Lee-Carterov model za projiciranje smrtnosti

Kostelac, Jelena Magdalena

Master's thesis / Diplomski rad

2017

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:217:847528

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-28**

Repository / Repozitorij:

Repository of the Faculty of Science - University of Zagreb

SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO–MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Jelena Magdalena Kostelac

LEE - CARTEROV MODEL ZA PROJICIRANJE SMRTNOSTI

Diplomski rad

Voditelj rada:
izv.prof.dr.sc. Miljenko Huzak

Zagreb, rujan, 2017

Ovaj diplomski rad obranjen je dana _____ pred ispitnim povjerenstvom u sastavu:

1. _____, predsjednik
2. _____, član
3. _____, član

Povjerenstvo je rad ocijenilo ocjenom _____.

Potpisi članova povjerenstva:

1. _____
2. _____
3. _____

Sadržaj

Sadržaj **iii**

Uvod **1**

1 Preliminarni dio **2**
1.1 Aktuarska notacija 2
1.2 Dekompozicija singularnih vrijednosti 5
1.3 ARIMA model 5
1.4 Newton-Raphsonova metoda 6

2 Opis modela **9**
2.1 Lee - Carterov model 9
2.2 Metoda najmanjih kvadrata 10
2.3 Poissonova metoda maksimalne vjerodostojnosti 15
2.4 Izglađivanje u Lee-Carterovom modelu 16
2.5 Odabir optimalnog perioda opažanja i analiza reziduala 17
2.6 Projiciranje smrtnosti 19

3 Cairns-Blake-Dowdov model i proširenje Lee-Carterovog modela **26**
3.1 Cairns-Blake-Dowd model 26
3.2 Lee - Carterov dobno-vremensko-kohortni model projekcije smrtnosti . . 28

4 Praktična primjena **34**
4.1 Primjer 34
4.2 Implementacija u R-u 40

Bibliografija **41**

Uvod

U ovom diplomskom radu detaljno ću opisati Lee - Carterov model za projiciranje smrtnosti. Glavna motivacija za projiciranje smrtnosti u novije doba su svakako životna osiguranja i mirovine. Aktuari zaposleni u osiguravajućim društvima i mirovinskim fondovima su nekoliko desetljeća kod projiciranja smrtnosti koristili tablice smrtnosti (životne tablice). Problem s kojim se trenutno suočavaju je dugovječnost ljudi, odnosno, činjenica da ljudi žive duže nego što je predviđeno tablicama smrtnosti. Ovaj trend rasta se događa u razvijenim zemljama, kao što su Velika Britanija i Belgija, sukladno s poboljšanjem uvjeta života, napredovanjem medicine i poboljšanjem standarda života u tim zemljama. Prognozama smrtnosti je, prije svega, nedostajala procjena brzine poboljšanja smrtnosti, tj. produljenja životne dobi. Prosječni životni vijek je u realizaciji nadmašio sve prognoze nastale polovicom prošlog stoljeća, unatoč brojnim pokušajima poboljšanja modela. Rezultat toga su gubici na poslovnom planu za osiguravajuća društva i same države (odnosno njihove mirovinske fondove), prisiljavajući ih na neplanirano alociranje kapitala u ta područja i dovodeći do smanjenja profitabilnosti. Glavni problem nastaje zbog činjenice da se rizik poboljšanja životnog vijeka ne može diverzificirati, te se stoga ne može kontrolirati standardnim mehanizmima, kao što je prodaja velikog broja polica, zbog činjenice da one nisu nezavisne. Pristup tom problemu je pronalazak boljeg modela za poboljšanje smrtnosti, onoga koji daje više pozornosti nivoima neizvjesnosti u prognozi. Jedan od takvih modela je svakako Lee - Carterov model. Njega su Ronald D. Lee i Lawrence Carter predstavili 1992. godine u članku "Modeliranje i predviđanje vremenskog niza smrtnosti SAD-a". Model je postao najrašireniji svjetska tehnika za predviđanje smrtnosti, koriste ga čak i Ujedinjeni Narodi te Cenzus Biro Sjedinjenih Američkih Država.

Poglavlje 1

Uvodni dio

1.1 Aktuarska notacija

Za početak uvodimo neke osnovne pojmove. Uzmimo osobu staru x godina. Tada je T_x njen budući životni vijek, to jest $x + T_x$ je godina života u kojoj osoba umire. T_x je slučajna varijabla s vjerojatnosnom funkcijom distribucije $G_x(t) = P(T_x \leq t),\ t \geq 0$. Funkcija $G_x(t)$ predstavlja vjerojatnost da će osoba umrijeti unutar t godina za svaki fiksni t. Pretpostavimo da je G_x apsolutno neprekidna vjerojatnosna funkcija distribucije. Neka je $g_x(t)$ njena vjerojatnosna funkcija gustoće. Stoga sa

$$\mathbb{P}\left(t < T_x \leq t + h\right) = \int_t^{t+h} g_x(s)\, ds,\ t \geq 0 \tag{1.1}$$

možemo označiti vjerojatnost da će osoba umrijeti u vremenskom intervalu od t do $t + h$, to jest, da će joj dob smrti biti između $x + t$ i $x + t + h$. Općeprihvaćena aktuarska notacija je

$${}_tq_x = G_x(t) \tag{1.2}$$

čime označavamo vjerojatnost da će osoba stara x godina umrijeti unutar narednih t godina. Sa

$${}_tp_x = 1 - G_x(t) \tag{1.3}$$

označimo vjerojatnost da će osoba koja je doživjela x godina preživjeti još barem t godina. Očekivani preostali životni vijek osobe stare x godina označavamo sa $\mathring{e}_x$. Definiramo ga sa:

$$\mathring{e}_x := \mathbb{E}\left(T_x\right) = \int_0^\infty t\, g_x(t)\, dt = \int_0^\infty \left[1 - G_x(t)\right] dt = \int_0^\infty {}_tp_x\, dt\,. \tag{1.4}$$

Intenzitet hazarda ili smrtnosti u dobi $x + t$ definiramo kao

$$\mu_{x+t} = \frac{g_x(t)}{1 - G_x(t)} = -\frac{d}{dt} \ln\left[1 - G_x(t)\right]\,. \tag{1.5}$$

Očekivano trajanje života pri rođenju se sada, kada smo definirali intenzitet, može zapisati i kao:

$$\mathring{e}_0 = \mathbb{E}(T_0) = \int_0^\infty t\, g_0(t)\, dt = \int_0^\infty t\, {}_tp_0\, \mu_t\, dt\,. \tag{1.6}$$

Također, sada iz 1.1, 1.3 i definicije intenizteta hazarda 1.5, vjerojatnost smrti u intervalu između t i $t + h$ možemo zapisati kao

$$\mathbb{P}(t < T_x \leq t + h) = \int_t^{t+h} {}_sp_x\, \mu_{x+s}\, ds\,. \tag{1.7}$$

Intenzitet hazarda također možemo definirati pomoću vjerojatnosti 1.3

$$\mu_{x+t} = -\frac{d}{dt} \ln {}_tp_x\,. \tag{1.8}$$

Integracijom ove tvrdnje (1.8) dobijemo

$${}_tp_x = e^{-\int_0^t \mu_{x+s}\, ds}\,. \tag{1.9}$$

[3]

Stope smrtnosti i izloženost riziku

Sada kada smo definirali T_x dodajemo indeks t koji predstavlja kalendarsko vrijeme. Dakle, ako je $x + T_x(t)$ dob u kojoj osoba umire, onda je $t + T_x(t)$ kalendarska godina u kojoj će umrijeti. Intenzitet hazarda u dobi x i vremenu t definiramo kao

$$\mu_x(t) = \lim_{\Delta \searrow 0} \frac{\mathbb{P}(x < T_0(t-x) \leq x + \Delta \,|\, T_0(t-x) > x)}{\Delta}\,. \tag{1.10}$$

Sada za svaku dob x i kalendarsku godinu t pretpostavimo da vrijedi

$$\mu_{x+\xi_1}(t + \xi_2) = \mu_x(t), \text{ za } 0 \leq \xi_1, \xi_2 < 1\,. \tag{1.11}$$

Ovu pretpostavku najbolje ilustrira Lexisov dijagram, koordinatni sistem kojemu je x-os kalendarsko vrijeme, a y-os dob. Obje vremenske skale se dijele u godišnje grupe, što particionira Lexisovu ravninu na kvadratne elemente. Pretpostavka 1.11 pretpostavlja da su stope smrtnosti konstantne unutar svakog kvadrata, ali mogu varirati među kvadratima. Uz 1.11 slijedi

$$p_x(t) = \exp\left(-\int_0^1 \mu_{x+\xi}(t + \xi)\, d\xi\right) = \exp(-\mu_x(t))\,. \tag{1.12}$$

Ovdje je $p_x(t)$ vjerojatnost da osoba stara x godina u kalendarskoj godini t, doživi dob od $x + 1$ godine. Sukladno tome, $q_x(t)$ je vjerojatnost da osoba stara x godina umre u kalendarskoj godini t prije nego doživi $x + 1$ godinu. Točnije

$$q_x(t) = \mathbb{P}\left(T_x(t) \leq 1\right) \text{ i } p_x(t) = 1 - q_x(t) = \mathbb{P}\left(T_x(t) > 1\right). \tag{1.13}$$

Izloženost riziku, ETR_{xt} (eng. *exposure − to − risk*) u dobi x, uz uvjet da je zadnji rođendan osoba imala u kalendarskoj godini t, je ukupno trajanje života osoba starih x godina sa rođendanom u kalendarskoj godini t. Sljedeća formula za izloženost riziku vrijedi uz uvjet 1.11. Neka je L_{xt} broj osoba starih x godina na zadnji rođendan 1. siječnja godine t. Slijedi

$$\mathbb{E}\left[ETR_{xt} \,|\, L_{xt} = l\right] = l \int_{\xi=0}^{1} (p_x(t))^{\xi} d\xi = -\frac{l}{\mu_x(t)} \left(1 - p_x(t)\right) = \frac{-l\, q_x(t)}{\ln(1 - q_x(t))}. \tag{1.14}$$

Dakle, aproksimativno je

$$ETR_{xt} \approx \frac{-L_{xt}\, q_x(t)}{\ln(1 - q_x(t))}. \tag{1.15}$$

Sada definirajmo funkciju δ_i kao

$$\delta_i = \begin{cases} 1, & \text{ako osoba } i \text{ umre u dobi } x \\ 0, & \text{inače} \end{cases} \tag{1.16}$$

za $i = 1, 2, \ldots, L_{xt}$. Nadalje, neka je τ_i dio godine t u kojoj je živjela osoba i te neka je D_{xt} broj zabilježenih smrti osoba dobi x koje su zadnji rođendan imale u kalendarskoj godini t. Točnije, D_{xt} obuhvaća sve osobe rođene od 1. siječnja kalendarske godine $t - x - 1$ do 31. prosinca kalendarske godine $t - x$. Slijedi

$$D_{xt} := \sum_{i=1}^{L_{xt}} \delta_i \tag{1.17}$$

i

$$ETR_{xt} := \sum_{i=1}^{L_{xt}} \tau_i. \tag{1.18}$$

Pod pretpostavkama 1.11 i 1.12 slijedi da je vjerodostojnost od parametra $\mu_x(t)$ na osnovi podataka (τ_i, δ_i) jednaka

$$p_x(t) = \exp(-\mu_x(t)) \tag{1.19}$$

ako osoba i preživi te

$$p_x(t)^{\tau_i}\, \mu_{x+\tau_i}(t + \tau_i) = \exp(-\tau_i\, \mu_x(t))\mu_x(t) \tag{1.20}$$

ako osoba i umre u vremenu τ_i tokom godine t. Iz 1.19 i 1.20 slijedi da je vjerodostojnost

$$\exp\left(-\tau_i\,\mu_x(t)\right)\left(\mu_x(t)\right)^{\delta_i}. \tag{1.21}$$

Ako su vremena života pojedinaca nezavisna, vjerodostojnost od $\mu_x(t)$ na osnovi opažanja grupe L_{xt} osoba starih x godina jednaka je

$$L(\mu_x(t)) = \prod_{i=1}^{L_{xt}} \exp\left(-\tau_i\,\mu_x(t)\right)\left(\mu_x(t)\right)^{\delta_i} = \exp\left(-\mu_x(t)\,ETR_{xt}\right)\left(\mu_x(t)\right)^{D_{xt}}. \tag{1.22}$$

Sada nađemo točku maksimuma dane funkcije vjerodostojnosti od L, u oznaci $\hat{\mu}_x(t)$, koji je procjenitelj od $\mu_x(t)$

$$\hat{\mu}_x(t) = \frac{D_{xt}}{ETR_{xt}} = \hat{m}_x(t). \tag{1.23}$$

$\hat{m}_x(t)$ su neizglađene stope smrtnosti za dob x u kalendarskoj godni t. [2]

1.2 Dekompozicija singularnih vrijednosti

Dekompozicija singularnih vrijednosti (eng. *Singular Value Decomposition*), ili skraćeno SVD, je dekompozicija realne ili kompleksne matrice koja ima široku primjenu. Koristi se kod rangiranja web stranica (Google), kompresije slika, problema najmanjih kvadrata i u brojnim drugim područjima. Osnovni teorem o ovoj dekompoziciji glasi

Teorem 1. *Neka je* $\mathbf{A}$ *proizvoljna matrica tipa mxn , uz* $m \geq n$ *.* $\mathbf{A}$ *se može dekomponirati kao* $\mathbf{A} = \hat{\mathbf{U}}\begin{bmatrix}\mathbf{\Sigma}\\ 0\end{bmatrix}\mathbf{V}^* = U\mathbf{\Sigma}\mathbf{V}^*$*, gdje je* $\Sigma = diag(\sigma_1, \cdots, \sigma_n)$*, uz* $\sigma_1 \geq \cdots \geq \sigma_n \geq 0$*, a* $\hat{\mathbf{U}} = [U, U_0]$ *je unitarna matrica reda m, a* $\mathbf{V}$ *je unitarna matrica reda n.*

Stupce matrice $\hat{\mathbf{U}}$ zovemo lijevi singularni vektori (oznaka u_i), a stupce matrice $\mathbf{V}$ desni singularni vektori (oznaka v_i). Dijagonalne elemente σ_i matrice Σ zovemo singularne vrijednosti. [5]

1.3 ARIMA model

ARIMA (eng. *Autoregressive Integrated Moving Average*) model je generalizacija ARMA (eng. *Autoregressive Moving Average*) modela. ARMA(p, q) model definiramo eksplicitno sa:

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \epsilon_t + \beta_1\epsilon_{t-1} + \cdots + \beta_q\epsilon_q, \tag{1.24}$$

gdje je ϵ_t bijeli šum, nekoreliran vremenski niz s očekivanjem 0 i konstantnom varijancom σ^2, u oznaci $WN(0, \sigma^2)$. Formulu ARMA procesa (X_t) možemo zapisati i kao

$$\left(1 - \sum_{k=1}^{p} \alpha_k L^k\right) X_t = \left(\sum_{k=1}^{q} \beta_q L^k\right) \epsilon_t \tag{1.25}$$

gdje je L operator pomaka definiran sa

$$L\, X_t = X_{t-1}\,. \tag{1.26}$$

Potencije operatora L definiramo sa

$$L^k\, X_t = X_{t-k}\,. \tag{1.27}$$

ARIMA model se korisiti prilikom analize vremenskih nizova, pretežito u statistici i ekonometriji, kod određivanja modela za dane podatke te predviđanja budućih vrijednosti danog vremenskog niza. Parametri ARIMA(p, d, q) modela su redom p koji određuje stupanj autoregresivnog (AR) modela, parametar q koji je stupanj moving-average (MA) modela te d koji je stupanj diferenciranja podataka u modelu. Model je dan formulom koja slijedi iz 1.25 :

$$\left(1 - \sum_{k=1}^{p} \alpha_k L^k\right) (1 - L)^d\, X_t = \left(\sum_{k=1}^{q} \beta_q L^k\right) \epsilon_t\,. \tag{1.28}$$

Operator diferenciranja definiramo sa

$$\Delta X_t = X_t - X_{t-1} = (1 - L)\, X_L \tag{1.29}$$

te za svaki $d \in \mathbb{N}$

$$\Delta^d\, X_t = (1 - L)^d\, X_t\,. \tag{1.30}$$

[1]

1.4 Newton-Raphsonova metoda

Newton-Raphsonova metoda (poznata kao i Newtonova metoda) koristi se u numeričkoj analizi kako bismo pronašli aproksimacije nultočki realnih funkcija, odnosno našli x koji zadovoljava

$$f(x) = 0. \tag{1.31}$$

Ime je dobila po matematičaru i fizičaru Isaacu Newtonu te matematičaru Josephu Raphsonu. Metoda kreće od inicijalne točke x_0, koja je procjena nultočke, odnosno, točke za

koju procijenimo da je blizu nultočke funkcije. Neka je r nultočka funkcije f i neka je $h = r - x_0$ veličina koja pokazuje koliko je procjena x_0 daleko od nultočke r.

$$0 = f(r) = f(x_0 + h) \approx f(x_0) + h\, f'(x_0). \tag{1.32}$$

Jednadžba 1.32 slijedi iz Taylorovog teorema.

