejsce zamieszkania w momencie badania	0 = miasto 1 = wieś	395 192
wykształcenie	poziom wykształcenia	1 = podstawowe 2 = zasadnicze 3 = średnie 4 = wyższe	224 121 197 45
zatrudnienie	sytuacja na rynku pracy w momencie badania	1 = pracująca zawodowo 2 = niepracująca zawodowo 3 = nigdy niepracująca zawodowo	445 119 23
związek	fakt pozostawania w związku małżeńskim	0 = nie 1 = tak	12 575
wyznanie	wyznanie	0 = inne niż rzymskokatolickie 1 = rzymskokatolickie	15 572

Źródło: obliczenia własne

W demografii za wiek rozrodczy kobiet przyjmuje się wiek od 15 do 49 roku życia, co oczywiście wynika z uwarunkowań biologicznych. W pracy [60], poświęconej metaanalizie trendów zmian płodności, autor rozważał liczbę dzieci kiedykolwiek urodzonych (CEB – Children Ever Born). Miernik ten może być używany w przypadku dowolnych grup wiekowych, jednak ze względu na cel badania autorzy uwzględnili tylko kobiety powyżej 40 roku życia. W naszym badaniu również uwzględniamy tylko kobiety, które okres prokreacyjny mają już za sobą. Na podstawie wstępnych obliczeń dla całego zbioru danych otrzymano, że wiek kobiety w chwili urodzenia ostatniego dziecka wynosi maksymalnie 44 lata. Uwzględniając to, postanowiono badaniu poddać 587 kobiet w wieku powyżej 44 lat. W badaniu zmienna objaśniana wyrażająca ogólną liczbę posiadanych dzieci przez kobiety przyjmuje wartości od 0 do 8.

3. METODA BADANIA

3.1. PODEJŚCIE BAYESOWSKIE

Podjęcie bayesowskie stanowi metodę analiz statystycznych alternatywną w stosunku do analiz częstościowych i jest obecnie stosowane na równi z podejściem kla-

sycznym. Łączenie subiektywnej wiedzy a priori z informacją pochodzącą z próby stanowi podstawę metod bayesowskich. Twierdzenie Bayesa przekształca posiadaną wiedzę przed obserwacją oraz informację zawartą w empirycznych danych w prawdopodobieństwo a posteriori. Wnioskowanie w statystyce bayesowskiej bazuje na tak otrzymanych rozkładach a posteriori.

Niech \mathbf{Y} oznacza przestrzeń prób, Θ przestrzeń parametrów, a $\{P_\theta : \theta \in \Theta\}$ rodzinę możliwych rozkładów prawdopodobieństwa na przestrzeni \mathbf{Y} . W estymacji bayesowskiej dodatkowo zakłada się istnienie pewnego rozkładu prawdopodobieństwa na pewnej klasie podzbiorów mierzalnych zbioru Θ . Rozkład ten nazywany jest rozkładem a priori parametru⁷ θ , opisuje on stopień naszej wiary w różne wartości parametru zanim zostaną wykonane odpowiednie obserwacje ([59]). Główną wadą analiz bayesowskich jest brak informacji dotyczących wyboru rozkładu a priori, o wyborze odpowiedniego rozkładu mogą decydować zdobyte doświadczenia oparte na wcześniejszych badaniach lub pewne logiczne przesłanki⁸. Warto podkreślić, że jeśli w wyniku doświadczenia otrzymuje się dostatecznie dużo informacji, to nawet dość znaczne zmiany rozkładu a priori nie pociągają za sobą większych zmian w rozkładach a posteriori ([59]).

Rozważmy estymację parametru θ na podstawie danych $\mathbf{y} = \{y_1, \dots, y_n\}$. Jeśli przyjmiemy, że $p(\theta)$ oznacza rozkład a priori parametru θ , $p(\mathbf{y}|\theta)$ łączną funkcję prawdopodobieństwa lub gęstości zależną od parametru θ , to ze wzoru Bayesa warunkowy rozkład $p(\theta|\mathbf{y})$ parametru θ (rozkład a posteriori) wyznaczany jest w następujący sposób

$$p(\theta|\mathbf{y}) = \frac{p(\mathbf{y}|\theta) p(\theta)}{p(\mathbf{y})},$$

gdzie $p(\mathbf{y})$ oznacza rozkład brzegowy. Powyższy wzór z dokładnością do normalizującej stałej można zapisać w następującej postaci $p(\theta|\mathbf{y}) \propto L(\theta; \mathbf{y}) p(\theta)$, gdzie $L(\theta; \mathbf{y})$ oznacza funkcję wiarygodności parametru θ .

3.2. UOGÓLNIONY MODEL LINIOWY

Wykorzystany w pracy model regresji Poissona należy do klasy uogólnionych modeli liniowych. Uogólnione modele liniowe umożliwiają testowanie wpływu predyktorów jakościowych oraz ciągłych na zmienną zależną, której rozkład należy do rodziny rozkładów wykładniczych, ponadto zmienne niezależne mogą wpływać na zmienną zależną w sposób nieliniowy. Omawiane modele zostały po raz pierwszy przedstawione w pracach [51] i [53], dodatkowe informacje oraz przykłady zastosowań można również znaleźć m.in. w [15], [32] oraz [62].

⁷ Parametr θ traktowany jest jako zmienna losowa lub wektor losowy o określonym rozkładzie.

⁸ Przykłady zastosowań subiektywnych rozkładów a priori można znaleźć m.in. w pracach [8], [64] oraz [65].

