

**Jakub Bijak**

Środkowoeuropejskie Forum Badań Migracyjnych i Ludnościowych, Warszawa

**Barbara Więckowska**

Szkoła Główna Handlowa w Warszawie

## **PROGNOZOWANIE PRZECIĘTNEGO DALSZEGO TRWANIA ŻYCIA NA PODSTAWIE MODELU LEE I CARTERA – WYBRANE ZAGADNIENIA**

### **1. Wstęp**

Postępujący proces starzenia się ludności, zwłaszcza w krajach rozwiniętych, sprawia, że ryzyko starości zaczyna być postrzegane przede wszystkim przez pryzmat jego drugiej fazy – odpowiedniego rozłożenia konsumpcji zgromadzonych oszczędności na cały okres emerytalny bez względu na realny czas życia [28, s. 221]. Wymusza to reformowanie systemów emerytalnych – odchodzenie od systemów o zdefiniowanym świadczeniu na rzecz systemów o zdefiniowanej składce. Wskutek tych zmian wysokość otrzymywanego świadczenia emerytalnego będzie zależeć m.in. od założonego (prognozowanego) przeciętnego dalszego trwania życia. Szczególnie istotne są w tym kontekście takie tendencje, jak jednoczesne wydłużanie się życia ludzkiego i kumulowanie się zgonów w ostatnich jego latach, obserwowane w postaci tzw. rektangularyzacji i ekspansji krzywej przeżycia [29, s. 251; 22, s. 7].

Obliczenia wysokości indywidualnego świadczenia emerytalnego oparte na bieżących tablicach trwania życia (TTŻ) z założenia nie doszacowują ryzyka dożycia przez przyszłych emerytów kolejnych lat po przejściu na emeryturę. Wynika to z faktu, że TTŻ są tablicami przekrojowymi i nie uwzględniają procesu wydłużania się przeciętnego dalszego trwania życia poszczególnych kohort (grup osób urodzonych w tym samym roku).

Niniejsza praca ma na celu przedstawienie metodologii prognozowania przeciętnego dalszego trwania życia z zastosowaniem zmodyfikowanego modelu autorstwa R.D. Lee i L.R. Cartera [16]. Prognoza taka może pozwolić na uwzględnienie efektu starzenia się populacji, związanego z rokiem kalendarzowym (*period effect*), wiekiem (*age effect*), a także, aczkolwiek w sposób pośredni, ze starzeniem się poszczególnych kohort (*cohort effect*). Metodologia zaproponowana przez Lee i Cartera ma charakter stochastyczny i uwzględnia niepewność prognozowanego zjawiska. W niniejszej pracy uwzględniona została niepewność wynikająca z trzech czynników: zmienności liczby zgonów podlegającej rozkładowi Poissona, niepewności oszacowań parametrów modelu oraz ekstrapolacji trendu. Rozważania teoretyczne zostały zilustrowane na przykładzie prognozy przeciętnego dalszego trwania życia w Finlandii, wyznaczonej na lata 2005-2050.

Praca składa się z czterech części. W pierwszej przedstawiono pokrótce historię prognozowania umieralności, wraz z przedstawieniem wybranych modeli, w tym modelu Lee i Cartera w wersji oryginalnej i zmodyfikowanej (punkt 2). Następnie omówione zostały praktyczne aspekty prognozowania umieralności za pomocą modelu Lee i Cartera (punkt 3). Część trzecia pracy dotyczy prognozy przeciętnego (dalszego) trwania życia dla Finlandii otrzymanej na podstawie opisanej wcześniej metodologii (punkt 4). Pracę kończy krótkie podsumowanie wyników i przedstawienie najważniejszych wniosków wraz z rekomendacjami dla dalszych badań (punkt 5).

## 2. Wybrane modele prognozowania umieralności – rys historyczny

Pionierskie próby określenia przeciętnego oraz przeciętnego dalszego trwania życia datuje się na XVI w., a jako pierwsze praktyczne wykorzystanie tych miar wskazuje się wyliczenie wysokości świadczeń emerytalnych przez J. de Witta, premiera Danii, w 1671 r. [10, s. 6]<sup>1</sup>. Przez stulecia obliczenia de Witta pozostały jednak niezauważone i niedocenione, gdyż dotyczyły tablic hipotetycznych, nie opartych na zaobserwowanych w rzeczywistości liczbach zgonów.

Pierwsze tablice umieralności stworzone na podstawie obserwacji zostały wyliczone w 1662 r. przez J. Graunta dla Londynu (za: [31]) oraz, niezależnie, w 1693 r. przez brytyjskiego matematyka E. Halleya dla Wrocławia. Metodologia konstrukcji tablic na podstawie liczby osób żyjących oraz liczby osób zmarłych w danym roku stosowana jest do dziś. Z kolei pierwsza praca na temat długości trwania życia opublikowana została przez N. Bernoulliego w 1709 r. [10, s. 7].

---

<sup>1</sup> Literatura wskazuje również na wcześniejsze próby oszacowania długości trwania życia (np. tablice ulpiańskie z II w. n.e. (za: [31]), jednakże nie miały one bezpośredniego związku z ubezpieczeniami.

Początkiem nowej ery w modelowaniu umieralności było opracowanie B. Gompertza [9]. Zaproponował on analityczny zapis prawa dotyczącego natężenia liczby zgonów<sup>2</sup>,  $\mu_t$ , w postaci następującej funkcji wykładniczej [6, s. 78]:

$$\mu_t = B \cdot c^t, \quad (1a)$$

gdzie  $B$  oraz  $c$  oznaczają pewne stałe,  $t$  – czas.

Gompertz twierdził, że istnieją dwie główne przyczyny zgonów: skutek starzenia się, zależny od wieku, oraz niezależna od wieku wypadkowa intensywność zgonów, co jednak nie wpłynęło na zapis jego prawa. Pomysł Gompertza wykorzystał W. Makeham [19], modelując intensywność umieralności jako [6, s. 78]:

$$\mu_t = A + B \cdot c^t, \quad (1b)$$

gdzie  $A$ ,  $B$  oraz  $c$  są parametrami modelu,  $t$  oznacza indeks czasu.

Sformułowanie praw Gompertza i Makehama wniosło istotny wkład do prób ujęcia funkcji stosowanych w tablicach trwania życia w postaci analitycznej oraz prób modelowania intensywności umieralności na postawie innych funkcji – potęgowej (typu Weibulla) czy wielomianowej (np. [11], za: [10, s. 7; 8, s. 18]).

Analizując rozwój modeli opisujących długość trwania życia, można zdefiniować kilka płaszczyzn zmian tych modeli [24]:

- od modeli dyskretnych do ciągłych,
- od modeli dwustanowych do wielostanowych,
- od modeli homogenicznych do heterogenicznych,
- od modeli statycznych do dynamicznych,
- od modeli deterministycznych do stochastycznych.

Pierwsze modele długości trwania życia zakładały wiek jako zmienną dyskretną. Jak już wspomniano, przełomem w tym podejściu było wyprowadzenie ogólniejszych praw dotyczących intensywności umieralności. Jednakże model ciągły – zakładający wypłaty w chwili śmierci oraz renty ciągłe – został wprowadzony w 1869 r. przez W. Woolhouse'a [24, s. 9].

Rozwój produktów ubezpieczeniowych również determinował charakter modeli. Pierwotnie, ze względu na rodzaj wypłacanych świadczeń, istotne było tylko to, jak długo osoba będzie żyła oraz kiedy umrze – modelowano jedynie stany życia i śmierci. Wraz z rozwojem świadczeń ubezpieczeniowych, związanych z uzależnieniem wypłaty świadczenia od stanu zdrowia (np. w ubezpieczeniu uszczerbku na zdrowiu czy ubezpieczeniu na wypadek poważnego zachorowania) albo od sytuacji życiowej ubezpieczonego (np. w ubezpieczeniu od bezrobocia), wprowadzono modele wielostanowe.

<sup>2</sup> Pierwszą pracę dotyczącą intensywności umieralności opublikował A. de Moivre.