Teorem 2. *Neka je f funkcija koja je neprekidno diferencijabilna k+1 puta na otvorenom intervalu I. Za bilo koje točke x i x+h tada postoji točka w između x i x+h takva da vrijedi* $f(x + h) = f(x) + f'(x)\, h + \frac{1}{2} f''(x)\, h^2 + \ldots + \frac{1}{k!} f^{[k]}(x)\, h^k + \frac{1}{(k+1)!} f^{[k+1]}(w)\, h^{k+1}$.

Slijedi da za male vrijednosti h vrijedi:

$$f(x + h) \approx f(x) + f'(x)\, h\,. \tag{1.33}$$

Stoga, osim ako je $f'(x_0)$ blizu 0, slijedi da je

$$r = x_0 + h \approx x_0 - \frac{f(x_0)}{f'(x_0)}\,. \tag{1.34}$$

Ako je početna točka blizu nultočke, odnosno metoda konvergira na intervalu na kojem se nalazi početna točka, te je zadovoljen uvjet da je derivacija u toj točki različita od nule, bolja aproksimacija x_1 je dana sa

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}\,. \tag{1.35}$$

Metodu iterativno ponavljamo dok ne dostignemo prethodno zadanu točnost:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \tag{1.36}$$

Za funkciju više varijabli, odnosno $f : \mathbf{R}^n \to \mathbf{R}^n$ iterativna formula metode je:

$$x_{n+1} = x_n + [J_f(x_n)]^{-1} f(x_n)\,, \tag{1.37}$$

gdje je J_f Jacobijeva matrica čiji su elementi dani sa

$$J_{f_{ij}} = \frac{\partial f_i}{\partial y_j}. \tag{1.38}$$

Kod optimizacije funkcije, to jest pronalaska točke minimuma funkcije, postupak je sljedeći. Znamo da je nužan uvjet minimuma

$$\nabla f(x) = 0 \tag{1.39}$$

te se tada provodi Newton-Raphsonovu metoda na $\nabla f(x)$. Iterativna formula metode je

$$x_{n+1} = x_n - [\nabla^2 f(x_n)]^{-1} \nabla f(x_n)\,, \tag{1.40}$$

gdje je $\nabla^2 f(x_n)$ Hesseova matrica funkcije f, to jest,

$$[\nabla f]_{ij} = \frac{\partial^2 f}{\partial y_i \partial y_j}. \tag{1.41}$$

Newton-Raphsonova metoda se u statistici često koristi za optimizaciju funkcije vjerodostojnosti, to jest, prilikom nalaženja točke maksimuma funkcije vjerodostojnosti (eng. *Maximum Likelihood Estimation*, MLE). Neka je $y = (y_1, y_2, \ldots, y_n)$ vektor nezavisnih jednako distribuiranih varijabli i neka je θ $p-$dimenzionalni parametar, $\theta = (\theta_1, \ldots, \theta_p)$, gdje je $p \leq n$. Neka je $f(y \,|\, \theta)$ funkcija gustoće od y. Tada je funkcija vjerodostojnosti od θ dana sa:

$$L(\theta) = \prod_{i=1}^{n} f(y_i \,|\, \theta). \tag{1.42}$$

Logaritamska funkcija vjerodostojnosti je dana sa:

$$l(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log(f(y_i \,|\, \theta)). \tag{1.43}$$

Neka je u $(p \times 1)-$dimenzionalni vektor definiran sa:

$$u(\theta) = \frac{\partial l(\theta)}{\partial \theta} = \begin{bmatrix} \frac{\partial l(\theta)}{\partial \theta_1} \\ \vdots \\ \frac{\partial l(\theta)}{\partial \theta_p}. \end{bmatrix} \tag{1.44}$$

Vektor $u(\theta)$ zovemo eng. *score vector* logaritamske funkcije vjerodostojnosti. Očekivanje $u(\theta)$ u odnosu na y je jednako nuli i varijanca od $u(\theta)$ je jednaka:

$$\mathbf{V}ar(u(\theta)) = \mathbf{E}\left[u(\theta)\, u(\theta)^T\right] = -\mathbf{E}\left[\frac{\partial^2 l(\theta)}{\partial \theta_j \, \partial \theta_k}\right]. \tag{1.45}$$

Matricu

$$I(\theta) = -\mathbb{E}\left[\frac{\partial^2 l(\theta)}{\partial \theta_j \partial \theta_k}\right] \tag{1.46}$$

zovemo Fisherova informacijska matrica. Sada se Newton-Raphsonova metoda primjenjuje kako bismo našli točku maksimuma $\hat{\theta}$ funkcije vjerodostojnosti $L(\theta)$ koristeći 1.40:

$$\hat{\theta}_{n+1} = \hat{\theta}_n - H(\hat{\theta}_n)^{-1} u(\hat{\theta}_n)\,, \tag{1.47}$$

gdje je H Hesseova matrica. Formulu 1.47 možemo zamijeniti i sa

$$\hat{\theta}_{n+1} = \hat{\theta}_n - I(\hat{\theta}_n)^{-1} u(\hat{\theta}_n). \tag{1.48}$$

[4]

Poglavlje 2

Lee - Carterov model

2.1 Model

Lee - Carterov model je značajan primjer stohastičkog pristupa predviđanju smrtnosti. Samom modelu su dodana brojna poboljšanja, kao što su, na primjer, obuhvaćanje korelacije muške i ženske populacije. Neka je $m_x(t)$ središnja stopa smrtnosti za dob x i vrijeme t. Model ima log-bilinearnu (logaritamsku-bilinearnu po bazi e ; ln-bilinearnu) formu oblika:

$$\ln m_x(t) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t} \tag{2.1}$$

gdje α_x predstavlja dobni obrazac smrtnosti uprosječen po vremenu (dobne stope smrtnosti), a β_x opisuje odstupanja od prosjeka kada κ_t varira. Promjenu u stopi smrtnosti kroz vrijeme opisuje univarijatni indeks smrtnosti κ_t. $\epsilon_{x,t}$ opisuje šum (grešku) koja je normalno distribuirana sa očekivanjem 0 i varijancom σ^2 i označava povijesni utjecaj koji nije obuhvaćen modelom (gaussovski bijeli šum). [2]
U procjeni modela metodom najmanjih kvadrata koristi se SVD dekompozicija (1.2) tako što uzimamo prvi element dekompozicije kao rješenje. Sustav 2.1 je neodređen ako ne uvedemo dodatna ograničenja. Lee i Carter stoga predlažu normalizaciju danu sa:

$$\sum_x \beta_x = 1 \text{ i } \sum_t \kappa_t = 0 \tag{2.2}$$

zbog čega je svaki α_x prosjek log-središnjih stopa smrtnosti tokom kalendarskih godina. Jednom kada procijenimo parametre α_x, β_x i κ_t i dobijemo procijenjene vrijednosti $\hat{\alpha}_x, \hat{\beta}_x$ i $\hat{\kappa}_t$ za svaki t smrtnost predviđamo modelirajući vrijednosti κ_t pomoću vremenskih nizova. Na primjer, pretpostavimo da su one slučajna šetnja s pomakom. Počevši od danog t' za svaki $t > t'$ predviđamo stope smrtnosti kao u nastavku:

$$m_x(t) = \exp\left(\hat{\alpha}_x + \hat{\beta}_x \kappa_t\right) = m_x(t') \exp\left[\hat{\beta}_x(\kappa_t - \hat{\kappa}_{t'})\right]. \tag{2.3}$$

Bitno je napomenuti da se $m_x(t)$ modelira kao stohastički proces jer proizlazi iz stohastičkih procesa κ_t te stoga za projicirane vrijednosti stopa smrtnosti možemo izračunati i pripadne intervale predviđanja. Interpretacija parametara je sljedeća: $\exp \alpha_x$ je opći oblik kretanja smrtnosti, dok se stvarni zakon promjene smrtnosti mijenja prema indeksu smrtnosti κ_t sukladno s parametrom β_x (ovisno o dobi). Promjena β_x nam govori koje stope opadaju brzo, a koje sporije tokom vremena kao reakcija na promjenu parametra κ_t. Dakle, β_x predstavlja obrazac promjene smrtnosti ovisno o dobi i ukazuje na osjetljivost logaritma zakona smrtnosti za dob x za varijacije u vremenu indeksa κ_t. β_x čak može biti i negativan za neke dobi x, ukazujući da smrtnost u toj dobi raste, dok je za neke dobi isto tako moguće da opada. U praksi, ovo se događa rijetko, većinom za najstarije životne dobi ili izvanredne slučajeve, kao što je na primjer negativan β_x za mlade muškarce u određenim industrijaliziranim zemljama izazvan povećanom stopom smrtnosti uzrokovanom AIDS-om u kasnim 80-im i 90-im godinama prošlog stoljeća [2].
Pretpostavimo da imamo opažene podatke za niz kalendarskih godina $t = t_1, t_2, \ldots, t_n$ i za niz dobi $x = x_1, x_2, \ldots, x_m$. Za dana opažanja želimo procijeniti odgovarajuće parametre α_x, β_x i κ_t. Za početak moramo uvesti dodatna ograničenja na parametre. U 2.1 parametre α_x razlikujemo za aditivnu konstantu, β_x za multiplikativnu konstantu, a κ_t do na sve linearne transofrmacije. To može dovesti do toga da ne razlikujemo dvije različite parametrizacije, odnosno, možemo dobiti isti rezultat za različite parametre. Na primjer, neka je, $\alpha'_x = \alpha_x - 2\beta_x$, $\beta'_x = 2\beta_x$ i $\kappa'_t = \frac{\kappa_t}{2} + 1$. Iako parametri nisu jednaki kao u 2.1 dobit ćemo iste rezultate za stope smrtnosti nakon kraćenja. Stoga se parametri normaliziraju kao 2.2. Iako je to najraširenija pristup, ponekad se koriste djelomično izmijenjena ograničenja kao što su

$$\kappa_{t_n} = 0 \text{ i } \sum_{x=x_1}^{x=x_m} \beta_x^2 = 1 \,. \tag{2.4}$$

Bitno je napomenuti da ovo nije stvarni problem, jer parametrizacije koje daju jednake rezultate samo dovode do beskonačnog broja ekvivalentnih maksimuma. Svaki od tih maksimuma nam daje identičnu projekciju. Stoga nam prihvaćanje ograničenja 2.2 samo omogućava da izaberemo jedan od ekvivalentnih maksimuma, radi jednostavnosti. Izbor ograničenja nema utjecaja na kvalitetu predviđanja smrtnosti, odnosno 2.2 je nad ostalim ograničenjima (npr. 2.4) izabran zbog jednostavnosti.

2.2 Procjena metodom najmanjih kvadrata

Model koji se uobičajeno koristi za procjenu α_x, β_x i κ_t je

$$\ln \hat{m}_x(t) = \alpha_x + \beta_x \, \kappa_t + \epsilon_x(t) \tag{2.5}$$

za $x = x_1, x_2, \ldots, x_m$ i $t = t_1, t_2, \ldots, t_n$, gdje $\hat{m}_x(t)$ označava ranije spomenute opažene stope smrtnosti u dobi x tokom godine t (1.23), a $\epsilon_x(t)$ homoskedastične centrirane greške. Ho-

moskedastičnost znači da sve varijable u nizu imaju jednaku konačnu varijancu, odnosno $\mathbb{V}ar\left[\epsilon_x(t)\right] = \sigma_\epsilon^2$, za svaki t. Greška je i centrirana pa vrijedi $\mathbb{E}\left[\epsilon_x(t)\right] = 0$. Ponekad je pretpostavka o homoskedastičnosti grešaka upitna jer je prirodni logaritam opaženih stopa smrtnosti uobičajeno više varijabilan u starijoj, nego u mlađoj dobi. Homoskedastičnost ne predstavlja problem ako aktuari ograniče raspon godina za modeliranje do 50 i više, kako bi se izbjegla nestabilnost oko slučajnih skokova. Ovaj model je posebno koristan kada aktuari imaju samo skup stopa smrtnosti $\hat{m}_x(t)$ na raspolaganju za predviđanje. Kada su im dostupni dodatni podaci, kao što su opažanja o broju smrti i izloženost riziku, Poissonov model opisan u idućem poglavlju uspješno ubacuje dodatne podatke u model. Model 2.5 prilagođavamo matrici opaženih stopa smrtnosti po godinama pomoću SVD (1.2). $\hat{\alpha}_x, \hat{\beta}_x$ i $\hat{\kappa}_x$ su takvi da minimiziraju danu kvadratnu funkciju:

$$\theta_{LS}(\alpha, \beta, \kappa) = \sum_{x=x_1}^{x_m} \sum_{t=t_1}^{t_n} \left(\ln \hat{m}_x(t) - \alpha_x - \beta_x \kappa_t\right)^2 , \tag{2.6}$$

gdje je LS engleska kratica za Least Squares, odnosno najmanje kvadrate. Minimizacija ove funkcije ekvivalentna je procjeni maksimalne vjerodostojnosti, pod uvjetom da $\epsilon_x(t)$ dolaze iz normalne distribucije, odnosno greške su gaussovske. 1993. godine Wilmoth je predložio nadopunu metode predlažući težinsku funkciju najmanjih kvadrata umjesto obične. Funkcija je tada oblika

$$\theta_{WLS}(\alpha, \beta, \kappa) = \sum_{x=x_1}^{x_m} \sum_{t=t_1}^{t_n} w_{xt} \left(\ln \hat{m}_x(t) - \alpha_x - \beta_x \kappa_t\right)^2 . \tag{2.7}$$

Empirijske studije su pokazale da kada se za težine w_{xt} uzmu d_{xt}, odnosno broj opaženih smrti za dane x i t, procijenjeni parametri su jako blizu parametrima dobivenih Poissonovom procjenom. Doduše, ovo nije idealna opcija jer je broj smrti slučajna varijabla.
Sljedeći korak kod računanja procijenjenih parametara je pronalazak točke minimuma dane funkcije. Ona mora zadovoljavati jednadžbu

$$\frac{\delta}{\delta \alpha_x} \theta_{LS} = 0 , \tag{2.8}$$

odnosno

$$\sum_{t=t_1}^{t_n} \ln \hat{m}_x(t) = (t_n - t_1 + 1)\alpha_x + \beta_x \sum_{t=t_1}^{t_n} \kappa_t . \tag{2.9}$$

Iz uvjeta normalizacije 2.2 mora biti $\sum_{t=t_1}^{t_n} \kappa_t = 0$ pa stoga slijedi

$$\hat{\alpha}_x = \frac{1}{t_n - t_1 + 1} \sum_{t=t_1}^{t_n} \ln \hat{m}_x(t) . \tag{2.10}$$

Dakle, minimizacija 2.6 se sastoji od uzimanja prosjeka po x-tom retku $\ln \hat{m}_x(\cdot)$ za procijenjeni $\hat{\alpha}_x$., dolje navedene matrice $\ln \mathbf{M}$. Kada model 2.5 prilagođavamo običnom metodom najmanjih kvadrata procijenjena vrijednost α_x je jednaka prosjeku $\ln \hat{m}_x$ kroz vrijeme t. Tada $\exp \alpha_x$ predstavlja opći oblik raspodjele smrtnosti. Sada $\hat{\beta}_x$ i $\hat{\kappa}_t$ možemo dobiti pomoću dekompozicije singularnih vrijednosti matrice $\ln \hat{\mathbf{M}} - \hat{\alpha}$ definirane kako slijedi. Stope smrtnosti povežemo u matricu

$$\mathbf{M} := \begin{bmatrix} m_{x_1}(t_1) & \dots & m_{x_1}(t_n) \\ \vdots & \ddots & \vdots \\ m_{x_m}(t_1) & \dots & m_{x_m}(t_n) \end{bmatrix}. \tag{2.11}$$

Matrica $\mathbf{M}$ ima dimenziju $(x_m - x_1 + 1) \times (t_n - t_1 + 1)$. Sljedeći korak je kreiranje matrice $\mathbf{Z}$, definirane sa:

$$\mathbf{Z} = \ln \hat{\mathbf{M}} - \hat{\alpha} = \begin{bmatrix} \ln \hat{m}_{x_1}(t_1) - \hat{\alpha}_{x_1} & \dots & \ln \hat{m}_{x_1}(t_n) - \hat{\alpha}_{x_1} \\ \vdots & \ddots & \vdots \\ \ln \hat{m}_{x_m}(t_1) - \hat{\alpha}_{x_m} & \dots & \ln \hat{m}_{x_m}(t_n) - \hat{\alpha}_{x_m} \end{bmatrix}. \tag{2.12}$$

