

Modelowanie dzietności: adaptacja modelu Lee-Cartera¹

Streszczenie

Praca poświęcona jest modelowaniu dzietności kobiet przy pomocy zmodyfikowanych modeli typu Lee-Cartera. Materiał empiryczny poddany analizie to dane pochodzące z GUS o urodzeniach ogółem według wieku kobiet w Polsce w latach 1971-2018, oraz o urodzeniach pierwszego dziecka w okresie 1971-2017. Modyfikacje modelu Lee-Cartera – pierwotnie zaprojektowanego do modelowania i prognozowania umieralności – zostały poczynione w taki sposób, aby uwypuklić bezpośredni związek parametrów modelu z ewolucją współczynników dzietności brutto oraz średniego wieku rodzenia dzieci, tak w ujęciu przekrojowym jak i kohortowym.

Słowa kluczowe

przekrojowy współczynnik dzietności brutto, kohortowy współczynnik dzietności brutto, bezdzietność, model Lee-Cartera, model Renshaw'a-Habermana

1. Wprowadzenie

Spadek dzietności jest zjawiskiem powszechnie występującym w krajach wysoko rozwiniętych od wielu dekad. W Polsce i krajach naszego regionu Europy wystąpił nieco później niż w Europie Zachodniej i Ameryce Północnej, przybierając jednak bardziej dramatyczny charakter. Spadek wydaje się być głębszy, a okres, w którym do zasadniczej zmiany doszło, wydaje się być krótszy, tak jakbyśmy mieli do czynienia ze zjawiskiem „doganiania” w tym względzie krajów gospodarczo wyżej od Polski rozwiniętych. Na poziomie tak ogólnie opisanego zjawiska nie ma między badaczami dzietności sporu. Oczywiście dzieje się dziś tak dlatego, że okres najbardziej dramatycznych zmian mamy już w Polsce za sobą. Zanim proces ten przeszedł w fazę zmian stosunkowo niewielkich i powolnych, rozstrzygnięcie sporów czy i na ile mamy do czynienia z prawdziwym spadkiem docelowej dzietności, a na ile z procesem odkładania w czasie planów prokreacyjnych, okazywało się nieomal niemożliwe. Jednak nawet i teraz dokładniejszy pomiar, umiejscowienie w czasie, i analiza owego spadku napotyka na istotne trudności. Problem w tym, iż generalnemu spadkowi dzietności towarzyszy znaczny wzrost przeciętnego wieku kobiet rodzących dzieci. W rezultacie przekrojowe

¹ Pierwszym efektem badań w tym zakresie była praca licencjacka Jagody Niemczynowicz „Wpływ efektów kalendarzowych i kohortowych na depresję urodzeniową w Polsce”, przygotowanej pod kierunkiem Wojciecha Otto i obronionej na Wydziale Nauk Ekonomicznych Uniwersytetu Warszawskiego w czerwcu 2018 roku. Niniejszy artykuł jest efektem kontynuacji tych badań, prowadzonych po części wspólnie, po części zaś samodzielnie przez Wojciecha Otto.

współczynniki dzietności brutto, które miałyby ilustrować przeciętną liczbę urodzeń żywych przypadających na jedną kobietę w danym wieku, zsumowane po grupach wiekowych, reprezentują łączną dzietność nieistniejącej populacji, o cechach odległych od jakiegokolwiek rzeczywistej kohorty kobiet. Z drugiej strony powszechne stosowanie przekrojowych współczynników dzietności mimo powyższych wątpliwości ma istotne uzasadnienie, jako że wartość takiego współczynnika za dany rok można wyznaczyć już w lecie następnego roku, kiedy zazwyczaj pojawiają się dane o dzietności kobiet według wieku. Znacznie bardziej interesujące są kohortowe współczynniki dzietności, niestety możemy je wyznaczyć w prosty sposób i z sensowną precyzją tylko dla tych kohort, które lwia część aktywności prokreacyjnej mają już za sobą. Na przykład grupując razem urodzenia żywe kobiet w wieku 45 lat i więcej, możemy wyznaczyć współczynniki kohortowe maksymalnie dla kohort kobiet urodzonych 45 lat temu, a których szczyt aktywności prokreacyjnej miał miejsce 15-20 lat wcześniej, niż nasze najświeższe dane.

Pewną odpowiedzią na niedoskonałości prostej analizy przekrojowej jest metoda dekompozycji współczynnika dzietności na tzw. efekt *quantum* oraz efekt *tempo*, jak również różne modyfikacje tej metody. Metodę tę zaproponowali w 1998 roku Bongaarts i Feeney. Metoda ma tę zaletę, że efekt *quantum* można traktować jako swoistą aproksymację rzeczywistej (w domyśle – ukrytej) dzietności w roku kalendarzowym, wyczyszczony (przynajmniej częściowo) ze zniekształcającego tę dzietność wpływu zmian przeciętnego wieku rodzenia dzieci (efektu *tempo*). Nie znosi to jednak nadal podstawowej wątpliwości – „kohorta syntetyczna”, której tak wyliczona wielkość odpowiada, nie istnieje. Natomiast oczywistą praktyczną zaletą tej metody jest możliwość wyznaczenia skorygowanego przekrojowego współczynnika dzietności za dany rok na podstawie dodatkowych danych jedynie za jeden następny, oraz jeden poprzedni rok.

Jeszcze inne podejście można oprzeć na modelach, w ramach których oddzielamy od siebie tę część zmian rozkładu urodzeń według wieku kobiet, które mają przynajmniej po części charakter systematyczny, od zmian przypadkowych i nietrwałych. Takie podejście sprzyja pewnemu syntetycznemu oglądowi ewolucji dzietności w okresie, z którego posiadamy obserwacje, ale także może posłużyć do ekstrapolacji poza ten okres w przeszłość i na przyszłość. Przykładowo, jeśli ustalimy dość wyraźne prawidłowości w okresie próby, możemy podjąć próbę przewidzenia, jak kształtować się będzie aktywność prokreacyjna kobiet po przekroczeniu wieku 35 lat z kohorty, którą właśnie do osiągnięcia wieku 35 lat obserwowaliśmy. Narzędziem w ramach tego podejścia są przede wszystkim różne rozwinięcia modelu, który zaproponowali Ronald D. Lee oraz Lawrence Carter w roku 1992. Model Lee-

Cartera w pionierskiej wersji skonstruowany został do celów analizy i predykcji umieralności, i liczne prace poświęcone różnym jego modyfikacjom wiążą się z ogromnym zapotrzebowaniem, jakie na predykcje tablic trwania życia zgłasza sektor komercyjnych ubezpieczeń na życie oraz kapitałowych funduszy emerytalnych. Mimo to pierwsze próby zastosowań narzędzia tego typu do analizy dzietności, pojawiły się już dawno (por. praca samego Ronalda Lee z roku 1993, i późniejsze prace np. Williama Bella z roku 1997), a – choć nieliczne – można także napotkać zastosowania na przykład do modelowania imigracji i emigracji. Choć jakiegokolwiek systematyczne zmiany są w tym ostatnim przypadku są znacznie mniej stabilne i uchwytne, to (pewnie nie bez powodu) ten temat pojawia się w ramach prac nad projekcjami populacji, wymagającymi uwzględnienia wszystkich trzech procesów (umieralności, płodności i migracji) równocześnie.

W dalszych podrozdziałach tej pracy przedstawimy cały szereg modyfikacji, jakie wprowadziliśmy do podstawowej konstrukcji modelu Lee-Cartera, oraz wyniki zastosowania zmodyfikowanego modelu (a właściwie zmodyfikowanych modeli) do badania populacji kobiet polskich. Natomiast nie będziemy odwoływać się do prób wnioskowania o przyszłości dla danej populacji (np. polskich kobiet) na podstawie wiedzy o przeszłości innych populacji, które można z jakichś powodów uznać za wzorzec, powielany z pewnym opóźnieniem przez polskie kobiety. Wiedza o ewolucji w innych populacjach jest bardzo często wykorzystywana w analizach jakościowych, rzadko jednak sposób jej wykorzystania daje się wtłoczyć w ramy ścisłego wnioskowania statystycznego. Nie będziemy się także odwoływać do metod stosowanych w badaniach opartych na innych źródłach danych, niż statystyka państwowa.

Na koniec uwag wstępnych garść informacji o wykorzystywanych danych. Są to prawdopodobieństwa urodzenia żywego dziecka przez kobietę w wieku ukończonych x lat w ciągu roku t , oznaczanych przez $q_{x,t}$ w przypadku gdy chodzi o urodzenie każdego dziecka, zaś przez $q1_{x,t}$, gdy chodzi o urodzenie pierwszego dziecka. Tak zdefiniowane prawdopodobieństwo wyznacza się empirycznie jako iloraz liczby urodzeń żywych (wszystkich bądź pierwszych) podzielonej przez liczbę kobiet w badanej populacji kobiet, którym takie zdarzenie mogło się przytrafić.

Należy zwrócić uwagę na to, iż niekiedy ta sama kobieta może w liczniku ilorazu $q_{x,t}$ pojawić się więcej niż raz, o ile np. urodziła żywe bliźniaki (trojaczki, czworaczki itd.) i/lub urodziła w ciągu roku t żywe dzieci pochodzące z dwóch kolejnych ciąż, ale w obu przypadkach ciążę te swój finał znalazły w czasie, gdy wiek tej kobiety (liczony w ukończonych latach) wciąż wynosił x . Prościej jest w przypadku współczynnika $q1_{x,t}$ gdzie w liczniku każda kobieta

może pojawić się co najwyżej raz. Oznacza to oczywiście, iż ściśle rzecz biorąc jedynie współczynnik $q1_{x,t}$ jest rzeczywiście miarą prawdopodobieństwa odpowiedniego zdarzenia, podczas gdy współczynnik $q_{x,t}$ jest empirycznym odpowiednikiem oczekiwanej liczby dzieci urodzonych w roku t przez kobietę w wieku x . Dla uproszczenia będziemy się posługiwać jednak krótszym terminem prawdopodobieństwo, z uwagi na to że w praktyce różnice pomiędzy tymi dwiema wielkościami są pomijalne.

Mianownik obu tych ułamków jest wyznaczany w szczególny sposób. Jest on sumą wag – liczb z przedziału $(0,1)$ – przypisanych wszystkim kobietom urodzonym w latach $(x - t - 1)$ oraz $(x - t)$, a żyjącym wciąż na początku roku t . Waga przypisana danej kobiecie mierzy długość tej części roku t , w której kobieta ta nadal żyje i jej wiek mierzony w ukończonych latach wynosi x . W języku angielskim otrzymaną w ten sposób sumę wag określamy terminem *exposure*, co nie ma dobrego tłumaczenia na język polski. Chodzi o liczbę jednostek ryzyka, którą w naszym przypadku można by określić precyzyjnie, acz topornie mianem liczby kobieto-lat, podobnie jak w ubezpieczeniach komunikacyjnych można mówić o liczbie samochodo-lat. Od razu należy tu zaznaczyć, że w przypadku współczynnika $q1_{x,t}$ analizowaliśmy w badaniu prawdopodobieństwa bezwarunkowe, nie eliminując z mianownika liczby tych kobiet, które już swoje pierwsze dziecko urodziły. W wielu badaniach (szczególnie, gdy interesują nas szczegóły dotyczące urodzeń pierwszych, drugich trzecich, itd., odstępy w czasie między kolejnymi urodzeniami itp.) badacze analizują prawdopodobieństwa warunkowe urodzenia k -tego dziecka, które oblicza się biorąc pod uwagę liczbę tylko tych kobiet, które już urodziły $(k - 1)$ dzieci. Wyboru prawdopodobieństw bezwarunkowych dokonaliśmy z uwagi na pewną ich oczywistą własność. Chodzi o to, że współczynnik zagregowany $q_{x,t}$ jest z definicji zwyczajną sumą współczynników $q1_{x,t}, q2_{x,t}, q3_{x,t}, \dots$ tylko wtedy, kiedy te ostatnie to prawdopodobieństwa bezwarunkowe. Jeśli w tej roli obsadzimy prawdopodobieństwa warunkowe, prostota tego związku znika. Drugim, bardziej subtelnym powodem jest to, że GUS liczy *exposure* dla prawdopodobieństw warunkowych opierając się na rzeczywistych danych jedynie raz na wiele lat, a w latach pośrednich na zgrubnych szacunkach i interpolacjach. Sytuację komplikuje konieczność uwzględnienia procesów emigracji i imigracji, które w istotny sposób modyfikują zarówno liczbę kobiet rodzących, jak i liczbę kobiet nie rodzących dzieci w kolejnych latach kalendarzowych i w różnych grupach wiekowych. Korekty te mają dość skomplikowany charakter, a przy tym są zawodne, co związane jest z trudnościami ustalenia w ramach systemów statystyki państwowej, jaka liczba wyjazdów lub powrotów z zagranicy ma charakter trwały. To właśnie stanowi argument

zwolenników analizowania prawdopodobieństw warunkowych – niektórzy badacze (szczególnie ci związani z GUS-em) twierdzą, że błędy pomiaru efektów emigracji i imigracji są mniejsze w przypadku danych służących do oszacowania prawdopodobieństw warunkowych. Nam argumenty przemawiające za wyższą wiarygodnością danych GUS o prawdopodobieństwach bezwarunkowych niż warunkowych wydały się bardziej przekonujące, na co oczywiście żadnych twardych dowodów nie mamy. Oceną zasadności takiego poglądu utrudnia fakt, że dla oceny błędów pomiaru ilorazu kluczowe jest to, czy błędy pomiaru licznika i mianownika są silnie dodatnio skorelowane, czy nie. O ile można znaleźć opinie pracowników merytorycznych GUS o skali tych błędów z osobna, to na ocenę stopnia ich skorelowania nigdzie natrafić nie zdołaliśmy.

