

APPENDIX A

Technische beschrijving van Prognosetafel AG2014

1 Definities

De prognosetafel geeft per geslacht¹ voor leeftijden $x \in \mathcal{X} := \{0, 1, 2, \dots, 120\}$ en jaren $t \in \mathcal{T} := \{2014, 2015, \dots, 2064\}$ de best estimate voor de éénjarige sterftkansen $q_x(t)$. Dat is de kans dat iemand die op 1 januari van jaar t leeft en op 1 januari van jaar $t - x$ geboren werd op 1 januari van jaar $t + 1$ gestorven zal zijn. Het model stelt de gebruiker overigens in staat ook voor de jaren na 2064 een prognose op te stellen.

De sterftkansen worden niet direct gemodelleerd maar in plaats daarvan specificeren we de bijbehorende force of mortality (of 'hazard rate') $\mu_x(t)$. We veronderstellen dat $\mu_{x+s}(t+s) = \mu_x(t)$ voor alle $0 \leq s < 1$ en dan geldt

$$q_x(t) = 1 - \exp(-\mu_x(t)). \quad (1)$$

Elk dynamisch model dat wordt beschreven in termen van de force of mortality $\mu_x(t)$ kan dus met bovenstaande formule omgeschreven worden in termen van éénjarige sterftkansen.

2 Dynamisch Model

Voor leeftijden tot en met 90 jaar, (x, t) in $\tilde{\mathcal{X}} \times \mathcal{T}$ met $\tilde{\mathcal{X}} = \{0, 1, 2, \dots, 90\}$, wordt het Li-Lee model gebruikt²:

$$\ln \mu_x(t) = \ln \mu_x^{\text{EU}}(t) + \ln \mu_x^{\text{NL}}(t) \quad (2)$$

$$\ln \mu_x^{\text{EU}}(t) = A_x + B_x K_t \quad (3)$$

$$\ln \mu_x^{\text{NL}}(t) = \alpha_x + \beta_x \kappa_t \quad (4)$$

met een dynamica die voor elke $x \in \tilde{\mathcal{X}}$ en $t \geq 2013$ gegeven wordt door

$$K_{t+1} = K_t + \theta + \epsilon_{t+1} \quad (5)$$

$$\kappa_{t+1} = a \kappa_t + \delta_{t+1} \quad (6)$$

Dit betekent dat een random walk with drift model verondersteld wordt voor West-Europese veranderingen en een eerste orde autoregressief model voor specifiek Nederlandse sterftedynamica.

3 Sluiting van de Tafel

Voor leeftijden boven de 90 jaar, (x, t) in $\tilde{\mathcal{X}} \times \mathcal{T}$ met $\tilde{\mathcal{X}} = \{91, 92, \dots, 120\}$, wordt de sluitingsmethode van Kannistö gebruikt die gebaseerd is op een logistische regressie aan de hand van de tafel voor leeftijden $y \in \mathcal{X}^k = \{80, 81, \dots, 90\}$. Het aantal leeftijden y_k waar de regressie op gebaseerd wordt is dus $n = 11$, het gemiddelde van die leeftijden is $\bar{y} = (1/n) \sum_{k=1}^n y_k = 85$ en de kwadraatsom van de afwijking is $\sum_{k=1}^n (y_k - \bar{y})^2 = 110$.

Sluiting middels Kannistö betekent dat voor $x \in \tilde{\mathcal{X}}$

$$\mu_x(t) = L \left(\sum_{k=1}^n w_k(x) L^{-1}(\mu_{y_k}(t)) \right).$$

Hierin zijn L en L^{-1} respectievelijk de logistische en inverse logistische functies:

$$L(x) = 1/(1 + \exp(-x))$$

$$L^{-1}(x) = -\ln((1/x) - 1),$$

en de regressiegewichten worden gegeven door

$$w_k(x) = \frac{1}{n} + \frac{(y_k - \bar{y})(x - \bar{y})}{\sum_j (y_j - \bar{y})^2} = \frac{1}{11} + \frac{(y_k - 85)(x - 85)}{110}.$$

¹Alle modellen en parameters worden voor mannen en vrouwen los van elkaar geschat.

²Li, N. and Lee, R. (2005) Coherent Mortality Forecasts for a Group of Populations: An Extension of the Lee-Carter Method. *Demography* 42(3), pp. 575-594.