$\mathbf{Z}$ ima dimenziju istu kao $\mathbf{M}$, tj. $(x_m - x_1 + 1) \times (t_n - t_1 + 1)$. Pretpostavljamo nepostojanje dobno-vremenske interakcije što znači da su β_x fiksni kroz vrijeme i κ_t fiksni kroz godine. Slijedi, $\hat{\beta}_x$ i $\hat{\kappa}_t$ su takvi da minimiziraju

$$\tilde{\theta}_{LS}(\beta, \kappa) = \sum_{x=x_1}^{x_m} \sum_{t=t_1}^{t_n} (z_{xt} - \beta_x \kappa_t)^2 \ . \tag{2.13}$$

Rješenje dobivamo dekompozicijom singularnih vrijednosti na sljedeći način. Neka je $\mathbf{Z}\mathbf{Z}^\mathbf{T}$ kvadratna matrica dimenzije $(t_n - t_1 + 1) \times (t_n - t_1 + 1)$ i $\mathbf{Z}^\mathbf{T}\mathbf{Z}$ kvadratna matrica dimenzije $(x_m - x_1 + 1) \times (x_m - x_1 + 1)$. Neka je $\mathbf{u}_1$ svojstveni vektor koji odgovara najvećoj svojstvenoj vrijednosti matrice $\mathbf{Z}^\mathbf{T}\mathbf{Z}$. Isto tako, neka je $\mathbf{v}_1$ svojstveni vektor koji odgovara najvećoj svojstvenoj vrijednosti matrice $\mathbf{Z}\mathbf{Z}^\mathbf{T}$. Najbolja aproksimacija $\mathbf{Z}$ u smislu metode najmanjih kvadrata je

$$\mathbf{Z} \approx \mathbf{Z}^* = \sqrt{\lambda_1} \mathbf{v}_1 \mathbf{u}_1^T \, , \tag{2.14}$$

gdje je λ_1 najveća svojstvena vrijednost pripadne dekompozicije. Iz 2.14 slijedi

$$\hat{\beta} = \frac{\mathbf{v}_1}{\sum_{j=1}^{x_m - x_1 + 1} v_{1j}} \tag{2.15}$$

i

$$\hat{\kappa} = \sqrt{\lambda_1} \left(\sum_{j=1}^{x_m - x_1 + 1} v_{1j} \right) \mathbf{u}_1 \tag{2.16}$$

uz uvjet da je $\sum_{j=1}^{x_m - x_1 + 1} v_{1j} \neq 0$. Ograničenja 2.2 su tada zadovoljena za $\hat{\beta}_x$ i $\hat{\kappa}_t$. Drugi najveći svojstveni vektor (onaj uz drugu najveću svojstvenu vrijednost) i ostali redom, kod dekompozicije, sadrže reziduale. Kod populacija s niskim mortalitetom 2.14 objašnjava oko 95% varijance $\ln \hat{m}_x(t)$.

Newton-Raphsonova metoda

U primjeni na Lee-Carterov model prvi korak Newton-Raphsonove metode je parcijalno deriviranje funkcije $\theta_{LS}(\alpha, \beta, \kappa)$ (2.6) po varijablama α_x, β_x i κ_t kao što je prethodno objašnjeno u opisu metode (1.4) :

$$\begin{aligned} 0 &= \sum_{t=t_1}^{t_n} (\ln \hat{m}_x(t) - \alpha_x - \beta_x \kappa_t), \;\; x = x_1, x_2, \ldots, x_m \\ 0 &= \sum_{x_1}^{x_m} \beta_x \, (\ln \hat{m}_x(t) - \alpha_x - \beta_x \kappa_t), \;\; t = t_1, t_2, \ldots, t_n \\ 0 &= \sum_{t=t_1}^{t_n} \kappa_t \, (\ln \hat{m}_x(t) - \alpha_x - \beta_x \kappa_t), \;\; x = x_1, x_2, \ldots, x_m. \end{aligned} \tag{2.17}$$

Svaka od ovih jednadžbi oblika je $f(\xi) = 0$, gdje ξ predstavlja jedan od parametara α_x, β_x i κ_t. Počevši od inicijalne vrijednosti ξ^0, $k+1$- vu iteraciju dobijemo iz formule:

$$\xi^{(k+1)} = \xi^k - \frac{f(\xi^{(k)})}{f'(\xi^k)}. \tag{2.18}$$

Dakle, parametre dobijemo iz sljedećih iterativnih relacija:

$$\begin{aligned} \hat{\alpha}_x^{(k+1)} &= \hat{\alpha}_x^{(k)} + \frac{\sum_{t=t_1}^{t_n} (\ln \hat{m}_x(t) - \hat{\alpha}_x^{(k)} - \hat{\beta}_x^{(k)} \, \hat{\kappa}_t^{(k)})}{t_n - t_1 + 1} \\ \hat{\kappa}_t^{(k+1)} &= \hat{\kappa}_t^{(k)} + \frac{\sum_{x=x_1}^{x_m} \hat{\beta}_x^{(k)} \, (\ln \hat{m}_x(t) - \hat{\alpha}_x^{(k+1)} - \hat{\beta}_x^{(k)} \, \hat{\kappa}_t^{(k)})}{\sum_{x=x_1}^{x_m} (\hat{\beta}_x^{(k)})^2} \\ \hat{\beta}_x^{(k+1)} &= \hat{\beta}_x^{(k)} + \frac{\sum_{t=t_1}^{t_n} \hat{\kappa}_t^{(k+1)} \, (\ln \hat{m}_x(t) - \hat{\alpha}_x^{(k+1)} - \hat{\beta}_x^{(k)} \, \hat{\kappa}_t^{(k)})}{\sum_{t=t_1}^{t_n} (\hat{\kappa}_t^{(k+1)})^2}. \end{aligned} \tag{2.19}$$

Procjene parametara dane sa 2.10, 2.15 i 2.16 pomoću metode najmanjih kvadarta ili 2.19 pomoću Newton-Raphsonove metode ne zadovoljavaju 2.2. Stoga, mijenjamo parametar $\hat{\alpha}_x$ sa $\hat{\alpha}_x + \hat{\beta}_x \, \bar{\kappa}$, parametar $\hat{\kappa}_t$ sa $(\hat{\kappa}_t - \bar{\kappa}) \, \hat{\beta}_.$ i naposljetku parametar $\hat{\beta}_x$ sa $\hat{\beta}_x / \hat{\beta}_.$. Ovdje je $\hat{\beta}_.$ suma $\hat{\beta}_x$-eva koji proizlaze iz dekompozicije singularnih vrijednosti ili Newton-Raphsonove metode, a $\bar{\kappa}$ aritmetička sredina $\hat{\kappa}_t$-ova koje dobijemo navedenim metodama. Još jedan od

prijedloga Lee-Carterovog modela je da se κ_t ne uzimaju iz Newton-Raphsonove metode niti pomoću dekompozicije singularnih vrijednosti, kao α_x i β_x, već da se prilagode opaženom broju smrti $\sum_{x=x_1}^{x_m} D_{xt}$ u godini t. Kako je poželjno da je razlika između stvarnog i očekivanog broja ukupnih smrti u svakoj godini jednaka 0, κ_t su takvi da zadovoljavaju jednadžbu

$$\sum_{x=x_1}^{x_m} D_{xt} = \sum_{x=x_1}^{x_m} ETR_{xt} \exp(\hat{\alpha}_x + \hat{\beta}_x \zeta) \tag{2.20}$$

po varijabli ζ. To znači da κ_t ponovno procjenjujemo pomoću prethodno procijenjenih α_x i β_x na način da dobivene stope smrtnosti primijenjene na stvarnu izloženost riziku daju ukupan broj smrti koji se podudara sa opaženim brojem smrti za godinu t. Jednadžba 2.20 nema eksplicitno rješenje pa je rješavamo numerički (na primjer, Newton-Raphsonovom metodom).

Poboljšanja metode

Jedno od poboljšanja metode je prilagodba κ_t kako bi se bolje reproducirao opaženi period očekivanog trajanja života u izabranoj dobi, umjesto predložene prilagodbe ukupnom broju zabilježenih smrti u danoj godini. U praksi, aktuari prvo odabiru početnu dob x_0. U projekcijama smrtnosti za x_0 se često uzimaju dobi od 60 ili 65 godina, suprotno uobičajenom $x_0 = 0$. Zatim se κ_t prilagođava kako bi odgovarali opaženom očekivanom trajanju života u dobi x_0 u godini t uz α_x i β_x koji su procijenjeni nekom od gore navedenih metoda. Slijedi da $\hat{\kappa}_t$ rješavaju jednadžbu:

$$\begin{aligned} e^{\uparrow}_{x_0}(t) = {} & \frac{1 - \exp\left(-\exp\left(\hat{\alpha}_{x_0} + \hat{\beta}_{x_0}\zeta\right)\right)}{\exp\left(\hat{\alpha}_{x_0} + \hat{\beta}_{x_0}\zeta\right)} + \sum_{k\geq 1}\left(\prod_{j=0}^{k-1} \exp\left(-\exp\left(\hat{\alpha}_{x_0+j} + \hat{\beta}_{x_0+j}\zeta\right)\right)\right) \\ & \times \frac{1 - \exp\left(-\exp\left(\hat{\alpha}_{x_0+k} + \hat{\beta}_{x_0+k}\zeta\right)\right)}{\exp\left(\hat{\alpha}_{x_0+k} + \hat{\beta}_{x_0+k}\zeta\right)} \end{aligned} \tag{2.21}$$

po varijabli ζ. Ovdje je $e^{\uparrow}_{x_0}(t)$ očekivano trajanje života u dobi x u kalendarskoj godini t, čija je formula:

$$\begin{aligned} \bar{e}^{\uparrow}_{x_0}(t) &= \int_{\zeta\geq 0} \exp\left(-\int_0^{\zeta} \mu_{x+\eta}(t)d\eta\right)d\zeta \\ &= \frac{1 - \exp\left(-\mu_x(t)\right)}{\mu_x(t)} + \sum_{k\geq 1}\left(\prod_{j=0}^{k-1} \exp\left(-\mu_{x+j}(t)\right)\right)\frac{1 - \exp\left(-\mu_{x+k}(t)\right)}{\mu_{x+k}(t)}, \end{aligned} \tag{2.22}$$

pri čemu se $\bar{e}^{\uparrow}_{x_0}(t)$ računa iz periodnih tablica smrtnosti za kalendarsku godinu t i dani niz intenziteta smrtnosti $\mu_{x+k}(t)$, za $k = 0, 1, \ldots$ U danoj formuli 2.22 omjer $\frac{1-\exp\left(-\mu_{x+k}(t)\right)}{\mu_{x+k}(t)}$ je prosjek dijelova godine koji su proživjele osobe dobi x+k godina, a produkt $\prod_{j=0}^{k-1} \exp\left(-\mu_{x+j}(t)\right)$

predstavlja vjerojatnost ${}_kp_x^{\uparrow}(t)$ doživljenja dobi $x + k$ izračunatu iz tablice smrtnosti. Prednost ovog poboljšanja je da ne zahtijeva podatke o izloženosti riziku ni broju smrti u godini t.

Iduće poboljšanje je također vezano uz κ_t. Prijedlog je prilagodba vremenskog indeksa opaženoj distribuciji smrti po godinama. Ovo poboljšanje pretpostavlja da distribucije smrti po godinama D_{xt} imaju Poissonovu distribuciju za broj smrti po godinama i koristi devijaciju kako bi izmjerio preciznost prilagodbe. Za fiksnu kalendarsku godinu t, D_{xt} su nezavisne slučajne varijable iz Poissonove distribucije sa očekivanjem $ETR_{xt} \exp(\hat{\alpha}_x + \hat{\beta}_x \kappa_t)$. Ovdje su $\hat{\alpha}_x$ i $\hat{\beta}_x$ procijenjeni pomoću gore navedenih metoda, odnosno, ili metodom dekompozicije singularnih vrijednosti ili Newton-Raphsonovom metodom. Kako bi se κ_t što bolje prilagodio, opaženi D_{xt} mora minimizirati Poissonovu logaritamsku funkciju vjerodostojnosti

$$\sum_{x=x_1}^{x_m} (D_{xt} \ln (ETR_{xt} \exp(\hat{\alpha}_x + \hat{\beta}_x \zeta)) - ETR_{xt} \exp(\hat{\alpha}_x + \hat{\beta}_x \zeta)) \tag{2.23}$$

za varijablu ζ. Ekvivalentno tome je da κ_t minimizira po ζ funkciju:

$$\mathfrak{D}(\zeta) = 2 \sum_{x=x_1}^{x_m} \left(D_{xt} \ln \frac{D_{xt}}{\hat{D}_{xt}} - (D_{xt} - \hat{D}_{xt}) \right) \tag{2.24}$$

gdje je $\hat{D}_{xt} = ETR_{xt} \exp(\hat{\alpha}_x + \hat{\beta}_x \zeta)$ očekivani broj smrti, uz dane $\hat{\alpha}_x$ i $\hat{\beta}_x$. Također je potrebno prilagoditi dobivene κ_t kako bi zadovoljavali 2.2. To postižemo mijenjanjem $\hat{\kappa}_t$ sa $\hat{\kappa}_t - \bar{\kappa}$ i $\hat{\alpha}_x$ sa $\hat{\alpha}_x + \hat{\beta}_x \bar{\kappa}$. $\bar{\kappa}$ je aritmetička sredina prilagođenih $\hat{\kappa}$-ova.

2.3 Procjena Poissonovom metodom maksimalne vjerodostojnosti

Ova metoda se koristi kada na raspolaganju imamo i opaženi broj smrti D_{xt} i odgovarajuće rizike izlaganja ETR_{xt}. Tada možemo primijeniti i metodu najmanjih kvadrata na omjer broja smrti i izloženosti riziku (odnosno $\hat{m}_x(t) = D_{xt}/ERT_{xt}$). No, metoda koja slijedi ipak bolje iskorištava dane informacije. Također, ne pretpostavlja da je varijabilnost $\hat{m}_x$ jednaka bez obzira na dob x. Pretpostavimo da broj smrti dobi x u godini t ima Poissonovu slučajnu razdiobu. Fiksirajmo sljedeći par, dob x i kalendarsku godinu t. D_{xt} tada broji opažene smrti među L_{xt} osoba starih x godina 1. siječnja godine t. Pretpostavimo da je preostali životni vijek tih osoba nezavisan i jednako distribuiran. Funkcija vjerodostojnosti je proporcionalna Poissonovoj funkciji vjerodostojnosti. To vrijedi uz uvjet da je očekivanje Poissonove razdiobe jednako $ETR_{xt}\,\mu_x(t) = ETR_{xt} \exp(\alpha_x + \beta_x \kappa_t)$. Parametre α_x, β_x i κ_t

procjenjujemo tako da maksimiziramo sljedeću funkciju:

$$L(\alpha,\beta,\kappa) = \sum_{x=x_1}^{x_m}\sum_{t=t_1}^{t_n}\left(D_{xt}(\alpha_x+\beta_x\kappa t) - ETR_{xt}\exp(\alpha_x+\beta_x\kappa_t)\right)+c\,, \tag{2.25}$$

gdje je c konstanta. Ekvivalentno, parametri se mogu procijeniti i minimizacijom devijacije definirane sa

$$D = -2(L(\alpha,\beta,\kappa) - L_f) \tag{2.26}$$

gdje je L_f logaritamska vjerodostojnost punog modela. Puni model karakteriziramo izjednačavanjem prilagođenog i stvarnog broja smrti. Predloženi model ne možemo procijeniti pomoću standardnih paketa koji implementiraju Poissonovu regresiju. Razlog tome je bilinearni izraz $\beta_x\,\kappa_t$ koji je prisutan u modelu. Jednadžba vjerodostojnosti se može riješiti pomoću Newton-Raphsonove metode (2.19 kao kod metode najmanjih kvadrata). Početni uvjeti su $\hat{\alpha}_x^{(0)} = 0, \hat{\beta}_x^{(0)} = 0$ i $\hat{\kappa}_t^{(0)} = 0$. Koeficijente dobivamo sljedećom rekurzivnom formulom:

$$\hat{\alpha}_x^{(k+1)} = \hat{\alpha}_x^{(k)} - \frac{\sum_{t=t_1}^{t_n}\left(D_{xt} - ETR_{xt}\,\exp\left(\hat{\alpha}_x^{(k)}+\hat{\beta}_x^{(k)}\,\hat{\kappa}_t^{(k)}\right)\right)}{-\sum_{t=t_1}^{t_n} ETR_{xt}\,\exp\left(\hat{\alpha}_x^{(k)}+\hat{\beta}_x^{(k)}\,\hat{\kappa}_t^{(k)}\right)} \tag{2.27}$$

$$\hat{\kappa}_t^{(k+1)} = \hat{\kappa}_t^{(k)} - \frac{\sum_{x=x_1}^{x_m}\left(D_{xt} - ETR_{xt}\,\exp\left(\hat{\alpha}_x^{(k+1)}+\hat{\beta}_x^{(k)}\,\hat{\kappa}_t^{(k)}\right)\right)\hat{\beta}_x^{(k)}}{-\sum_{x=x_1}^{x_m} ETR_{xt}\,\exp\left(\hat{\alpha}_x^{(k+1)}+\hat{\beta}_x^{(k)}\,\hat{\kappa}_t^{(k)}\right)\left(\hat{\beta}_x^{(k)}\right)^2} \tag{2.28}$$

$$\hat{\beta}_x^{(k+1)} = \hat{\beta}_x^{(k)} - \frac{\sum_{t=t_1}^{t_n}\left(D_{xt} - ETR_{xt}\,\exp\left(\hat{\alpha}_x^{(k+1)}+\hat{\beta}_x^{(k)}\,\hat{\kappa}_t^{(k+1)}\right)\right)\hat{\kappa}_t^{(k+1)}}{-\sum_{t=t_1}^{t_n} ETR_{xt}\,\exp\left(\hat{\alpha}_x^{(k+1)}+\hat{\beta}_x^{(k)}\,\hat{\kappa}_t^{(k+1)}\right)\left(\hat{\kappa}_t^{(k+1)}\right)^2}\,. \tag{2.29}$$