Ostatnia rzecz wymagająca wyjaśnienia, to użycie przez nas terminu „kohorta”. Będziemy mianowicie używać umownego terminu „kohorta kobiet z roku t ” dla określenia tych kobiet, których dotyczy szereg prawdopodobieństw $q_{15,t+15}, q_{16,t+16}, q_{17,t+17}, \dots$. Sposób wyznaczania tych ilorazów oznacza, iż w skład tej umownej kohorty wchodzić będą (częściowo) zarówno kobiety urodzone w roku t , jak i w roku $(t - 1)$.

Ostatecznie wykorzystaliśmy dane których źródłem pierwotnym jest GUS, choć bezpośrednio z tego źródła skorzystaliśmy tylko w odniesieniu do lat 2017 i 2018. Za lata 1971-2016 pobraliśmy dane z Human Fertility Database², choć pierwotnym źródłem ich pochodzenia w odniesieniu do Polski są oczywiście dane GUS. Rzecz w tym, że bezpośrednio na stronie GUS można znaleźć takie dane (w formie zdigitalizowanej) za lata 1981 – 2018. Metodologia obróbki danych przez HFD we współpracy z GUS-em zapewnia jednak, że starsze dane (pochodzące oczywiście też z GUS-u), zostały przed dołączeniem do bazy doprowadzone (co do szczegółów metodologii) do porównywalności z danymi późniejszymi, czego sami z pewnością nie byłibyśmy w stanie rzetelnie dokonać korzystając z danych pochodzących ze starych publikacji GUS. Porównanie danych dostępnych na stronie GUS za lata 1981-2016 z danymi z HFD upewniło nas, w której tabeli GUS szukać danych identycznych do tych zaczerpniętych z HFD, by następnie w oparciu o tę samą tabelę uzupełnić dane o lata 2017 i 2018.

Co ważne, rok 2017 był ostatnim rokiem, kiedy GUS gromadził i publikował dane o urodzeniach według kolejności. Dane na ten temat zniknęły nie w samym GUS-ie, ale u źródła (notując zajście urodzenia żywego zaniechano notowania, z którym z kolei urodzeniem żywym

² Wykorzystaliśmy dane z pliku POLasfrRRbo.txt, pobranego z witryny <https://www.humanfertility.org/cgi-bin/main.php> w dniu 21 kwietnia 2019 roku

danej kobiety mamy do czynienia). Dlatego dokonana przez nas analiza ogranicza się w przypadku współczynników $q_{1,x,t}$ do próbki za lata 1971-2017, podczas gdy analizą współczynników $q_{x,t}$ objęliśmy także rok 2018.

Badaniem objęliśmy grupy wiekowe 15 – 45 lat, które z wyjątkiem grup skrajnych są grupami jednorocznymi. Współczynnik oznaczany przez nas jako $q_{15,t}$ jest tak naprawdę sumą współczynników jednorocznych grupy 15 lat, 14 lat, itd. Współczynnik oznaczany zaś jako $q_{45,t}$ jest sumą współczynników grup 45, 46, 47 itd. W zbiorach oryginalnych danych rolę grup skrajnych pełni grupa wiekowa 12 lat i mniej oraz grupa wiekowa 55 lat i więcej. Uznaliśmy, iż szkoda w postaci utraty informacji jest w obu przypadkach mniejsza, niż skutek w postaci zaciemnienia obrazu przez zaśmiecenie analizy nieistotnie małymi współczynnikami, bez znaczenia dla oceny całkowitej dzietności.

2. Model Lee-Cartera

W roku 1992 Ronald D. Lee oraz Lawrence Carter zaproponowali nowy sposób modelowania śmiertelności. Wersja oryginalna i jej późniejsze modyfikacje stały się fundamentem badań i prognoz w tym zakresie, między innymi nakierowanych na konstrukcję prognozowanych (niekiedy używa się określenia projektowanych lub dynamicznych) tablic trwania życia. Postać modelu w wersji pierwotnej przedstawia się następująco:

$$\ln \mu_{x,t} = \alpha_x + \beta_x \kappa_t + \varepsilon_{x,t}, \quad x \in \{x_{min}, \dots, x_{max}\}, \quad t \in \{t_{min}, \dots, t_{max}\}$$

gdzie:

- (x, t) – to odpowiednio wiek w latach i rok kalendarzowy,
- $\mu_{x,t}$ – to współczynnik natężenia zgonów osób w wieku x w roku kalendarzowym t ,
- α_x – to średni poziom logarytmu współczynnika zgonów osób w wieku x w całym okresie od t_{min} do t_{max} ,
- $\beta_x \kappa_t$ - to systematyczna część odchylenia $\ln \mu_{x,t}$ od α_x , często nazywana w skrócie „efektem kalendarzowym”,
- $\varepsilon_{x,t}$ - to losowa część odchylenia $\ln \mu_{x,t}$ od α_x , często nazywana efektem indywidualnym
- $\beta_{x_{min}}, \beta_{x_{min}+1}, \dots, \beta_{x_{max}}$ – to ciąg współczynników wyznaczających strukturę wiekową efektów kalendarzowych,
- $\kappa_{t_{min}}, \kappa_{t_{min}+1}, \dots, \kappa_{t_{max}}$ – to szereg czasowy reprezentujący dynamikę efektów kalendarzowych.

Dodatkowo autorzy przyjęli założenie, iż składniki losowe $\varepsilon_{x,t}$ są niezależnymi zmiennymi losowymi o rozkładach normalnych z zerową wartością oczekiwaną i jednością

wariancją. Dopełnieniem modelu (niezbędnym, o ile zamierzamy użyć modelu do generowania prognoz), jest model szeregu czasowego κ_t , w najprostszej wersji w postaci modelu błędzenia przypadkowego z dryfem.

Dla celów uzyskania jednoznacznej identyfikacji nieznanymi parametrami modelu przyjmuje się jeszcze dwa założenia normujące, najczęściej w postaci $\sum_x \beta_x^2 = 1$, $\sum_t \kappa_t = 0$.

Przy powyższych założeniach estymacja nieznanymi parametrami modelu jest stosunkowo prosta – zastosowanie metody największej wiarygodności prowadzi do estymatorów równoważnych tym, które otrzymamy minimalizując sumę kwadratów reszt, a rezultat jest następującej postaci:

- estymatorami współczynników α_x są zwykle średnie arytmetyczne z obserwacji $\ln \mu_{x,t}$ po czasie t ,

- estymatory wektorów struktury wiekowej β oraz dynamiki w czasie κ są równoważne wektorowi ładunków czynnikowych oraz przebiegu w czasie wartości czynnika głównego, które otrzymujemy w wyniku analizy składowych głównych zastosowanych do naszych obserwacji po odjęciu wyestymowanych wstępnie parametrów α_x . Obliczeniowo stosuje się tutaj metodę tzw. *Single Value Decomposition* (SVD).

Obliczeniowo zadanie ciągle jest stosunkowo proste, o ile wzbogacimy systematyczną część modelu o dodatkowy (drugi, trzeci, ...) efekt kalendarzowy. Ma to sens na przykład w przypadku, gdy spodziewamy się iż dynamika zmian śmiertelności w starszych przedziałach wiekowych okaże się istotnie inna (w badanym okresie czasu) niż dynamika zmian w młodszych grupach wiekowych. Wtedy zadanie rozwiązujemy niemal tak samo, wyławiamy jednak nie tylko pierwszą składową główną, ale także składową drugą, ewentualnie trzecią itd.

Zadanie komplikuje się w sposób istotny, jeśli poddamy pierwotną postać modelu jednej (lub większej liczbie) innym modyfikacji. Jedną z takich modyfikacji to odstępianie od założenia o rozkładzie normalnym składników losowych, najczęściej na rzecz założenia o rozkładzie Poissona. Ta modyfikacja jest zasadna wtedy, gdy analizujemy próbki o ograniczonej liczebności, pochodzące z badań reprezentacyjnych bądź z bazy danych o klientach zakładu ubezpieczeń. Nie dotyczą więc one naszego przypadku. Inne modyfikacje, które wystąpią w modelach w tym badaniu zastosowanych, omówimy w następnych podrozdziałach.

3. Modyfikacje pierwotnej postaci modelu Lee-Cartera – składnik systematyczny

Oszacowaliśmy w sumie trzy pary modeli o różnej strukturze, każda w wersji dla danych o dzietności ogółem $q_{x,t}$ i w wersji dla danych o urodzeniach pierwszego dziecka $q1_{x,t}$. Struktura tych modeli (w wersji dla dzietności ogółem) jest postaci:

$$q_{x,t} = \alpha_x + \beta1_x \kappa_t + \beta2_x \gamma_{t-x} + \varepsilon_{x,t} \quad [1]$$

$$q_{x,t} = \beta1_x + \beta2_x \kappa2_t + \varepsilon_{x,t} \quad [2]$$

$$q_{x,t} = \beta1_x \gamma1_{t-x} + \beta2_x \gamma2_{t-x} + \varepsilon_{x,t} \quad [3]$$

Modele te różnią się więc specyfikacją części systematycznej. Pierwszy z nich jest uzupełnieniem pierwotnej wersji modelu Lee-Cartera o efekt kohortowy w postaci $\beta_x \gamma_{t-x}$, gdzie współczynniki wektora β reprezentują strukturę wiekową, zaś wyrazy szeregu czasowego γ_t reprezentują zmienność w ciągu kolejnych kohort (a nie lat kalendarzowych, jak szereg κ_t). Taka specyfikacja pojawia się m.in. w pracach Renshaw'a i Habermana z 2003 roku i późniejszych. Model drugi zakłada współistnienie dwóch efektów kalendarzowych (i brak efektów kohortowych), zaś model trzeci – odwrotnie – wyłącznie dwa efekty kohortowe.

Na parametry modelu 1 narzucono najbardziej standardowy zestaw ograniczeń. Założono więc że sumy wyrazów obu szeregów $\sum_{t=1971}^{2018} \kappa_t$ oraz $\sum_{t=1926}^{2003} \gamma_t$ równe są zero, przyjęto również założenia normujące skalę elementów wektora $\beta1$ oraz $\beta2$.

Modele 2 i 3 w odróżnieniu od modelu 1, skonstruowane są w taki sposób, aby za ich pomocą można było przeprowadzić dekompozycję zmian dzietności na dwa odrębne efekty. W modelu 2 przyjęto, iż jedynie suma wyrazów szeregu $\kappa2_t$ wynosi zero, suma elementów wektora $\beta2$ też wynosi zero, zaś suma elementów wektora $\beta1$ wynosi jeden. Jednoznaczność identyfikacji parametrów wymaga dodatkowego założenia normującego skalę elementów wektora $\beta2$. Taki zestaw ograniczeń uzupełniony przez usunięcie z modelu wektora α gwarantuje jednoznaczną interpretację dwóch efektów. Pierwszy efekt kalendarzowy odpowiada mianowicie w pełni i wyłącznie za zmiany dzietności całkowitej, zaś drugi efekt kalendarzowy odpowiada w pełni i wyłącznie za zmiany struktury wiekowej dzietności. Parametr $\kappa1_t$ możemy jednoznacznie interpretować jako przekrojowy współczynnik dzietności w roku t . Natomiast przyrosty wyrazów szeregu $\kappa2_t$ odpowiadają za zmiany, jakim podlega struktura wiekowa dzietności, kiedy wzrostowi dzietności w jednych grupach wiekowych towarzyszą kompensujące spadki dzietności w innych grupach. Dekompozycja zmian dzietności na zmiany jej poziomu i zmiany jej struktury wiekowej jest klarowna i pełna, z jednym zastrzeżeniem – zmiany które nie dają się w ten klarowny sposób zdekomponować zawarte są w składniku indywidualnym $\varepsilon_{x,t}$, i im skala jego wahań jest większa, tym mniej doskonała jest przeprowadzona dekompozycja. Analogiczne ograniczenia narzucono na

parametry modelu 3, z tą jednak różnicą, że tym razem zarówno zmiany dzietności ogółem jak i zmiany jej struktury wiekowej model ten wiąże z różnicami pomiędzy kolejnymi kohortami kobiet, a nie z kolejnymi latami kalendarzowymi. W szczególności estymowane w tym modelu elementy szeregu $\gamma 1_t$ interpretujemy jako oszacowania kohortowych współczynników dzietności. Także i w tym wypadku mamy do czynienia z dekompozycją tym bardziej doskonałą, im mniejsza jest skala wahań nie dających się zdekomponować składników losowych. Pewne dodatkowe źródło niedoskonałości dekompozycji w modelu 3 wiąże się jednak z faktem, iż dla kohort najstarszych i najmłodszych mamy niekompletne obserwacje. Dokładniej temu problemowi przyjrzymy się nieco dalej.

Dociekliwy Czytelnik zapewne już się zorientował, dlaczego przedmiotem modelowania w naszym przypadku są oryginalne wartości współczynników $q_{x,t}$ oraz $q1_{x,t}$, a nie – jak w tradycyjnych badaniach śmiertelności – ich logarytmy. Powód jest prosty: jeśli dokonamy wstępnego zlogarytmowania naszych danych, cały urok interpretacji wyników uzyskanych w modelach 2 i 3 w jednej chwili prysnie. Uwaga ta oczywiście nie dotyczy modelu 1, w którym każdy ze współistniejących efektów – tak kalendarzowy jak i kohortowy – reprezentuje potencjalnie oba rodzaje zmian – zarówno zmiany dzietności ogólnej, jak i zmiany jej struktury wiekowej.