Uogólniony model liniowy można zapisać w postaci macierzowej następującym wzorem:

$$g(\mathbf{E}(\mathbf{Y})) = \mathbf{X}\beta,$$

gdzie \mathbf{Y} oznacza wektor wartości zmiennej zależnej, $\mathbf{E}(\mathbf{Y})$ wektor wartości oczekiwanych zmiennej zależnej, \mathbf{X} macierz wartości zmiennych niezależnych, β wektor współczynników, a g funkcję wiążącą (np. [62]).

Podstawą definicji uogólnionych modeli liniowych jest założenie, że zmienna zależna ma rozkład należący do rodziny rozkładów wykładniczych. Jeśli rozkład prawdopodobieństwa zmiennej zależnej y ma rozkład prawdopodobieństwa należący do rodziny rozkładów wykładniczych, to można go przedstawić w następujący sposób:

$$f(y; \theta, \varphi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\varphi)} + c(y, \varphi)\right\},$$

gdzie a , b i c są z góry ustalonymi funkcjami, θ jest kanonicznym parametrem, φ jest zaś parametrem skali, jeśli jest on nieznan, to traktowany jest jako parametr rozproszenia.

Istnieje wiele rozkładów prawdopodobieństwa, które można przedstawić w powyższej postaci. Przyjmując $\theta = \ln \lambda$, $\varphi = 1$, $a(\varphi) = \varphi$, $b(\theta) = \lambda$ oraz $c(y, \varphi) = -\ln(k!)$ otrzymujemy rozkład Poissona:

$$f(k; \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, 3, \dots, \quad \lambda > 0.$$

W zależności od rozkładu zmiennej zależnej można użyć różnych funkcji wiążących, dla rozkładu Poissona standardową funkcją wiążącą jest funkcja logarytmiczna.

3.3. MODEL REGRESJI POISSONA – PODEJŚCIE BAYESOWSKIE

Przedstawimy teraz model regresji Poissona w ujęciu bayesowskim. Rozważania dotyczące bayesowskich modeli regresji można znaleźć m.in. w [8], [14], [19] oraz [26].

Założmy, że zmienna losowa y_i , $i = 1, 2, \dots, n$ ma rozkład Poissona z parametrem⁹ λ_i . Wybierając za funkcję wiążącą funkcję logarymiczną otrzymujemy $\ln(\lambda_i) = \sum_{j=1}^k x_{ij}\beta_j$, gdzie x_{ij} są zmiennymi objaśniającymi, $\beta = (\beta_1, \dots, \beta_k)$ wektorem nieznanymi parametrów. Stąd $\lambda_i = \exp\left(\sum_{j=1}^k x_{ij}\beta_j\right)$. Przy przyjętych oznaczeniach funkcja wiarygodności dla wektora parametrów β ma następującą postać:

$$L(\beta; \mathbf{y}) = \left(\prod_{i=1}^n y_i!\right)^{-1} \exp\left(-\sum_{i=1}^n \exp\left(\sum_{j=1}^k x_{ij}\beta_j\right) + \sum_{j=1}^k \beta_j \sum_{i=1}^n x_{ij}y_i\right).$$

⁹ λ_i jest średnią i wariancją rozkładu Poissona.

Specyfikację rozkładów a priori dla parametrów regresji uogólnionych modeli liniowych można znaleźć w [14], zazwyczaj wybieramy rozkłady normalne ze średnią zero i pewną wariancją. We wnioskowaniu bayesowskim szczególną rolę odgrywają nieinformacyjne rozkłady a priori, niosące ze sobą minimalną wiedzę statystyczną, ale często zamiast nich stosujemy rozkłady wystarczająco mało informacyjne¹⁰. W pracy [38], autorzy pokazują, że użycie nieinformacyjnego jednostajnego rozkładu a priori w estymacji uogólnionych modeli liniowych może prowadzić do niewłaściwych rozkładów a posteriori, w tym przypadku alternatywnym wyjściem może być wykorzystanie reguły Jeffreysa ([39], [40]).

Przyjmując dla niezależnych parametrów β_1, \dots, β_k rozkłady normalne a priori ze średnią m_j i wariancją σ_j^2 , otrzymujemy następujący wzór na łączną gęstość a priori

$$p(\beta_1, \dots, \beta_k) = \prod_{j=1}^k \frac{1}{\sigma_j \sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_j^2} (\beta_j - m_j)^2\right).$$

Wówczas rozkład a posteriori można zapisać, z dokładnością do normalizującej stałej, w następującej postaci

$$p(\beta_1, \dots, \beta_k | \mathbf{y}) \propto \exp\left(-\sum_{j=1}^k \frac{1}{2\sigma_j^2} \beta_j^2 + \sum_{j=1}^k \left(\frac{m_j}{\sigma_j^2} + \sum_{i=1}^n x_{ij} y_i\right) \beta_j - \sum_{i=1}^n \exp\left(\sum_{j=1}^k x_{ij} \beta_j\right)\right).$$

Wnioskowanie o dowolnej współrzędnej wektora parametrów $\beta = (\beta_1, \dots, \beta_k)$ odbywa się z brzegowego rozkładu a posteriori, otrzymanego poprzez całkowanie po pozostałych współrzędnych łącznego rozkładu a posteriori i tylko w podręcznikowych przykładach może opierać się jedynie na metodach analitycznych ([55]). W praktyce często wykorzystujemy metody symulacyjne, które umożliwiają generowanie prób losowych z dowolnej gęstości a posteriori. Obecnie dużą popularnością cieszą się metody Monte Carlo oparte na łańcuchach Markowa (Markov Chain Monte Carlo Method – MCMC). Podstawą metod MCMC jest generowanie ergodycznego łańcucha Markowa, który po upływie odpowiednio długiego czasu osiąga rozkład stacjonarny zwany w podejściu bayesowskim rozkładem a posteriori. Najbardziej znanym algorytmem tych metod obliczeniowych jest algorytm Metropolis oraz jego uogólnienie – algorytm Metropolis-Hastingsa. Natomiast szczególnym przypadkiem algorytmu Metropolis-Hastingsa jest próbnik Gibbsa.