Początkowo w modelowaniu długości trwania życia zakładano homogeniczność badanej populacji, zwracając przy tym uwagę na możliwość występowania zjawiska antyselekcji ryzyka (np. w pracy J. de Witta). Dopiero F. Corbaux w 1833 r. założył heterogeniczność populacji – rozważał tablice trwania życia jako wynikową nakładania się różnych tablic życia dla homogenicznych podgrup wyodrębnionych na podstawie pewnych czynników, takich jak płeć czy zawód (za: [24, s. 10-17]). Na podstawie tego podejścia powstała idea budowy tablic selektywnych. Naukowcy nadal wykazują duże zainteresowanie tym podejściem, szczególnie w modelowaniu długości trwania życia osób w wieku starszym, wykorzystując koncepcję tzw. *frailty-based modelling* [14].

Pierwsze modele trwania życia miały charakter deterministyczny. Mimo tego, że modelowały zmienną losową, jaką jest długość trwania życia, w czasach E. Halleya rozpatrywana ona była w kategoriach szans przeżycia. Zmienna losowa opisująca wartość obecną świadczenia ubezpieczeniowego, która jest funkcją dalszego trwania życia, wprowadzona została dopiero w 1950 r. w pracach B. de Finettiego [7] i E. Sverdrupa [27], (za: [24, s. 4]).

Dopóki nie były dostępne dostatecznie długie szeregi czasowe, na podstawie których można było zaobserwować trend umieralności, wszystkie modele były statyczne. Dopiero na początku XX w. zaczęto rozważać modele dynamiczne i, co za tym idzie, prognozować poziom umieralności. Obecnie jest to najważniejszy kierunek modelowania aktuarialnego, wykorzystywany przede wszystkim w szacowaniu wysokości świadczeń długoterminowych.

W rachunku aktuarialnym ubezpieczeń osobowych modelowane są przede wszystkim miary związane z umieralnością zależne od wieku, jak np. intensywność umieralności  $\mu$  czy prawdopodobieństwo zgonu  $q$ . Dlatego też w ogólnym zapisie te wielkości ujmowane są jako funkcja dwóch zmiennych: wieku  $x$  i roku kalendarzowego  $t$ , np.:  $\mu = \mu(x, t)$ . Pozwala to uwzględnić efekt zmian w umieralności związanej zarówno z rokiem kalendarzowym (*period effect*) oraz wiekiem (*age effect*), jak i z poszczególnymi generacjami (*cohort effect*) [23, s. 282].

W prognozowaniu demograficznym najczęściej wykorzystywane jest podejście związane z rokiem kalendarzowym (zwane przekrojowym lub poprzecznym), a także z poszczególnymi kohortami (zwane kohortowym lub wzdłużnym), przy czym oba traktowane są jako nawzajem się uzupełniające [21, s. 73-75]. Analiza kohortowa odnosi się do badania zjawisk demograficznych dla poszczególnych generacji (kohort) badanej zbiorowości, czyli grup osób urodzonych w tym samym roku (latach). Analiza przekrojowa natomiast dotyczy wybranego momentu w czasie, dla którego badane są zjawiska dla hipotetycznych kohort, łączących cechy różnych generacji zaobserwowane w danym roku.

W ujęciu stochastycznym zaobserwowane wartości intensywności umieralności stanowią realizacje pewnej zmiennej losowej. Dlatego też, oprócz prognozy punktowej (wartości oczekiwanej) zmiennej w danym roku, ważne jest również

określenie zakresu możliwych zmian wyznaczonych przez rozkłady predykcyjne (rozkłady prawdopodobieństwa przyszłych wartości danej zmiennej wyznaczone z modelu prognostycznego przy danych obserwacjach z próby). Przykładem modelu uwzględniającego niepewność w ujęciu stochastycznym, a jednocześnie pozwalającego na jednoczesną analizę efektów związanych z wiekiem i rokiem kalendarzowym, jest model autorstwa R.D. Lee oraz L.R. Cartera [16]. Model ten, określony przez N. Brouhns, M. Denuita i J.K. Vermunta [5, s. 376] jako „potężne i eleganckie narzędzie prognozowania umieralności”, stanowi jedną z najważniejszych i najczęściej stosowanych zaawansowanych metod prognozowania umieralności zaproponowanych w ciągu ostatnich dwudziestu lat.

W oryginalnej postaci modelu Lee i Cartera [16] zmienną objaśnianą jest logarytm centralnego natężenia umieralności (rocznego współczynnika natężenia zgonów) osób wieku  $x$  w roku  $t$ , będącego tu zmienną dyskretną, oznaczaną przez  $m(x, t)$ . Postać modelu dana jest wzorami (za: [17, s. 21-22]):

$$\ln[m(x, t)] = a_x + b_x \cdot \kappa_t + \xi_{x,t}, \quad (2a)$$

gdzie:

$$\kappa_t = \kappa_{t-1} + c + \varepsilon_t. \quad (2b)$$

W (2a)  $a_x$  opisuje umieralność w grupach wieku uśrednioną względem czasu,  $\kappa_t$  – zmiany w poziomie umieralności w czasie, tu modelowane za pomocą procesu błędzenia losowego z dryfem (2b),  $b_x$  zaś opisuje odchylenia od uśrednionego poziomu umieralności związane z wpływem zmian  $\kappa_t$ . Dodatkowo zakłada się, że  $\sum_x b_x = 1$  oraz  $\sum_t \kappa_t = 0$ , przy czym założenia te, nie wprowadzające *de facto* niczego ponad normalizację zmiennych  $b_x$  oraz  $\kappa_t$ , gwarantują identyfikowalność modelu [5, s. 376]. Składniki losowe  $\xi_{x,t}$  oraz  $\varepsilon_t$  opisują niepewność odpowiednich zmiennych objaśnianych nie wyjaśnioną przez model. Zakłada się, że są one niezależnymi zmiennymi losowymi o rozkładach normalnych z wartością oczekiwaną równą zeru i stałą wariancją (Gaussowski „biały szum”).

W przypadku samej struktury modelu krytyce zostało poddane modelowanie zmiennej nieobserwowalnej – logarytmu centralnego natężenia umieralności [15, s. 4]. Do obliczeń aktuarialnych niezbędna jest prognoza liczby zgonów, dlatego modyfikacje modelu są oparte na modelowaniu właśnie tej wartości. Jeden z wariantów zmodyfikowanego modelu Lee i Cartera zakłada, że liczba zgonów osób w wieku  $x$  w roku  $t$ ,  $D(x, t)$ , jest zmienną losową o rozkładzie Poissona (np. [5, s. 378]):

$$D(x, t) \sim \text{Poisson}[E(x, t) \cdot \mu(x, t)], \quad (3a)$$

gdzie:

$$\ln[\mu(x, t)] = a_x + b_x \cdot \kappa_t, \quad (3b)$$

oraz:

$$\kappa_t = \kappa_{t-1} + c + \varepsilon_t. \quad (3c)$$

We wzorze (3a)  $\mu(x, t)$  oznacza intensywność umieralności, będącą ciągłym odpowiednikiem dyskretnej zmiennej  $m(x, t)$  w modelu (2),  $E(x, t)$  zaś – centralną liczbę osób narażonych na ryzyko zgonu (*exposure to risk*). Tak jak poprzednio zakłada się ponadto, że  $\sum_x b_x = 1$  oraz  $\sum_t \kappa_t = 0$  oraz że składniki losowe  $\varepsilon_t$  są Gaussowskim białym szumem.

### 3. Szacowanie przekrojowych i kohortowych tablic trwania życia na podstawie modelu Lee i Cartera

Aby móc uwzględnić w modelowaniu i prognozowaniu umieralności zarówno podejście przekrojowe (poprzeczne), jak i kohortowe (wzdłużne), potrzebne są dane o stanach ludności i zdarzeniach demograficznych, dla których wiek osób składających się na badaną zbiorowość wyrażany jest w latach **osiągniętych** w ciągu roku kalendarzowego. Załóżmy, że dostępne są roczne dane dla lat  $t = 1, 2, \dots, T$ , jednorocznych grup wieku  $x = 0, 1, \dots, x_{\max}$ , składające się z: liczby zgonów w danym roku kalendarzowym i grupie wieku,  $D(x, t)$ , liczby urodzeń w ciągu roku,  $B(t)$ , oraz liczby ludności w danym wieku na początku danego roku kalendarzowego  $P(x, t)$ , przy czym ta ostatnia wielkość jest dostępna również dla  $t = T + 1$ .