Wypada wreszcie wyartykułować oczekiwania i możliwe interpretacje, które skłonni bylibyśmy nadać wynikom, gdyby się na przykład okazało że jeden z tych modeli jest wyraźnie lepiej lub wyraźnie gorzej dopasowany do danych, niż pozostałe. Jednym słowem, na jakie pytania badawcze zaproponowane modele miałyby udzielać odpowiedzi. Gdyby model 1 okazał się wyraźnie lepszy od modeli 2 i 3, oznaczałoby to że w rzeczywistości występują silne efekty zarówno kohortowe, jak i kalendarzowe, i nakładanie się ich na siebie utrudnia dokonanie prostej i klarownej dekompozycji zmian na zmiany dzietności ogólnej i zmiany jej struktury wiekowej. Z kolei porównanie wyników dopasowania modeli 2 i 3 może pomóc odpowiedzieć na ciekawe pytanie – czy dramatyczny spadek dzietności, który dokonał się w środkowej części badanego okresu, miał charakter przede wszystkim zmiany pokoleniowej, czy też raczej wiązać go możemy ze zjawiskami, które wpływały w podobny sposób na aktywność prokreacyjną kobiet w różnym wieku w tych samych latach kalendarzowych. Akceptowalne wyniki oszacowań modeli 2 i/lub 3 zarówno w odniesieniu do modelowania współczynników $q_{x,t}$ jak i $q1_{x,t}$, stwarzają możliwości udzielenia odpowiedzi na jeszcze jedno ciekawe pytanie. Wyznaczywszy bowiem współczynnik dzietności ogółem (ile finalnie dzieci przypada na jedną kobietę) oraz prawdopodobieństwo posiadania dziecka (dopełnienie odsetka kobiet finalnie bezdzietnych), możemy pierwszy wskaźnik podzielić przez drugi, i otrzymać w rezultacie

finalną liczbę dzieci przypadającą na kobietę, która w ogóle dzieci posiada. Krótko mówiąc, nasze pytania badawcze:

- dotyczą – w zakresie materii badanej – mechanizmu spadku dzietności, z jakim mieliśmy do czynienia w Polsce w badanym okresie, w szczególności zaś odpowiedzi na pytania:
 - na ile ostateczne efekty można przypisać spadkowi liczby dzieci przypadających na jedną kobietę która w ogóle dzieci posiada, a na ile wzrostowi odsetka kobiet bezdzietnych
 - jak na zachodzące zmiany dzietności nakładał się proces przesuwania się wieku rodzenia dzieci
 - jak ten proces przebiegał w czasie, w szczególności zaś czy miał w większej mierze charakter zmiany pokoleniowej, czy też raczej dokonywał się w rytmie kolejnych lat kalendarzowych
- dotyczą – w zakresie metodologii – tego, czy zaproponowane modele wraz ze wszystkimi szczegółami estymacji ich parametrów mogą z sukcesem posłużyć do odpowiedzi na tego rodzaju pytania merytoryczne, jakie w badaniu sobie postawiliśmy.

4. Modyfikacje pierwotnej postaci modelu Lee-Cartera – składnik losowy

Składniki losowe w modelach typu Lee-Cartera mogą zarówno wykazywać heteroscedastyczność, jak i autokorelacje. Pierwsze z tych zjawisk przynajmniej pośrednio bywa uwzględniane za pośrednictwem założeń o rozkładzie składników losowych. Przykładowo, stała wariancja odchyłeń w modelu na logarytmach oznacza de facto, iż w odniesieniu do surowych danych zakładamy stały stosunek wariancji do kwadratu wartości oczekiwanej naszych obserwacji. Z kolei założenie o rozkładzie Poissona dla surowych danych oznacza, iż przyjmujemy stałość stosunku wariancji odchyłeń do wartości oczekiwanej w pierwszej potęgze.

Co ciekawe, drugie z tych zjawisk praktycznie nie ma literatury (jeśli pominąć wzmianki o referatach na konferencjach prezentowanych przez autora, Wojciecha Otto (por. ASMDA 2019, book of abstracts)). Tymczasem pierwszy rzut oka na reszty z dowolnego modelu tej klasy, zastosowanego do danych o śmiertelności czy dzietności pochodzących ze statystyk narodowych, pozwala dostrzec „duże plamy” sąsiadujących ze sobą reszt dodatnich lub reszt ujemnych.

Efekt „dużych plam” można interpretować dwojako. Zapewne korelacje w jakiejś części są skutkiem charakteru błędów pomiaru popełnianych w statystykach ludności (na przykład na

skutek stopniowo narastającego błędu szacunków dokonywanych w kolejnych latach od ostatniego spisu powszechnego). W pewnej jednak części (wydaje się że w przeważającej) są one refleksem rzeczywistych prawidłowości, których systematyczna część modelu nie pokrywa. W szczególności każdy z trzech zaproponowanych modeli zakłada, iż zmiany w czasie mają charakter niejako globalny, oddziałując (w proporcjach zadanych stałą strukturą wektora β) na wszystkie grupy wiekowe kobiet czy też wszystkie lata życia (od 15 do 45) danej kohorty. Tymczasem można się spodziewać występowania także zmian w aktywności prokreacyjnej o charakterze „lokalnym”, dotyczącym kilku sąsiadujących lat obserwacji oraz grup kobiet w zbliżonym wieku. Ogólnie chodzi więc o występowanie silnych autokorelacji $corr(\varepsilon_{x,t}, \varepsilon_{x-i,t-j})$, gdy wskaźniki odstepu i oraz j są niewielkie.

W szczególnych przypadkach autokorelacom tym można nadać konkretniej brzmiącą interpretację, która też może być pomocna przy diagnostyce poprawności specyfikacji systematycznej części modelu. Jeśli na przykład obserwujemy (oczywiście za pośrednictwem obserwacji reszt) silne autokorelacje $corr(\varepsilon_{x,t}, \varepsilon_{x-k,t-k})$, możemy je przypisać zjawisku efektów kohortowych. Jeśli autokorelacja dla $k = 1$ jest silna, ale wraz z rosnącym odstępem k jej siła szybko zanika, możemy mówić konkretniej o zjawisku lokalnych efektów kohortowych. Jeśli jednak kolejne autokorelacje są silne, dodatnie, i szybko nie zanikają, możemy podejrzewać istnienie globalnych efektów kohortowych, nie uwzględnionych w systematycznej części modelu. Błąd zbyt ubogiej specyfikacji może polegać na tym, iż występujących w rzeczywistości globalnych efektów kohortowych w ogóle w modelu nie uwzględniliśmy (casus modelu 2), lub też że uwzględniliśmy pojedynczy efekt kohortowy, a powinniśmy uwzględnić dwa takie efekty (casus modelu 1). Podobnie może wyglądać nasza interpretacja, jeśli zauważymy silne dodatnie autokorelacje $corr(\varepsilon_{x,t}, \varepsilon_{x-k,t})$. Oznaczają one występowanie lokalnych efektów kalendarzowych, ale także – o ile korelacje nie znikają szybko wraz ze wzrostem odstepu k – mogą sygnalizować zbyt ubogą specyfikację części systematycznej, w której uwzględniliśmy zbyt małą liczbę globalnych efektów kalendarzowych.

Ostatecznie we wszystkich trzech modelach i ich dwóch wersjach dla danych $q_{x,t}$ oraz $q1_{x,t}$, wykorzystaliśmy jednolitą specyfikację autoregresyjnego modelu składników losowych, z funkcją autoregresji rozciągniętą na dwa wymiary (poprzednie lata w zbliżonych grupach wiekowych, oraz młodsze grupy wiekowe w tym samym roku):

$$\varepsilon_{x,t} = \sum_{i=-3}^3 \sum_{j=1}^3 \rho_{i,j} \varepsilon_{x-i,t-j} + \sum_{i=1}^3 \rho_{i,0} \varepsilon_{x-i,t} + \eta_{x,t}, \quad [4]$$

gdzie założenie o niezależności utrzymujemy już jedynie dla innowacji $\eta_{x,t}$. Stosując ten wzór dla obserwacji na brzegach obszaru próbki przyjęliśmy założenie techniczne, iż składniki losowe $\varepsilon_{x,t} = 0$ we wszystkich przypadkach gdy wiek x jest mniejszy niż 15 lub większy niż 45, jak również gdy $t < 1971$. Wobec tego estymacja funkcji autoregresji po części opiera się na obserwacjach ułomnych. Dotyczy to obserwacji dla wieku 15, 16 i 17 lat oraz 43, 44 i 45 lat, jak również obserwacji z lat 1071, 1972, oraz 1973. Przyjęte z góry założenie, iż ignorujemy zależności składników losowych oddalonych od siebie o więcej niż trzy lata (kalendarzowe, bądź wiekowe) wynikało właśnie z przekonania, iż w przeciwnym przypadku udział w próbie obserwacji „ułamnych” byłby zbyt wielki.

Poza niezależnością wzajemną innowacji $\eta_{x,t}$ przyjęliśmy także, iż mają one rozkłady normalne o zerowej wartości oczekiwanej, oraz wariancjach stałych w przekroju lat kalendarzowych, zależnych jednak od wieku. Nie znajdując empirycznego potwierdzenia żadnego z typowych założeń dotyczących tego, jak owe wariancje zależą od wartości oczekiwanych, w każdym przypadku dokonywaliśmy ich oszacowania, zakładając jedynie iż kolejne wartości odchyłeń standardowych dla wieku $x = 15, 16, 17, \dots, 45$ leżą na linii łamanej złożonej z trzech prostych odcinków, co wymagało estymacji 6 parametrów (po 2 na każdy odcinek prostej linii łamanej).

5. Estymacja parametrów

Estymację parametrów zaproponowanych modeli przeprowadziliśmy ważoną metodą największej wiarygodności, obciążając funkcję wiarygodności odpowiednio skonstruowaną funkcją kary. Dokładniej, metodami iteracyjnymi znajdowaliśmy minimum następującej funkcji:

$$f(\alpha, \beta_1, \beta_2, \kappa_1, \kappa_2, \gamma_1, \gamma_2, \rho, \nu) = \sum_{x,t} \left\{ \ln \sigma_x^2 + (\eta_{x,t} / \sigma_x)^2 \right\} \omega_x + \text{Kara}(\beta_1, \beta_2, \gamma_1, \gamma_2, \rho) \quad [5]$$

- gdzie pierwszy składnik to pomnożony przez (-2) logarytm ważonej funkcji wiarygodności z pominięciem stałej,
- ω_x to wagi nadane obserwacjom dotyczącym poszczególnych grup wiekowych,
- ν to parametry łamanej, na której leżą wartości odchyłeń standardowych σ_x rozkładu innowacji $\eta_{x,t}$,
- „Kara” to funkcja kary nałożona na współczynniki wektorów struktury wiekowej, elementy wektora dynamiki efektów kohortowych, oraz współczynniki funkcji autoregresji składników losowych ukryte razem za symbolem ρ ,

- $\eta_{x,t} = \varepsilon_{x,t} - \sum_{i=-3}^3 \sum_{j=1}^3 \rho_{i,j} \varepsilon_{x-i,t-j} - \sum_{i=1}^3 \rho_{i,0} \varepsilon_{x-i,t}$ to innowacje
- $\varepsilon_{x,t} = q_{x,t} - \hat{q}_{x,t}$, gdzie współczynnik $\hat{q}_{x,t}$ oznacza wyliczoną z systematycznej części modelu wartość współczynnika $q_{x,t}$, zgodnie ze wzorem [1], [2] bądź [3], zaś
- $\varepsilon_{x,t} = q_{1x,t} - \widehat{q}_{1x,t}$, jest analogiczną wielkością w modelu urodzeń pierwszego dziecka
- na parametry modeli narzucone są odpowiednie ograniczenia normujące

Oczywiście zadanie maksymalizacji rozwiązywane było dla każdego z trzech podstawowych modeli przy dodatkowych oczywistych ograniczeniach: narzucono ograniczenia zerowe na elementy wektorów κ_2 oraz γ_2 w modelu [1], na elementy wektorów $\alpha, \gamma_1, \gamma_2$ w modelu [2], oraz na elementy wektorów $\alpha, \kappa_1, \kappa_2$ w modelu [3].

W rzeczywistości dokonaliśmy 12 różnych oszacowań, ponieważ dla każdego z 6 dotychczas omówionych modeli dokonaliśmy estymacji parametrów dopuszczając niezerowe wartości współczynników autoregresji ρ , oraz estymacji z narzuconymi na wszystkie elementy zbioru parametrów ρ ograniczeniami zerowymi.

Głębszego wyjaśnienia wymaga użycie przez nas wag ω_x . Typowe okoliczności, kiedy ważymy poszczególne obserwacje, związane są w gruncie rzeczy z występującą w modelach heteroscedastycznością składników losowych. Tymczasem my heteroscedastyczność już uwzględniliśmy bezpośrednio w postaci zróżnicowanych parametrów σ_x^2 , zaś wagi ω_x wprowadziliśmy z zupełnie innego powodu. Aby te powody wyjaśnić, przyjrzyjmy się najpierw pewnym dwóm skrajnym podejściom do budowania naszych modeli.