Kriterij zaustavljanja je porast u logaritamskoj funkciji vjerodostojnosti koji je manji od prethodno odabranog dovoljno malog fiksnog broja. Dakle, kriterij zaustavljanja određujemo na početku procedure. Preostaje još namjestiti parametre dobivene u 2.27, 2.28 i 2.29 kako bi zadovoljavali uvjet 2.2. To postižemo zamjenom $\hat{\kappa}_t$ sa $(\hat{\kappa}_t - \bar{\kappa})\sum_{x=x_1}^{x_m}\hat{\beta}_x$, koeficijent $\hat{\beta}_x$ zamijenimo sa $\hat{\beta}_x / \sum_{x=x_1}^{x_m}\hat{\beta}_x$ i naposljetku $\hat{\alpha}_x$ sa $\hat{\alpha}_x + \hat{\beta}_x\,\bar{\kappa}$. Ovdje $\bar{\kappa}$ predstavlja aritmetičku sredinu vrijednosti $\hat{\kappa}_t$.

2.4 Izglađivanje u modelu

Parametri α_x su uglavnom već glatki, jer predstavljaju prosječni utjecaj smrtnosti u dobi x. Također, procijenjeni parametri κ_t često nisu glatki, ali projicirani κ_t jesu jer ih dobijemo pomoću modela vremenskih nizova. Dakle, u modelu je potrebno izgladiti procijenjene β_x.

Ovo postižemo pomoću metode najmanjih kvadrata ili metodom maksimalne vjerodostojnosti. Ako koristimo metodu najmanjih kvadrata, koristimo sljedeću funciju:

$$\theta(\alpha,\beta,\kappa) = \sum_{x=x_1}^{x_m}\sum_{t=t_1}^{t_n}(\ln\hat{\mu}_x(t) - \alpha_x - \beta_x\kappa_t)^2 + \pi_\beta\sum_{x=x_1}^{x_m}(\beta_{x+2} - 2\beta_{x+1} + \beta_x)^2\,, \tag{2.30}$$

gdje je π_β parametar izglađivanja za β_x. Parametar izglađivanja nam daje ravnotežu između toga da procjena prati dane podatke (preciznosti) i glatkoće. Preciznost dobijemo sumom kvadriranih reziduala, dok nam glatkoću daje drugi dio izraza, "kazneni" dio (eng. *penalty term*). Taj dio u danoj formuli je suma kvadrata izraza $(\beta_{x+2} - 2\beta_{x+1} + \beta_x)$. Izbor parametra izglađivanja je jako bitan jer dobijemo drugačije prilagođene rezultate svaki put kada ga mijenjamo. Izbor optimalnog π_β je zasnovan na opaženim podacima. Prilikom njegovog određivanja koriste se tehnike unakrsne validacije (eng. *cross − validation*). Ako parametre procjenjujemo Poissonovom metodom maksimalne vjerodostojnosti, za izglađivanje koristimo logaritamsku funkciju:

$$L(\alpha,\beta,\kappa) - \frac{1}{2}\pi_\beta\sum_{x=x_1}^{x_m}(\beta_{x+2} - 2\beta_{x+1} + \beta_x)^2 \tag{2.31}$$

gdje je L funkcija definirana u 2.25. Parametar izglađivanja π_β opet dobijemo pomoću unakrsne validacije.

2.5 Odabir optimalnog perioda podešavanja i analiza reziduala

Mnoga aktuarska istraživanja baziraju projekcije smrtnosti na statističkim podacima za godine od 1950. godine do danas. Postavlja se pitanje zašto podaci iz perioda nakon 1950. godine bolje predstavljaju očekivanja za buduće podatke od nekog drugog vremenskog perioda, na primjer, perioda nakon 1800. ili 1900. godine. Postoje brojni razlozi za to, jedan od njih je činjenica da je stopa opadanja smrtnosti bila ujednačenija od 1950. do 2000. godine nego što je bila u periodu od 1900. do 2000. godine. Drugi razlog za to je kvaliteta podataka o smrtnosti, pogotovo za stariju dob, u periodu od 1900. do 1950. godine. Sljedeći razlog su uzroci smrti. Nakon 1950. godine smanjio se broj smrti uzrokovanih zaraznim bolestima, dok su smrti čiji su uzrok srčane bolesti ili oblici tumora bile najčešće, kao što je slučaj i danas. To utječe na stope smrtnosti uzrokovane bolestima, čiji se omjer prije 1950. godine ne može primijeniti na buduća očekivanja smrtnosti, posebno zbog napredovanja medicine. Prema Lee-Carterovom radu duljina vremenskih nizova smrtnosti nije problem dok god se radi o periodu dužem od 10-20 godina. Daljnja istraživanja (Lee i Miller 2001. godine) pokazuju da se točniji rezultati dobiju ako se kao početak perioda

opažanja uzime 1950. godina. U svojem ocjenjivanju Lee-Carterovog modela Lee i Miller su 2001. godine opazili da je za podatke o smrtnosti iz SAD-a predviđanje za period od 1900. do 1989. godine odstupalo od stvarnih podataka više nego kada se koristio period od 1990. do 1997. godine. Glavni izvor greške je bilo nesrazmjer između prilagođenih stopa za zadnju godinu perioda opažanja (1989. u njihovom istraživanju) i stvarnih stopa u toj godini. Stoga je provedena prilagodba odstupanja. Također je zamijećeno da parametri β_x nisu bili stabilni kroz cijelo 20. stoljeće. Kako bi dobili stabilne parametre β_x Lee i Miller predlažu 1950. godinu kao prvu godinu perioda opažanja. Zaključili su da je na ovaj način moguće izbjeći velika odstupanja (*outliere*). Slično njihovom istraživanju, Lundtröm i Ovist su 2004. godine skratili period opažanja koji se prethodno kretao od 1901. pa sve do 2001. godine na period od zadnjih 25 godina za švedske podatke. Također, Baran i Pap su 2007. godine primijenili Lee-Carterov model kako bi predvidjeli stope smrtnosti u Mađarskoj za period od 2004. do 2040. godine. Koristili su jedan set podataka opaženih od 1949. do 2003. godine i drugi set podataka od 1989. do 2003. godine. Model u kojem su koristili opažanja od 1949. do 2003. godine predvidio je rast stopa smrtnosti za muškarce stare 45 do 55 godine. Tada se postavilo pitanje je li Lee-Carterov model primjenjiv na zemlje koje pokazuju neobične trendove smrtnosti, kao što je Mađarska. Međutim, model u kojem su koristili opažanja u periodu od 1989. do 2003. godine predvidio je trend opadanja smrtnosti za muškarce i žene. Ovo pokazuje da je odabir optimalnog perioda opažanja također bitna stavka prilikom istraživanja. 2002. godine Booth, Maindonald i Smith su dizajnirali proceduru za odabir optimalnog perioda opažanja koji odgovara najdužem periodu za koji su procijenjeni indeksi smrtnosti κ_t linearni. Posebno, ovi autori žele maksimizirati prilagodbu cijelog modela ograničavanjem perioda opažanja uz pretpostavku linearnosti. Odabir perioda opažanja se bazira na omjeru srednjih odstupanja prilagodbe Lee-Carterovog modela u odnosu na linearnu prilagodbu podataka. Ovaj omjer računamo tako da početna godina varira i odabrani period opažanja je onaj za koji je omjer značajno manji nego periodi koji počinju prethodnih godina. Autori pretpostavljaju da je trend prilagođenih $\hat{\kappa}_t$ linearan, a pretpostavku baziraju na jedinstvenom obrascu opadanja smrtnosti. Opadanje stopa smrtnosti je spomenuto u brojnim istraživanjima, uključujući ono Lee i Cartera 1992. godine te Tuljapurkara i Boea 2000. godine. Kada $\hat{\kappa}_t$ odstupaju od linearnosti, period opažanja se aproksimativno smanjuje. Kao zadnja godina opažanja uzima se t_n tako da je period opažanja određen početnom godinom t_{start}. Ograničavanje perioda opažanja na najdulji (nedavni) period od t_{start} do t_n za koji prilagođeni κ_t ne odstupaju značajno od linearnosti ima brojne prednosti. Na ovaj način izbjegavaju se promjene u trendu $\hat{\kappa}_t$ što smanjuje nepouzdanost predviđanja. Također, povećava se vjerojatnost da će β_x zadovoljiti pretpostavku vremenske nepromjenjivosti. Denuit i Goderniaux su 2005. godine predložili proceduru za odabir kalendarske godine t_{start}. Godinu t_{start} uz uvjet $t_{start} \geq t_1$ biramo tako da zadovoljava uvjet je niz $\hat{\kappa}_t$ za $t = t_{start}, t_{start} + 1, \ldots, t_n$ najbolje aproksimiran pravcem. Kriterij za odabir je koeficijent determinacije R^2. Koeficijent determinacije R^2

nam zapravo daje informaciju o tome koliko rasipanja izlaznih podataka potječe od funkcijske ovisnosti, a koliko otpada na takozvano rezidualno ili neobjašnjeno rasipanje. Želimo ga maksimizirati (kao funkciju broja opažanja). Odabir optimalnog perioda opažanja se našao i na udaru kritičara. Neki stručnjaci tvrde da bi mogao dovesti do podcjenjivanja nepouzdanosti predviđanja. [2]
Slijedi analiza reziduala. Varijance grešaka (koje moraju biti slučajnog karaktera, inače ukazuju na interakciju između dobi i godina) moraju biti jednake, odnosno greške moraju biti homogene. Greške također moraju biti nezavisne. U Lee-Carterovom modelu prilagodba $\hat{\kappa}_t$-ova može utjecati na reziduale pa se oni analiziraju uz $\hat{\kappa}_t$. Kada su parametri procijenjeni metodom najmanjih kvadrata analiziraju se Pearsonovi reziduali. Pearsonovi reziduali su reziduali podijeljeni korijenom funkcije varijance $V(\mu)$. Dani su formulom

$$r_{Pi} = \frac{r_i}{\sqrt{V(\hat{\mu}_i)}}. \tag{2.32}$$

U Lee-Carterovom modelu Pearsonovi reziduali su dani formulom

$$r_{xt} = \frac{\hat{\epsilon}_x(t)}{\sqrt{\frac{1}{(x_m - x_1)(t_n - t_1)} \sum_{x=x_1}^{x_m} \sum_{t=t_1}^{t_n} (\hat{\epsilon}_x(t))^2}}, \tag{2.33}$$

gdje su

$$\hat{\epsilon}_x(t) = \ln \hat{m}_x(t) - (\hat{\alpha}_x + \hat{\beta}_x \hat{\kappa}_t). \tag{2.34}$$

Ako r_{xt} zadovoljavaju sve navedene uvjete (homogenost, nezavisnost) možemo zaključiti da model dobro opisuje vremenske trendove.

2.6 Projiciranje smrtnosti

Jedna od bitnih značajki Lee-Carterovog modela je pretpostavka da je vremenski faktor κ_t stohastički proces. Pomoću Box-Jenkinsove metode procjenjujemo i predviđamo κ_t. U ovom modelu pretpostavka je da je κ_t realizacija vremenskog niza kojeg modeliramo ARIMA modelom (1.3). Dinamika κ_t opisana je ARIMA (p, d, q) procesom ako su κ_t stacionarani i ako vrijedi;

$$\Delta^d \kappa_t = \phi_1 \Delta^d \kappa_{t-1} + \ldots + \phi_p \Delta^d \kappa_{t-d} + \epsilon_t + \psi_1 \epsilon_{t-1} + \ldots + \psi_q \epsilon_{t-q} \tag{2.35}$$

uz uvjete $\phi_p \neq 0$ i $\psi_q \neq 0$ i gdje je ϵ_t gusovski bijeli šum takav da $\sigma_\epsilon^2 > 0$ te je

$$\Delta \kappa_t = \kappa_t - \kappa_{t-1} \tag{2.36}$$

operator diferenciranja. κ_t je stacionaran ako vrijedi:
1. $\mathbb{E} \mid \kappa_t \mid^2 < \infty, \forall t \in \mathbb{Z}$

2. $\mathbb{E}[\kappa_t] = \mu, \forall t \in \mathbb{Z}$ i
3. $\gamma_\kappa(r, s) = \gamma_\kappa(r + t, s + t), \forall r, s, t \in \mathbb{Z}$.

$\gamma_\kappa(t) := \gamma_\kappa(t, t + s)$ je autokovarijacijska funkcija. Riječima, vremenski niz je (slabo) stacionaran ako su mu očekivanje i varijanca konstantni kroz vrijeme i kovarijanca bilo koja dva vremenska perioda, t i $t + k$, ovisi samo o duljini intervala između ta dva vremenska perioda k, a ne o početnom vremenu t.

Slijedi opis osnovnih koraka kod prilagođavanja ARIMA modela vremenskom nizu. Podsjetimo se, p je stupanj autoregresivnog procesa, d je stupanj diferenciranja, a q stupanj modela pomičnih prosjeka (eng. *moving average*) (1.3). Nestacionarnost niza može imati nekoliko uzroka. Jedan od njih može biti prisutnost determinističke komponente. Takav vremenski niz se može svesti na stacionaran postavljanjem regresije na vrijeme i nastavljanjem rada na dobivenim rezidualima. Ovakve nizove zovemo stacionarnima po trendu. Nestacionarnost može nastati i zanemarivanjem nakupljenih stohastičkih šokova. Tada trend možemo ukloniti diferenciranjem (uz $d = 1$). Ako vremenski niz ni nakon toga nije stacionaran, povećavamo stupanj diferenciranja na $d = 2$ i tako sve dok ne dobijemo stacionaran niz. Ovakav niz je diferencirano stacionaran. Prva provjera stacionarnosti je vizualna provjera prilikom koje na grafu pokušavamo odrediti prisutnost trenda. Još jedna od karakteristika nestacionarnog niza je i jaka korelacija koja polako opada s porastom vremenskog pomaka (eng. *lag*). To znači da kod stacionarnog vremenskog niza autokorelacijska funkcija geometrijski opada s vremenom. Dakle, proces ima, takozvano, kratko pamćenje (eng. *short-memory*) jer su opažanja koja su jako udaljena u vremenu u principu nezavisna. Kada vremenski niz jednom diferenciramo, autokorelacijska funkcija se smanjuje na linearnu funkciju te udaljena opažanja više nisu nezavisna. Osim grafičkih procedura postoje i formalni testovi za testiranje nul hipoteze da je vremenski niz stacionaran. Jedan od poznatijih testova je Kwiatowski–Philips–Schmidt–Shinov test (KPSS test). Ako jednom diferenciramo niz, možemo provjeriti je li ga potrebno još jednom diferencirati Dickey-Fullerovim testom o jediničnom korijenu. Sada želimo odrediti parametre p i q, respektivno AR i MA procesa. p i q određujemo proučavanjem autokorelacijske (acf) i parcijalne autokorelacijske funkcije (pacf) parametara κ_t. Kod AR(p), odnosno autoregresivnog procesa acf ima skok u lagu p i nakon toga geometrijski opada, dok pacf ima nagli pad nakon laga p. Kod MA(q) acf ima nagli pad nakon laga q, a pacf ima skok u lagu q i nakon toga geometrijski opada. Iste kriterije primjenjujemo na ARMA(p, q) i ARIMA(p, q) kako bi odredili parametre p i q. Kako bismo odredili broj i tip parametara, najbolji model odabiremo pomoću AIC kriterija. AIC (eng. *Akaike information criterion*) je mjera relativne kvalitete statističkog modela za dani set podataka. Za dani skup mogućih modela AIC procjenjuje kvalitetu svakoga u odnosu na ostale modele. Biramo model s najmanjim AIC kriterijem, to jest, onaj koji minimizira gubitak informacija. U R-u postoje gotove procedure koje će nam pomoću AIC kriterija i Yule-Walker procedure dati broj parametara i njihovu procjenu. Općenito, parametre ARIMA(p, q) modela uz poznate stup-

njeve p i q možemo procijeniti pomoću procedure `arima(x, order = c(0, 0, 0), include.mean = TRUE, ...)`. Nakon što odredimo parametre p i q bitno je provjeriti normalnost reziduala vremenskog niza. Za to postoje brojni testovi: Ljung-Box-Pierce, Shapiro-Wilk, Jarque-Bera itd. Za većinu ovih testova, također, postoje gotove procedure u R-u.
Jedan od čestih modela za κ_t je slučajna šetnja s pomakom. U tom slučaju κ_t dani su sa

$$\kappa_t = \kappa_{t-1} + d + \xi_t\,, \tag{2.37}$$

gdje su elementi ξ_t gaussovski, odnosno nezavisni i normalno distribuirani s očekivanjem 0 i varijancom σ^2 i gdje je d poznat kao parametar pomaka (otklona). Slijedi

$$\kappa_{t_n+k} = \kappa_{t_n} + kd + \sum_{j=1}^{k} \xi_{t_n+j}. \tag{2.38}$$

Predviđanje vremenskog indeksa u vremenu $t_n + k$ je dano sa

$$\dot{\kappa}_{t_n+k} = \mathbb{E}[\kappa_{t_n+k} \,|\, \kappa_{t_1}, \kappa_{t_2}, \ldots, \kappa_{t_n}] = \kappa_{t_n} + kd. \tag{2.39}$$