Przy powyższych założeniach oraz przyjętej definicji wieku spełnione są następujące równania bilansowe ludności, w których zmienną bilansującą są salda migracji netto (napływy minus odpływy ludności) dla poszczególnych grup wieku,  $I(x, t)$ :

$$P(0, t) = B(t) - D(0, t) + I(0, t), \text{ dla } x = 0; \quad (4a)$$

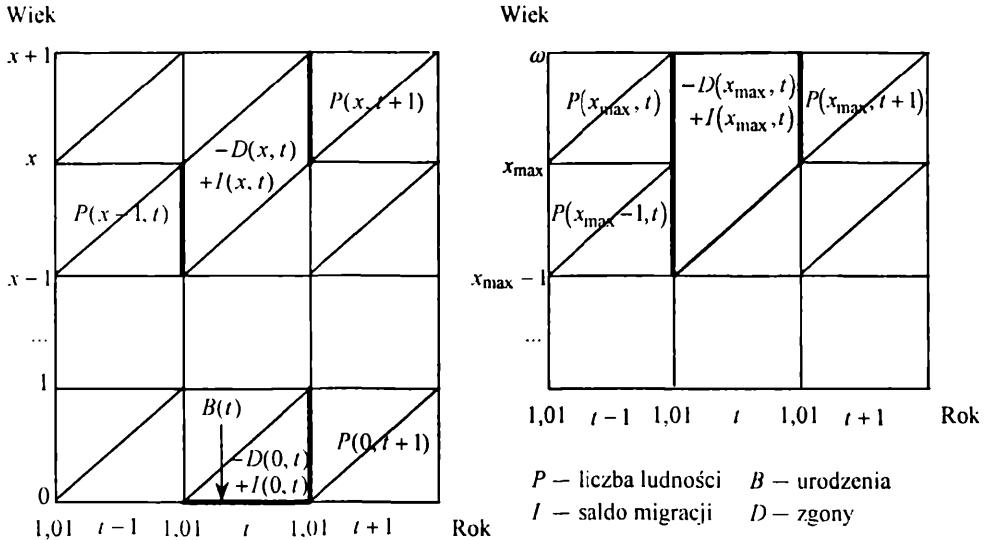
$$P(x, t) = P(x - 1, t - 1) - D(x, t) + I(x, t), \text{ dla } 0 < x < x_{\max}; \quad (4b)$$

$$P(x_{\max}, t) = P(x_{\max} - 1, t - 1) + P(x_{\max}, t - 1) - D(x_{\max}, t) + I(x_{\max}, t), \text{ dla } x = x_{\max}. \quad (4c)$$

Równania (4a)-(4c) mogą zostać zilustrowane z wykorzystaniem popularnej w demografii tzw. siatki Lexisa<sup>3</sup>, dla której przyjęte powyżej definicje i założenia

<sup>3</sup> Siatka (diagram) Lexisa pozwala na analizowanie zjawisk i zbiorowości demograficznych w kilku różnych perspektywach: według roku kalendarzowego (oznaczanego na osi poziomej), wieku (oś pionowa), lub poszczególnych kohort (diagonalnie). Szerzej na ten temat pisze np. J. Paradysz [21, s. 60-68].

wyznaczają przekrojowo-kohortowy plan obserwacji (*period-cohort observation plan*), pokazany na rys. 1. Taka perspektywa pozwala na konstruowanie syntetycznych mierników demograficznych w ujęciu zarówno przekrojowym, jak i kohortowym, poprzez agregowanie informacji odpowiednio dla kohort syntetycznych (pionowo) lub rzeczywistych (diagonalnie).



Rys. 1. Założenia dla równań bilansowych ludności w planie przekrojowo-kohortowym

Źródło: opracowanie własne.

Za E. van Imhoffem [12], przyjmijmy dodatkowo, że natężenie umieralności  $\mu(x, t)$  jest stałe dla każdego z jednostkowych obszarów planu okresowo-kohortowego (wyznaczonych przez odpowiednie trójkąty, równoległoboki lub trapezy na siatce Lexisa, pokazane na rys. 1) oraz że dla każdej grupy wieku saldo migracji  $I(x, t)$  jest rozłożone równomiernie w ciągu roku. Zachodzi wówczas [13, s. 6-7]:

$$P(0, t) = \{1 - \exp[-\mu(0, t)]\} \cdot [B(t) + I(0, t)] / \mu(0, t), \text{ dla } x = 0; \quad (5a)$$

$$P(x, t) = P(x-1, t-1) \cdot \exp[-\mu(x, t)] + \{1 - \exp[-\mu(x, t)]\} \cdot I(x, t) / \mu(x, t), \\ \text{dla } 0 < x < x_{\max}; \quad (5b)$$

$$P(x_{\max}, t) = [P(x_{\max}-1, t-1) + P(x_{\max}, t-1)] \cdot \exp[-\mu(x_{\max}, t)] + \\ + \{1 - \exp[-\mu(x_{\max}, t)]\} \cdot I(x_{\max}, t) / \mu(x_{\max}, t), \text{ dla } x = x_{\max}. \quad (5c)$$

Równania (5a)-(5c) są nieliniowe względem  $\mu(x, t)$ , więc natężenie umieralności może być szacowane z wykorzystaniem metod numerycznych, przy danych

$P(x, t)$  oraz wyliczonych z równań bilansowych (4a)-(4c) saldach migracji  $I(x, t)$ . Na podstawie tak uzyskanych wartości  $\mu(x, t)$ , parametry  $a_x, b_x$  oraz  $\kappa_t$  modelu Lee i Cartera postaci (3) mogą być następnie szacowane na kilka sposobów. Najpopularniejsza metoda, zaproponowana oryginalnie przez Lee i Cartera [16], jest oparta na rozkładzie wartości osobliwych macierzy (*singular value decomposition*, SVD) i odpowiada klasycznej metodzie najmniejszych kwadratów. Jej zastosowanie wymaga jednak przyjęcia dość silnego założenia o braku autokorelacji oraz o homoskedastyczności składnika losowego, co oznacza, że jego macierz kowariancji jest macierzą diagonalną o jednakowych elementach na przekątnej. Założenie o homoskedastyczności często było poddawane krytyce jako mało realistyczne (zob. np. [1; 5]), co znalazło potwierdzenie w badaniach empirycznych<sup>4</sup>. Alternatywę stanowi tu zastosowanie ważonej metody najmniejszych kwadratów [13; 15], lub metody największej wiarygodności (MNW, [5]). Przegląd metod został zamieszczony w pracy [15].

W niniejszej pracy parametry modelu (3) zostały oszacowane za pomocą MNW. Z założeniem, że zmienność liczby zgonów ma rozkład Poissona, maksymalizowany logarytm funkcji wiarygodności,  $LL$ , ma postać [5, s. 378]:

$$LL = \sum_x \sum_t \{ D(x, t) \cdot (a_x + b_x \cdot \kappa_t) - E(x, t) \cdot \exp(a_x + b_x \cdot \kappa_t) - \ln[D(x, t)!] \}. \quad (6)$$

$E(x, t)$  oznacza tu centralną liczbę osób narażonych na ryzyko zgonu, czyli przeciętną liczbę osobolat przeżywanych przez osoby w wieku  $x$  w roku  $t$ , równą  $D(x, t)/\mu(x, t)$ . Ostatni człon każdego składnika sumy (6),  $-\ln[D(x, t)!]$  nie zależy od estymowanych parametrów modelu i może zostać pominięty przy obliczaniu oszacowań  $a_x, b_x$  oraz  $\kappa_t$  poprzez maksymalizację  $LL$ .