- Pierwsze podejście prowadzi do zastosowania zwykłej metody największej wiarygodności, gdzie zróżnicowane traktowanie poszczególnych obserwacji wynika wyłącznie ze skali wariancji składników losowych (wtedy w naszym modelu przyjęlibyśmy wszystkie wagi ω_x równe jeden). Takie podejście znajduje uzasadnienie w pewnych ogólniejszych pryncypiach wnioskowania statystycznego. Zasadzają się one na przekonaniu, że nasza specyfikacja systematycznej części modelu jest w pełni poprawna, i że istnieje obiektywnie coś takiego, jak prawdziwe wartości parametrów modelu. My tylko szukamy informacji na ich temat, ukrytej za naszymi danymi z próbki, które ku naszemu utrapieniu obarczone są zaburzeniami losowymi. Wobec tego jasne jest, że obserwacje obarczone zaburzeniem losowym o małej wariancji zawierają więcej informacji o nieznanymi parametrach, niż obserwacje obarczone zaburzeniem losowym o dużej wariancji.
- Drugie podejście prowadzi do zignorowania ewentualnego zróżnicowania wariancji reszt. Głównym celem badania może być bowiem skonstruowanie takiego indeksu dzietności, który najlepiej jak to możliwe odzwierciedli zmiany w czasie (lub w przekroju kolejnych

kohort), jakim całkowita dzietność podlega. Nie chcemy więc, aby tendencje obserwowane w grupach wiekowych, które mają znikomy udział w ogólnej dzietności, zanadto wpłynęły na nasz indeks. Ma on odzwierciedlać zmiany dzietności reprezentatywne przede wszystkim dla tych grup wiekowych, które tę dzietność w znaczącej mierze kształtują. Do takiej narracji znacznie mniej pasuje opowieść o obiektywnie istniejących acz nieznanach parametrach modelu, których wartości chcemy z jak największą precyzją oszacować. Na miejscu jest tu raczej pogląd, że nie ma dobrych modeli rzeczywistości, a jedynie mniej złe, i – mając zawsze na uwadze postawione pytania badawcze – takich właśnie „mniej złych” modeli powinniśmy szukać.

Nasz pomysł na uwzględnienie zróżnicowanych wag niezależnie od tego, że uwzględniamy zróżnicowanie wariacji składników losowych, bierze się z przekonania, że za każdą z przedstawionych powyżej skrajności przemawiają rozsądne argumenty. Przyznanie się bowiem do tego, że pojęcia lokalnych i globalnych efektów kalendarzowych i kohortowych to bardziej twory naszego umysłu niż obiektywne byty, nie przeszkadza w ich używaniu w opisie i analizie rzeczywistości. Równocześnie uzasadniona jest obawa, że bez przydania odpowiednio niskich wag obserwacjom w tych grupach wiekowych, które w znikomym stopniu przyczyniają się do ogólnej dzietności, możemy otrzymać taki indeks dzietności, który w bardzo niedoskonały sposób odzwierciedla tempo tych zmian, które naprawdę mają znaczenie.

Kompromis, jaki przyjęliśmy, to oparcie się na estymacji metodą największej wiarygodności, uzupełnioną jednak o wagi. Stosownie do zadania, jakie mają spełnić wagi, przyjęto w badaniu takie ich wartości, które odzwierciedlają udział średniej (w przekroju lat obserwacji) dzietności w danej grupie wiekowej w dzietności ogółem. Dokładniej, w tych grupach wiekowych, w których wskaźnik udziału \bar{q}_x w sumie $\sum_{k=15}^{45} \bar{q}_k$ wyrażony w punktach procentowych przekracza wartość 1, przyjęliśmy wagę równą 1. Jeśli wspomniany wskaźnik był mniejszy od jedynki, przyjęliśmy wagę w wysokości tego wskaźnika zaokrąglonego z dokładnością do 0,1, a jeśli miałyby to zwrócić wagę zero, zaokrąglaliśmy z dokładnością do 0,01. W rezultacie otrzymaliśmy ciągi wag $\{\omega_{15}, \omega_{16}, \dots, \omega_{45}\}$ o postaci:

- $\left\{ \frac{1}{10}, \frac{3}{10}, \frac{8}{10}, 1, 1, \dots, 1, \frac{7}{10}, \frac{5}{10}, \frac{3}{10}, \frac{2}{10}, \frac{1}{10}, \frac{1}{10} \right\}$ dla modeli dzietności ogółem, i analogicznie:
- $\left\{ \frac{2}{10}, \frac{6}{10}, 1, 1, \dots, 1, \frac{9}{10}, \frac{7}{10}, \frac{5}{10}, \frac{4}{10}, \frac{3}{10}, \frac{2}{10}, \frac{2}{10}, \frac{1}{10}, \frac{15}{100}, \frac{1}{100}, \frac{1}{100} \right\}$ dla urodzeń 1-szego dziecka.

Wyraźne różnice w strukturze wag w pierwszym i drugim przypadku wynikają z prostego faktu, że aby urodzić dziecko drugie, trzecie itd., z reguły trzeba urodzić najpierw dziecko pierwsze.

Wyjaśnienia wymaga także funkcja kary występująca we wzorze [5]. W istocie jest ona sumą pięciu składników, dotyczących osobno grup parametrów $\beta_1, \beta_2, \gamma_1, \gamma_2, \rho$. Oczywiście w modelu [1] składnik odpowiadający za γ_2 nie występuje, w modelu [2] nie występują składniki odpowiadające za γ_1, γ_2 , a w modelach w których założono brak autokorelacji nie występuje składnik odpowiadający za współczynniki autoregresji ρ .

Stosunkowo proste jest wyjaśnienie dwóch pierwszych składników. Związane są one z założeniem, iż elementy wektorów struktury wiekowej powinny mieć względnie „gładki” przebieg. Zabieg dobrze znany w modelach stosowanych do badania ewolucji śmiertelności, w naszych modelach ma postać zastosowanej do wektorów β_1 i β_2 funkcji kary o postaci:

- $10000 \sum_{x=15}^{42} (\beta_x - 3\beta_{x+1} + 3\beta_{x+2} - \beta_{x+3})^2$.

Sens tego wzoru jest taki, że karane jest zbytne odstępstwo lokalne szeregu współczynników każdego z wektorów od przebiegu takiego, w którym współczynniki te leżą na tej samej paraboli. Wartość parametru skali kary (10 tysięcy) została tak dobrana, że we wszystkich szacowanych modelach wartość funkcji kary okazała się znacznie mniejsza od dziesięciu. W istocie wyniki nawet bez obarczenia estymacji funkcją kary wykazują akceptowalny poziom „gładkości”, podobny do tak samo rozumianej „gładkości” przebiegu elementów wektora α , estymowanego w niewymuszony sposób.

Znacznie bardziej wyszukana jest konstrukcja funkcji kary dotyczącej szeregów dynamiki efektów kohortowych. Chcielibyśmy aby estymatory elementów tych szeregów, podobnie jak szeregów dynamiki efektów kalendarzowych, odzwierciedlały możliwie najlepiej rzeczywiste prawidłowości, a nie nasze założenia *a priori*. Problem jednak tkwi w tym, że w przypadku efektów kohortowych kompletne dane dotyczą tylko kohort 1956, 1957, ..., 1973 w przypadku danych o dzietności całkowitej, zaś 1957, ..., 1972 w przypadku urodzeń pierwszego dziecka. Im kohorta starsza od kohorty 1956, lub młodsza od kohorty 1972, tym większej liczby obserwacji z końca lub z początku listy grup wiekowych brakuje. Wychodząc więc z założenia, że im mniej mamy obserwacji, tym mniej podstaw do twierdzenia, że kolejne wyrazy szeregu różnią się od siebie nawzajem, zastosowaliśmy funkcję kary, która dla danych dotyczących dzietności ogółem ma postać:

- $20 \sum_{t=1926}^{2002} w_t (\gamma_t - \gamma_{t+1})^2$, gdzie: [6]
 - $w_t = 0$ dla $t \in \{1955, 1956, \dots, 1973\}$,
 - $w_t = (\sqrt{2})^k$ dla $t = 1954 - k$ oraz $t = 1974 + k$,

z odpowiednim przesunięciem w wersji dla danych o urodzeniach pierwszego dziecka. Sądzimy, że wagi rosnące w miarę zbliżania się do obu krańców szeregu γ_t są odpowiednie,

choć oczywiście nie jesteśmy w stanie obronić arbitralnego założenia, że wzrost ten ma charakter postępu geometrycznego, i że iloraz postępu wynosi akurat $\sqrt{2}$. Obronę przyjętego poziomu parametru skali równego 20 przeprowadzimy dalej, w podrozdziale poświęconym prezentacji wyników estymacji.

Jeszcze bardziej wyszukaną postać ma składnik funkcji kary dotyczący zbioru współczynników autoregresji. Dołączenie tego składnika do maksymalizowanej funkcji motywowane jest chęcią uniknięcia problemu niestacjonarności. Chodzi o problem, który w klasycznym przypadku, gdy szereg czasowy indeksowany jest w jednym wymiarze, związany jest z ewentualnym wystąpieniem niestacjonarnej dynamiki odchyłeń losowych. Niestacjonarny proces odchyłeń może wygenerować narastające odchylenia szeregu od jakiegokolwiek trajektorii, którą chcielibyśmy interpretować jako „systematyczną” część modelu szeregu czasowego, tym samym stawiając w wątpliwość samo istnienie tejże „systematycznej” części. Istnieją jednak gotowe narzędzia, którymi w analizie możemy się posłużyć (np. estymacja największego co do modułu pierwiastka charakterystycznego w modelu autoregresyjnym, testowanie hipotezy że równy jest on jedynce). W naszych modelach mamy do czynienia z danymi indeksowanymi w dwóch wymiarach (u nas jest to czas kalendarzowy i wiek, w modelach przestrzennych może być to np. szerokość i długość geograficzna). W tym przypadku można mieć obawy o podobnym charakterze, ale gotowych narzędzi analitycznych nie ma. Dokładniej, chcielibyśmy aby funkcja autokorelacji uchwyciła efekty sąsiedztwa obserwacji (przede wszystkim lokalne efekty kohortowe i kalendarzowe, o ile takie w procesie generującym dane występują). Nie chcemy jednak, aby funkcja autoregresji reszt przejęła rolę wyjaśnienia efektów globalnych, skrzywiając przy tym oceny parametrów części systematycznej modelu, o ile została ona oczywiście prawidłowo wyspecyfikowana. Nie jest to problem wydumany. W modelu, w którym część systematyczna składa się wyłącznie z jednego efektu kalendarzowego (jak w pierwotnym modelu Lee-Cartera), uzupełnionym o model autoregresji składników losowych, swobodna estymacja parametrów ρ zwraca proces przejawiający ewidentne symptomy niestacjonarności. Chcielibyśmy jednak, aby wzbogacenie specyfikacji o istotne dodatkowe efekty globalne natychmiast usunęło symptomy niestacjonarności. Aby sobie to zapewnić, powinniśmy (choćby w sposób niedoskonały) zmierzyć odległość danego modelu od modelu niestacjonarnego, i wprowadzić do procesu estymacji funkcję kary rosnącą, gdy owa odległość maleje.

Funkcję kary, która zachowuje się w sposób opisany powyżej, skonstruowaliśmy w oparciu o ciąg macierzy, które tworzymy wykorzystując wzór rekurencyjny:

$$\varepsilon_{x,t}(k+1) = \sum_{i=-3}^3 \sum_{j=1}^3 \rho_{i,j} \varepsilon_{x-i,t-j}(k) + \sum_{i=1}^3 \rho_{i,0} \varepsilon_{x-i,t}(k), \text{ gdzie:}$$

- $\varepsilon_{x,t}$ to elementy a wiersza x i kolumny t macierzy o wymiarach 31×31 ,
- $(k+1)$ to numer macierzy wyliczonej w $k+1$ -szym kroku
- dokonując wyliczenia w danym kroku przyjmujemy po prawej stronie wyliczone w poprzednim kroku wartości macierzy, uzupełnione o zerowe wartości elementów o indeksach spoza zakresu, a więc dla x lub t równych $-2, -1, 0$ oraz x równych $32, 33, 34$.
- Startowe wartości macierzy (zaburzenie początkowe) przyjęto postaci $\varepsilon_{x,t}(0) = 1$ dla wszystkich (x, t) z zakresu od 1 do 31
- efekt zaburzenia początkowego po k krokach mierzymy jako $S_k = \sum_{x=1}^{31} \sum_{t=1}^{31} \varepsilon_{x,t}^2(k)$
- naszą miarą, której wzrost sygnalizuje zbliżanie się do modelu niestacjonarnego, jest suma $S = S_6 + S_7 + S_8 + S_{97} + S_{10}$
- wartości S przekraczające 5×31^2 wyraźnie wskazują niestacjonarność, wartości mniejsze od 31^2 stacjonarność, zaś wartości pomiędzy tymi liczbami nie są rozstrzygające
- funkcja kary dołączona do funkcji f we wzorze [5] to $\frac{1}{10} S$.

Oczywiście, podobnie jak w przypadku poprzednio omówionych składników funkcji kary, nie sposób obronić różne szczegóły sposobu mierzenia „bliskości do modelu niestacjonarnego”. Uznaliśmy jednak, że kierunkowo miara ta jest rozsądna.