Predviđanje prati pravac nagiba d gdje je varijabla k. Uvjetna varijanca predviđanja iznosi

$$\mathbb{V}ar[\kappa_{t_n+k} \,|\, \kappa_{t_1}, \kappa_{t_2}, \ldots, \kappa_{t_n}] = k\sigma^2. \tag{2.40}$$

Stoga se uvjetne standardne pogreške predviđanja povećavaju s korijenom udaljenosti k. Model slučajne šetnje s pomakom za predviđanje κ_t ekvivalentan je predviđanju smanjenja svake stope smrtnosti jedinstvene po dobi. Iz 2.37 slijedi da su razlike u očekivanim logaritamskim stopama smrtnosti između vremena $t + 1$ i t dane sa

$$\ln \mu_x(t+1) - \ln \mu_x(t) = \beta_x\, \mathbb{E}[\kappa_{t+1} - \kappa_t] = \beta_x\, d. \tag{2.41}$$

Omjer stopa smrtnosti između dvije uzastopne godine predviđanja je jednak $\exp(\beta_x\, d)$ te je invarijantan kroz vrijeme. Produkt $\beta_x\, d$ je jednak promjeni stope smrtnosti kroz vrijeme x. Parametar β_x u ovom slučaju možemo interpretirati kao normalizirani obrazac promjene stopa smrtnosti kroz vrijeme specifičnu po dobi x. Budućim profilima smrtnosti po godinama koje daje Lee-Carterov model se glatkoća uvijek smanjuje kroz vrijeme. To je zabilježeno u radu Girosija i Kinga iz 2007. godine. To objašnjava zašto je ovaj pristup posebno namijenjen predviđanju agregiranih demografskih indikatora, kao što je životno očekivanje ili vrijednosti anuiteta, ali ne baš životnim tablicama po dobnim skupinama. To proizlazi iz činjenice da je predviđanje logaritamskih stopa smrtnosti linearno kroz vrijeme (2.41). Kako β_x varijaju s dobi x glatkoća logaritamskih stopa smrtnosti po dobi se smanjuje s vremenom jer se udaljenost između logaritamskih stopa smrtnosti u različitim

dobnim skupinama povećava. Svaka razlika u β_x je pojačana kako predviđanje seže dalje u budućnost. 2.37 omogućuje da su $\kappa_t - \kappa_{t-1}$, $t = t_2, t_3, \ldots, t_n$ nezavisni i normalno distribuirani sa očekivanjem d i varijancom σ^2. Procjena metodom maksimalne vjerodostojnosti za d i σ^2 nam daje sljedeće procijenjene parametre:

$$\hat{d} = \frac{1}{t_n - t_1} \sum_{t=t_2}^{t_n} (\hat{\kappa}_t - \hat{\kappa}_{t-1}) = \frac{\hat{\kappa}_{t_n} - \hat{\kappa}_{t_1}}{t_n - t_1} \tag{2.42}$$

i

$$\hat{\sigma}^2 = \frac{1}{t_n - t_1} \sum_{t=t_2}^{t_n} (\hat{\kappa}_t - \hat{\kappa}_{t-1} - \hat{d})^2. \tag{2.43}$$

Ovaj pristup je poznat kao "metoda ravnala" predviđanja jer spaja prvu i zadnju točku dostupnih podataka pravcem ("ravnalom") i onda produžuje taj pravac u budućnost kako bi dobio predviđene točke. Zbog oblika izraza za $\hat{d}$ bitno je provjeriti vrijednosti κ_{t_n} do na razumne vrijednosti. Model slučajne šetnje s pomakom pokazao se dobrim modelom za κ_t na brojnim primjerima. Tuljapurkar u radu iz 2000. godine smatra da je opadanje κ_t-ova u skladu s slučajnom šetnjom s pomakom dobar model za zemlje G7 (SAD, Kanda, Japan, Njemačka, Italija, Francuska, Ujedinjeno kraljevstvo i Rusija). Čak i složeniji modeli daju rezultate koji su jako blizu rezultatima koje daje model slučajne šetnje s pomakom.

Predikcijski intervali

Prilikom predviđanja bitno se osvrnuti i na nepouzdanost vezanu za buduću smrtnost, odnosno dati informacije o greškama za predviđene vrijednosti. Tradicionalni demografski pristup je odrediti najviši i najniži mogući scenarij za srednje predviđanje koje se smatra najboljom procjenom. Za ovo nam je potrebna vjerojatnosna distribucija. Kako bismo dobili margine demografskih indikatora potrebni su nam predikcijski intervali. Izračunati relevantne predikcijske intervale analitički je teško jer imamo dva različita izvora nepouzdanosti. Prvi su uzoračke pogreške za parametre α_x, β_x i κ_t. Drugi su predikcijske greške za projicirane κ_t. Još jedan od razloga je činjenica da su stope smrtnosti dobivene složenom nelinearnom funkcijm ovisnom o više parametara, α_x, β_x i κ_t te ARIMA parametrima p i q (2.1). Ideja bootstrap metode je presložiti (eng. *resample*) originalne podatke kako bismo stvorili replikacije skupa podataka pomoću kojih možemo pronaći varijabilnost životnog očekivanja. Kako ovaj pristup zahtijeva ponavljanje procedura i analiza, koje smo provodili na originalnom skupu podataka na mnogim replikacijama skupa podataka, ova metoda zahtijeva jako računalo i vrijeme za računanje i provođenje svih operacija (eng. *computer − intensive*). Bootstrap metoda se općenito koristi kada je, kao u ovom slučaju, teoretsko računanje sa prilagođenim modelom previše složeno. Ako zanemarimo ostale

izvore pogreške, možemo koristiti granice pouzdanosti budućih κ_t kako bismo dobili predikcijske intervale. Kod dugoročnih prognoza (25 godina i više) pogreške nastale prilikom predviđanja vremenskog indeksa κ_t dominiraju nad pogreškama nastalim prilikom predviđanja stopa smrtnosti. Unatoč tome, predikcijski intervali koji se baziraju samo na pogreškama vezanim za predviđanje κ_t poprilično podcjenjuju greške u predviđanju za kratkoročne vremenske periode predviđanja. Prema Lee i Carteru predikcijski intervali koji se zasnivaju samo na greškama vezanim za κ_t su razumne aproksimacije samo za periode predviđanja veće od 10-25 godina. Ako želimo predviđati životno očekivanje za kraći vremenski rok ne možemo dobiti preciznu analizu temeljenu samo na greškama predviđanja. Prvo ću definirati period očekivanog trajanja života u dobi x u kalendarskoj godini t, odnosno $e^{\uparrow}_{x_0}(t)$ kao u 2.22. Da bismo izbjegli pretpostavku normalnosti koristimo bootstrap postotni interval kako bismo konstruirali interval pouzdanosti za predviđeno životno očekivanje. Bootstrap procedura daje B uzoraka α_x, β_x i κ_t parametara koje označavamo sa α^b_x, β^b_x i κ^b_t za $b = 1, 2, \ldots, B$. Proceduru možemo provesti na nekoliko načina. Prvi način su Monte Carlo simulacije. Uzorke izvlačimo direktno iz aproksimativne multivarijatne normalne distribucije parametara procijenjenih Poissonovom metodom maksimalne vjerodostojnosti $(\hat{\alpha}_x, \hat{\beta}_x, \hat{\kappa}_t)$. Ako uzmemo u obzir svojstva velikih uzoraka procjenitelja maksimalne vjerodostojnosti znamo da su $(\hat{\alpha}_x, \hat{\beta}_x, \hat{\kappa}_t)$ asimptotski multivarijatno normalno distribuirani sa očekivanjem (α, β, κ) i kovarijacijskom matricom koja je jednaka inverzu Fisherove informacijske matrice. Elementi Fisherove matrice su jednaki sljedećim izrazima: minus vrijednosti očekivanih vrijednosti druge derivacije logaritamske funkcije vjerodostojnosti. U matematičkoj statistici Fisherova informacija je mjera količine informacija koju opažena slučajna varijabla X nosi o nepoznatom parametru distribucije kojoj X pripada, odnosno θ. Službeno, to je varijanca rezultata ili očekivana vrijednost opažene informacije. Kada postoji N parametara, odnosno θ je vektor dimenzije $N \times 1$, to jest $\theta = [\theta_1, \theta_2, \ldots, \theta_N]^T$, Fisherova informacija ima oblik matrice dimenzije $N \times N$. Ovu matricu zovemo Fisherova informacijska matrica (eng. *Fisher information matrix*, FIM). Fisherova informacijska matrica je oblika

$$[J(\theta)]_{i,j} = \mathbb{E}_\theta \left[\left(\frac{\partial}{\partial \theta_i} \log f(X; \theta) \right) \left(\frac{\partial}{\partial \theta_j} \log f(X; \theta) \right) \right]. \tag{2.44}$$

FIM je pozitivno definitna simetrična matrica. FIM za $N-$ varijatnu multivarijatnu normalnu distribuciju, $X \sim N(\mu(\theta), \Sigma(\theta))$ ima poseban oblik. Neka je $\theta = [\theta_1, \ldots, \theta_K]^T$ K-dimenzionalni vektor parametara i $X = [X_1, \ldots, X_N]^T$ $N-$dimenzionalni vektor slučajnih varijabli. Neka je očekivanje slučajnog vektora dano sa $\mu(\theta) = [\mu_1(\theta), \ldots, \mu_N(\theta)]^T$ i neka je $\Sigma(\theta)$ kovarijacijska matrica. Tada je za $1 \leq m, n \leq K$, $(m, n)-$ti element FIM dan sa:

$$J_{m,n} = \frac{\partial \mu^T}{\partial \theta_m} \Sigma^{-1} \frac{\partial \mu}{\partial \theta_n} + \frac{1}{2} \mathtt{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_m} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_n} \right). \tag{2.45}$$

Drugi način je Poissonov bootstrap. Počevši od opažanja (ETR_{xt}, D_{xt}) stvara se B bootstrap

uzoraka (ETR_{xt}, D^b_{xt}) za $b = 1, \ldots, B$. Ovdje je D^b_{xt} realizacija iz Poissonove distribucije sa očekivanjem $ETR_{xt}\,\hat{\mu}_x(t) = D_{xt}$. Bootstrap broj smrti D^b_{xt} se dobije primjenom Poissonovog šuma na opaženi broj smrti. Za svaki bootstrap uzorak procjenjuju se α_x, β_x i κ_t. U Poissonovom procesu, broj opaženih događaja flukturira (varira) oko očekivanja λ sa standardnom devijacijom $\sigma_k = \sqrt{\lambda}$. Ove fluktuacije nazivamo Poissonov šum (eng. *Poisson noise*).
Još jedna od mogućnosti za bootstrap metodu je bootstrap reziduala. Reziduali moraju biti nezavisni i jednako distribuirani. Tada možemo rekonstituirati bootstrap reziudale i pomoću njih bootstrap podatke o smrtnosti. Dobra prilagodba modela rezultira nizom slučajnih reziduala koje možemo uzastopno replicirati (odnosno uzorkovati, *resample*). Dakle, dobar model je osnovni uvjet ovog pristupa. U suprotnom, može doći do deformacija na simuliranog histogramu veličine koju promatramo, odnosno očekivanog trajanja života. Proces funkcionira na sljedeći način. Stvorimo matricu $\mathbf{R}$ reziduala sa elementima r_{xt} definiranima kao u 2.32. Sada generiramo B bootstrap replikacija $\mathbf{R}^b$ za $b = 1, \ldots, B$, odnosno, uzorkujemo podatke pomoću jednostavnog slučajnog odabira s ponavljanjem iz skupa elemenata matrice $\mathbf{R}$. Zatim koristimo inverznu formulu za reziduale kako bismo dobili odgovarajuću matricu broja smrti D^b_{xt}. Sada možemo izračunati B skupova procijenjenih parametara $\hat{\alpha}^b_x$, $\hat{\beta}^b_x$ i $\hat{\kappa}^b_t$. Kada smo napravili bootstrap proceduru procjenjujemo model vremenskog niza koristeći κ^b_t kao točke podataka. Dobivamo novi skup procijenjenih ARIMA parametara p i q. Nakon toga možemo generirati projekcije κ^b_t za $t \geq t_n + 1$ koristeći ARIMA parametre. Buduće greške ξ^b_t uzorkujemo iz univarijatne normalne distribucije s očekivanjem 0 i standardnom devijacijom σ^b_ϵ. Dakle, κ_t se projiciraju koristeći ponovno procijenjeni ARIMA model. To ne znači da tražimo novi ARIMA model, nego zadržavamo ARIMA model kojeg smo izabrali na temelju originalnih podataka i parametre tog modela ponovno procjenjujemo pomoću bootstrap podataka. Prvi korak uzima u obzir neodređenost parametara α_x, β_x i κ_t. Drugi korak rješava problem neodređenosti ARIMA parametara koja ovisi o neodređenosti parametara α_x, β_x i κ_t. Treći korak osigurava da neodređenost u predviđanju parametara κ_t ovisi samo o standardnoj pogrešci u ARIMA procesu i o neodređenosti samih ARIMA parametara p i q. Konačno, u četvrtom koraku pri računanju tražene (relevantne) mjere, očekivanog trajanja života, uzimaju se u obzir svi izvori nepouzdanosti predviđanja. Na ovaj način dobivamo B realizacija α^b_x, β^b_x i κ^b_t temeljem kojih možemo izračunati traženi $e^{\uparrow}_x(t)$. Pretpostavimo da smo izračunali B bootstrap procjena $e^b_x(t)$, $b = 1, \ldots B$. Tada je $(1 - 2\alpha) \cdot 100$ postotni predikcijski interval za $e^{\uparrow}_x(t)$ dan sa $(e^{b(\alpha)}_x(t), e^{b(1-\alpha)}_x(t))$, gdje je $e^{b(\xi)}_x(t)$ *jednaka* $(B + 1) \times \xi$−-oj vrijednosti u uređenom uzorku što je jednako uzoračkoj ξ-kvantilnoj bootstrap-distribuciji od $e^{\uparrow}_x(t)$. To je jednako $(B + 1) \times \xi$−toj vrijednosti u nizu replikacija $e^b_x(t)$ za $b = 1, \ldots, B$. Na primjer, u slučaju da imamo $B = 999$ bootstrap uzoraka 0.95 i 0.05 uzorački kvantili su, respektivno, 950. i 50. vrijednost u rastućem nizu 999 replikacija $e^{\uparrow}_x(t)$. Za ove bootstrap metodee bitne su nepouzdanost parametara i arrowianska neodređenost (eng. *Arrowian uncertainty*). Ona je poznata i kao rizik u kojem je poznat skup budućih ishoda te

možemo dodijeliti vjerojatnost svakom mogućem ishodu. Distribuciju budućih ishoda generiramo pomoću poznatih modela sa poznatim parametrima. S druge strane, Knightovska neodređenost (eng. *Knightian uncertainty*) pridodaje jednaku važnost neodređenosti modela i parametara. Ona nije mjerljiva i ne može se izračunati. Kada bismo dodali važnost neodređenosti modela morali bismo uzeti u obzir nekoliko modela projiciranja smrtnosti i odrediti njihove vjerojatnosti i relativne vjerodostojnosti. Empirijske studije (Renshaw i Haberman 2008. godine) su pokazale varirajuće raspone pouzdanih i predikcijskih intervala baziranih na Monte Carlo metodi za različite skupove identifikacijskih ograničenja (npr. 2.2). Autori smatraju da ove razlike u rezultatima proizlaze iz prevelike parametrizacije (eng. *over parametrization*), a ne nelinearnosti parametarske strukture. [2]