Jak zauważyli N. Keilman i D.Q. Pham [13], niepewność w modelu Lee i Cartera wynika z co najmniej trzech czynników. Po pierwsze, zmienność liczby zgonów jest losowa i wyrażana przez rozkład Poissona. Po drugie, oszacowania parametrów modelu (3) z użyciem dowolnej metody estymacji są obciążone błędem. Wreszcie, model zawiera komponent zależny od czasu ( $\kappa_t$ ), którego ekstrapolacja stanowi trzecie źródło niepewności.

W przypadku ekstrapolacji zakłada się najczęściej, jak zostało wspomniane w punkcie 3, że proces losowy  $\kappa_t$  ma charakter błędzenia przypadkowego z dryfem  $c$  o niezależnych składnikach losowych  $\varepsilon_t$  mających jednakowe rozkłady normalne  $N(0, \sigma^2)$ . Wobec tego, za pracą N. Li, R. Lee i S. Tuljapurkara [17, s. 22], można przyjąć, że estymatory parametrów modelu dane są wzorami:

<sup>4</sup> Natężenie umieralności jest bardziej zmienne w najstarszych grupach wieku niż w młodszych ze względu na mniejszą liczbę obserwacji, która sprawia, że liczba zgonów podlega w większym stopniu losowym fluktuacjom. Do innych przyczyn można tu zaliczyć błędy pomiaru, związane z niewłaściwym określaniem wieku osoby zmarłej i nieprawidłowym wypełnieniem aktu zgonu [4, s. 129-130].



$$c_{\text{est}} = (\kappa_T - \kappa_1)/(T - 1) \quad (7)$$

oraz:

$$\sigma_{\text{est}}^2 = 1/(T - 1) \cdot \sum_{t=2}^T [\kappa_t - \kappa_{t-1} - c_{\text{est}}]^2. \quad (8)$$

Estymator błędu oszacowania stałej  $c$ ,  $sc_{\text{est}}$  jest równy:  $sc_{\text{est}} \approx \sigma_{\text{est}}/(T - 1)^{0.5}$ . Wzór na ekstrapolację  $\kappa_t$  dla  $t > T$ , ma wtedy postać [17, s. 23]:

$$\kappa_t = \kappa_T + [c_{\text{est}} + sc_{\text{est}} \cdot \eta] \cdot (t - T) + \sigma_{\text{est}} \cdot \sum_{\tau=T+1}^t \varepsilon_\tau, \quad (9)$$

gdzie  $\eta \sim N(0, 1)$ . Dodatkowo, co nie zostało uwzględnione w oryginalnym sformułowaniu modelu, aby uniknąć skoków między obserwowanymi a prognozowanymi wartościami  $\mu(x, t)$ , należy zakotwiczyć prognozę w ostatnim,  $T$ -tym okresie obserwacji dla każdej grupy wieku  $x$  (por. np. [3, s. 16]). W tym celu, zamiast  $\kappa_T$  w formule (9) można umieścić wyrażenie  $\kappa'_T(x) = \{\ln[\mu(x, t)] - a_x\}/b_x$ , wyznaczone dla każdego  $x$  z osobna.

Podstawiając we wzorze (9)  $\varepsilon_\tau = 0$  dla  $\tau = T + 1, \dots, t$ , otrzymujemy wartości prognoz punktowych dla  $\kappa_t$ . Wygenerowanie  $N$  (pseudo)losowych prób Monte Carlo dla składników  $\{\eta, \varepsilon_{T+1}, \dots, \varepsilon_t\}$  z niezależnych rozkładów  $N(0, \sigma^2)$  pozwala na przeprowadzenie w oparciu o (9) estymacji rozkładów predykcyjnych z uwzględnieniem jednego źródła niepewności: ekstrapolacji  $\kappa_t$ . Wyznaczone na ich podstawie przedziały predykcyjne, np. oparte na kwantylach rzędu  $\alpha/2$  i  $(1 - \alpha/2)$ , pozwalają na kwantyfikację tej niepewności.

Niepewność estymacji parametrów  $a_x$ ,  $b_x$  oraz  $\kappa_t$  modelu (3) można oszacować, wykorzystując np. metody bootstrapowe [13; 15]. W niniejszej pracy zastosowano metodę nieparametryczną zaproponowaną przez N. Keilmana i D.Q. Phama [13, s. 11-12], opartą na losowaniu  $S$  razy ze zwracaniem błędów dla poszczególnych  $\mu(x, t)$ , czyli różnic między wartościami zaobserwowanymi w próbie a wyznaczonymi z modelu. Na podstawie wylosowanych dla każdej pary  $(x, t)$  wartości tych błędów oraz oceny  $\mu(x, t)$  wygenerowanych przez model wyznaczane są nowe wielkości  $\mu^s(x, t)$ , dla  $s = 1, 2, \dots, S$ , na podstawie których liczone są symulowane wartości  $D^s(x, t)$ ,  $I^s(x, t)$  oraz  $E^s(x, t)$ . Aby uzyskane wielkości były zgodne z zaobserwowanymi stanami ludności, zakłada się tu nie zmienione wartości  $P(x, t)$  oraz  $B(t)$ .

Dla każdej próby bootstrapowej  $s$  estymowane są parametry modelu (3),  $a_x^s$ ,  $b_x^s$  oraz  $\kappa_t^s$ . Następnie dla każdego szeregu oszacowanych wartości  $\kappa_t^s$ , na podstawie

wzorów (7)-(9), wyznaczane są wartości prognozowane wraz z estymatorami rozkładów predykcyjnych, obliczanymi każdorazowo na podstawie  $N$  losowań Monte Carlo. Uzyskane w ten sposób  $S \cdot N$  prób pozwala na wyznaczenie estymatorów przedziałów predykcyjnych uwzględniających oba źródła niepewności (błędy oszacowań parametrów oraz ekstrapolacji  $\kappa_t$ ), np. na podstawie kwantyli rzędu  $\alpha/2$  oraz  $(1 - \alpha/2)$  z uzyskanych rozkładów.

Na podstawie oszacowanych i prognozowanych wartości centralnych  $\mu(x, t)$  oraz odpowiednich kwantyli z rozkładów predykcyjnych uzyskanych łącznie metodami bootstrapowymi i Monte Carlo, można wyznaczyć odpowiednie tablice trwania życia w ujęciu przekrojowym i kohortowym. Za pracą N. Keilmana i D.Q. Phama [13, s. 12], niech prawdopodobieństwo zgonu osoby w wieku  $x$  w ciągu roku  $t$  wynosi  $q(x, t) = 1 - \exp[-\mu(x, t)]$  dla  $0 < x < x_{\max}$ ,  $q(0, t) = 1 - \exp[-\mu(x, t)/2]$  dla osób w wieku 0 lat, a  $q(x_{\max}, t) = 1$  dla ostatniej, otwartej grupy wieku. Przyjmując  $l(x, t) = l(x - 1, t) \cdot [1 - q(x, t)]$  oraz  $l(0, t) \equiv 1$ , przeciętne dalsze trwanie życia osoby w wieku  $x$  lat w roku  $t$ , w ujęciu przekrojowym,  $e^p(x, t)$ , dane jest wzorem:

$$e^p(x, t) = \sum_{y=x}^{x_{\max}} [l(y, t) \cdot q(y, t) / \mu(y, t)] / l(x, t). \quad (10a)$$

Analogicznie można wyprowadzić wzór dla przeciętnego dalszego trwania życia osoby w wieku  $x$  lat w roku  $t$  w ujęciu kohortowym,  $e^c(x, t)$ . Dla  $q(x, t)$  zdefiniowanego tak samo jak uprzednio oraz  $l(x, t + 1) = l(x - 1, t) \cdot [1 - q(x, t)]$  i  $l(0, t) \equiv 1$  otrzymujemy wówczas:

$$e^c(x, t) = \sum_{y=x}^{x_{\max}} [l(y, t + y - x) \cdot q(y, t + y - x) / \mu(y, t + y - x)] / l(x, t). \quad (10b)$$

Dla współczynników kohortowych  $e^c(x, t)$  sumowanie odbywa się więc po kolejnych grupach wieku dla danej kohorty, a nie dla danego roku kalendarzowego, co implikuje konieczność sięgania „w przyszłość” i korzystania z prognozowanych wartości  $\mu$ . Trzeba jednak zaznaczyć, że w zastosowaniach praktycznych, ze względu na otwarty charakter ostatniej, najwyższej grupy wieku ( $x_{\max}$ ), wzór (10b) ma charakter przybliżony. Zakłada się tu również brak czynników perturbacyjnych w horyzoncie prognozy (takich, jak epidemie czy inne skokowe zmiany umieralności), które mogłyby zmodyfikować przewidywaną trajektorię przemian natężenia zjawiska (por. np. [21, s. 103]).