4 Benodigde Grootheden

Voor dit model zijn de volgende parameters nodig:

- $A_x, B_x, \alpha_x, \beta_x$ zijn leeftijdsafhankelijke parameters die voor elke leeftijd (en apart voor elk geslacht) $x \in \tilde{\mathcal{X}}$ gegeven worden, en
- de parameters a en θ en de startwaarden voor de twee tijdreeksen, K_{2013} en κ_{2013} , worden ook expliciet (en apart per geslacht) gegeven.

De stochastische variabelen $E_t := (\epsilon_t, \delta_t)$ zijn onderling onafhankelijk en hebben voor elke waarde van $t \in \mathcal{T}$ dezelfde verdeling: de bivariate normale verdeling met gemiddelde $(0, 0)$ en een covariantiematrix C die expliciet gegeven is. Deze covariantiematrix is niet nodig om de best estimates voor sterftkansen te genereren, maar wel nodig om simulaties uit te voeren die kunnen helpen om de onzekerheid rond deze sterftkansen in kaart te brengen.

Figuur 1 laat het resultaat van een stochastische simulatie zien. Voor verschillende leeftijden zijn de sterftkansen getoond tot en met 2013 en de best estimates en het 95% betrouwbaarheidsinterval voor die kansen vanaf 2013.

5 Best estimate, Gemiddelde, Mediaan en Kwantielen

Omdat we de *best estimate* voor toekomstige waarden van de tijdreeksen identificeren met de *most likely* uitkomst, komt die overeen met de reeks voor $(K_t^{\text{BE}}, \kappa_t^{\text{BE}})$ die verkregen wordt door voor alle $t \in \mathcal{T}$ de waarden $\epsilon_t = \delta_t = 0$ in te vullen. De sterftetafel die met die tijdreeks gegenereerd wordt is de prognosetafel 2014-2064. Als men die tafel gebruikt om de waarde van een portefeuille door te rekenen vindt men de best estimate (in de zin van: most likely) waarde van die portefeuille.

We merken op dat hetzelfde niet gezegd kan worden van, bijvoorbeeld, het gemiddelde, de mediaan of kwantielen. Immers, de waarde van de portefeuille is een niet-lineaire functie f van de stochastische sterftkansen. Gemiddelde, mediaan en kwantielen voor de uitkomsten van f wanneer de sterftkansen ingevuld worden kunnen niet bepaald worden door gemiddelden, medianen en kwantielen van de sterftkansen in te vullen in die functie. Voor het eenvoudige voorbeeld $f(x) = x^2$ weten we bijvoorbeeld dat voor stochastische X de waarden van $\mathbb{E}f(X)$ en $f(\mathbb{E}X)$ niet hetzelfde zijn; het verschil is immers de variantie van X . Wat geldt voor de niet-lineaire transformatie f van stochastische X , geldt ook voor de niet-lineaire transformatie van stochastische sterftkansen naar portefeuillewaarden: door simulatie van stochastische uitkomsten en het vervolgens bepalen van portefeuillewaarden kan men de verdeling, en dus gemiddelde, mediaan en kwantielen, bepalen. Hoewel het gewoon invullen van gemiddelden, medianen etc. dus niet de exacte uitkomst oplevert blijkt het verschil vaak klein te zijn als de portefeuillewaarden strikt stijgende of strikt dalende functies van sterftkansen zijn.