W niniejszej pracy przy estymacji modeli skorzystano z losowania Gibbsa¹¹, którego schemat przedstawiamy poniżej. Rozważmy łączny rozkład a posteriori o gęstości $p(\beta | \mathbf{y})$, gdzie $\beta = (\beta_1, \dots, \beta_k)$, $k > 1$, jest wektorem parametrów.

¹⁰ W przypadku rozkładów normalnych dla średniej przyjmujemy wartość zero, a dla wariancji pewną dużą liczbę.

¹¹ Przykłady wnioskowania bayesowskiego dla uogólnionych modeli liniowych z wykorzystaniem próbnika Gibbsa można również znaleźć w pracy [11].

Kroki algorytmu Gibbsa (np. [11], [24], [58]):

1. Przyjmujemy $t = 0$. Wybieramy dowolne wartości początkowe

$$\beta^{(0)} = (\beta_1^{(0)}, \beta_2^{(0)}, \dots, \beta_k^{(0)}).$$

2. Wygenerowanie jednej realizacji $\beta^{(t+1)} = (\beta_1^{(t+1)}, \beta_2^{(t+1)}, \dots, \beta_k^{(t+1)})$ wymaga następujących losowań z pełnych warunkowych rozkładów a posteriori:

$$\beta_1^{(t+1)} \sim p(\beta_1 | \beta_2^{(t)}, \dots, \beta_k^{(t)}, \mathbf{y})$$

$$\beta_2^{(t+1)} \sim p(\beta_2 | \beta_1^{(t+1)}, \beta_3^{(t)}, \dots, \beta_k^{(t)}, \mathbf{y})$$

...

$$\beta_k^{(t+1)} \sim p(\beta_k | \beta_1^{(t+1)}, \dots, \beta_{k-1}^{(t+1)}, \mathbf{y}).$$

3. Przyjmujemy $t = t + 1$ i wracamy do kroku drugiego, aż do uzyskania zbieżności algorytmu. Następnie krok drugi powtarzamy, aż do uzyskania pożądanej liczby realizacji.

Przedstawiony algorytm przy pewnych ogólnych warunkach jest zbieżny, zatem rozkład losowanego wektora $\beta^{(t)}$ dąży do rozkładu $p(\beta | \mathbf{y})$, gdy $t \rightarrow \infty$.

Istotną część analiz bayesowskich stanowi ocena zbieżności wygenerowanych łańcuchów. W teorii próbnik może działać w nieskończoność, wówczas zbieżność jest osiągnięta, w praktyce konieczne jest ustalenie pewnej liczby cykli spalonych (wstępnych), a do wnioskowania a posteriori wykorzystujemy tylko te właściwe realizacje. W ocenie zbieżności łańcucha Markowa istnieją pewne trudności, ponieważ brakuje decydujących testów, które rozstrzygają, czy łańcuch dąży do rozkładu stacjonarnego, dostępne testy sprawdzają na ogół tylko warunek konieczny zbieżności. Do najbardziej znanych testów, które umożliwiają ocenę zbieżności łańcucha Markowa w różnych aspektach należą: test Gelmana i Rubina ([10], [27]), test Geweke'a ([29]), test Heidelbergera i Welcha ([33], [34]) oraz test Raftery'a i Lewisa ([56], [57]).

4. BUDOWA, ESTYMACJA I OCENA MODELU

Estymacja i weryfikacja wszystkich modeli została przeprowadzona przy użyciu systemu SAS, w estymacji bayesowskiej wykorzystano metody Monte Carlo oparte na łańcuchach Markowa. Metody te są obecnie coraz częściej wykorzystywane w statystyce bayesowskiej do wyznaczania rozkładów a posteriori, jednak w literaturze rzadko wspomina się o badaniu zbieżności wygenerowanych łańcuchów. Wnioskowanie w estymacji bayesowskiej przy niesprawdzonych zbieżnościach dla wszystkich parametrów

modelu może prowadzić do błędnych wniosków. Ponieważ różne testy badają różne aspekty zbieżności łańcuchów i brakuje testów jednoznacznie rozstrzygających, czy wygenerowany łańcuch osiągnął rozkład stacjonarny, w niniejszej pracy wykorzystano trzy metody oceny zbieżności.

Budowę modelu wykorzystanego w badaniu rozpoczęto od oszacowania kilku modeli regresji Poissona metodą klasyczną. Następnie korzystając z podstawowych miar dobroci dopasowania do obserwowanych danych (np. [62]) oraz badając korelację między zmiennymi wybrano najlepszy model.

W estymacji bayesowskiej, w celu porównania wrażliwości wyników na zmiany założeń dotyczących rozkładów a priori, estymowano dwa modele. W pierwszym modelu przyjęto nieinformacyjne rozkłady a priori dla wszystkich parametrów modelu, w drugim modelu oprócz nieinformacyjnych rozkładów a priori wykorzystano również informacyjny rozkład a priori (por. [8]). W symulacji jako wartości inicjujące dla pierwszego łańcucha przyjęto wartości parametrów estymowanych metodą największej wiarygodności. Ponadto, aby zminimalizować wpływ oddziaływania tych wartości początkowych na wnioskowanie a posteriori przyjęto, że liczba cykli spalonych ma wynosić 2000, a liczba właściwych realizacji do wnioskowań a posteriori 10000.