## 4. Ilustracja empiryczna – przykład Finlandii

Przedstawiony zmodyfikowany model Lee i Cartera został użyty do sporządzenia prognozy przeciętnego (dalszego) trwania życia w Finlandii na lata 2006-2050. W modelu wykorzystano dane pochodzące z *Human Mortality Database* ([www.mortality.org](http://www.mortality.org)). Jest to baza danych zawierająca informacje dotyczące ludności oraz zgonów dla ponad 30 krajów świata, stworzona na uniwersytecie w Berkeley oraz w Instytucie Badań Demograficznych im. Maksa Plancka w Rostocku na potrzeby analiz ilościowych związanych z długowiecznością. Na potrzeby modelu użyto następujących danych w jednorocznych grupach wieku do  $x_{\max} = 100$  lat, w latach osiągniętych, dla obu płci łącznie:

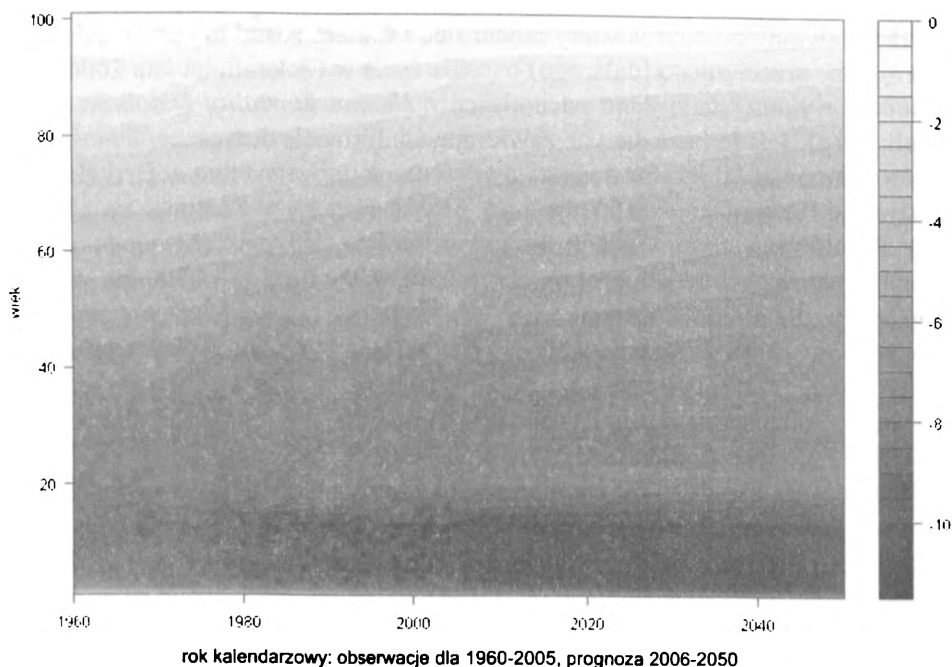
- zgonów według wieku i kohort ( $D$ ), za lata 1960-2005 (wielkość próby:  $T = 46$ ),
- liczba ludności na początku roku ( $P$ ), 1960-2006,
- liczba urodzeń w ciągu roku ( $B$ ), 1960-2005.

Obliczeń dokonano z użyciem własnego kodu stworzonego w oprogramowaniu R ([www.r-project.org](http://www.r-project.org))<sup>5</sup>. Przyjęto przy tym postać (3) modelu Lee i Cartera, opartą na rozkładzie Poissona, dla której przeprowadzono estymację metodą największej wiarygodności (6), przy czym dane wejściowe zostały przygotowane na podstawie wzorów (4)-(7). Ekstrapolacji (9) oraz obliczeń wartości przeciętnego (dalszego) trwania życia (10) dokonano: raz na podstawie prognozowanych wartości centralnych  $\mu(x, t)$ ,  $N = 1000$  razy na podstawie symulacji Monte Carlo dla ekstrapolacji trendu  $\kappa_t$ , a następnie  $S \cdot N = 100\,000$  razy dla stu prób bootstrapowych uwzględniających dodatkowo niepewność oszacowań parametrów modelu. W związku z taką strukturą zagadnienia, cała procedura obliczeniowa okazała się dość czasochłonna (154 godziny obliczeń z wykorzystaniem komputera z procesorem 2,2 GHz).

Wyniki obliczeń dla logarytmów natężenia umieralności  $\mu(x, t)$  dla okresu próby (1960-2005) oraz okresu prognozy (2006-2050) przedstawione są na rys. 2 i 3, odpowiednio w postaci dwuwymiarowej „mapy hipsometrycznej”, której wymiarami są czas  $t$  i wiek  $x$ , oraz wzorców natężenia umieralności dla lat 1960, 2005 i prognozy na rok 2050.

Zgodnie z „mapą hipsometryczną”  $\ln[\mu(x, t)]$ , przedstawioną na rys. 2, dla całego okresu prognozy we wszystkich grupach wieku będzie następował stopniowy spadek umieralności. Poprawa współczynników jest tu wyraźnie skorelowana z wiekiem – silniejsza w młodszych grupach wieku, słabsza w starszych. Największe zmiany logarytmów  $\mu(x, t)$  prognozowane są dla dzieci w wieku 1-15 lat (rys. 3). Można to wyjaśniać przewidywaną coraz większą wykrywalnością i wyleczalnością

<sup>5</sup> R jest systemem przeznaczonym do obliczeń statystycznych, analogicznym do środowiska S opracowanego w Laboratorium Bell, jednakże, w przeciwieństwie do tamtego, opartym na filozofii wolnego oprogramowania według licencji GNU.



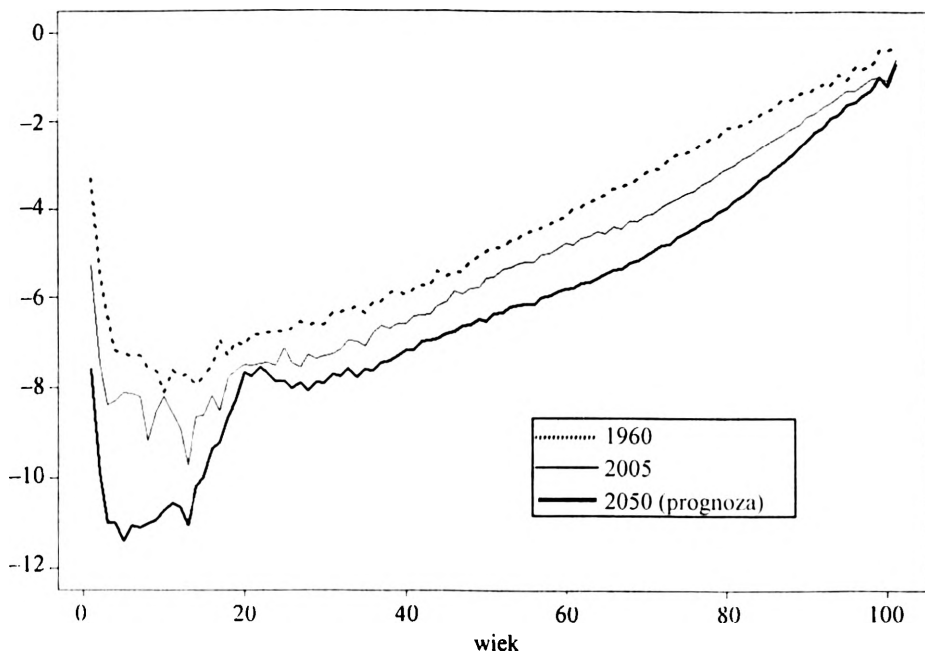
Rys. 2. Mapa hipsometryczna logarytmów współczynników  $\mu(x, t)$  dla obu płci łącznie w Finlandii, 1960-2050

Źródło: obliczenia własne na podstawie danych z *Human Mortality Database* ([www.mortality.org](http://www.mortality.org)).

chorób wieku dziecięcego oraz coraz powszechniejszym stosowaniem szczepień ochronnych.