Poglavlje 3

Cairns-Blake-Dowdov model i proširenje Lee-Carterovog modela

3.1 Cairns-Blake-Dowd model

U tipičnoj populaciji, stope smrtnosti po dobi rastu i padaju zajedno kroz vrijeme, odnosno promjene kroz vrijeme se događaju jednako za sve dobi. Stoga su u ovom modelu promjene kroz vrijeme određene skalarnim faktorom κ_t što implicira da su stope smrtnosti savršeno korelirane za dob. Stope opadanja $\ln \mu_x(t)$ za različite dobi su dane sa $\beta_x(\kappa_t - \kappa_{t-1})$ kako bi uvijek zadržale jednak omjer kroz vrijeme. U tome je snaga, ali ujedno i mana ovog modela jer u praksi relativna brzina opadanja za različite dobi može varirati. Jedna od jakih alternativa Lee-Carterovom modelu je model za predviđanje smrtnosti kojeg su predložili Andrew Cairns, David Blake i Kevin Dowd 2006. godine. Model uključuje dva vremenska faktora za razliku od Lee-Carterovog modela gdje samo jedan vremenski faktor određuje buduće stope smrtnosti. Dva vremenska faktora omogućuju modelu obuhvaćanje korelacija između stopa smrtnosti u različitim godinama, od jedne godine do iduće. Model možemo gledati kao kompromis između generalizirane linearne regresije i Lee-Carterovog modela. Izglađivanje u ovom modelu vrši se pomoću logit transformacije jednogodišnjih vjerojatnosti smrti. Logit funkciju modeliramo kao linearnu funkciju zavisnu o godinama. Parametri logit funkcije su modelirani kao stohastički procesi. Model se prilagođava povijesnim podacima, a predviđanje za parametre ovisne o vremenu se vrši pomoću bivarijatnog modela vremenskih nizova. Iz ove projekcije budućih koeficijenata linearne funkcije računaju se buduće jednogodišnje vjerojatnosti smrti u kombinaciji sa efektom linearnosti godina. Kod Cairns-Blake-Dowdovog modela koristi pretpostavka da je $\ln(q_x(t)/p_x(t))$ linearna po varijabli x za fiksni t, osim u mlađoj dobi. Ovu tvrdnju pokazuju empirijske

analize. Formula modela je sljedeća:

$$\ln \frac{p_x(t)}{q_x(t)} = \kappa_t^{(1)} + \kappa_t^{(2)} x\,, \tag{3.1}$$

što je ekvivalentno:

$$q_x(t) = \frac{\exp\left(\kappa_t^{(1)} + \kappa_t^{(2)} x\right)}{1 + \exp\left(\kappa_t^{(1)} + \kappa_t^{(2)} x\right)}. \tag{3.2}$$

$\kappa_t^{(1)}$ i $\kappa_t^{(2)}$ su stohastički procesi i čine bivarijatan vremenski niz čije buduće kretanje određuje projiciranje tablice smrtnosti. $\kappa_t^{(1)}$ se općenito smanjuje kroz vrijeme, čemu u prilog ide činjenica da stope smrtnosti opadaju kroz vrijeme za sve dobi. Ponekad, logaritam vjerojatnosti smrti $q_x(t)$ s obzirom na varijablu x ima blagu zakrivljenost nakon dobi umirovljenja. Ova zakrivljenost može se modelirati uključivanjem kvadratnog izraza dobi x u Cairns-Blake-Dowdov model. Kod ovog pristupa problem nastaje jer nije jasan odnos dinamike, odnosno kretanja, vremenskog faktora u odnosu na kvadrirani efekt. Dakle, kvadriranje dobi x poprilično povećava složenost modela. Vremenski nizovi $\kappa_t^{(1)}$ i $\kappa_t^{(2)}$ utječu na različite dobi na različite načine. Ovo je osnovna razlika u odnosu na Lee-carterov model gdje samo jedan vremenski niz objašnjava korelaciju među stopama smrtnosti za različite dobi x od jedne godine do iduće. Postoje neki empirijski dokazi koji idu u prilog Cairns-Blake-Dowdovom modelu, to jest, pokazuju da su promjene u stopama smrtnosti nesavršeno korelirane. Cairns-Blake-Dowdov model dopušta promjene osnovnih stopa smrtnosti koje su nesavršeno korelirane kroz dob x. Što je duži period za koji su dostupni podaci, odnosno opažanja, to model s dva vremenska niza daje bolje rezultate u odnosu na Lee-Carterov model s jednim vremenskim nizom. Na primjer, ako uzmemo u obzir cijelo 20. stoljeće poboljšanja vezana uz smrtnost se vežu uz mlađe dobi za prvu polovicu stoljeća te uz stariju dob u drugoj polovici 20. stoljeća. Model s dva vremenska niza ovdje bolje obuhvaća promjenu u dinamici između ta dva perioda. Doduše, zbog restrikcije optimalnog perioda prilagodbe na period nedavne prošlosti, odnosno na drugu polovicu 20. stoljeća, u Lee-Carterovom modelu drugi vremenski niz nije potreban. No, prijelaz s modela s jednim vremenskim nizom na onaj s dva vremenska niza ima šire posljedice u drugim područjima. Na primjer, kod sekuritizacije postojanje nesavršene korelacijske strukture implicira da investicijske obaveze (pasiva) vezane na dugovječnost zahtijevaju više od jednog investicijskog instrumenta. Sada pretpostavimo da imamo skup opaženih podataka za niz kalendarskih godina $t = t_1, t_2, \ldots, t_n$ i skup dobi $x = x_1, x_2, \ldots, x_m$. Pomoću ovih podataka želimo procijeniti parametre $\kappa_t^{(1)}$ i $\kappa_t^{(2)}$. To, kao i kod Lee-Carterovog modela, možemo napraviti pomoću metode najmanjih kvadrata. To znači da regresijski model:

$$\ln \frac{p_x(t)}{q_x(t)} = \kappa_t^{(1)} + \kappa_t^{(2)} x + \epsilon_x(t) \tag{3.3}$$

prilagođavamo opažanjima za kalendarske godine t, gdje su $q_x(t)$ jednogodišnje vjerojatnosti smrti, a greške $\epsilon_x(t)$ su gaussovske. Sada funkciju

$$\theta_t(\kappa) = \sum_{x=x_1}^{x_m} \left(\ln \frac{p_x(t)}{q_x(t)} - \kappa_t^{(1)} - \kappa_t^{(2)} x \right)^2 \tag{3.4}$$

želimo minimizirati za svaku kalendarsku godinu t te procijeniti parametre $\kappa_t^{(1)}$ i $\kappa_t^{(2)}$. Za razliku od Lee-Carterovog modela, gdje procijenjeni vremenski indeks κ_t ovisi o periodu opažanja, kod Cairns-Blake-Dowdovog modela vremenski indeksi $\kappa_t^{(1)}$ i $\kappa_t^{(2)}$ se procjenjuju zasebno za svaku kalendarsku godinu t.
Što se tiče optimalnog vremenskog perioda opažanja opet nalazimo razlike između Lee-Carterovog i Cairns-Blake-Dowdovog modela. Kako parametre ovisne o vremenu procjenjujemo zasebno za svaku kalendarsku godinu, na njih ne utječe promjena raspona kalendarskih godina prilikom opažanja. Uobičajeno je da se kao početna godina uzima maksimum početnih godina za svaki vremenski faktor.
Razliku nalazimo i u rezidualima modela. Kod Cairns-Blake-Dowdovog modela oni su dani formulom:

$$r_{xt} = \frac{\hat{\epsilon}_x(t)}{\sqrt{\frac{1}{(x_m - x_1 - 1)(t_n - t_1 + 1)} \sum_{x=x_1}^{x_m} \sum_{t=t_1}^{t=t_n} (\hat{\epsilon}_x(t))^2}} \,. \tag{3.5}$$

3.2 Lee - Carterov dobno-vremensko-kohortni model projekcije smrtnosti

Lee-Carterov model se može proširiti i na projekciju po dobnim skupinama. Uz notaciju kao do sada, D_{xt} predstavlja broj smrti u populaciji dobi x i u vremenu t. Dostupan je niz podataka (d_{xt}, ETR_{xt}) gdje je d_{xt} stvarni (opaženi) broj smrti te ETR_{xt} odgovarajuća izloženost riziku, kao što je objašnjeno u ranijim poglavljima (1.1). Intezitet smrtnosti i empirijske stope smrtnosti su označene sa $\mu_x(t)$ te $m_x(t)$ respektivno, gdje je $m_x(t) = d_{xt}/ETR_{xt}$. Unakrsna klasifikacija (eng. $cross - classification$) se vrši ili za svaku kalendarsku godinu $t \in [t_1, t_n]$ po dobi $x \in [x_1, x_m]$ ili je grupirana u m kategorija za svaku dob x. U drugom slučaju godina rođenja (početna godina kohorte koja se prati) ili kohortna godina $z = t - x \in [t_1 - x_m, t_n - x_1]$ (gdje je rang $n + m - 1$) je dobro definirana. Pretpostavimo da vrijedi drugi slučaj. Jednadžba Lee-Carterovog modela (2.1) je:

$$\mu_x(t) = \exp\left(\alpha_x + \beta_x \kappa_t\right) \tag{3.6}$$

uz normalizaciju parametara (2.2):

$$\sum_{t=t_1}^{t_n} \kappa_t = 0, \ \sum_{x=x_1}^{x_m} \beta_x = 1 \,. \tag{3.7}$$

Jednadžba (3.6) se raspisuje u sljedeći oblik:

$$\mu_x(t) = \exp(\alpha_x + RF(x,t)), \tag{3.8}$$

gdje je RF redukcijski faktor smrtnosti dan sa:

$$\ln RF(x,t) = \beta_x \kappa_t . \tag{3.9}$$

Ograničenja pod 3.7 se prilagode tako da

$$\ln RF(x,t_n) = 0, \ \forall x . \tag{3.10}$$

Slijedi generalizacija strukture modela kako bi se dobio izraz koji odgovara pojedinim dobnim skupinama. Potrebno je razmotriti APC (eng. *age − period − cohort*) verziju modela koju se 2006. godine predložili Renshaw i Haberman:

$$\ln \mu_x(t) = \alpha_x + \beta_x^0 \iota_{i-x} + \beta_x^1 \kappa_t . \tag{3.11}$$

Redukcijski faktor smrtnosti je sada dan sa:

$$RF(x,t) = \exp(\beta_x^0 i_{i-x} + \beta_x^1 \kappa_t), \tag{3.12}$$

gdje bilinearan izraz $\beta_x^0 i_{i-x}$ predstavlja efekt za različite dobne skupine.
Iz jednadžbi 3.11 i 3.12 možemo dobiti i osnovni Lee-Carterov model ako je $\beta_x^0 = 0$ te AC (eng. *age − cohort*) model po dobnim skupinama ako je $\beta_x^1 = 0$.
Slijedi analiza pogrešaka modela. Neka je D_{xt} slučajan broj smrti, varijabla iz poissonovske distribucije. Sada definiramo prva dva momenta varijabli odziva Y_{xt},

$$Y_{xt} = D_{xt}, \tag{3.13}$$

kao:

$$\mathbb{E}[Y_{xt}] = ETR_{xt}\, \mu_x(t) = ETR_{xt} \exp(\alpha_x) RF(x,t) \tag{3.14}$$

te

$$\mathbb{V}ar[Y_{xt}] = \phi \mathbb{E}[Y_{xt}], \tag{3.15}$$

gdje je ϕ neki skalar. Sada je nelinearni prediktor η_{xt} definiran sa:

$$\eta_{xt} := \ln \mathbb{E}[Y_{xt}] = \ln ETR_{xt} + \alpha_x + \ln RF(x,t) . \tag{3.16}$$

Neka su težine $w_{xt} = 1$ osim u slučaju kada nema podatka u ćeliji. Tada je težina jednaka nuli. Alternativna struktura pogreške je dana sa:

$$Y_{xt} = \ln\left(\frac{D_{xt}}{ETR_{xt}}\right). \tag{3.17}$$

Sada je

$$\mathbb{E}[Y_{xt}] = \alpha_x + \ln RF(x,t) \tag{3.18}$$

te

$$\mathbb{V}ar[Y_{xt}] = \frac{\phi}{w_{xt}}. \tag{3.19}$$

Slijede dva alternativna modela prilagodbe za nelinearni prediktor η_{xt}. Prvu metodu je predložio Wilmoth 1993. godine, a drugu metodu su predložili James i Segal 1982. godine. Druga metoda se temelji na metodi analize smrtnosti koja uključuje interakcije između dobi i vremena te ima primjenu kod medicinskih statističkih istraživanja. Slijedi prikaz navedenih metoda u kontekstu Lee-Carterovog modela te opis procedura prilagodbe za APC i AC verzije modela objašnjenje u ovom poglavlju.
U prvoj proceduri koriste se procjene parametara pomoću metode najmanjih kvadarata objašnjene u 2.2 te jednadžbe 3.15 i 3.16. Iterativna formula procjene metodom najmanjih kvadrata objašnjena je sljedećim koracima:

1. Određivanje početnih vrijednosti $\hat{\alpha}_x, \hat{\beta}_x$ i $\hat{\kappa}_t$ pomoću kojih se izračuna $\hat{Y}_{xt}$.

2. Ažuriranje parametra $\hat{\alpha}_x$ te ponovno računanje $\hat{Y}_{xt}$.
   Ažuriranje parametra $\hat{\kappa}_t$ i prilagodba uvjetu normalizacije $\sum_{t=t_1}^{t_n} \kappa_t = 0$ te ponovno računanje $\hat{Y}_{xt}$.
   Ažuriranje parametra $\hat{\beta}_x$ te ponovno računanje $\hat{Y}_{xt}$.
   Računanje $D(Y_{xt}, \hat{Y}_{xt})$.

3. Ponavljanje ciklusa; ciklus staje kada $D(Y_{xt}, \hat{Y}_{xt})$ konvergira.

Ovdje su

$$Y_{xt} = \ln \hat{m}_x(t) \;\text{ i }\; \hat{Y}_{xt} = \hat{\alpha}_x + \hat{\beta}_x \hat{\kappa}_t \tag{3.20}$$

te

$$D(Y_{xt}, \hat{Y}_{xt}) = \sum_{x,t} 2\, w_{xt} \int_{\hat{Y}_{xt}}^{Y_{xt}} \frac{Y_{xt} - u}{V(u)} du = \sum_{x,t} w_{xt} (Y_{xt} - \hat{Y}_{xt})^2 \tag{3.21}$$

sa težinama

$$w_{xt} = \begin{cases} 1, & ETR_{xt} > 0 \\ 0, & ETR_{xt} = 0 \end{cases} \tag{3.22}$$

i funkcijom $V(u) = 1$. Ažuriranje nekog parametra θ provodi se Newton-Raphsonovom metodom (2.2):

$$\theta' = u(\theta) = \theta - \frac{\frac{\partial D}{\partial \theta}}{\frac{\partial^2 D}{\partial \theta^2}} \tag{3.23}$$

gdje je D devijacija modela, a θ' ažurirani parametar. Uobičajene svojstvene vrijednosti koje zadovoljavaju 3.7 su $\hat{\kappa}_t = 0$ i $\hat{\beta}_x = \frac{1}{k}$, gdje je k broj dobi x, a parametar $\hat{\alpha}_x$ se procjenjuje pomoću SVD dekompozicije:

$$\hat{\alpha}_x = \frac{1}{t_n - t_1 + 1} \sum_{t=t_1}^{t_n} \ln \hat{m}_x(t). \tag{3.24}$$

$\hat{\alpha}_x$ se procjenjuje logaritmom geometrijske sredine empirijskih stopa smrtnosti.
Druga iterativna formula generira procjene metode najmanjih kvadrata uz poissonovsku strukturu pogrešaka objašnjenu u 3.14 i 3.15 na bazi:

$$Y_{xt} = D_{xt} \tag{3.25}$$

i

$$\hat{Y}_{xt} = \hat{D}_{xt} = ETR_{xt} \exp(\hat{\alpha}_x + \hat{\beta}_x \hat{\kappa}_t) \tag{3.26}$$

što slijedi iz 1.23. Slijedi:

$$D(Y_{xt}, \hat{Y}_{xt}) = \sum_{x,t} 2\, w_{xt} \int_{\hat{Y}_{xt}}^{Y_{xt}} \frac{Y_{xt} - u}{V(u)} du = \sum_{x,t} 2\, w_{xt} \left\{ Y_{xt} \ln\left(\frac{Y_{xt}}{\hat{Y}_{xt}}\right) - (Y_{xt} - \hat{Y}_{xt}) \right\}, \tag{3.27}$$

gdje je $V(u) = u$.
U drugoj proceduri iterativna formula je sljedeća:

1. Postavljanje početnih vrijednosti $\hat{\beta}_x$.

2. Za dane $\hat{\beta}_x$ ažuriraju se $\hat{\alpha}_x$ i $\hat{\kappa}_t$.
   Za dane $\hat{\kappa}_t$ ažuriraju se $\hat{\alpha}_x$ i $\hat{\beta}_x$.
   Računanje $D(Y_{xt}, \hat{Y}_{xt})$.