Dla wstępnego modelu zadano nieinformacyjny rozkład normalny a priori ze średnią równą 0 i wariancją 10^6 dla wszystkich parametrów regresji: $p(\beta) \sim N_9(\mathbf{0}, 10^6 \mathbf{I})$.

Wyniki estymacji modelu ze wszystkimi zmiennymi zawiera tabela 2.

Tabela 2.

Parametr	Wartości oczekiwane a posteriori	Odchylenia standardowe a posteriori	Obszary największej gęstości a posteriori ¹²	
Wyraz wolny	-0.9426	0.3618	-1.6958	-0.2489
m.zamieszkania	0.3097	0.0588	0.1932	0.4232
wykształcenie1	0.3835	0.1211	0.1600	0.6441
wykształcenie2	0.2126	0.1276	-0.0285	0.4787
wykształcenie3	0.2054	0.1222	-0.0256	0.4568
zatrudnienie1	0.1164	0.1342	-0.1609	0.3601
zatrudnienie2	0.1428	0.1440	-0.1459	0.4128
zwiazek	1.0258	0.3079	0.4710	1.7063
wyznanie	0.2812	0.1575	-0.0447	0.5758

Źródło: obliczenia własne

Na podstawie otrzymanych obszarów największej gęstości a posteriori można stwierdzić ([8]), że zmienne *zatrudnienie* oraz *wyznanie* są statystycznie nieistotne,

¹² Obszary największej gęstości a posteriori (HPD) zostały wyznaczone przy $\alpha = 0.05$ dla wszystkich parametrów modelu, w kolejnych estymacjach wartość ta pozostaje bez zmian.

zatem oszacowano kolejny model bez uwzględnienia tych zmiennych. Ponownie zadano nieinformacyjny rozkład normalny a priori ze średnią równą 0 i wariancją 10^6 dla wszystkich parametrów regresji: $p(\beta) \sim N_6(\mathbf{0}, 10^6\mathbf{I})$. Wyniki estymacji nowego modelu zostały przedstawione w tabeli 3.

Tabela 3.

Parametr	Wartości oczekiwane a posteriori	Odchylenia standardowe a posteriori	Obszary największej gęstości a posteriori	
Wyraz wolny	-0.5530	0.3270	-1.2238	0.0386
m_zamieszkania	0.2928	0.0579	0.1872	0.4134
wykształcenie1	0.3856	0.1248	0.1541	0.6422
wykształcenie2	0.2178	0.1323	-0.0272	0.4930
wykształcenie3	0.2047	0.1262	-0.0359	0.4587
związek	1.0337	0.3083	0.4857	1.7081

Źródło: obliczenia własne

Następnie estymowano kolejny model, do którego włączono dodatkową wiedzę spoza próby. Subiektywny rozkład a priori przyjęto w oparciu o informacje zawarte w publikacji [2] oraz na podstawie danych GUS. Dla parametru *m_zamieszkania* zadano rozkład normalny a priori ze średnią równą 0.398645 oraz wariancją równą 0.00374. Dla wszystkich pozostałych współczynników regresji przyjęto nieinformacyjny rozkład normalny a priori ze średnią równą 0 i wariancją 10^6 . Niestety wygenerowane łańcuchy dla niektórych parametrów modelu charakteryzowały się wysoką autokorelacją i nie były zbieżne, zatem estymowano kolejny model znacznie zwiększając liczbę iteracji. Charakterystyki pięćdziesięciotysięcznych prób a posteriori przedstawiono w tabeli 4.

Tabela 4.

Parametr	Wartości oczekiwane a posteriori	Odchylenia standardowe a posteriori	Obszary największej gęstości a posteriori	
Wyraz wolny	-0.5049	0.3143	-1.1470	0.1015
m_zamieszkania	0.3169	0.0277	0.2644	0.3551
wykształcenie1	0.3781	0.1236	0.1359	0.6173
wykształcenie2	0.2152	0.1313	-0.0411	0.4736
wykształcenie3	0.2062	0.1265	-0.0385	0.4561
związek	0.9787	0.2989	0.4179	1.5929

Źródło: obliczenia własne

Dla dwóch ostatnich modeli otrzymano podobne wartości charakterystyk oraz własności wygenerowanych prób a posteriori z niewielkim wskazaniem na ostatni mo-

del. Zatem w pracy poddano interpretacji i ocenie ostatni model, również z powodu uwzględnienia w nim dodatkowej wiedzy a priori.

W badaniach zbieżności wygenerowanych łańcuchów w przypadku modeli wielu parametrów często stosowany jest test Gelmana i Rubina ([10], [27]), który bazuje na wielokrotnych symulacjach łańcuchów MCMC dla tych samych parametrów przy wyborze różnych wartości inicjujących. Istota tego testu polega na porównaniu wariancji wewnątrz łańcucha z wariancją między tymi łańcuchami. Duże różnice pomiędzy tymi dwiema wariancjami wskazują na brak zbieżności łańcucha. Przedstawione w tabeli 5 wartości wskazują na zbieżności wszystkich wygenerowanych łańcuchów.

Tabela 5.

Parametr	Oszacowanie	97.5% ograniczenie górne
Wyraz wolny	1.0183	1.0465
m_zamieszkania	1.0000	1.0001
wykształcenie1	1.0011	1.0040
wykształcenie2	1.0010	1.0036
wykształcenie3	1.0008	1.0028
zwiazek	1.0172	1.0421

Źródło: obliczenia własne

Najprostszym testem, pozwalającym wnioskować, czy wygenerowany łańcuch osiągnął rozkład stacjonarny jest test Geweke'a ([29]). Oparty jest on na porównaniu wartości średniej dla początkowych realizacji łańcucha z wartością średniej dla dalszych realizacji łańcucha. Statystyka testowa ma asymptotyczny standardowy rozkład normalny. Na podstawie wyników przedstawionych w tabeli 6 stwierdzono brak podstaw do odrzucenia hipotezy, że wygenerowane łańcuchy dla poszczególnych parametrów modelu są zbieżne, przy dowolnym poziomie istotności.