Na koniec warto zauważyć, że efekty obarczenia minimalizowanej funkcji karami moglibyśmy osiągnąć równie dobrze stosując podejście bayesowskie, i w ramach tego podejścia odpowiednio konstruując szczegóły rozkładu a priori. Arbitralność szczegółów konstrukcji poszczególnych składników funkcji kary zastąpilibyśmy wtedy arbitralnością konstrukcji i doboru wartości parametrów rozkładu a priori. W pewnym stopniu byłaby to zmiana opakowania, w którym umieszczamy ten sam w istocie produkt.

6. Kryteria ocena porównawczej alternatywnych modeli

Dla celów porównania jakości dopasowania poszczególnych modeli do danych zastosowaliśmy kryterium Akaike poddane pewnym modyfikacjom. Często stosowana postać kryterium dana jest wzorem:

- $AIC = -2 \ln L + 2k$, gdzie k to liczba swobodnie szacowanych parametrów.

Samo kryterium, podobnie jak metoda największej wiarygodności, z którą kryterium jest związane, posiada pożądane własności asymptotyczne. Jeśli jednak liczba obserwacji jest skończona, a liczba szacowanych parametrów duża, można obronić modyfikację postaci:

- $AIC = -2\ln L + 2k \frac{n}{n-k}$, gdzie n to liczba obserwacji.

Uznaliśmy, że właśnie taka sytuacja w omawianych modelach zachodzi, skoro wraz z rosnącą liczbą lat obserwacji rośnie liczba estymowanych elementów wektorów $\kappa_1, \kappa_2, \gamma_1, \gamma_2$, a przy znacznym wzroście liczby lat obserwacji można łatwiej znaleźć uzasadnienie dołączenia do modelu jeszcze jednego efektu kalendarzowego czy kohortowego.

Potrzebę dodatkowej modyfikacji możemy sobie uświadomić, gdy zauważymy, że nie wszystkie obserwacje niosą pełnowartościową informację o estymowanych parametrach. Uwzględniliśmy to już przypisując niektórym obserwacjom wagę mniejszą od jedynki w funkcji wiarygodności. Skoro tak, to powinniśmy nie tylko zwykłą funkcję wiarygodności występującą w kryterium Akaike zastąpić ważoną funkcją wiarygodności, ale także uwzględnić to w funkcji kary, zastępując zwykłą liczbę obserwacji n sumą wag, które obserwacjom przypisaliśmy.

Skrótu AIC będziemy dalej używać mając na myśli kryterium Akaike poddane powyżej opisanym modyfikacjom. Natomiast skrót $AIC+Kary$ używać będziemy dla oznaczenia kryterium AIC powiększonemu dodatkowo o wartość funkcji kary występującej we wzorze [5]. Uzasadnienie do stosowania tej ostatniej modyfikacji wiąże się z tym, że już w procesie estymacji parametrów wyraziliśmy gotowość obniżenia funkcji wiarygodności, o ile zostanie to skompensowane odpowiednim obniżeniem funkcji kary. Jeżeli było to uzasadnione w procesie estymacji parametrów, to nie ma powodu, aby tego nie uwzględnić na tym etapie, kiedy wybranym parametrom modelu przypisujemy wartości zerowe lub narzucamy na nie inaczej wyspecyfikowane ograniczenia.

7. Generalna ocena i selekcja modeli do głębszej analizy

W Tabeli 1 zamieszczono oceny dla modeli w których założono zerowe współczynniki funkcji autoregresji składników losowych. Rzut oka na zawartość Tabeli 1 pozwala od razu na wyciągnięcie zasadniczego wniosku:

Tabela 1. Oceny modeli zakładających zerowe współczynniki autoregresji

	Dietność ogółem			Urodzenia pierwsze		
	Model [1]	Model [2]	Model [3]	Model [1]	Model [2]	Model [3]
AIC	-13808,5	-13138,7	-12379,7	-13046,6	-11704,4	-11961,5
$AIC+Kary$	-13729,4	-13134,9	-12374,4	-13026,1	-11696,9	-11943,7

Źródło: obliczenia własne

- pomysł konstrukcji modeli [2] i [3] nie sprawdził się, ponieważ ograniczenie części systematycznej wyłącznie do efektów kalendarzowych bądź wyłącznie kohortowych powoduje dramatyczne pogorszenie dopasowania, i to zarówno w przypadku modeli dzietności ogólnej, jak i modeli urodzeń pierwszego dziecka
- pozostałe porównania (modelu [2] z modelem [3], miar dopasowania AIC z miarami $AIC+Kary$) nie mają sensu, ponieważ w każdym z tych przekrojów różnice są małe, w porównaniu do różnic pomiędzy modelem [1] a pozostałymi.

Wydawać by się więc mogło, że pomysł konstrukcji modeli [2] i [3] po prostu się nie sprawdził, i próba w miarę precyzyjnej dekompozycji zmian dzietności w ramach tych modeli na zmiany ogólnego poziomu dzietności i zmiany jej struktury wiekowej będzie dalece niedoskonała z uwagi na nakładanie się na siebie istotnych efektów kalendarzowych i kohortowych.

Obraz diametralnie się jednak zmieni, gdy przyjrzymy się zamieszczonym w Tabeli 2 ocenom jakości dopasowania modeli uwzględniających autoregresję składników losowych.

Tabela 2. Oceny modeli uwzględniających mechanizm autoregresji składników losowych

	Dzietność ogółem			Urodzenia pierwsze		
	Model [1]	Model [2]	Model [3]	Model [1]	Model [2]	Model [3]
AIC	-15881,9	-15673,7	-15809,4	-14125,8	-14088,4	-14121,3
$AIC+Kary$	-15766,9	-15563,1	-15675,9	-14110,1	-14008,2	-14061,0

Źródło: obliczenia własne

Wnioski które wyciągnąć można z tych ocen, a także z ich porównania z ocenami z Tabeli 1, są następujące:

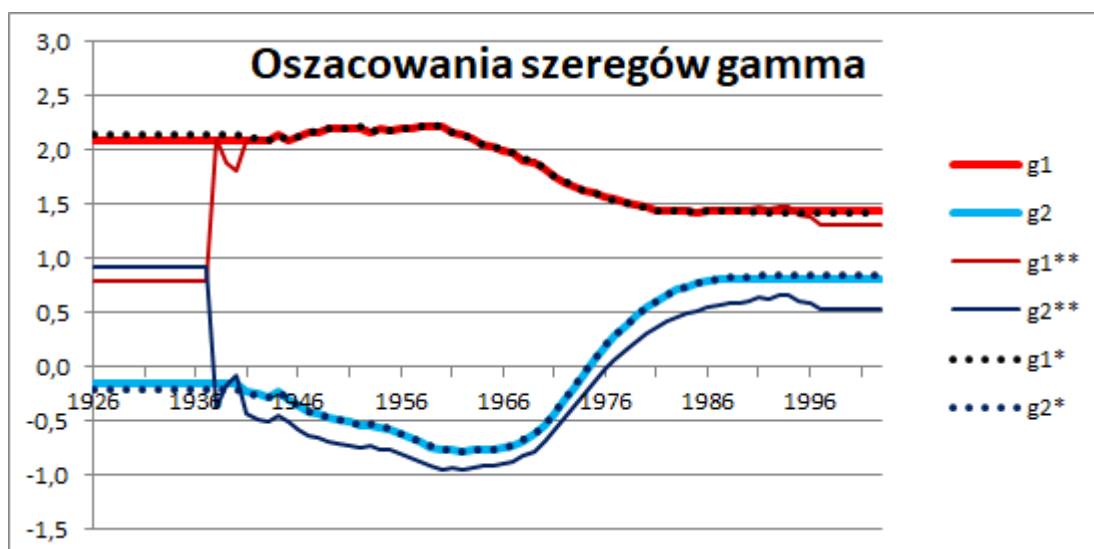
- Mechanizm generowania naszych danych zawiera silne lokalne i globalne efekty kalendarzowe i kohortowe. Wszystkie modele [1], [2] i [3] mają sens, choć występuje umiarkowana przewaga dopasowania modelu mieszanego [1] nad modelem z efektami kohortowymi [3], i tego ostatniego nad modelem z efektami kalendarzowymi [2]. Ten ranking jest taki sam dla modeli dzietności ogółem i dla modeli urodzeń pierwszych, choć w tym pierwszym przypadku różnice są wyraźniejsze.
- Autoregresja składników losowych jest na tyle silna, że pewnie w każdym z modeli obejmuje nie tylko lokalne efekty sąsiedztwa, ale także w przybliżony sposób odzwierciedla pominięte w części systematycznej tych modeli znaczące efekty globalne;
- Pomysł konstrukcji modeli [2] oraz [3] i użycia ich do analizy zmian dzietności można więc uznać ostatecznie za trafiony, z zastrzeżeniem iż analizę tę należy oprzeć na modelach

w wersji uwzględniającej autoregresję składników losowych. Sens tego postępowania można streścić następująco:

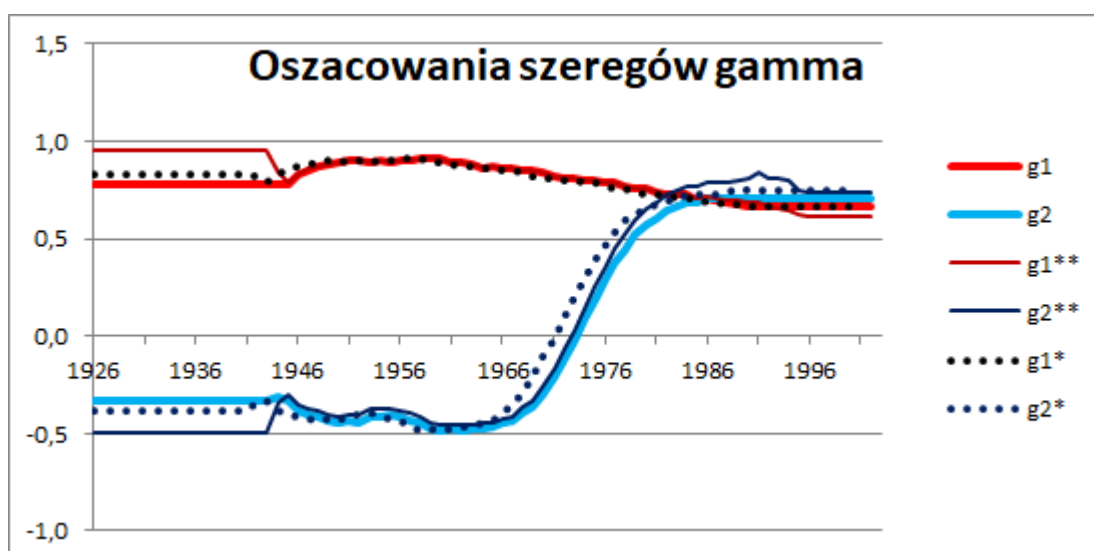
- Autoregresja składników losowych w modelu [2] pokrywa w znaczącym stopniu zmiany o charakterze kohortowym, pomagając w wydobyciu z naszych danych systematycznych efektów kalendarzowych w stosunkowo czystej postaci
- Autoregresja składników losowych w modelu [3] pokrywa w znacznym stopniu zmiany o charakterze kalendarzowym, pomagając w wydobyciu z naszych danych systematycznych efektów kohortowych w czystszej postaci

8. Metodyka estymacji i dodatkowe ograniczenia na parametry modelu

Często w celu ograniczenia błędów estymacji wprowadzamy ograniczenia a priori na wartości parametrów, niekoniecznie wierząc w ich zgodność z prawdą. Typowa motywacja to duża wariancja swobodnych ocen, w związku z czym jest sens, aby się jej pozbyć, nawet za cenę wprowadzenia obciążenia. Ma to szczególne uzasadnienie wtedy, kiedy duże wariancje wynikają z tego, iż wyjściowa specyfikacja modelu zawiera dużą liczbę parametrów. Jest to dokładnie nasz casus. Toteż tam, gdzie sugerowały to wyniki wstępnych ocen, takie ograniczenia wprowadziliśmy. Dotyczy to zarówno elementów szeregów $\kappa_1, \kappa_2, \gamma_1, \gamma_2$, jak i wektorów struktury wiekowej β_1, β_2 oraz elementów zbioru parametrów autoregresji ρ . Zachodzi jednak istotna różnica pomiędzy przypadkiem narzucania ograniczeń na elementy szeregów γ_1, γ_2 oraz pozostałymi przypadkami. W pozostałych przypadkach był wybór – mogliśmy takie ograniczenia narzucać lub nie. W przypadku szeregów γ_1, γ_2 wiadomo było z góry, że kilka pierwszych i kilka ostatnich wyrazów tych szeregów występuje jedynie w kilku obserwacjach, i wielka wahliwość ich ocen wynika z przypadkowości odchyłeń indywidualnych, będących składnikami tych obserwacji. Problem w szczególnym nasileniu dotyczy modelu [3], w którym występuje ponadto zjawisko znane ze zwykłej regresji liniowej jako problem współliniowości zmiennych objaśniających. Łatwo to pokazać na przykładzie, który zarazem zilustruje to, na czym problem polega, jak i to, jak z tym problemem sobie poradziliśmy. Rolę przykładu spełnią wyniki estymacji szeregów γ_1, γ_2 w modelu [3] w trzech wersjach ograniczeń. Wyniki dla przypadku badania dzietności całkowitej naniesiono na Wykresie 1, zaś wyniki dla badania urodzeń pierwszych dzieci na Wykresie 2.