Wyliczone na podstawie oszacowanych i prognozowanych wartości  $\mu(x, t)$  przeciętne trwanie życia w Finlandii dla obu płci łącznie w ujęciu przekrojowym,  $e^P(0, t)$ , zostało zaprezentowane na rys. 4. Do oceny niepewności przyjęto 80-procentowe przedziały ufności z rozkładów predykcyjnych<sup>6</sup>. Węższe przedziały odnoszą się tylko do jednego źródła niepewności prognozy: ekstrapolacji  $\kappa_t$ , szersze przedziały obejmują natomiast dodatkowo niepewność estymacji parametrów modelu. W obu wypadkach wzięta została pod uwagę zmienność liczby zgonów modelowana za pomocą rozkładu Poissona.

<sup>6</sup> 80-procentowe przedziały predykcyjne stanowią standardową rozpiętość przyjmowaną w prognozowaniu demograficznym. Uzasadnieniem jest tu potrzeba unikania niepotrzebnego zwiększania oceny niepewności, która i tak w zagadnieniach ludnościowych jest bardzo duża. W. Lutz, W.C. Sanderson i S. Scherbov [18, s. 37] argumentują, że 80-procentowe przedziały są dużo bardziej odporne niż np. 95-procentowe.

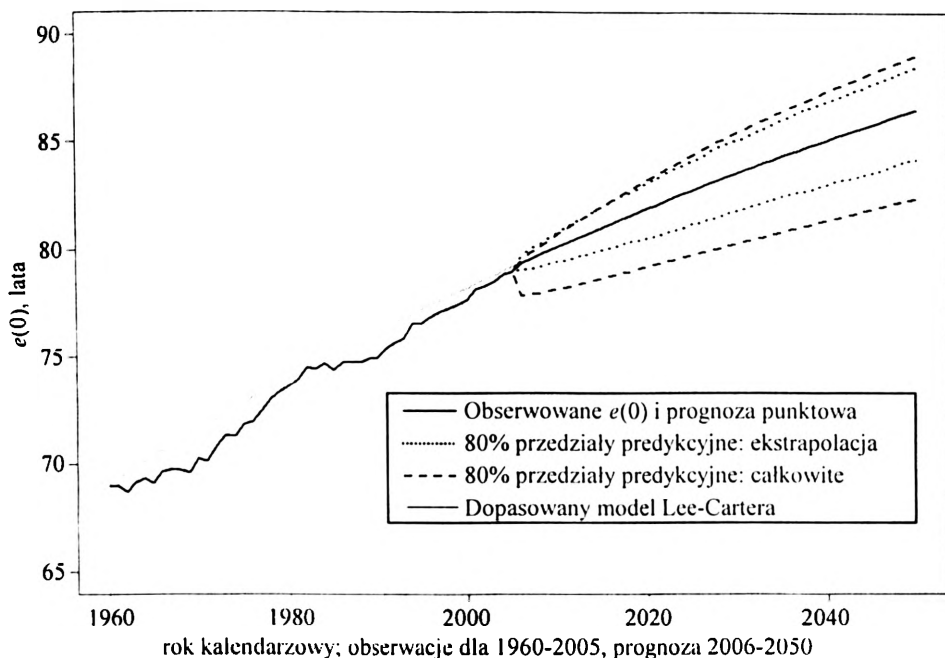


Rys. 3. Wzorce umieralności: logarytmy współczynników  $\mu(x, t)$  dla obu płci łącznie w Finlandii, 1960, 2005 i 2050

Źródło: obliczenia własne na podstawie danych z *Human Mortality Database* ([www.mortality.org](http://www.mortality.org)).

Z przedstawionych oszacowań wynika, że przeciętne trwanie życia w Finlandii dla obu płci łącznie wzrosło od poziomu około 69 lat w 1960 r. do 79 lat w 2005 r. Na podstawie uzyskanej prognozy spodziewany jest dalszy wzrost, do 86,5 roku na koniec horyzontu prognozy, z 80-procentowym przedziałem predykcyjnym wyznaczonym dla obu źródeł niepewności obejmującym wartości między 82,4 a 89 lat.

Na rysunku 4 widać, że dopasowanie modelu Lee i Cartera dla Finlandii wskazuje na przeszacowanie wartości  $e^P(0, t)$  w analizie *ex post*. Z drugiej strony warto nadmienić, że inne badania empiryczne wskazują, że model Lee i Cartera w ogólności raczej nie doszacowuje wartości przeciętnego trwania życia [2, s. 8]. Wspomniane rozbieżności mogą prawdopodobnie wynikać z dość dużej wrażliwości modelu Lee i Cartera na dobór oraz wielkość próby, na co zwrócili uwagę np. N. Keilman i D.Q. Pham [13]. Warto nadmienić, że z tego też względu do oceny niepewności prognoz bardziej odpowiednie są przedziały szersze, które uwzględniają błędy zarówno estymacji, jak i ekstrapolacji [13].

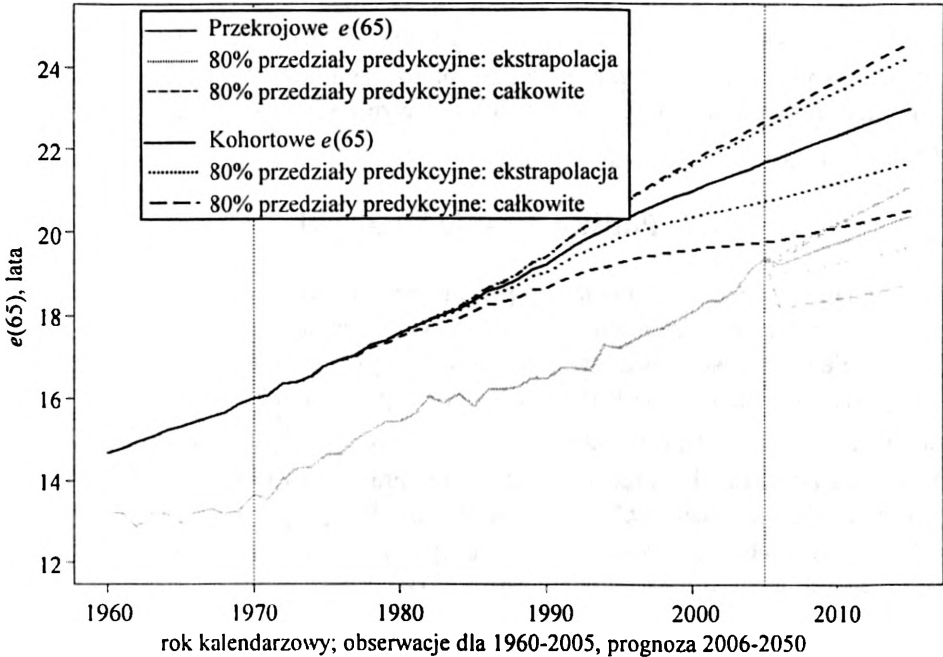


Rys. 4. Przekrojowe wskaźniki  $e^p(0, t)$  dla obu płci łącznie w Finlandii. 1960-2050

Źródło: obliczenia własne na podstawie danych z *Human Mortality Database* ([www.mortality.org](http://www.mortality.org)).