Tabela 6.

Parametr	z	Pr > z
Wyraz wolny	1.1757	0.2397
m_zamieszkania	1.3187	0.1873
wykształcenie1	0.2703	0.7869
wykształcenie2	0.2815	0.7783
wykształcenie3	0.3184	0.7501
zwiazek	-1.3635	0.1727

Źródło: obliczenia własne

Test Geweke'a bada zbieżność wygenerowanych łańcuchów, test Heidelberga-Welcha ([33], [34]) ocenia również, czy otrzymany łańcuch jest wystarczający do estymacji średniej z zadaną dokładnością. Test Heidelberga-Welcha składa się z dwóch części: testu stacjonarności (stationarity test) i testu półszerokości (halfwidth test). Test stacjonarności testuje hipotezę, że łańcuch Markowa jest zbieżny do rozkładu stacjonarnego, natomiast test półszerokości sprawdza, czy liczebność próby a posteriori jest wystarczająca, aby estymować wartość średnią z zadaną dokładnością. Otrzymano (tabela 7), że brak jest podstaw do odrzucenia hipotezy o zbieżności łańcuchów dla wszystkich parametrów modelu, przy dowolnym poziomie istotności, ponadto liczebności prób a posteriori są wystarczające do oszacowania średniej z dokładnością 0.05.

Reasumując można przyjąć, że otrzymane próby a posteriori są odpowiednie do wnioskowań statystycznych.

Tabela 7.

Parametr	Test stacjonarności		Test półszerokości		
	Statystyka Cramera-von Misesa		Połowa szerokości przedz. ufności	Średnia	Względna półszerokość
	Stat.	Poziom p			
Wyraz wolny	0.0956	0.6071	0.0441	-0.5049	-0.0874
m.zamieszkania	0.1336	0.4435	0.000269	0.3169	0.000849
wykształcenie1	0.2798	0.1543	0.00548	0.3781	0.0145
wykształcenie2	0.3140	0.1235	0.00536	0.2152	0.0249
wykształcenie3	0.2907	0.1436	0.00539	0.2062	0.0262
zwiazek	0.0761	0.7150	0.0428	0.9787	0.0437

Źródło: obliczenia własne

5. ANALIZA WYNIKÓW ESTYMACJI I WNIOSKI

Rozważania na temat płodności można prowadzić w wielu aspektach, w niniejszej pracy w modelowaniu liczby posiadanych dzieci przez kobiety uwzględniono następujące determinanty: miejsce zamieszkania, wykształcenie, fakt pozostawania w związku małżeńskim, zatrudnienie oraz wyznanie. Otrzymano, że dwie ostatnie zmienne oraz dwa poziomy zmiennej wykształcenie nie mają statystycznego wpływu na modelowaną zmienną.

Bardziej szczegółowa analiza otrzymanych rezultatów wymaga dodatkowych obliczeń, których wyniki przedstawiono w tabeli 8.

Analizując od roku 1950 wartości współczynników dzietności ([2], [36], [44]) można zaobserwować znaczące różnice w płodności pomiędzy miastem i wsią. Uzyskane w niniejszej pracy wyniki potwierdzają wspomniane różnice w zachowaniach

prokreacyjnych pomiędzy kobietami zamieszkującymi miasta i wsie. Otrzymano, że kobiety mieszkające na wsi posiadają o około 37% więcej dzieci niż kobiety mieszkające w miastach. Natomiast bez uwzględnienia dodatkowej wiedzy a priori (por. tabela 3) otrzymano, że ta różnica jest mniejsza i wynosi blisko 34%. Uzyskany wynik jest zgodny z założeniami ekonomicznej teorii płodności zarówno Lebensteina jak i Beckera. W badanym okresie dzieci na wsi pracowały, zatem o ich liczbie mogły decydować spodziewane dochody rodziny uzyskane z przyszłej pracy dziecka.

Tabela 8.

Parametr	Wartości oczekiwane a posteriori ¹³	Exp(Wartości oczekiwane a posteriori)
m_zamieszkania	0.3169	1.3729
wykształcenie1	0.3781	1.4595
wykształcenie2	0.2152	1.2401
wykształcenie3	0.2062	1.2290
zwiazek	0.9787	2.6610

Źródło: obliczenia własne

Zgodne z teorią Beckera ([5]) oraz innymi badaniami dotyczącymi Polski są również otrzymane wyniki dotyczące wpływu wykształcenia na płodność – im wyższy poziom wykształcenia, tym mniejsza liczba posiadanych dzieci (np. [2]). Zatem potwierdziło się wcześniejsze przypuszczenie, że niższy poziom wykształcenia w porównaniu z wykształceniem wyższym skutkuje większą liczbą posiadanych dzieci, różnica ta pomiędzy kobietami z wykształceniem podstawowym i wyższym wynosi około 46%. Ponadto otrzymaliśmy, że dwa poziomy wykształcenia: zasadnicze i średnie nie mają istotnego wpływu na liczbę posiadanych dzieci. Odwołując się do koncepcji drugiego przejścia demograficznego można stwierdzić, że coraz większe pozytywne zmiany w aspiracjach edukacyjnych społeczeństwa niestety skutkują obniżeniem płodności.