Wykres 1. Wyniki oszacowań szeregów γ_1, γ_2 w trzech wariantach narzuconych ograniczeń dla modelu diety całkowitej. Źródło: opracowanie własne



Wykres 2. Wyniki oszacowań szeregów γ_1, γ_2 w trzech wariantach narzuconych ograniczeń dla przypadku urodzenia pierwszego dziecka. Źródło: opracowanie własne

- Pierwsze z tych oszacowań naniesione są na wykresach 1 oraz 2 pod nazwą $g1^{**}$ oraz $g2^{**}$. Uzyskane zostały one w wyniku estymacji parametrów modelu bez ograniczeń elastycznych i przy narzuceniu pewnej liczby ograniczeń sztywnych narzuconych na γ_1, γ_2 . Brak ograniczeń elastycznych oznacza, że z funkcji celu [5] usunięto ten składnik funkcji kary, który odpowiada szeregom γ_1, γ_2 , opisany wzorem [6]. Ograniczenia sztywne polegały na przyjęciu, iż dla kohort starszych od kohorty 1932 wartości czynników kohortowych są równe $\gamma_{1,1932}, \gamma_{2,1932}$ zaś dla kohort młodszych od kohorty 1997 są równe $\gamma_{1,1997}, \gamma_{2,1997}$. Okazało się, że otrzymane tak oceny γ_1 przyjmują dla najstarszych kohort absurdalne wartości. W modelu diety całkowitej za absurdalne uznaliśmy wartości

ujemne, zaś w modelu urodzeń pierwszego dziecka wartości ujemne oraz większe od jedynki. Krok po kroku przesuwaliśmy więc granicę, poniżej której wszystkie wartości γ_{1x} , γ_{2x} uznajemy za równe. Okazało się, że granicę dla kohort najstarszych granicę tę z kohorty 1932 przesunąć trzeba było do kohorty 1937 (dla dzietności całkowitej) oraz aż do kohorty 1943 w modelu urodzeń pierwszych. Pierwszy rzut oka na wykresy pokazuje, że wyeliminowane zostały oceny absurdalne, ale nadal występują oceny niewiarygodne. Przy tym występuje silna ujemna korelacja błędów estymacji tych szeregów dla kohort najstarszych, i dodatnia dla najmłodszych. Jest to prostą konsekwencją faktu, iż dla kohort najstarszych trudno odróżnić od siebie dwie sytuacje:

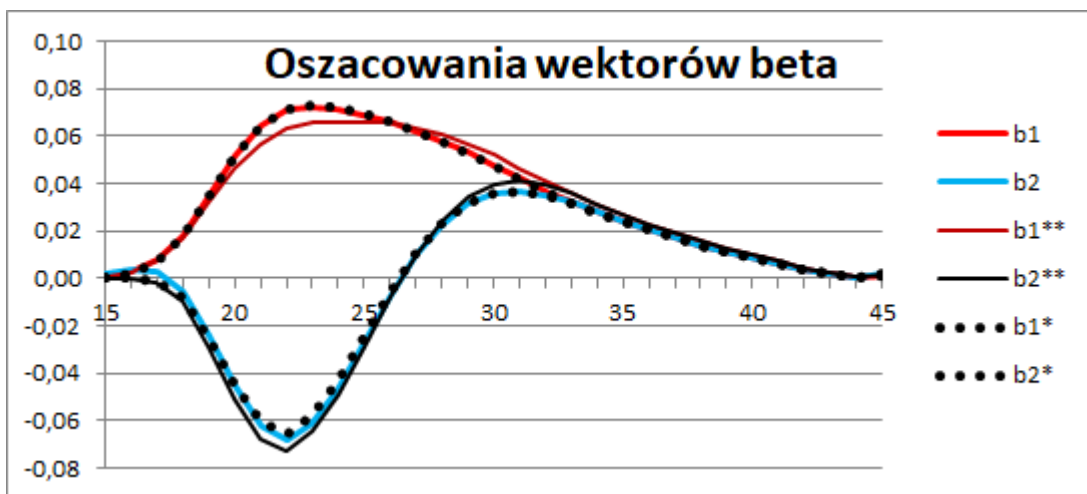
- względnie dużą częstość urodzeń późnych przy niskiej dzietności całkowitej,
- względnie niską częstość urodzeń późnych przy wysokiej dzietności całkowitej

Dla kohort najmłodszych, które obserwujemy jedynie w pierwszych latach aktywności prokreacyjnej, zależność błędów jest oczywiście odwrotna.

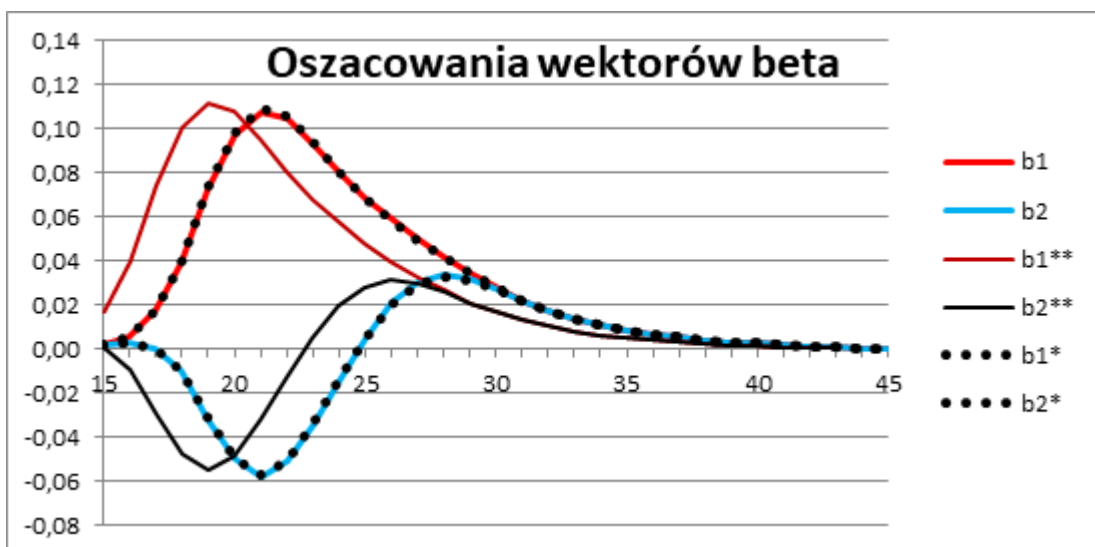
- Następne oszacowania naniesione są na wykresach 1 oraz 2 pod nazwą $g1^*$ oraz $g2^*$. Uzyskane one zostały przy uwzględnieniu tej samej listy ograniczeń sztywnych, co poprzednio, ale przy równoczesnym uwzględnieniu ograniczeń elastycznych w postaci [6].
- Finalną wersję oszacowań naniesiono na wykresy 1 oraz 2 pod nazwą $g1$ oraz $g2$. Powstały one w wyniku ewentualnego dołączania dodatkowych ograniczeń powyżej granicy wyznaczonej przez kohortę 1937 (lub 1943) oraz poniżej granicy wyznaczonej przez kohortę 1997, o ile w wyniku takiego przesunięcia malała wartość kryterium $AIC+Kary$. Przesuwanie tych granic tym razem dokonywane było niezależnie dla szeregów γ_1 oraz γ_2 .

Możemy teraz podsumować rolę, jaką w procesie estymacji spełniły ograniczenia elastyczne, wprowadzone za pośrednictwem funkcji kary. Przede wszystkim przywróciły one zgodny z intuicją związek, który w normalnych sytuacjach zachodzi pomiędzy tym, jak duże są różnice wyestymowanych przyrostów analizowanych szeregów z tym, na ile te różnice są istotne statystycznie. Taki związek zachodzi oczywiście w przypadku estymacji szeregów κ_1, κ_2 , gdzie problem niekompletnych obserwacji na krańcach nie występuje. Ułatwia to znakomicie intuicyjne zapanowanie nad procesem stopniowego rozszerzania listy przyjętych ograniczeń o nowe ograniczenia. Bardzo duże podobieństwo przebiegu szeregów $g1^*$ i $g2^*$ do szeregów $g1$ i $g2$ potwierdza, iż cel ten został osiągnięty, a co za tym idzie że występujący we wzorze [6] czynnik skalujący równy 20 nie jest zbyt mały. O tym, że wartość tego czynnika nie jest z kolei za duża, świadczy to, iż proces dołączania dodatkowych ograniczeń na ostatnim etapie, nakierowany na minimalizację kryterium $AIC+Kary$ przyniósł w efekcie także (łącznie, niekoniecznie w każdym kroku z osobna) zmniejszenie wartości kryterium AIC .

Wyniki estymacji w powyżej opisanych wariantach ograniczeń różniły się także pod względem oszacowań innych parametrów, mimo iż manipulacje ograniczeniami dotyczyły tylko szeregów γ_1, γ_2 . Na wykresach 3 i 4 naniesiono wyestymowane wartości elementów wektorów β_1, β_2 w ramach każdego z wariantów, zachowując zgodność z wykresami 1 i 2 w odniesieniu do konwencji oznaczeń. Widać wyraźnie, że ogromne błędy estymacji które obarczają wyniki pierwszego wariantu ograniczeń rozciągają się także na błędy estymacji wektorów β_1, β_2 . Tym razem bardziej dramatyczne są różnice w oszacowaniach dla modelu urodzeń pierwszego dziecka (wykres 4) niż dla diety całkowitej (wykres 3).



Wykres 3. Wyniki oszacowań wektorów β_1, β_2 w trzech wariantach ograniczeń narzucanych na szeregi γ_1, γ_2 w modelu diety całkowitej. Źródło: *opracowanie własne*



Wykres 4. Wyniki oszacowań wektorów β_1, β_2 w trzech wariantach ograniczeń narzucanych na szeregi γ_1, γ_2 w modelu urodzeń pierwszego dziecka. Źródło: *opracowanie własne*.

Przy okazji prezentacji wykresów 3 i 4 można łatwo wyjaśnić, w jaki sposób unormowaliśmy skalę elementów wektora β_2 , a w rezultacie tym samym skalę wahań szeregu czasowego γ_2 . Przyjeliśmy mianowicie iż wartość pewnego wybranego elementu wektora β_2 równa jest analogicznemu elementowi wektora β_1 . Narzucenie tego ograniczenia dla którejś ze starszych grup wiekowych pozwoliło dostrzec, iż proporcje elementów tych wektorów są bardzo podobne, i rozciągnąć (z korzyścią w postaci redukcji zarówno oceny $AIC+Kary$ jak i oceny AIC) to ograniczenie na wiele elementów dotyczących starszych grup wiekowych. Interpretacja wyników estymacji wektora β_2 jest bardzo klarowna. Ilustruje on przesuwanie się aktywności prokreacyjnej ku starszym grupom wiekowym, jeśli wyrazy szeregu γ_2 rosną, i ruch odwrotny, jeśli maleją. W obu przypadkach ten klarowny obraz jest nieco zaburzony w najmłodszych grupach wiekowych, ale skala tego zaburzenia jest znikoma.

W procesie narzucania ograniczeń zerowych na parametry autoregresji składników losowych nie stosowaliśmy żadnego szczególnego klucza, dorzucając sukcesywnie te ograniczenia zerowe, które pozwalały zredukować ocenę $AIC+Kary$. Oszacowane w finalnej wersji modelu [3] wartości parametrów autoregresji zamieszczone są w Tabelach 3 i 4. Pola czerwone w tych tabelach odpowiadają pozycji tego składnika losowego, który pełni rolę zmiennej objaśnianej w modelu autoregresji. Pola niebieskie odpowiadają efektom kalendarzowym, ceglaste efektom kohortowym, żółte pozostałym efektom sąsiedztwa, zaś białe tym współczynnikom, które w toku procedury uznano za nieistotnie różne od zera.

Tabela 3. Oceny współczynników autoregresji $\rho_{i,j}$ w modelu [3] dzietności całkowitej.

(i, j)	3	2	1	0
3		0,092	-0,122	-0,091
2	-0,081	-0,151		0,129
1	0,079		-0,211	0,791
0		0,051	0,201	
-1			0,210	
-2	0,053		0,126	
-3	-0,124		-0,080	

Źródło: obliczenia własne

Można zauważyć, że w składnikach losowych obu modeli tkwią silne efekty kalendarzowe o charakterze lokalnym, i być może globalnym (szczególnie w modelu dzietności całkowitej). Ponadto daje się zauważyć, że uwzględnienie w części systematycznej globalnych efektów kohortowych spotyka się w funkcji autokorelacji ze swoistą korektą w postaci lokalnej

ujemnej autokorelacji (niezbyt silnej). Względnie wyraźna jest także zależność bieżącego składnika losowego od swego rodzaju średniej wartości składników losowych charakteryzujących wiele zbliżonych grup wiekowych w poprzednim roku.

Tabela 4. Oceny współczynników autoregresji $\rho_{i,j}$ w modelu [3] urodzeń pierwszych.

(i, j)	3	2	1	0
3		0,053	-0,118	
2		-0,181		
1		0,113	-0,158	0,723
0			0,233	
-1			0,249	
-2			0,052	
-3	-0,072	-0,068		

Źródło: obliczenia własne

Zastosowana metodyka estymacji modelu [2] z efektami kalendarzowymi jest znacznie prostsza, przede wszystkim dlatego iż nie występuje tam specyficzny dla modelu kohortowego problem niekompletnych informacji na krańcach próbki. Nie ma więc szczególnych problemów metodologicznych, którym warto by było poświęcić tutaj uwagę. Warto jednak zaprezentować oceny współczynników autoregresji w związku z wyartykułowanymi wcześniej oczekiwaniami co do tego, iż efekty nieobecne w systematycznej części modelu powinny znaleźć swoje odzwierciedlenie w funkcji autoregresji składników losowych. W kolejnych tabelach 5 i 6 zamieszczono te oceny, dla przypadku modelu dzietności całkowitej i urodzeń pierwszych, odpowiednio.