Porównanie wyników prognozy przeciętnego dalszego trwania życia osoby w wieku 65 lat w Finlandii dla lat 1960-2015, w ujęciu przekrojowym i kohortowym, zaprezentowane zostało na rys. 5 i w tab. 1. Ograniczenie horyzontu prezentacji do 2015 r. wynika z tego, że jest to ostatni rok, dla którego można wyliczyć współczynniki kohortowe na podstawie przekrojowych wielkości  $\mu(x, t)$  prognozowanych do 2050 r., w sytuacji, gdy  $x \leq 100$ . Ze względu na otwartą ostatnią grupę wieku (100 lat i więcej), uwaga o przybliżonym charakterze  $e^c(x, t)$ , poczyniona w poprzedniej sekcji, pozostaje w mocy. Warto zauważyć, że wielkości kohortowe pokazane na rys. 5 cechuje dużo bardziej regularne zachowanie niż ich podlegające większym periodycznym fluktuacjom przekrojowe odpowiedniki (szerzej na ten temat piszą np. F. van Poppel, E. Tabeau i F. Willekens, [25]).

Jak wynika z tab. 1, różnica pomiędzy  $e^c(65, t)$  a  $e^p(65, t)$  w 2005 r. sięga 2,3 roku, natomiast prognozowana na 2015 r. – ponad 2,6 roku przeciętnego dalszego trwania życia. Jak nietrudno zauważyć, wspomniane różnice mają fundamentalne znaczenie w kontekście zarządzania finansami zakładów emerytalnych. Co więcej, nawet „pesymistyczna” prognoza kohortowego przeciętnego dalszego trwania życia



Rys. 5. Przekrojowe i kohortowe wskaźniki  $e^P(65, t)$  oraz  $e^c(65, t)$  dla obu płci łącznie w Finlandii, 1960-2015

Źródło: obliczenia własne na podstawie danych z *Human Mortality Database* ([www.mortality.org](http://www.mortality.org)).

Tabela 1. Przekrojowe i kohortowe wskaźniki  $e^P(65, t)$  oraz  $e^c(65, t)$  dla obu płci łącznie w Finlandii: lata 1970, 1990, 2005 i 2015

Rok	Przekrojowe $e^P(65, t)$			Kohortowe $e^c(65, t)$		
	10%	centralne	90%	10%	centralne	90%
1970	–	13,67	–	–	15,98	–
1990	–	16,45	–	18,64*	19,21*	19,39*
2005	–	19,37	–	19,77*	21,67*	22,68*
2015	18,75*	20,38*	21,10*	20,51*	23,00*	24,58*

\* Gwiazdką oznaczono wartości wykorzystujące prognozy z modelu Lee i Cartera.

80-procentowe przedziały predycyjne odnoszą się do dwóch źródeł niepewności (estymacji i ekstrapolacji).

Źródło: obliczenia własne na podstawie danych z *Human Mortality Database* ([www.mortality.org](http://www.mortality.org)).

dla 2015 r., czyli dolna granica 80-procentowego przedziału predykcyjnego,  $e^c(65, 2015) = 20,51$  roku życia, jest wyższa niż prognozowana wartość centralna odpowiedniego współczynnika przekrojowego,  $e^P(65, 2015) = 20,38$  roku. Można zatem wnioskować, że z dużym prawdopodobieństwem mieszkańcy Finlandii będą w ciągu najbliższych lat żyć jeszcze dłużej, niż wynika z prognozy przekrojowych wartości  $e^P(x, t)$ .

## 5. Podsumowanie i wnioski

Porównanie wyliczonych wielkości przeciętnego dalszego trwania życia w ujęciu przekrojowym i kohortowym pozwala na stwierdzenie, że w sytuacji, gdy mamy do czynienia z obserwowanym i prognozowanym spadkiem umieralności, jak w pokazanym przykładzie dla Finlandii, z dużym prawdopodobieństwem zachodzi nierówność  $e^c(x, t) > e^P(x, t)$ . Dla obu mierników uwzględnienie dwóch źródeł niepewności, związanych z błędami oszacowań parametrów oraz prognoz czynnika  $\kappa_t$ , prowadziło do szerszych przedziałów predykcyjnych niż w przypadku uwzględnienia wyłącznie błędu wynikającego z ekstrapolacji. Potwierdza to zarówno wyniki wcześniejszych badań (np. [13]), jak i intuicję.

W przypadku kohortowych współczynników  $e^c(x, t)$  niepewność zaczyna odgrywać rolę jeszcze dla lat  $t$  z okresu próby. Wtedy to do obliczeń systematycznie zaczynają być wykorzystywane prognozowane wartości  $\mu(x, t)$  najpierw dla najstarszych, a potem również dla coraz młodszych grup wieku. W związku z tym kohortowe miary przeciętnego dalszego trwania życia charakteryzuje się większą niepewnością od przekrojowych, co jest nie tylko zgodne z intuicją, ale też zostało potwierdzone w pokazanym przykładzie empirycznym.

Z przedstawionych rozważań wypływają dwa podstawowe wnioski praktyczne:

1. Prognozowanie natężenia umieralności na podstawie odpowiednio zdefiniowanych danych pozwala na obliczenie przeciętnego dalszego trwania życia w ujęciu zarówno przekrojowym, jak i kohortowym. Do celów ubezpieczeniowych, w szczególności w przypadku produktów długoterminowych, wyznaczanie takich wartości, jak wysokość składek, wartość przyszłych świadczeń czy wysokość rezerw techniczno-ubezpieczeniowych, metodologicznie poprawne jest używanie mierników kohortowych, gdyż przekrojowe nie doszacowują rzeczywistego dalszego trwania życia w badanej zbiorowości w warunkach ciągłego spadku umieralności. Wiąże się to jednak, jak już wspomniano, ze zwiększeniem niepewności oszacowania.

2. Przedstawiony w pracy zmodyfikowany model Lee i Cartera pozwala na formalną analizę różnych źródeł niepewności co do przyszłych zmian przeciętnego (dalszego) trwania życia. Tworzenie przedziałów predykcyjnych ma bardzo istotne



znaczenie w zastosowaniach ubezpieczeniowych. Pozwala ono bowiem na wstępne oszacowanie antyselekcji ryzyka<sup>7</sup>, w tym przypadku przejawiającego się jako długowieczność, np. przez analizę różnicy między górną granicą przedziału predykcyjnego a prognozowaną wartością centralną.

Do dalszych kierunków badań związanych z prognozowaniem umieralności z wykorzystaniem modelu Lee i Cartera należą analizy innych postaci funkcji trendu  $\kappa_t$ . Ponadto warto również rozważyć zastosowanie modeli z bardziej skomplikowaną strukturą, np. z kilkoma komponentami typu  $b_x \cdot \kappa_t$  [2; 13; 26]. Tego typu analizy pozwoliłyby na częściowe uwzględnienie kolejnego źródła niepewności, odnoszącego się do specyfikacji modelu.

Ponadto, rozszerzenie niniejszej analizy w warstwie empirycznej stanowić mogą obliczenia przeciętnego dalszego trwania życia w Polsce, na podstawie dostępnych danych GUS pochodzących z rejestracji bieżącej. Analiza taka wymaga jednak przede wszystkim dokładnej weryfikacji jakości i kompletności danych, połączonej z ich ewentualnym uzupełnieniem i korektami, które to zagadnienia same w sobie były przedmiotem szerszego projektu badawczego. Ich wyniki zamieszczone zostały w raporcie B. Więckowskiej i J. Bijaka [30] przygotowanym pod redakcją naukową prof. T. Szumlicza.

## Podziękowania

Praca powstała w ramach projektu „Analiza ubezpieczeniowych implikacji wyników prognozy przeciętnego dalszego trwania życia uzyskanej metodą Lee i Cartera”, kierowanego przez prof. Tadeusza Szumlicza i przeprowadzanego na zlecenie Polskiej Izby Ubezpieczeń. Autorzy kierują wyrazy wdzięczności pod adresem prof. Juhy Alho oraz prof. Nico Keilmana za komentarze metodologiczne do wybranych fragmentów pracy.

## Literatura

- [1] Alho J., *Discussion of Lee*, „North American Actuarial Journal” 2000, vol. 4, no. 1, s. 91-93.
- [2] Alho J., *Methods used in drawing up mortality projections: predictive distributions of future mortality* [referat na „Fifteenth International Conference of Social Security Actuaries and Statisticians”, Helsinki, 23-25 maja 2007].
- [3] Andreev K.F., Vaupel J.W., *Forecasts of Cohort Mortality after Age 50*, MPIDR Working Paper 2006-012, Max Planck Institute for Demographic Research, Rostock 2006.