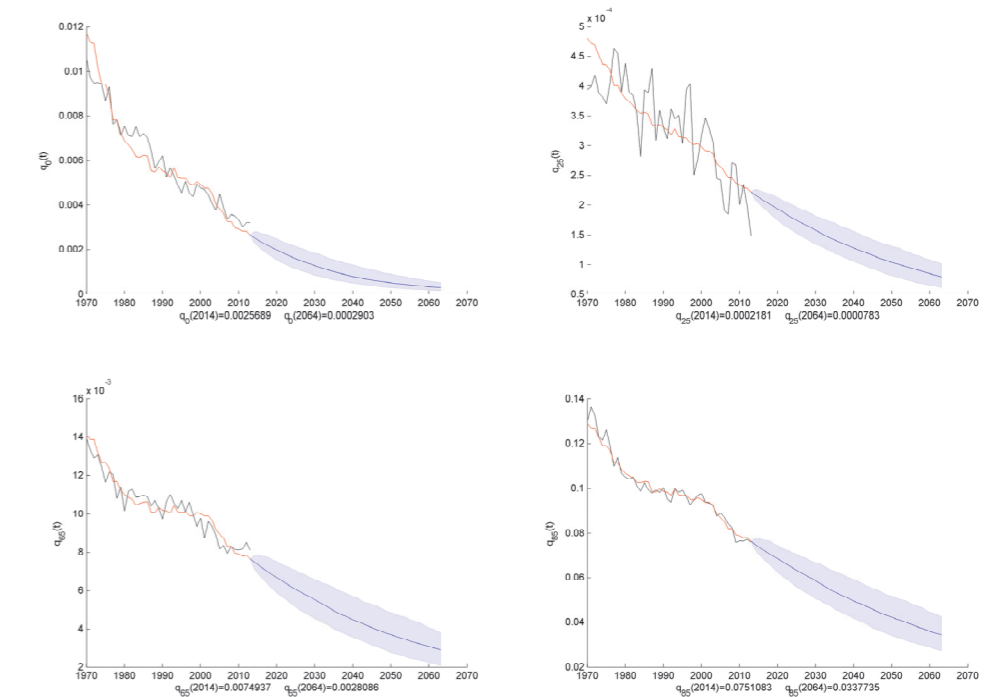


Fig 1. Sterftkansen, vrouwen

6 Best Estimate van Periodelevensverwachting & Cohortlevensverwachting

Als we de resterende levensverwachting van iemand willen bepalen op 1 januari van jaar t onder de aanname dat die persoon op 1 januari van jaar $t - x$ geboren werd (met $x \in \mathcal{X}$ en $t \in \mathcal{T}$) en aannemen dat iemand die binnen een kalenderjaar sterft gemiddeld nog de helft van dat kalenderjaar in leven is, dan vinden we voor de zogenaamde *cohortlevensverwachting*:

$$e_x^{\text{coh}}(t) = \frac{1}{2} + \sum_{k=0}^{\infty} \prod_{s=0}^k (1 - q_{x+s}(t+s))$$

Wanneer een sterftekans $q_{x+s}(t+s)$ nodig is voor een leeftijd $x + s > 120$ dan kan men die gelijk veronderstellen aan $q_{120}(t+s)$. De tabel kent geen beperkingen meer wat betreft de tijdshorizon: men kan $q_{x+s}(t+s)$ bepalen voor $t + s > 2064$ door de best estimate berekeningen voort te zetten na tijdstip 2064. Om de levensverwachtingen in 2064 nog voor alle leeftijden goed te kunnen inschatten is het aan te raden om als horizon voor alle berekeningen 120 jaar na 2064 te nemen. Merk op dat we volgens bovenstaande formule "diagonaal door de prognosetafel lopen". De kans dat de persoon op tijdstip $t+k$ nog leeft is immers het product van overlevingskansen $1 - q_{x+s}(t+s)$ voor alle jaren s tussen 0 en k waarbij ieder jaar de persoon niet alleen een jaar ouder wordt, maar we ook telkens met een nieuwe kolom in de sterftetafel te maken krijgen.

Dit laatste effect wordt niet meegenomen in de *periodelevensverwachting*:

$$e_x^{\text{per}}(t) = \frac{1}{2} + \sum_{k=0}^{\infty} \prod_{s=0}^k (1 - q_{x+s}(t))$$

die suggereert dat de sterftekansen van vandaag (tijdstip t) niet meer zullen veranderen in de tijd. Dit leidt tot een verkeerd beeld van de levensverwachting en hoewel deze periodelevensverwachting vaak nog aangeduid wordt met "de levensverwachting" is dat onjuist.

Figuren 2 en 3 tonen het verloop tot 2013 en de projecties na 2013 van respectievelijk de periodelevensverwachting en de cohortlevensverwachting voor mannen. Wanneer we die met elkaar vergelijken vallen twee zaken op. Ten eerste blijkt de cohortlevensverwachting bij jonge leeftijden regelmatig te verlopen in de tijd. Omdat men bij de berekening van die verwachting diagonaal door de tabel loopt, wordt de dynamiek in latere jaren nu al meegenomen. Dit in tegenstelling tot de berekening van de periodelevensverwachting, die alleen kansen van het huidig jaar t gebruikt waardoor bij overgang naar de volgende waarde van t dus relatief grotere veranderingen optreden. Ten tweede zien we dat bij projecties van de levensverwachting in de toekomst de cohortlevensverwachting de onzekerheid over toekomstige sterftekansen direct zichtbaar maakt. Bij de periodelevensverwachting lijkt het alsof de onzekerheid slechts langzaam toeneemt in de tijd, omdat men de dynamica in latere jaren bewust niet meeneemt bij het bepalen van de toekomstige overlevingskansen.