Tabela 5. Oceny współczynników autoregresji $\rho_{i,j}$ w modelu [2] dzietności całkowitej.

(i, j)	3	2	1	0
3	0,142	-0,127	-0,088	
2	-0,152	0,187	-0,250	0,063
1	-0,103	-0,059	0,349	0,451
0			0,251	
-1	0,069		0,216	
-2			0,072	
-3			-0,123	

Źródło: obliczenia własne

Tabela 6. Oceny współczynników autoregresji $\rho_{i,j}$ w modelu [2] urodzeń pierwszych.

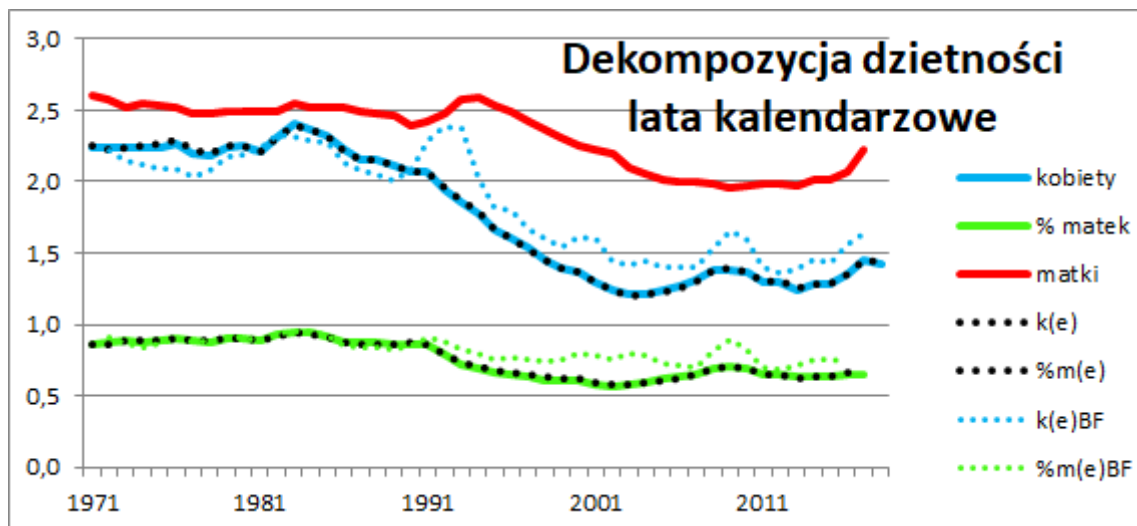
(i, j)	3	2	1	0
3	0,141	-0,076		-0,045
2	-0,123	0,058	-0,117	-0,068
1	-0,089		0,281	0,444
0			0,286	
-1			0,303	
-2				
-3		-0,122		

Źródło: obliczenia własne

I rzeczywiście wyniki są w miarę zgodne z oczekiwaniami. Efekty kalendarzowe nadal występują, ale są przede wszystkim słabsze niż w modelach [3], i sprawiają wrażenie efektów lokalnych. W obu modelach obecne są za to efekty kohortowe. Nie są one zbyt silne (moduły współczynników nie są bardzo wysokie), ale ze względu na ten sam znak współczynników reprezentują po części globalne efekty kohortowe, nieobecne w części systematycznej tych modeli.

9. Wyniki oszacowań modeli z efektami kalendarzowymi

Na wykresie 5 zamieszczone zostały wartości współczynników dzietności całkowitej, prawdopodobieństwa posiadania dzieci (w cyklu całego okresu aktywności prokreacyjnej), oraz współczynników dzietności matek, którym to terminem w określać będziemy w skrócie średnią liczbę dzieci przypadających na kobietę, która w ogóle dzieci posiada. Gruba linia niebieska oraz jasnozielona odpowiadają wyestymowanym wartościom elementów wektora κ_1 w modelu dzietności ogólnej oraz w modelu urodzeń pierwszego dziecka, odpowiednio. Widać że niemal idealnie pokrywają się one z liniami czarnych kropek, odpowiadającymi współczynnikom dzietności brutto (całkowitej i w odniesieniu do 1-go dziecka) wyliczonym na podstawie surowych danych. Linia czerwona zaznaczono dzietność matek. Widać, że zasadniczy spadek dzietności miał miejsce w ostatniej dekadzie XX wieku, zahaczając nieco na pierwsze lata wieku XXI. Pokrywa się to z latami głębokich przemian, związanych z przechodzeniem do gospodarki rynkowej. W następnych latach tempo zmian przygasa.



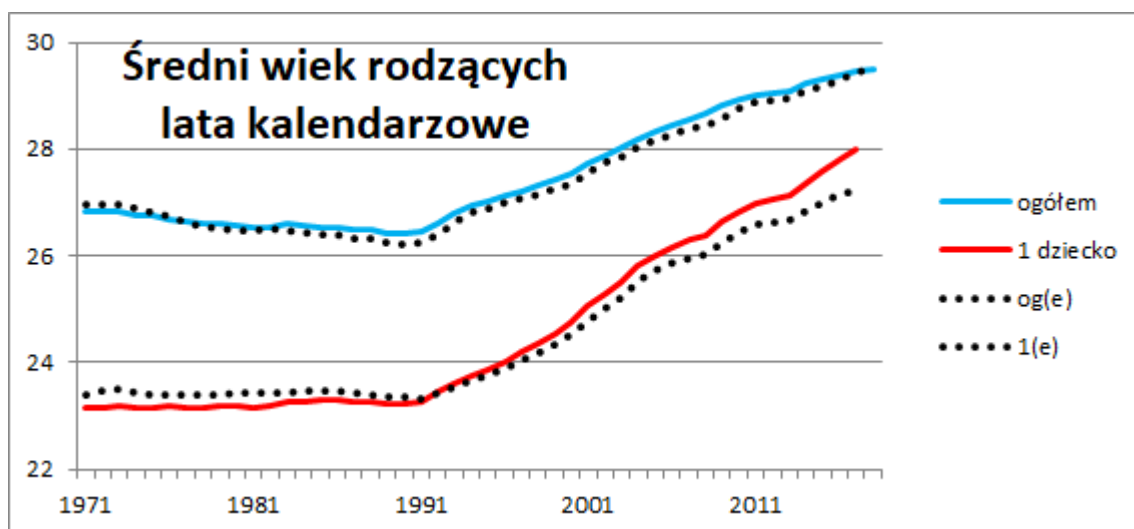
Wykres 5. Dzietność kobiet, prawdopodobieństwo posiadania dzieci, oraz dzietność matek. Źródło: opracowanie własne

Finalna skalę zmian można ocenić porównując na przykład dwa wybrane lata – jeden sprzed okresu zmian, a drugi po tym okresie. Na przykład porównując lata 1983 i 2003 obserwujemy spadek dzietności całkowitej z poziomu 2,40 do poziomu 1,21. W tym samym okresie odsetek kobiet rodzących co najmniej jedno dziecko spada z poziomu 94% do poziomu 58%. Wynikający z tego wskaźnik dzietności matek spada z poziomu 2,55 do poziomu 2,10.

W sumie oznacza to więc, że na spadek dzietności do 50% wyjściowego poziomu składa się spadek odsetka matek do 61% wyjściowego poziomu, oraz spadek dzietności matek do 82% wyjściowego poziomu. Widać więc, że problem depresji urodzeniowej to przede wszystkim wzrost odsetka kobiet bezdzietnych, a po wtóre dopiero, i w znacznie mniejszym stopniu, spadek liczby dzieci przypadających na kobietę, która w ogóle dzieci posiada. Należy przy tym przyznać, że dobór lat 1983 i 2003 został dokonany tak, aby wyniki porównania brzmiały najbardziej dramatycznie. Inny dobór lat wygenerowałby nieco mniej drastyczny, ale wciąż dramatyczny obraz zmian, do których doszło. Także efekty dekompozycji spadku dzietności przy innym doborze lat okazałyby się inne, ale w żadnym przypadku nie doszłoby do odwrócenia proporcji – zawsze efekt wzrostu odsetka kobiet bezdzietnych przeważa nad efektem spadku dzietności matek.

Na wykresie 5 zamieszczono dodatkowo szeregi współczynników składnika *quantum* wyodrębnionego z danych przekrojowych dzietności zgodnie z metodą Bongaartsa i Feeneya. Zanim je jednak zinterpretujemy, zapoznać się trzeba ze zmianami średniego wieku matek rodzących dzieci. Zmiany te naniesiono na wykresie 6, gdzie ponownie grubymi liniami niebieską zaznaczono średni wiek kobiet rodzących policzony na podstawie modelu [2], zaś linie czarnych kropek odpowiadają obliczeniom na podstawie surowych danych empirycznych.

W przypadku wieku zgodność wyników obliczeń na podstawie modeli i na podstawie surowych danych nie jest już tak doskonała, choć kierunki trendów, ich punkty zwrotne, momenty lokalnych odchyłeń – są w pełni zgodne. Obraz zmian jest bardzo prosty – do początku lat dziewięćdziesiątych XX wieku wiek rodzenia pierwszego dziecka pozostawał na stałym poziomie, a wiek rodzenia dzieci ogółem nieznacznie acz stabilnie się obniżał. Od tego momentu mamy wyraźny i dość stabilny wzrost obu wskaźników, nieco szybszy dla wieku rodzenia pierwszego dziecka, z niewielkimi lokalnymi przyspieszeniami około roku 2001 oraz 2009. Dokładnie wtedy, kiedy rośnie wiek rodzenia dzieci, obserwujemy przekraczanie zwykłych współczynników dzietności przez ich wartości skorygowane metoda Bongaartsa-Feeneya, z największym skokiem tej korekty około roku 1992, i mniejszymi skokami około roku 2001 oraz roku 2009.



Wykres 6. Średni wiek kobiet rodzących dzieci. Źródło: opracowanie własne

Pozwala to jeszcze dobitniej sens tych korekt zinterpretować – odpowiadają one w gruncie rzeczy na pytanie, jaką byśmy dzietność zaobserwowali w danym roku, gdyby tej dzietności nie zaciemnił efekt odłożenia w czasie (a nie redukcji) planów prokreacyjnych kobiet. Taki rzeczywiście był zamierzony efekt metody „poprawiania” surowego współczynnika dzietności, zaproponowany przez Bongaartsa i Feeneya. Niezamierzony efekt w postaci destabilizacji szeregu współczynników dzietności, często wskazywany w literaturze jako wada tej metody, rzeczywiście w naszym przypadku wystąpił, i jest wyraźnie widoczny na wykresie 5.

Chociaż skala zmian dzietności przed rokiem 1990 oraz po roku 2000 jest nieporównanie mniejsza, niż w ostatniej dekadzie XX wieku, warto na chwilę zatrzymać się nad lokalnymi

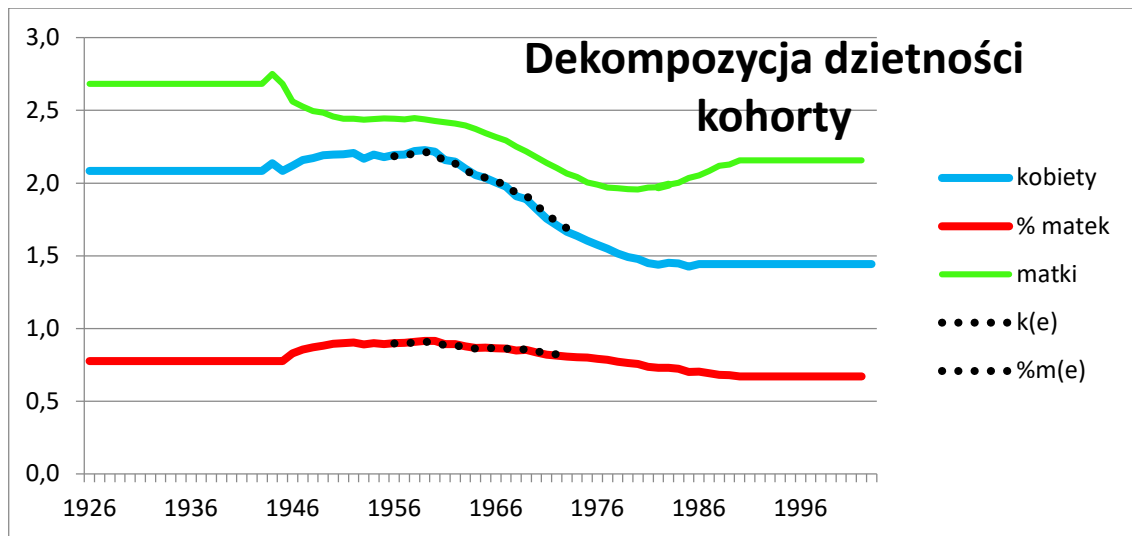
maksimummi dzietności około roku 1982, 2009 oraz w roku 2017. Dekompozycja pokazuje krańcowo odmienny charakter tych lokalnych szczytów. Lokalny wzrost dzietności po w stanie wojennym miał mieszany charakter – nastąpił wówczas zarówno wzrost odsetka kobiet rodzących swoje pierwsze dziecko, jak i wzrost dzietności matek. W okresie światowego kryzysu finansowego dochodzi natomiast do wzrostu dzietności ogólnej, którego niemal wyłącznym czynnikiem jest wzrost odsetka kobiet rodzących pierwsze dziecko. Wzrost dzietności całkowitej w roku 2017 (poprzedzony niewielkim wzrostem już w roku 2016) ma odwrotny charakter, ponieważ głównym jego czynnikiem jest wzrost dzietności matek. Ciśnię się wręcz interpretacja, że chodzi o efekty programu 500+, w początkowych latach skierowanego na wsparcie rodzin z większą liczbą dzieci. Dzietność całkowita w roku 2018 jest minimalnie mniejsza od tej z roku 2017, tak jakby siła rażenia tego programu zdążyła się już wyczerpać. Nie możemy jednak przyjrzeć się czynnikom tego wyhamowania bliżej, ponieważ ostatnie dane o urodzeniach 1-go dziecka dotyczą roku poprzedniego. Szkoda, bo właśnie w roku 2018 mogły się już ujawnić efekty objęcia programem 500+ także i pierwszych dzieci.