<sup>7</sup> Zgodnie z *Glossary of Insurance Terms* antyselekcja (selekcja negatywna) jest to tendencja do zawierania umów ubezpieczenia przez osoby, u których prawdopodobieństwo wystąpienia ubezpieczonego ryzyka jest wyższe niż przeciętne, tj. szkody występują częściej niż przeciętnie (*Glossary of Insurance Terms*, Merritt Publishing, Santa Monica, California 1997).

- [4] Boleslawski L., Tabeau E., *Comparing theoretical age patterns beyond the age of 80*. [w:] E. Tabeau, A. van den Berg Jeths, C. Heathcote (red.), *Forecasting Mortality in Developed Countries: Insights from a Statistical, Demographic and Epidemiological Perspective*, Kluwer Academic, Dordrecht 2001, s. 127-155.
- [5] Brouhns N., Denuit M., Vermunt J.K., *A Poisson log-bilinear regression approach to the construction of projected lifetables*, „Insurance: Mathematics and Economics 2002, vol. 31 no. 3, s. 373-393.
- [6] Bowers N.L. et al., *Actuarial Mathematics*, Society of Actuaries, USA, Schaumbury 1997.
- [7] Finetti B. de, *Matematica attuariale*, „Quaderni dell'Istituto per gli Studi Assicurativi”, Trieste 1950, vol. 5, s. 53-103.
- [8] Gerber H.U. *Life Insurance Mathematics*, Swiss Association of Actuaries, Springer, Berlin–Heidelberg 1997.
- [9] Gompertz B., *On the nature of the function expressive of the law of human mortality and on a new mode of determining life contingencies*, „Philosophical Transactions of the Royal Society of London” 1825, vol. 115, s. 513-585.
- [10] Haberman S., *Landmarks in the history of actuarial science (up to 1919)*, Actuarial Research Paper no. 84, Department of Actuarial Science and Statistics, City University, London 1996.
- [11] Heligman L., Pollard J.H., *The age pattern of mortality*, „Journal of the Institute of Actuaries” 1980, vol. 107, s. 49-75.
- [12] Imhoff E. van, *The exponential multidimensional demographic projection model*, „Mathematical Population Studies” 1990, vol. 2, no. 3, s. 171-182.
- [13] Keilman N., Pham D.Q., *Prediction intervals for Lee-Carter-based mortality forecasts*, referat na Europejską Konferencję Ludnościową w Liverpoolu, 21-24 czerwca 2006 r.
- [14] Kohler H.-P., Kohler I., *Frailty modelling for adult and old age mortality: the application of a modified de moivre hazard function to sex differentials in mortality*, „Demographic Research” 2000, vol. 3, no. 8, www.demographic-research.org.
- [15] Koissi M.-C., Shapiro A.F., Högnäs G., *Evaluating and extending the Lee-Carter model for mortality forecasting: Bootstrap confidence interval*, „Insurance: Mathematics and Economics” 2006, vol. 38, no. 1, s. 1-20.
- [16] Lee R.D., Carter L.R., *Modeling and forecasting the time series of U.S. mortality*, „Journal of the American Statistical Association” 1992, vol. 87, no. 419, s. 659-671.
- [17] Li N., Lee R.D., Tuljapurkar S., *Using the Lee-Carter method to forecast mortality for populations with limited data*, „International Statistical Review” 2004, vol. 72, no. 1, s. 19-36.
- [18] Lutz W., Sanderson W.C., Scherbov S. (red.), *The end of world population growth in the 21<sup>st</sup> century: new challenges for human capital formation and sustainable development*, Earthscan, London 2004.
- [19] Makcham W.M., *On the law of mortality*, „Journal of the Institute of Actuaries” 1860, vol. 13, s. 325-358.
- [20] Neill A., *Life Contingencies*, Heinemann, London 1989.
- [21] Paradysz J., *Ogólna charakterystyka zasad i metod analizy demograficznej. System współczynników demograficznych i zasady budowy tablic eliminacji*, [w:] M. Kędelski, J. Paradysz, Demografia, AE, Poznań 2006, s. 55-106.
- [22] Pitacco E., *Longevity risk in living benefits*, Working Paper 23/02, Center for Research on Pensions and Welfare Policies, Turin 2002.
- [23] Pitacco E., *Survival models in a dynamic context: a survey*, „Insurance: Mathematics and Economics” 2004, vol. 35, no. 2, s. 279-298.

- [24] Pitacco E., *From Halley to frailty: a review of survival models for actuarial calculations*. [manuskrypt], Department of Applied Mathematics, University of Trieste 2005, dostępny pod adresem: [ssrn.com/abstract=741586](http://ssrn.com/abstract=741586) (15 maja 2007).
- [25] Poppel F. van, Tabeau E., Willekens F., *Trends and sex-differentials in Dutch mortality since 1850: Insights from a cohort- and period-perspective*, *Genus* 1996, vol. LII, no. 3/4, s. 107-134.
- [26] Renshaw A.E., Haberman S., *A cohort-based extension to the Lee-Carter model for mortality reduction factors*, „Insurance: Mathematics and Economics” 2006, vol. 38, no. 3, s. 556-570.
- [27] Sverdrup E., *Basic concepts in life assurance mathematics*, „Skandinavisk Aktuarietidskrift” 1952, no. 3-4, s. 115-131.
- [28] Szumlicz T., *Ubezpieczenie społeczne: teoria dla praktyki*. Wydawnictwo Branta, Bydgoszcz–Warszawa 2005.
- [29] Tabeau E., *Czynniki kulturowe jako determinanty umieralności w demografii*, [w:] I.E. Kotowska (red.), *Przemiany demograficzne w Polsce w latach 90. w świetle drugiego przejścia demograficznego*. SGH, Warszawa 1999, s. 241-260.
- [30] Więckowska B., Bijak J., pod red. T. Szumlicza, *Analiza ubezpieczeniowych implikacji wyników prognozy przeciętnego dalszego trwania życia uzyskanej metodą Lee i Cartera*, Raport przygotowany na zlecenie Polskiej Izby Ubezpieczeń (PIU). Dodatek specjalny do „Wiadomości Ubezpieczeniowych” 2007, nr 11/12. Biblioteka Centrum Edukacji Ubezpieczeniowej, Fundacja Edukacji Ubezpieczeniowej, Warszawa 2007.
- [31] Wycinka E., *Tablica Ulpiana. Zapomniane początki tablic wymieralności*. „Studia Demograficzne” 2002, vol. 142, no. 2, s. 29-43.

## FORECASTING LIFE EXPECTANCY USING THE LEE–CARTER MODEL: SELECTED ISSUES

### Summary

The on-going process of population ageing, observed especially in the developed countries, is almost certainly going to enforce the reform of pension systems – a departure from the *pay-as-you-go* systems with defined benefits towards solutions based on defined contributions. Due to these changes, the height of the pension received will depend, among other factors, on the estimated life expectancy of the retired person.

Calculations of the height of an individual pension on the basis of current life tables, published by national statistical authorities, underestimate the risk of surviving subsequent years after retirement by the prospective pensioners. This is due to the fact that life tables are period-specific and do not take into account the lengthening of the human life-span for particular cohorts of population (groups of people born in the same year).

The paper aims at presenting the methodology of forecasting life expectancy using a modified version of the Lee and Carter (1992) model. Such forecast allows for considering the effect of population ageing related to the calendar year (period), age, as well as, albeit indirectly, the ageing of particular cohorts. In addition, the methodology proposed by Lee and Carter has a stochastic character and enables to accommodate the uncertainty of the forecasted phenomenon. The current paper takes into account uncertainty stemming from three sources: the Poisson variability of the numbers of deaths, estimation of the parameters of the model, as well as trend extrapolation. Theoretical discussion is illustrated with a forecast of life expectancy for Finland, prepared for the period 2005-2050.