Pomimo, że zdarzenia takie jak zawarcie małżeństwa i urodzenie dziecka niekoniecznie są ze sobą powiązane, to jednak w większości krajów ciągle wyższy współczynnik dzietności występuje wśród małżeństw. Również w Polsce w całym okresie powojennym płodność pozostaje w bardzo silnym związku z małżenością ([2]). Z drugiej strony od roku 1980 obserwowany jest ciągły wzrost urodzeń pozamałżeńskich, co niewątpliwie świadczy o zachodzących przemianach społecznych. W badaniu otrzymano, że kobiety będące kiedykolwiek w związku małżeńskim mają przeszło dwu i półkrotnie więcej dzieci niż pozostałe kobiety. Zatem można przypuszczać, że posiadanie kolejnego dziecka przez kobiety zamężne jest dużo bardziej prawdopodobne niż w przypadku pozostałych kobiet. Pozytywny wpływ związku małżeńskiego na

¹³ Wartości z tabeli 4.

dziatność jest również zauważalny w innych krajach, ale skala tego wpływu jest różna ([35]).

Analizy wzajemnych powiązań pomiędzy płodnością a aktywnością zawodową są przedmiotem wielu badań prowadzonych głównie dla krajów Europy zachodniej. Obserwowany w Polsce w latach 1960-1988 wzrost udziału kobiet w rynku pracy ([36]) i spadek płodności ([2]) sugerują konflikt pomiędzy macierzyństwem, a pracą. Potwierdza to raport ONZ (1985), według którego wzrost liczby kobiet pracujących jest powiązany ze spadkiem poziomu płodności. Wyniki przedstawione w tabeli 2 wskazują na nieistotną dodatnią zależność pomiędzy liczbą dzieci a sytuacją na rynku pracy, co może sugerować, że kobiety, które kiedykolwiek pracowały posiadają większą liczbę dzieci. Nieistotność rozważanej determinanty w Polsce można tłumaczyć tym, że w warunkach gospodarki centralnie planowanej rozwiązania w sferze polityki rodzinnej przenosiły role opiekuńczo-wychowawcze z rodziców na państwo poprzez szeroki dostęp do instytucji usług społecznych (żłobki, przedszkola, szkoły, zajęcia pozalekcyjne), zatem urodzenie dziecka nie musiało wiązać się z rezygnacją z pracy. Dodatnia zależność pomiędzy zatrudnieniem a płodnością dla innych krajów została również przedstawiona w publikacjach [9] i [61], szersze rozważania na ten temat można także znaleźć w pracy [35], w której uzyskano jednak odwrotną zależność.

Nieistotność zmiennej wyznaczenie nie potwierdza wcześniejszych przypuszczeń, że kobiety wyznania rzymskokatolickiego mogą posiadać większą liczbę dzieci. Może to wynikać z tego, że około 98% Polaków deklaruje, że jest wyznania rzymskokatolickiego ([25]), zatem pewne wzorce zachowań mogą być przyjmowane przez resztę społeczeństwa.

Przedstawione rezultaty określają wpływ różnych determinant na płodność w Polsce i stanowią uzupełnienie dotychczas prezentowanych w literaturze analiz. Wykorzystany w badaniu model regresji Poissona bada nie tylko kierunek zależności, ale również jej siłę, co umożliwia głębszą analizę rozważanych powiązań, natomiast podejście bayesowskie daje możliwość włączenia do modelu informacji pochodzących z innych badań oraz lepsze oszacowanie parametrów modelu ([19]). Przedstawione w literaturze wyniki (np. [35]) wskazują na to, że w poszczególnych krajach i okresach badane zależności między różnymi cechami a płodnością nie zawsze są identyczne, co więcej mogą mieć kierunek dodatni lub ujemny. Zatem warto badać te zależności dla Polski, ponieważ wyniki nie zawsze są oczywiste i zgodne z powszechnie przyjętymi przekonaniem.

Szkoła Główna Handlowa w Warszawie

LITERATURA

- [1] Adelman I., *An Econometric Analysis of Population Growth*. The American Economic Review, Vol. 54, No. 2, 314-339, 1963.
- [2] Balicki J., Frątczak E., Nam Ch.B., *Przemiany ludnościowe*. Fakty – Interpretacje – Opinie. Wydawnictwo Uniwersytetu Karola Stefana Wyszyńskiego w Warszawie, 2007.