10. Wyniki oszacowań modeli z efektami kohortowymi

Kluczowe wyniki oszacowań modeli [3] zamieszczone są na wykresie 7. Są to kohortowe współczynniki dzietności całkowitej, prawdopodobieństwa posiadania dzieci w ogóle, oraz wynikające z tych wielkości współczynniki dzietności matek. Tym razem ich wartości można jednoznacznie przypisać kolejnym kohortom kobiet. Dzięki temu ich nazwy znacznie trafniej odzwierciedlają ich rzeczywistą zawartość informacyjną. Równocześnie jednak widoczne na tym wykresie tendencje trzeba interpretować w zupełnie innych kategoriach niż te, obserwowane przez pryzmat współczynników przekrojowych.

Zanim jednak przejdziemy do interpretacji, parę uwag ogólnych. Po pierwsze, na wykresie zamieszczone zostały (uwidocznione w postaci czarnych kropek) te same współczynniki wyznaczone bez modelu, na podstawie surowych danych. Gołym okiem widać, jak krótka jest lista kohort, dla których takie – czysto empiryczne – współczynniki daje się policzyć. Unaocznia to jak duża jest wartość dodana wynikająca z zastosowania modelu. Z drugiej strony nie należy przesadzać – model nie tworzy informacji, których w danych nie ma – model pozwala jedynie „uzbrojonym okiem” na te dane spojrzeć, i wydobyć możliwie dużo informacji w nich ukrytych. Lokalnie płaskie przebiegi trendów naniesionych na wykresie odpowiadają kohortom nazbyt starym lub nazbyt młodym, aby skąpe informacje na ich temat zawarte w próbkę pozwoliły stwierdzić, czy w ciągu kolejnych kohort dany współczynnik raczej rósł, czy raczej malał. Do tych kohort, dla których informacje w próbkę są nazbyt skąpe,

trzeba zaliczyć co najmniej kohorty kobiet urodzonych przed i w trakcie drugiej wojny światowej, oraz urodzonych po roku 1990. Dla kolejnych kohort powojennych, aż do kohort kobiet urodzonych w latach 80-tych, można – zachowując odpowiednią dozę ostrożności – próbować nadawać jakąś interpretację wyliczonym współczynnikiem.



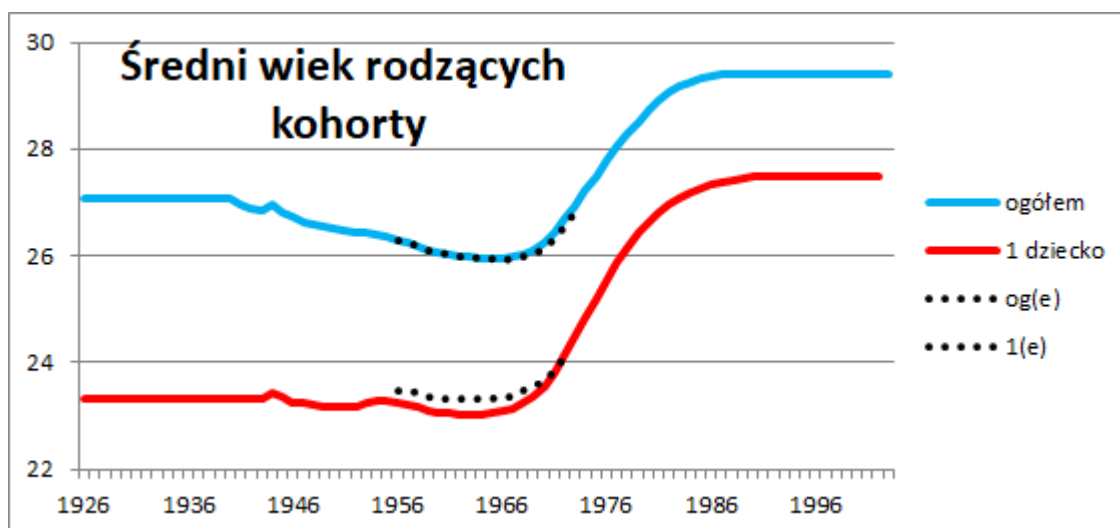
Wykres 7. Kohortowe współczynniki dzietności kobiet, prawdopodobieństwa posiadania dzieci i dzietności matek. Źródło: *opracowanie własne*

Interpretację zmian należy dokonywać w kategoriach zmian pokoleniowych. Istnieje bogata literatura socjologiczna, w której przedmiotem badań są zmiany pokoleniowe postaw i zachowań. Postawy i zachowania prokreacyjne są ich ważną częścią, i tak należałoby umiejscowić materię badaną za pośrednictwem modeli [3].

I tak w przekroju kohort kobiet urodzonych w pierwszych latach po wojnie nastąpił znaczny wzrost odsetka kobiet rodzących dzieci, przy równoczesnym znacznym spadku dzietności matek. Obie te tendencje niemal zanikły w przekroju kohort urodzonych w latach pięćdziesiątych, kiedy oba te wskaźniki ulegały jedynie nieznacznym wahaniom, składając się razem na względnie stabilny, stosunkowo wysoki wskaźnik dzietności kobiet. Z końcem lat pięćdziesiątych rozpoczęła się zmiana pokoleniowa, kiedy w przekroju kolejnych kohort wówczas urodzonych następował wyraźny wzrost odsetka kobiet bezdzietnych, i równocześnie znaczny spadek dzietności matek, co razem złożyło się na dramatyczny spadek dzietności całkowitej. Symptomy początku kolejnej zmiany pokoleniowej można przypisywać kohortom kobiet urodzonych około roku 1980. Spadkowy trend dzietności matek, który spowolnił już wcześniej, zmienił kierunek na trend rosnący, podczas gdy trend rosnącego odsetka kobiet bezdzietnych trwał dalej. Siła tych przeciwnych tendencji jest podobna, co powoduje, iż

dzielnosc ogólna kohort kobiet urodzonych w ciągu lat 80-tych waha się nieznacznie wokół stałego poziomu. Czy te tendencje utrzymają się dla kohort kobiet urodzonych po roku 1990, tego dopiero się dowiemy.

Zmiany pokoleniowe obserwowane przez pryzmat współczynników dzielnosci uzupełnia obraz zmian, któremu podlegał przeciętny wiek rodzenia dzieci. Zmiany te obserwować możemy na wykresie 8.



Wykres 8. Średni wiek kobiet rodzących dzieci. Źródło: *opracowanie własne*

W ciągu kolejnych kohort powojennych następował nieznaczny spadek wieku rodzenia pierwszego dziecka, przy nieco silniejszej tendencji spadku średniego wieku rodzenia wszystkich dzieci. Kryć się za tym może wyraźniejszy spadek średniego wieku rodzenia dzieci drugich, trzecich i następnych, choć może to także wynikać z nieznacznie malejącego udziału urodzeń dalszych w liczbie urodzeń wszystkich. Łączny efekt tych zmian to spadek średniego wieku rodzenia pierwszych dzieci od poziomu wieku około 23,5 do poziomu około 23 lat, oraz od poziomu około 27 lat do 26 lat dla dzieci wszystkich. Kohorty kobiet urodzonych w połowie lat sześćdziesiątych inicjują odwrócenie się dotychczasowych tendencji. Następujący potem trend wzrostu wieku rodzenia dzieci (tak pierwszych, jak i następnych) utrzymuje się aż do kohort kobiet urodzonych pod koniec lat osiemdziesiątych, kiedy wiarygodność szacunków spada do tego stopnia, że lepiej uznać że dalszych tendencji tymczasem nie znamy. Ocena łącznego efektu tych zmian jest więc niepewna, ale z tym zastrzeżeniem można uznać, że wiek średni rodzenia pierwszego dziecka wzrósł z poziomu 23 lat do poziomu 26,5 roku, zaś dla dzieci wszystkich z poziomu 26 lat do poziomu niespełna 29,5 roku. Ciąg kohort, w przekroju których dokonują się największe zmiany, to kohorty kobiet urodzonych w latach

siedemdziesiątych i na początku lat osiemdziesiątych. Na ile wygasanie tych zmian naprawdę dotyczy kohort urodzonych w drugiej połowie lat osiemdziesiątych, na ile zaś jest to efektem narastającej niekompletności danych, tego dokładnie nie wiemy. Z pewnością obie te przyczyny mają w tym swój udział, ale jakie są tutaj proporcje, tego dopiero dowiemy się w przyszłości.

11. Uwagi końcowe

Czytelnik może być nieco zaskoczony lakonicznym opisem uzyskanych wyników w podrozdziałach 9 oraz 10. I rzeczywiście, oszacowane modele pozwalają wyciągnąć znacznie bogatsze wnioski na temat zmian, jakim podlegała aktywność prokreacyjna Polek w badanym okresie. W szczególności analizie przekrojowej jak i kohortowej można było poddać ewolucję całych rozkładów wieku rodzenia dzieci, a nie tylko średniej wartości z tych rozkładów. Lektura podrozdziałów 9 i 10 może sprawiać wrażenie, że prezentacja wyników badania służy bardziej ilustracji efektów zastosowania szeregu innowacji metodologicznych, niż samej analizie zmian dzietności kobiet. Jeśli Czytelnik rzeczywiście takie wrażenie odniósł, to znaczy że trafnie odczytał intencje autorów. Chodziło nam w gruncie rzeczy (przynajmniej w tym tekście) o zaprezentowanie innowacji metodologicznych, zastosowanych przede wszystkim w procesie estymacji modeli kohortowych, ale w znacznej mierze dotyczącej także modeli przekrojowych. W szczególności obu przypadków dotyczy sposób uwzględnienia autokorelacji składników losowych. W modelach analizy przekrojowej chodziło o precyzyjniejsze wydobywanie globalnych efektów kalendarzowych spoza zaciemniających analizę efektów pozostałych. W modelach zaś analizy kohortowej chodziło o coś z grubsza odwrotnego – precyzyjniejszy pomiar globalnych efektów kohortowych, wydobytych spoza efektów pozostałych, tym razem traktowanych jako efekty zaciemniających analizę.

BIBLIOGRAFIA

- Bell W. R. 1997. „Comparing and assessing time series methods for forecasting age-specific fertility and mortality rates”. *Journal of Official Statistics*, vol. 13, no. 3. Sztokholm: Statistics Sweden, s. 279–303.
- Bijak J., Więckowska B. 2008. „Prognozowanie przeciętnego dalszego trwania życia na podstawie modelu Lee i Cartera: wybrane zagadnienia”. W: Ostasiewicz W. red. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, nr 1 (1201) Statystyka aktuarialna - teoria praktyka. Wrocław: Uniwersytet Ekonomiczny we Wrocławiu, s. 9-27.

- Bongaarts J., Feeney G. 1998. „On the Quantum and Tempo of Fertility”. *Population and Development Review*, vol. 24, no. 2. Nowy Jork: Population Council, Wiley-Blackwell, s. 271-291.
- Lee R. D., Carter L. R. 1992. „Modeling and Forecasting U.S. Mortality”. W: Little R. J. A. i in. red. *Journal of the American Statistical Association*, vol. 87, no. 419. Alexandria: American Statistical Association, s. 659-671.
- Lee R. D. 1993. „Modeling and Forecasting the Time Series of U.S. Fertility: Age Distribution, Range, and Ultimate Level”. W: Fildes R. red. *International Journal of Forecasting*, vol. 9, is. 2. Medford: International Institute of Forecasters, Elsevier, s. 187-202.
- Papież M. 2011. „Wykorzystanie modelu Lee-Cartera do prognozowania współczynników zgonów w Polsce”. W: Balcerowicz-Szkutnik M. red. *Studia Ekonomiczne / Uniwersytet Ekonomiczny w Katowicach*, nr 95 Współczesne problemy demograficzne w dobie globalizacji : aspekty pozytywne i negatywne. Katowice: Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, s. 83-91.
- Rossa A. 2009. „Dynamiczne tablice trwania życia oparte na metodologii Lee-Cartera i ich zastosowanie do obliczania wysokości świadczeń emerytalnych”. *Acta Universitatis Lodzensis. Folia Oeconomica*, t. 231. Łódź: Wydawnictwo Uniwersytetu Łódzkiego, s. 367-384.
- Sobotka T. i in. 2005. „‘Missing Births’: Decomposing the Declining Number of Births in Europe into Tempo, Quantum, and Age Structure Effects”. *European demographic research papers*, no. 2. Vienna: Vienna Institute of Demography.