3. Ponavljanje ciklusa; ciklus staje kada $D(Y_{xt}, \hat{Y}_{xt})$ konvergira.

Za dane $\hat{\beta}_x$ ili $\hat{\kappa}_t$ ažuriranje parametara se vrši odabirom nekog linearnog modela te prilagodbom prediktora, koji je linearan u odnosu na preostale parametre. Stoga su poissonovske varijable odaziva dane sa $Y_{xt} = D_{xt}$ poredane s ciljem generiranja istih rezultata kao u iterativnom procesu iz 2.2. Na početku modela navode se ograničenja

$$\kappa_{t_1} = 0, \quad \sum_x \beta_x = 1\,, \tag{3.28}$$

što je u suglasnosti s uvjetom normalizacije 3.7.

Prilagodba APC i AC modela

APC modeliranje predstavlja problem jer uključuje tri faktora: dob, vrijeme i kohorte, gdje je kohorta = vremenski period-dob. Kako bi bio osiguran jedinstveni set parametara primjenjuje se sljedeća taktika procjene parametra. Prvo se procjenjuju α_x pomoću SVD dekompozicije kao u 3.24. Onda se procjenjuju preostali parametri redukcijskog parametra RF pomoću drugu metode opisane u ovom poglavlju. Poissonove varijable odaziva su dane sa $Y_{xt} = D_{xt}$ te su dana ograničenja:

$$\sum_x \beta_x^0 = 1,\ \sum_x \beta_x^1 = 0 \text{ te ili } \iota_{t_1 - x_k} = 0 \text{ ili } \kappa_{t_1} = 0\,. \tag{3.29}$$

Efektivne početne vrijednosti su $\beta_x^0 = \beta_x^1 = \frac{1}{k}$, gdje je k broj dobi x. Jednom kada je procijenjen parametar α_x procjena preostalih parametara je moguća i prvom procedurom. U tom slučaju početne vrijednosti su $\beta_x^0 = \beta_x^1 = 1$.
Kod AC verzije Lee-Carterovog modela jedinstvenost modela se postiže ograničenjima

$$\iota_{t_1 - x_k} = 0 \text{ i } \sum_x \beta_x^0 = 1. \tag{3.30}$$

Prilagodba modela se tada vrši pomoću prve metode u terminima α_x, β_x^0 i $\iota_{t_1 - x_k}$. Prilagodba modela je moguća i pomoću druge metode zamjenom κ_t sa $\iota_{t_1 - t_k}$. Početne vrijednosti su $\beta_x^0 = 1$ te se sukladno tome prilagođavaju glavni efekti strukture modela $\alpha_x + \iota_{t-x}$ u skladu sa distribucijskim pretpostavkama u prvoj metodi.

Projekcije stopa smrtnosti

Projicirane stope smrtnosti:

$$\dot{m}_x(t_n + s) = \hat{m}_x(t_n)\, \dot{RF}(x, t_n + s),\ s > 0 \tag{3.31}$$

se izračunavaju poravnjanjem sa zadnjim dostupnim stopama smrtnosti $\hat{m}_x(t_n)$. Projicirani redukcijski faktor je dan formulom:

$$\dot{RF}(x, t_n + s) = \exp\left[\hat{\beta}_x^0(\tilde{\iota}_{t_n - x + s} - \hat{\iota}_{t_n - x}) + \hat{\beta}_x^1(\dot{\kappa}_{t_n + s} - \hat{\kappa}_{t_n})\right] \tag{3.32}$$

gdje je

$$\tilde{\iota}_{t_n - x + s} = \begin{cases} \hat{\iota}_{t_n - x + s},\ 0 < s \le x - x_1 \\ \dot{\iota}_{t_n - x + s},\ s > x - x_1 \end{cases}. \tag{3.33}$$

Redukcijski faktor se bazira na procjenama parametara $\hat{\beta}_x^i, \hat{\iota}_z, \hat{\kappa}_t^i$ te predviđanjima vremenskih nizova:

$$\begin{aligned} \{\hat{\iota}_z \in [t_1 - x_k, t_n - x_1]\} &\mapsto \{\dot{\iota}_{t_n - x_1 + s} : s > 0\} \\ \{\hat{\kappa}_t : t \in [t_1, t_n]\} &\mapsto \{\dot{\kappa}_{t_n + s} : s > 0\}\,. \end{aligned} \tag{3.34}$$

Vrijedi

$$\lim_{s\to 0} \dot{RF}(x, t_n + s) = 1. \tag{3.35}$$

Procjene vremenskog niza se uglavnom generiraju pomoću ARIMA procesa, najčešće pomoću slučajne šetnje s pomakom (ARIMA(0,1,0)). Ekstrapolirane stope smrtnosti se rastavljaju na dva faktora sljedećom formulom:

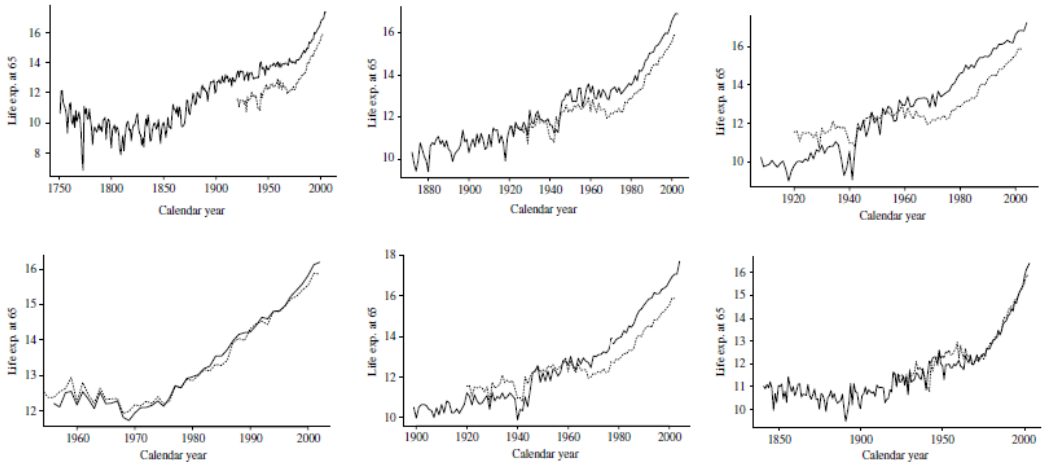
$$\dot{m}_x(t_n + s) = \exp(\,\hat{\alpha}_x^0 + \hat{\beta}_x^0 \hat{\iota}_{t_n - x} + \hat{\beta}_x^1 \hat{\kappa}_{t_n})\, \dot{RF}(x, t_n + s),\ \ s > 0. \tag{3.36}$$

Poglavlje 4

Praktična primjena

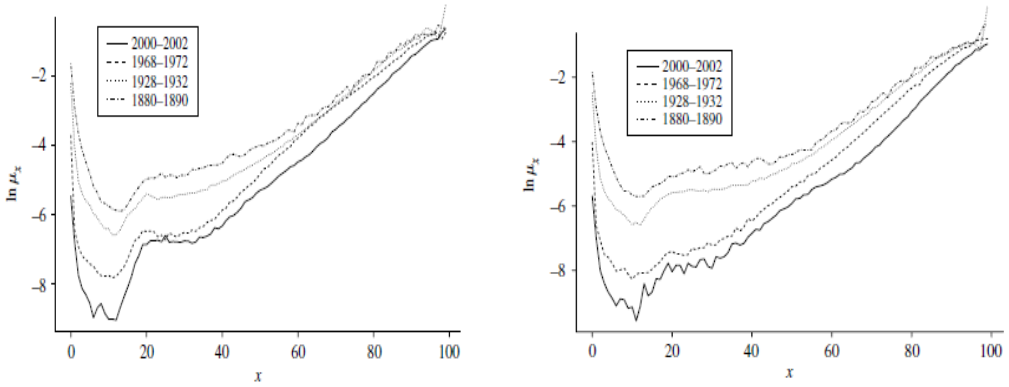
4.1 Primjer

Duljina trajanja života se povećavala od postanka čovjeka. Među ranim ljudima očekivani životni vijek je bio od 20 do 30 godina dok se u današnje doba, razvojem tehnologije, zdravstva i promjenom načina života, on skoro utrostručio. Oko 1750. godine prikupljeni su prvi državni podaci o populaciji u nordijskim zemljama. U to vrijeme, očekivano trajanje života pri rođenju bilo je od 35 do 40 godina u razvijenijim zemljama. Oko 1850. ta brojka je porasla na očekivano trajanje života pri rođenju od 40 do 45 godina. Nagle promjene počele su krajem 19. stoljeća te je već sredinom 20. stoljeća ta brojka iznosila od 60 do 65 godina. Do početka 21. stoljeća očekivano trajanje života je doseglo brojku od 70 godina, dakako u razvijenijim zemljama. Sada se uvodi pojam dugovječnosti. Najveći porast dugovječnosti se dogodio zadnjih 150 godina te nastavlja rasti. U prvoj polovici 20. stoljeća došlo je do značajnog smanjenja smrtnosti dojenčadi i djece te majki pri porodu. To je posljedica poboljšanja javnog zdravstva i prehrane koji su doveli do smanjenja zaraznih bolesti. Došlo je i do duljeg životnog vijeka kod starijih osoba zahvaljujući poboljšanju medicinske skrbi. Pojavila su se nova otkrića vezana za najčešće uzroke smrtnosti, kardiovaskularne bolesti i tumore. Stanovništvo je u prosjeku zadnjih godina starije životne dobi zbog dugovječnosti i smanjenog broja novorođenčadi. Kako je smrtnost dinamičan proces za predviđanje budućnosti je bitno pažljivo analizirati prošle trendove. Pitacco, Denuit, Haberman i Olivieri su proučavali smrtnost belgijske populacije kod koje je uočen problem dugovječnosti [2]. Također su usporedili životno očekivanje u Belgiji sa životnim očekivanjima nekih drugih zemalja Europe; Švedske, Italije, Španjolske, zapada Njemačke, Francuske, Engleske i Walesa (4.1). Iako je u tim zemljama trend usporediv, regionalne razlike u stopama smrtnosti stvaraju razlike među životnim očekivanjima pojedinih zemalja, posebno u dobi u kojoj se ide u mirovinu. Podatke za istraživanje omogućila im je službena belgijska statistička agencija Statistics Belgium. Jedan od zadataka Statistic Belgium je konstrukcija tablica



Slika 4.1: Usporedba Belgije (isprekidana linija) redom sa Švedskom, Italijom, Španjolskom,zapadnom Njemačkom, Francuskom te Engleskom i Walesom

smrtnosti po dobnim skupinama, posebno za muškarce i žene. U svakoj tablici smrtnosti belgijske populacije smrtnost je reprezentirana skupom jednogodišnjih vjerojatnosti smrti q_x. U ovom istraživanju korištene su tablice smrtnosti za periode od 1880.-1890. godine, 1928.-1932., 1968.-1972. i za period od 2000.-2002. godine kako bi se istražila dugoročna evolucija dugovječnosti u Belgiji. Slike u ovom poglavlju preuzete su iz literature [2]. Proces nastanka tablica smrtnosti je sljedeći. Tablice smrtnosti stvaraju hipotetsku dobnu

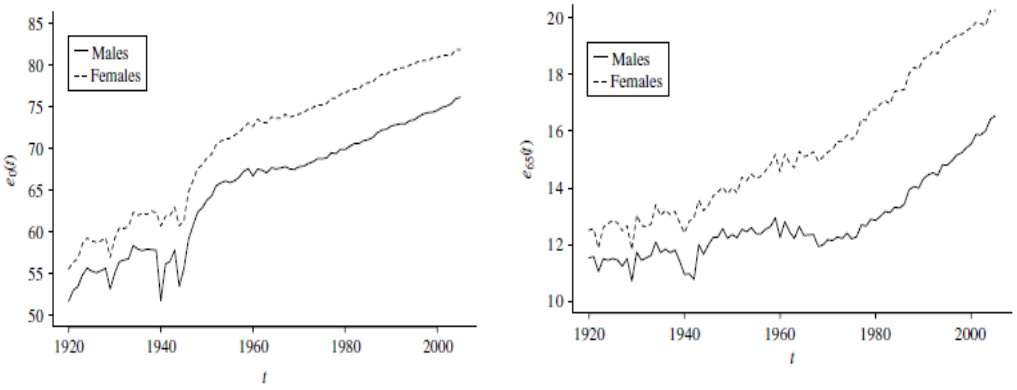


(a) Muškarci (b) Žene

Slika 4.2: Stope smrtnosti

skupinu od recimo 100 000 osoba dobi 0, obično za muškarce i žene posebno. Tada se

ti podaci upotpune dobno-spolnim specifičnim godišnjim vjerojatnostima smrti, odnosno brojem smrti za 1000, 10 000 ili 100 000 osoba određene dobi i spola opaženih u danoj populaciji. Na ovaj način analitičari mogu pratiti kako se broj od 100 000 hipotetskih osoba (zvanih generacijska kohortna skupina) smanjuje zbog smrti kako grupa stari. Postoje dvije osnovne vrste tablica smrtnosti, periodičke tablice smrtnosti i kohortne tablice smrtnosti. Periodične tablice smrtnosti reprezentiraju smrtnost populacije tijekom relativno kratkog vremenskog perioda, obično između jedne i tri godine. Tablice smrtnosti koje su bazirane na opažanjima neke populacije uglavnom se konstruiraju kao periodočne tablice smrtnosti jer su novi podaci o populaciji uglavnom dostupni periodično, na primjer u vrijeme popisa stanovništva. Ove tablice su korisne prilikom analize promjena u smrtnosti populacije tokom vremena te su stoga korištene u ovom istraživanju. Promjena u smrtnosti je funkcija dobi x i vremena t pa se ovaj pristup naziva dobno-vremenski pristup. Pretpostavka istraživanja je da su dobni parametri smrtnosti konstantni za svaki par (x, t), ali se razlikuju od jednog para do drugog. Slike 4.2 prikazuju stope smrtnosti za dane periode tablica smrtnosti za muškarce i žene. Slike 4.3 prikazuju trend u periodu očekivanog trajanja života pri