-
- [3] Becker G.S., *A treatise on the family*. Harvard University Press, Cambridge, Mass. 1981.
- [4] Becker G.S., *Ekonomiczna teoria zachowań ludzkich*. PWN, 1990.
- [5] Becker G.S., *Human capital: A theoretical and empirical analysis with special reference to education*. The University of Chicago Press, 1993.
- [6] Becker G.S., Barro R.J., *A Reformulation of the Economic Theory of Fertility*. The Quarterly Journal of Economics, Vol. 103, No.1, 1-25, 1988.
- [7] Bernardo J., Smith A., *Bayesian Theory*. A John Wiley & Sons, 2004.
- [8] Bolstad W.M., *Introduction to Bayesian statistics*. A John Wiley & Sons, 2007.
- [9] Brewster K.L., Rindfuss R.R., *Fertility and women's employment in industrialized nation*. Annual Review of Sociology, 26, 271-296, 2000.
- [10] Brookes S., Gelman A., *General methods for monitoring convergence of iterative simulations*. Journal of Computational and Graphical Statistics, 7, 434-455, 1998.
- [11] Casella G., George E.I., *Explaining the Gibbs sampler*. The American Statistician, 46, 167-74, 1992.
- [12] Coleman J.S., *Foundations of social theory*. Cambridge. Harvard University Press, 1990.
- [13] Dellaportas P., Smith A.F.M., *Bayesian inference for generalized linear models and proportional hazard models via Gibbs sampling*. Applied Statistics, Vol. 42, No. 3, 443-459, 1993.
- [14] Dey D.K., Ghosh S.K., Mallick B.K., *Generalized Linear Models: A Bayesian Perspective*, New York: Marcel Decker, 2000.
- [15] Dobson A.J., *An introduction to generalized linear models*. Chapman & Hall, 1991.
- [16] Doss H., Narasimhan B., *Bayesian Poisson regression using the Gibbs sampler: Sensitivity analysis through dynamic graphics*. Technical report, Penn State Erie, 1994.
- [17] Easterlin R.A., *The Economics and Sociology of Fertility: A Synthesis*. Tilly Ch. (Red.) Historical Studies of Changing Fertility, Princeton, 1979.
- [18] Easterlin R.A., *Birth and fortune: the impact of numbers on personal welfare*. Basic Books, 1987.
- [19] El-Sayyad G.M., *Bayesian and classical analysis of Poisson regression*. Journal of Royal Statistical Society. Series B. Vol. 35, No. 3, 445-451, 1973.
- [20] Frątczak E., *Modelowanie cyklu życia jednostki i rodziny teoria i praktyka*. Oficyna Wydawnicza SGH, 1999.
- [21] Frątczak E., Józwiak J., Paszek B., *Zastosowania analizy historii zdarzeń w demografii*. Oficyna Wydawnicza SGH, 1996.
- [22] Frątczak E., Kowalska I., Götz R., Drobnič S., Blossfeld H.P., *Polish family and fertility survey. A user' guide*. Warsaw and Bremen, 1996.
- [23] Frątczak E., Sobczak I. (Red.), *Dzietność kobiet polskich w okresie transformacji ustrojowej*. Polskie Towarzystwo Demograficzne, 2000.
- [24] Gamerman D., Lopes H., *Markov Chain Monte Carlo: stochastic simulation for Bayesian inference*. Chapman & Hall (CRC Press), 2006.
- [25] Gawryszewski A., *Ludność Polski w XX Wieku*. Instytut Geografii i Przestrzennego Zagospodarowania im. Stanisława Leszczyckiego, PAN, 2005.
- [26] Gelman A., Carlin J.B., Stern H.S., Rubin D.B., *Bayesian data analysis*. Chapman & Hall/CRC, 2000.
- [27] Gelman A., Rubin D., *Inference from iterative simulation using multiple sequences*. Statistical Science, 7, 457-72, 1992.
- [28] Geman S., Geman D., *Stochastic relaxation, Gibbs distributions and the Bayesian restoration of image*. IEEE Transaction on Pattern Analysis and Machine Intelligence, 6, 721-741, 1984.
- [29] Geweke J., *Evaluating the accuracy of sampling-based approaches to calculating posterior moments*. Bayesian Statistics, 4, 169-193, 1992.
- [30] Gilks W., Best N., Tan K., *Adaptive rejection Metropolis sampling with Gibbs sampling*. Applied Statistics, 44, 455-472, 1995.

- [31] Gilks W., Wild P., *Adaptive rejection sampling for Gibbs sampling*. Applied Statistics, 41, 337-348, 1992.
- [32] Hardin J.W., Hilbe J.M., *Generalized linear models and extensions*. Stata Press, 2007.
- [33] Heidelberger P., Welch P., *A spectral method for confidence interval generation and run length control in simulations*. Communication of the ACM, 24, 233-245, 1981.
- [34] Heidelberger P., Welch P., *Simulation run length control in the presence of an initial transient*. Operation Research, 31, 1109-1144, 1983.
- [35] Hilgeman Ch., Butts C.T., *Women's employment and fertility: A welfare regime paradox*. Social Science Research, Vol. 38, Issue 1, 103-117, 2009.
- [36] Holzer J.Z., *Demografia*. Polskie Wydawnictwo Ekonomiczne, Warszawa, 2003.
- [37] Holzer J.Z., Kowalska I., *Fertility and family surveys in countries of the ECE region*. Standard country report, Poland, UNECE, UNPF, 1997.
- [38] Ibrahim J.G., Laud P.W., *On Bayesian analysis of generalized linear models using Jeffreys's prior*. Journal of the American Statistical Association, Vol. 86, No. 416, 981-986, 1991.
- [39] Jeffreys H., *An invariant form for the prior probability in estimation problems*. Proceedings of the Royal Society of London, Ser. A 186, 453-461, 1946.
- [40] Jeffreys H., *Theory of probability*. New York: Oxford University Press, 1961.
- [41] Kaplan H., *Evolutionary wealth flows theories of fertility: empirical test and new models*. Population and development Review, 20 (4), 1994.
- [42] Klasen S., Launov A., *Analysis of the Determinants Fertility Decline in the Czech Republic*. Journal of Population Economy, Vol. 19, 25-54, 2006.
- [43] Kotowska I.E. (Red.), *Przemiany demograficzne w Polsce w latach 90.w świetle koncepcji drugiego przejścia demograficznego*. Oficyna Wydawnicza SGH, 1999.
- [44] Kotowska I.E., Józwiak J., Matysiak A., Baranowska A., *Poland: Fertility decline as a response to profound societal and labour market changes?* Demographic Research, Vol.19, Article 22, 795-854, 2008.
- [45] Kurkiewicz J., *Modele przemian płodności w wybranych krajach europejskich w świetle drugiego przejścia demograficznego*. Wydawnictwo Akademii Ekonomicznej w Krakowie, 1998.
- [46] Leibenstein H., *Rozwój i zacofanie gospodarcze*. Studia z teorii rozwoju gospodarczego, 1963.
- [47] Lesthaeghe R., *A century of demographic and cultural change in Western Europe, an exploration of underlying dimensions*. Population and Development Review, 9 (3), 411-435, 1983.
- [48] Lesthaeghe R., *The second demographic transition in Western Countries: An interpretation*. Interuniversity Programme in Demography. Working Paper, No. 2, 1991.
- [49] Lesthaeghe R., Surkyn J., *Cultural dynamics and economic theories of fertility change*. Population and Development Review, 14 (1), 1-45, 1988.
- [50] Lynch S.M., *Introduction to applied Bayesian statistics and estimation for social scientists*. Springer, 2007.
- [51] McCullagh P., Nelder J.A., *Generalized linear models*. Chapman & Hall, 1989.
- [52] Montgomery M.R., Casterline J.B., *Social Influence, Social Learning and New models of fertility*. Population and Development Review, Vol. 22, 151-175, 1996.
- [53] Nelder J.A., Wedderburn R.W.M., *Generalized linear models*. J. R. Statist. Soc. A, 135, 370-384, 1972.
- [54] Oláh L.Sz., Frątczak E., *Becoming a Mother in Hungary and Poland during State Socialism*. Demographic Research, Special Collection 3, Article 9, 211-244, 2004.
- [55] Osiewalski J., *Ekonometria bayesowska w zastosowaniach*. Wydawnictwo Akademii Ekonomicznej w Krakowie, 2001.
- [56] Raftery A., Lewis S., *One long run with diagnostics: implementation strategies for Markov Chain Monte Carlo*. Statistical Science, 7, 493-497, 1992.