7 Implementatie

Wanneer men niet alleen geïnteresseerd is in de huidige prognosetafel maar het model ook wil gebruiken om stochastische scenario's te genereren voor toekomstige sterftekansen, kan men het volgende schema aanhouden.

1. Simuleer de tijdreeksen (K_t, κ_t) volgens de vergelijkingen (5)-(6). De startwaarden in $t = 2013$ zijn steeds hetzelfde en expliciet gegeven. Om een stochastische variabele (ϵ_t, δ_t) met covariantiematrix

$$C = \begin{bmatrix} \sigma_\epsilon^2 & \rho\sigma_\epsilon\sigma_\delta \\ \rho\sigma_\epsilon\sigma_\delta & \sigma_\delta^2 \end{bmatrix}$$

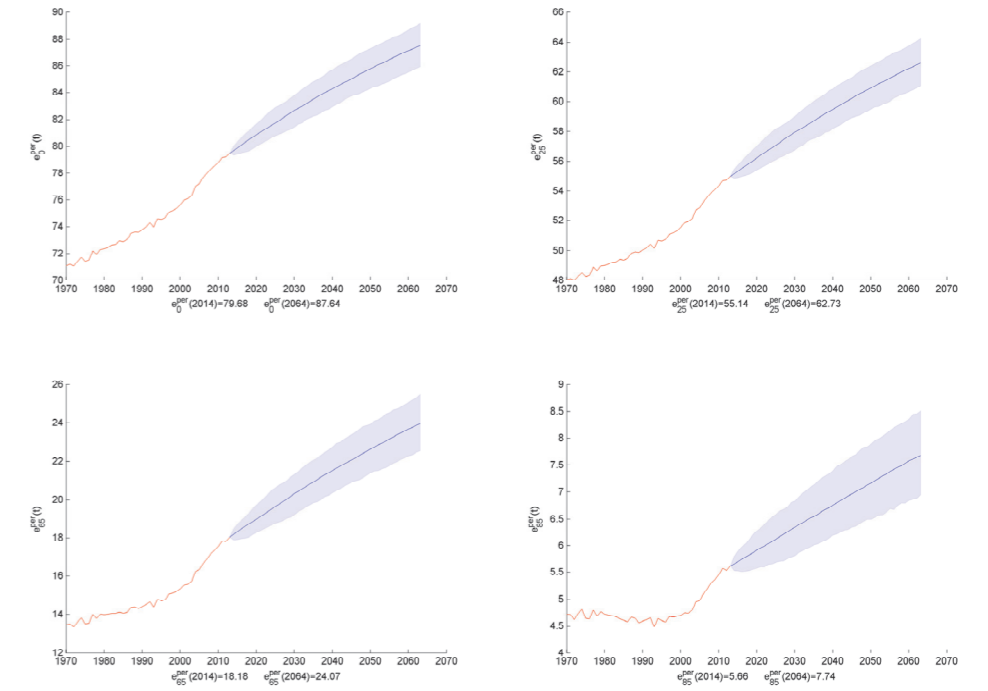


Fig 2. Periodelevensverwachting, mannen

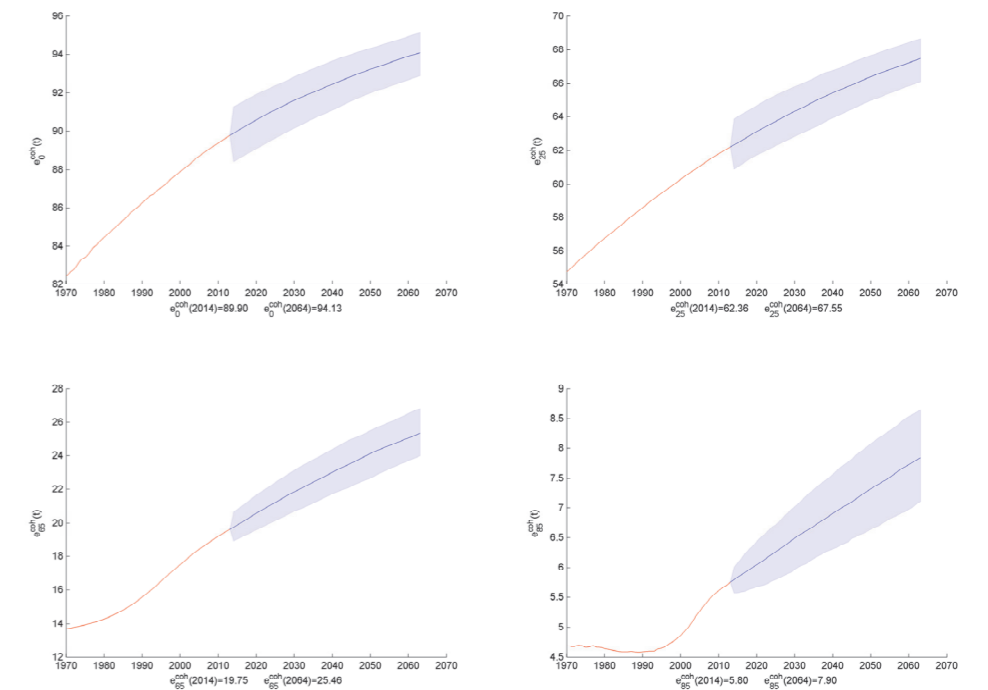


Fig 3. Cohortlevensverwachting, mannen

te genereren kan men twee onafhankelijke standaard-normaal verdeelde variabelen $Z_1 \sim N(0, 1)$ en $Z_2 \sim N(0, 1)$ nemen en vervolgens $\epsilon_t = \sigma_\epsilon Z_1$ en $\delta_t = \sigma_\delta(\rho Z_1 + \sqrt{1 - \rho^2} Z_2)$ uitrekenen. De waarden van σ_ϵ , σ_δ en ρ zijn makkelijk te bepalen als de covariantiematrix C gegeven is. Als men N simulaties uitvoert tot en met eindjaar T , genummerd met $i = 1, \dots, N$ dan heeft men dus voor $t = 2014, 2015, \dots, T$ nu tijdreeksrealisaties (K_t^i, κ_t^i) .

2. Door de uitkomsten in te vullen in (3)-(4) vindt men $\ln \mu_x^{i,EU}(t)$ en $\ln \mu_x^{i,NL}(t)$, voor alle x tot en met een leeftijd van $x = 90$ jaar, want alle benodigde tijdsafhankelijke parameters (A_x , α_x , B_x en β_x) liggen vast en zijn gegeven. Vervolgens is voor die leeftijden, alle tijden t en alle scenario's i de waarde van $\mu_x^i(t) = \exp(\ln \mu_x^{i,EU}(t) + \ln \mu_x^{i,NL}(t))$ uit te rekenen.