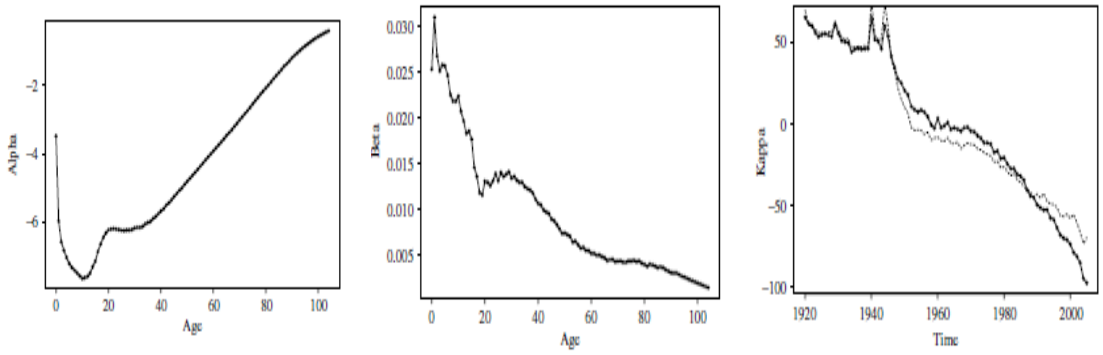


(a) Očekivano trajanje života pri rođenju

(b) Očekivano trajanje života u dobi od 65 godina

Slika 4.3: Očekivano trajanje života

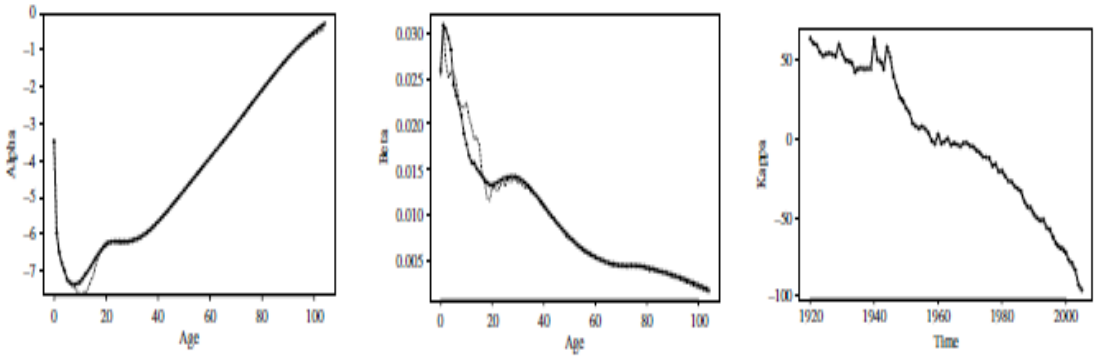
rođenju $e_0^{\uparrow}(t)$ i u dobi za mirovinu, odnosno 65 godina, $e_{65}^{\uparrow}(t)$. Period očekivanog trajanja života u određenoj dob x zasnovan je na stopama smrtnosti za tu i sve više godine života (dob) opažene u specifičnoj godini. U analizu su uključene kalendarske godine od 1920. do 2005. i dobi od 0 do 104 godine. Uzeta je gornja granica od 104 godine jer belgijska tablica smrtnosti za period od 2002. do 2004. godine, koja služi kao baza predviđanja, maksimalno doseže dob od 104 godine. Slika 4.4 prikazuje procijenjene parametre. Procijenjeni $\hat{\alpha}_x$ prikazuje tipičan oblik skupa logaritamskih stopa smrtnosti sa relativno visokim vrijednostima oko rođenja i padom u mlađoj dobi i rastom kako se dob povećava. Procijenjeni $\hat{\beta}_x$ se smanjuju s dobi sugerirajući tako da su padovi smrtnosti koncentrirani u mlađoj



(a) $\hat{\alpha}_x$ (b) $\hat{\beta}_x$ (c) $\hat{\kappa}_t$

Slika 4.4: Procijenjeni parametri

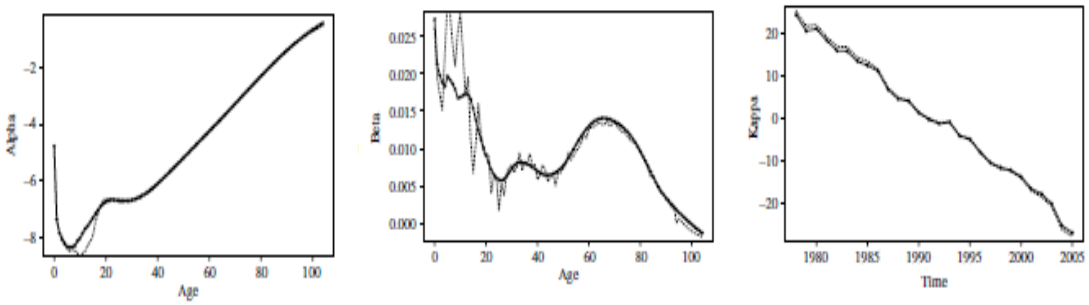
dobi. Procijenjeni $\hat{\kappa}_t$ su prilagođeni kako bi reproducirali opaženi period očekivanja života pri rođenju. Model objašnjava 92.9 posto varijance modela. Ovaj postotak dobijemo kao omjer prve svojstvene (singularne) vrijednosti SVD dekompozicije i zbroja svih svojstvenih vrijednosti. Sljedeći korak je izglađivanje stopa smrtnosti kao što je objašnjeno u poglavlju 2.4. Iz slike 4.5 vidimo da izglađivanje ne utječe toliko na parametre α_x i κ_t, ali utječe na parametar β_x. Ovaj model objašnjava 93.7 posto varijance. Sada slijedi određi-



(a) $\hat{\alpha}_x$ (b) $\hat{\beta}_x$ (c) $\hat{\kappa}_t$

Slika 4.5: Izglađeni procijenjeni parametri

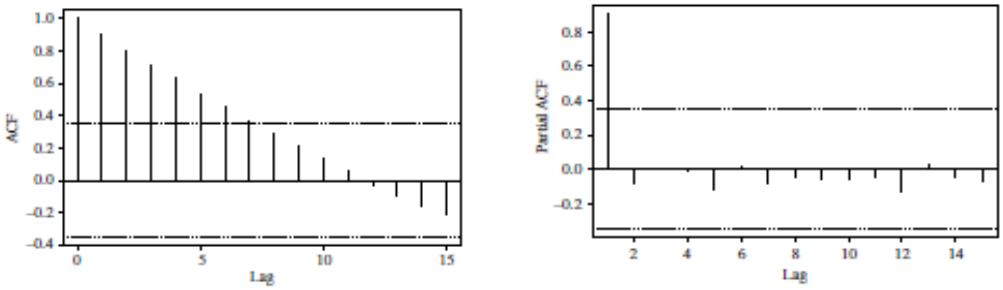
vanje vremenskog perioda. Iz procedure objašnjene u poglavlju 2.5 (Booth) slijedi da je $t_{start} = 1978$. Zbog toga se odbacuju opažanja od 1920. do 1977. godine. Na sljedećoj slici 4.6 su prikazani procijenjeni parametri kada ograničimo podatke na kalendarske godine od 1978. do 2005. godine. Sada model prije izglađivanja objašnjava 67.7 posto ukupne varijance (za muškarce), a izglađeni model 90.57 posto. Slijedi analiza reziduala koja je ta-



(a) $\hat{\alpha}_x$ (b) $\hat{\beta}_x$ (c) $\hat{\kappa}_t$

Slika 4.6: Procijenjeni parametri za optimalni period 1978.-2005. godine

kođer objašnjena u poglavlju 2.5 Reziduali otkrivaju sistematične uzorke i relativno velike vrijednosti za mlađe dobi. Prisutnost trenda kod reziduala narušava pretpostavku nezavisnosti. Također je upitna homoskedastičnost jer se na grafu primjećuje grupiranje uzoraka. Ako se uzorak ograniči samo na dob od 60 i više godina reziduali postaju nestrukturirani. Idući korak je modeliranje κ_t pomoću modela vremenskog niza. Iz slike 4.7, odnosno autokorelacijske i parcijalne autokorelacijske funkcije, čini se da ARIMA(0,1,0) odgovara ovom vremenskom nizu. Dodatna pojašnjenja su dana u poglavlju 2.6. Shapiro-Wilk test



(a) ACF (b) PACF

Slika 4.7: Autokorelacijska i parcijalna autokorelacijska funkcija

normalnosti daje p-vrijednost koja iznosi 0.2308. Stoga se ne može odbaciti nulta hipoteza o normalnosti reziduala. To potvrđuje Jarque-Bera test normalnosti sa p-vrijednošću koja iznosi 0.4827. Slijedi zaključak da je slučajna šetnja s pomakom dobar model za procjenu κ_t. Iz 2.42 i 2.43 slijedi da je $\hat{d} = -0.58677$ i $\hat{\sigma}^2 = 0.39858$ kod belgijske populacije muškaraca za optimalni period od 1974. do 2005. godine.

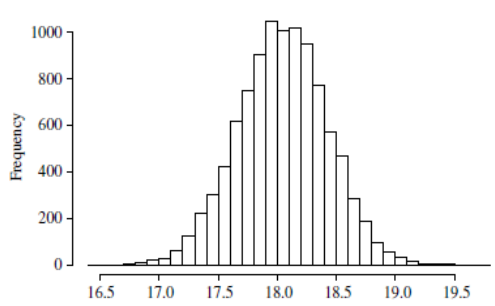
U Lee-Carterovom modelu buduće dobno specifične stope smrtnosti se dobivaju korištenjem ekstrapoliranih κ_t i fiksnih α_x i β_x. Točnije projekcije $\hat{\kappa}_{t_n+s}$ od κ_{t_n+s} za $s = 1, 2, \ldots$ se uvrštavaju u formulu za intenzitet smrtnosti

$$\hat{\mu}_x(t_n + s) = \exp(\hat{\alpha}_x + \hat{\beta}_x \dot{\kappa}_{t_n+s}) \,. \tag{4.1}$$

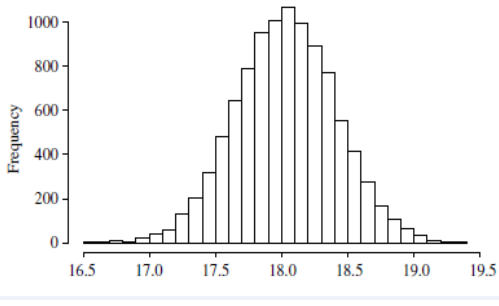
U ovom slučaju, odskočne stope (stope u zadnoj godini perioda prilagodbe) su prilagođene stope. Neki stručnjaci, kao što je Bell, predlažu da predviđanje treba započeti opaženim, a ne prilagođenim stopama kako bi se smanjio skok između opaženih i predviđenih stopa smrtnosti u prvoj godini predviđanja. Predviđene stope smrtnosti su povezane za zadnjim dostupnm empirijskim stopama smrtnosti idućom fomulom:

$$\dot{\mu}_x(t_n + s) = \hat{m}_x(t_n) \exp\left(\hat{\beta}_x(\dot{\kappa}_{t_n+s} - \hat{\kappa}_{t_n})\right) = \hat{m}_x(t_n)\, RF(x, t_n + s) \,. \tag{4.2}$$

RF ovdje označava redukcijski faktor dan formulom 3.9, a $\hat{m}_x(t_n)$ označava stope smrtnosti opažene u dobi x i vremenu t_n, a ne one prilagođene. U ovom primjeru korištena je belgijska tablica smrtnosti za period od 2002. do 2004. godine.
Sada slijedi bootstrap metoda. Sa 10 000 replikacija dobiju se histogrami na slici 4.8 za životno očekivanje po dobnim skupinama za muškarce $\bar{e}^{\nearrow}_{65}(t)$. Vrijednost točke predviđanja je 18.17. Očekivanje bootstrap vrijednosti je 18.05, a standardna devijacija je 0.3802. Bootstrap 90 postotni interval pouzdanosti je [17.41183, 18.6609]. Za Poissonov bootstrap rezultati su gotovo identični. Očekivanje iznosi 18.03, a standardna devijacija 0.3795. Bootstrap 90 postotni interval pouzdanosti iznosi [17.40993,18.6558]. Histogrami za obje bootstrap metode (4.8) se gotovo podudaraju.



(a) Histogram bootstrap - uzorka



(b) Histogram Poissonovog bootstrap - uzorka

Slika 4.8: Histogrami

4.2 Implementacija u R-u

Postoje i funkcije programskog jezika R koje olakšavaju modeliranje pomoću Lee-Carterovog modela. U paketu *demography* je nekoliko funkcija koje predviđaju stope smrtnosti pomoću Lee-Carterovog modela. Jedna od čestih funkcija je

```
forecast.lca
```

gdje se parametri κ_t procjenjuju pomoću slučajne šetnje s pomakom te se nakon toga množe sa parametrima β_x kako bismo dobili krivulju demografskih stopa. Funkcija je oblika

```
forecast(object, h = 50,
jumpchoice = c("fit", "actual"), level = 80, ...)
```

gdje je *object* izlaz funkcije *lca*, *h* broj godina za koje predviđamo stope, a *jumpchoice* metoda koju koristimo za stope zadnje godine, biramo između stvarnih stopa zadnje godine i prilagođenih. Argument *level* označava značajnost predikcijskog intervala. Ovdje je

```
lca(data, series=names(data$rate)[1], years=data$year,
ages=data$age, max.age=100,adjust = c("dt"))
```

funkcija Lee-Carterovog modela.

Bibliografija

[1] Basrak B., *Vremenski nizovi, predavanja*, https://web.math.pmf.unizg.hr/~bbasrak/, pristupljeno:13.03.2017.

[2] Denuit M. Olivieri A. Haberman S., Pitacco E., *Modelling Longevity Dynamics for Pensions and Annuity Business*, Oxford University Press, 2009.

[3] Gerber H.U., *Life Insurance Mathematics*, Springer, 1997.

[4] Marušić M., *Praktikum iz numeričkih metoda u statistici, predavanja*, https://web.math.pmf.unizg.hr/nastava/pnms/, pristupljeno 14.02.2017.

[5] Singer S., *Numerička analiza, predavanja*, https://web.math.pmf.unizg.hr/~singer/num_mat/index.html, pristupljeno: 15.05.2017.

Sažetak

U ovom diplomskom radu opisan je Lee-Carterov model za projiciranje smrtnosti. U preliminarnom dijelu pojašnjena je aktuarska notacija potrebna za razumijevanje modela, te metode koje se primjenjuju prilikom procjene parametara modela. Spomenute metode su dekompozicija singularnih vrijednosti, ARIMA model vremenskog niza te Newton-Raphsonova metoda rješavanja nelinearnih jednandžbi.
Lee-Carterov model je detaljno pojašnjen uz opise primjene metode najmanjih kvadrata te Poissonove metode maksimalne vjerodostojnosti prilikom određivanja parametara modela. Radi boljih projekcija stopa smrtnosti dan je i opis izglađivanja podataka koji se provodi u modelu te odabir optimalnog vremenskog perioda opažanja podataka. Analiza reziduala se provodi kako bi se provjerila učinkovitost modela. Također je dan i opis samog projiciranja smrtnosti.
Kako bi se dobio bolji uvid u model dana je njegova usporedba s Cairns-Blake-Dowdovim modelom te opis njegovog proširenja. Proširenje se provodi uključujući i kohorte u dobno-vremensku analizu modela.
Za kraj je dan praktičan primjer primjene Lee-Carterovog modela na podatke za belgijsku populaciju iz literature. Također je dodan kratak opis implementiranih procedura Lee-Carterovog modela u programskom alatu R.

Summary

In this master's thesis is described the Lee-Carter's model for the projection of the mortality. The preliminary part describes actuarial notation necessary to understand the model and the methods that are used during the estimation of the parameters of the model. Mentioned methods are decomposition of the singular values, ARIMA model of time series and Newton-Raphson method for solving non-linear equations.
Lee-Carter's model is described in detail with additional descriptions of the least squares method and Poisson's maximum likelihood method used to estimate the parameters of the model. In order to gain better projections of mortality rates there is also given the description of the smoothing of the data and selection of the optimal time period in which data is being collected. There is also the description of the projection of the mortality itself. To get a better insight of the model there is described its comparison to the Cairns-Blake-Dowd's model. There is also described the model's extension. It applies to the addition of the cohort effect to the model that includes time and age components.
To conclude, it is given the practical example of usage of the model on the Belgium population from the literature. There is also given short description of the implemented procedures of the Lee-Carter's model in the programming tool R.

Životopis

Jelena Magdalena Kostelac rođena je 20.04.1994. godine u Zagrebu. Trenutno prebivalište joj je na adresi Petra Preradovića 57 u Jasenovcu. U Jasenovcu je završila i Osnovnu školu Jasenovac. Daljnje školovanje je nastavila upisavši XV.gimaziju u Zagrebu, poznatu i kao MIOC. U 2012. godini je završila srednju školu te upisala smjer Matematika na Matematičkom odsjeku Prirodoslovno-matematičkog fakulteta. Preddiplomski studij završila je 2015. godine i postala sveučilišna prvostupnica (baccalaurea) matematike. Iste godine je upisala diplomski studij Matematička statistika, također na Matematičkom odsjeku Prirodoslovno-matematičkog fakulteta.
Tokom studiranja radila je nekoliko studentskih poslova, kao što je prikupljanje donacija za UNICEF te posao pomoćnog konobara u Le Bistrou hotela Esplanade. Stručnu praksu odradila je Hrvatskoj agenciji za nadzor financijskih institucija te u Privrednoj banci Zagreb. Također, na fakultetu je sudjelovala na natjecanju eStudenta Case Study Competition, gdje je tim čija je članica bila došao u finale radeći na slučaju Zagrebačke banke. Ove godine je radila i za tvrtku Nesi Projekt na projektu za lanac hotela Maistra Adris grupe.
Tečno govori engleski jezik i dobro vlada alatima Microsoft Office (Word, Excel i PowerPoint) te programima R, SAS i MATLAB. Ima osnovno znanje o programskim jezicima C,C++ i Python. U slobodno vrijeme druži se s bliskim ljudima, putuje te povremeno odlazi u teretanu.