- [57] Raftery A., Lewis S., *The number of iterations, convergence diagnostics and generic Metropolis Algorithms*. In Gilks W., Spiegelhalter D., Richardson S., editors, Practical Markov Chain Monte Carlo, London, UK: Chapman & Hall, 1995.
- [58] Robert C., Casella G., *Monte Carlo statistical methods*. Springer Texts in Statistics, 2004.
- [59] Silvey S.D., *Wnioskowanie statystyczne*, PWN, 1978.
- [60] Skirbekk V., *Fertility trends by social status*. Demographic Research. Vol. 18, Article 5, 145-180, 2008.
- [61] Sleebos J.E., *Low fertility rates in OECD countries*. Facts and Policy Responses. Organization for Co-operation and Development, 2003. <http://www.oecd.org/dataoecd/13/38/16587241.pdf>
- [62] Stanisław A., *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny*. Tom 2. StatSoft, 2007.
- [63] Szoltysek M., *Rozrodczość w perspektywie ekonomicznej i demografia przeszłości*. Roczniki Dziejów Społecznych i Gospodarczych. Vol. 65, 25-53, 2005.
- [64] Szreder M., *Informacje a priori w klasycznej i bayesowskiej estymacji modeli regresji*. Wydawnictwo Uniwersytetu Gdańskiego, 1994.
- [65] Szreder M., Osiewalski J., *Przykład bayesowskiej estymacji modelu rynku w nierównowadze z wykorzystaniem subiektywnych rozkładów a priori*. Przegląd Statystyczny, Nr 3-4, 1991.
- [66] Van de Kaa D.J., *Europe's second demographic transition*. Population Bulletin, Vol. 42, No. 1, 1987.
- [67] Van de Kaa D.J., *Post modern fertility preferences: from changing value orientation to new behaviour*. Population and Development Review (supplement), Vol.27, 290-331, 2001.
- [68] Wang W., Famoye F., *Modeling household fertility decisions with generalized Poisson regression*. Journal of Population Economics, 10, 273-283, 1997.

ANALIZA PŁODNOŚCI KOBIEC W POLSCE Z WYKORZYSTANIEM BAYESOWSKIEGO MODELU REGRESJI POISSONA

Streszczenie

Celem niniejszej pracy jest zbadanie zachowań prokreacyjnych Polek poprzez identyfikację czynników je determinujących z wykorzystaniem metod bayesowskich. W pracy zastosowano bayesowski model regresji Poissona. Wybrany model umożliwił określenie kierunku i skali wpływu wybranych czynników na liczbę dzieci posiadanych przez kobiety. Natomiast podejście bayesowskie dało możliwość włączenia do modelu informacji a priori oraz lepsze oszacowanie parametrów modelu. W estymacji wykorzystano metody Monte Carlo oparte na łańcuchach Markowa, a w szczególności próbnik Gibbsa.

Badanie przeprowadzono na podstawie danych indywidualnych pochodzących z polskiego badania retrospektywnego „Przemiany rodziny i wzorce dzietności w Polsce” (1991). W analizie płodności kobiet uwzględniono następujące czynniki: miejsce zamieszkania, wykształcenie, fakt pozostawania w związku małżeńskim, zatrudnienie oraz wyznanie. Otrzymane rezultaty porównano z dotychczasowymi wynikami badań dla Polski i innych krajów.

Słowa kluczowe: płodność, model regresji Poissona, wnioskowanie bayesowskie, metody Monte Carlo oparte na łańcuchach Markowa

FERTILITY ANALYSIS OF WOMEN IN POLAND USING BAYESIAN POISSON REGRESSION MODEL

A b s t r a c t

The primary objective of the work is to use Bayesian methods to investigate women fertility in Poland and identify key factors influencing it. Bayesian Poisson regression model has been used in the analysis. The model allows determining factors that have a significant impact on the number of children born. Moreover Bayesian approach makes it possible to incorporate a priori knowledge and improve the estimation of model parameters. The model has been estimated using Markov chain Monte Carlo method with Gibbs sampling.

The work has been based on the Polish study "Family changes and Fertility Patterns in Poland" (1991). The following attributes have been considered in the analysis of women fertility: place of living, education, marital status, employment and religion. The results have been compared with the results of related research for Poland and other countries.

Key words: fertility, Poisson regression model, Bayesian inference, Markov chain Monte Carlo method