3. Per scenario i dient de tafel vervolgens gesloten te worden volgens de procedure van Kannistö. Voor $x > 90$ bepalen we daarom $\mu_x^i(t) = L(\sum_{k=1}^n w_k(x) L^{-1}(\mu_{y_k}^i(t)))$. Het is handig om een paar expliciete tussenstappen te maken:

$$\begin{aligned} H_k^i(t) &= L^{-1}(\mu_{y_k}^i(t)) \\ G_x^i(t) &= \left(\frac{1}{11} - \frac{85(x-85)}{110} \right) \sum_{k=1}^{11} H_k^i(t) + \frac{(x-85)}{110} \sum_{k=1}^{11} y_k H_k^i(t) \\ \mu_x^i(t) &= L(G_x^i(t)) \end{aligned}$$

als men de berekeningen terug wil brengen tot een aantal simpele sommaties. Dit maakt implementatie in Excel bijvoorbeeld makkelijker met behulp van SUM en SUMPRODUCT.

4. Tot slot vindt men de prognosetafel voor elk scenario i middels (1): $q_x^i(t) = 1 - \exp(-\mu_x^i(t))$.

5. Toepassen van de prognosetafel voor elk scenario i op een portefeuille met verplichtingen levert N verschillende waarden als uitkomst op en dus een verdeling voor de stochastische waarde van de portefeuille. Die kan vervolgens gebruikt worden om mediaan, gemiddelde en kwantielen te bepalen. Het aantal simulaties bepaalt de nauwkeurigheid van de verdeling en bijbehorend gemiddelde, mediaan en kwantielen.

8 Kalibratie van het Model

De parameterwaarden in bovenstaand model zijn bepaald door maximum likelihood kalibratie. Daarbij is gebruik gemaakt van Deaths D_{xt} en Exposures E_{xt} zoals die te vinden zijn in de Human Mortality Database en Statline van het Centraal Bureau voor de Statistiek. Steeds is aangenomen dat voor gegeven exposures de geobserveerde sterftegevallen D_{xt} een Poissonverdeling hebben en dat de verwachting van D_{xt}/E_{xt} gelijk is aan de te modelleren force of mortality $\mu_x(t)$.

1. Uit de Human Mortality Database³ zijn exposures E_{xt}^{EU} en deaths D_{xt}^{EU} verkregen voor de relevante West-Europese landen, met $x \in \mathcal{X}^o := \{0, 1, \dots, 90\}$ en $t \in \mathcal{T}^o := \{1970, \dots, 2009\}$. Het betreft steeds de som van alle exposures en de som van alle sterftegevallen in de betreffende landen, inclusief Nederland. De parameters A_x , B_x en K_t zijn vervolgens zo bepaald dat de Poisson likelihoodfunctie voor de geobserveerde sterftes D_{xt}^{EU} zo groot mogelijk is bij de gegeven exposures E_{xt}^{EU} :

$$\max_{A_x, B_x, K_t} \prod_{x \in \mathcal{X}^o} \prod_{t \in \mathcal{T}^o} (E_{xt}^{EU} \mu_x^{EU}(t))^{D_{xt}^{EU}} \exp(-E_{xt}^{EU} \mu_x^{EU}(t)) / (D_{xt}^{EU}!).$$

Om een unieke specificatie van de drie vectoren te krijgen normaliseren we door te eisen dat de som van de elementen van K_t over $t \in \mathcal{T}^o$ gelijk is aan 1 en de som van de elementen van B_x over $x \in \mathcal{X}^o$ gelijk is aan 0.

2. Niet alle Europese landen hebben gegevens voor de periode 2010-2013 in de Human Mortality Database staan. Daarom zijn de waarden van K_t in de vorige stap bepaald tot en met 2009. Vervolgens wordt lineair geëxtrapoleerd voor 2010 tot en met 2013, dus

$$K_{2009+s} = K_{2009} + s(K_{2009} - K_{1970}) / (2009 - 1970),$$

voor $s \in \{1, 2, 3, 4\}$.

³De gebruikte gegevens zijn gedownload op 29 mei 2014.

3. Voor Nederland zijn Exposures en Deaths na 2009 uit de CBS database gehaald⁴. Dit levert voor alle $x \in \mathcal{X}$ en $t \in \{2010, 2011, 2012, 2013\}$ de waarden voor

- P_{xt} , de bevolking op 1 januari van jaar t met een leeftijd tussen x en $x + 1$, en
- C_{xt} , het aantal mensen dat overleden is binnen jaar t , die op 31 december van jaar t tussen x en $x + 1$ jaar oud zouden zijn.

Omzetting naar Exposures en Deaths vindt plaats met behulp van de methode die vastgelegd is in het protocol van de Human Mortality Database⁵. Dat betekent dat we uitrekenen voor $t \in \{2010, 2011, 2012, 2013\}$ en $x \in \{1, 2, \dots, 90\}$:

$$\begin{aligned} E_{xt}^{NL} &= \frac{1}{2}(P_{x,t} + P_{x,t+1}) + \frac{1}{6}(\frac{1}{2}C_{xt} - \frac{1}{2}C_{x+1,t}) \\ D_{xt}^{NL} &= \frac{1}{2}(C_{x,t} + C_{x+1,t}) \end{aligned}$$

en voor $x = 0$:

$$\begin{aligned} E_{0,t}^{NL} &= \frac{1}{2}(P_{0,t} + P_{0,t+1}) + \frac{1}{6}(C_{0,t} - \frac{1}{2}C_{1,t}) \\ D_{0,t}^{NL} &= C_{0,t} + \frac{1}{2}C_{1,t}. \end{aligned}$$

De Nederlandse Exposures and Deaths vóór 2010 zijn direct uit de Human Mortality Database te halen.

4. De maximum likelihood methode wordt nu toegepast op alleen de Nederlandse data om α_x , β_x en κ_t te bepalen, middels

$$\max_{\alpha_x, \beta_x, \kappa_t} \prod_{x \in \mathcal{X}^o} \prod_{t \in \bar{\mathcal{T}}} (E_{xt}^{NL} \mu_x(t))^{D_{xt}^{NL}} \exp(-E_{xt}^{NL} \mu_x(t)) / (D_{xt}^{NL}!),$$

met $\bar{\mathcal{T}} = \{1970, \dots, 2013\}$ en \mathcal{X}^o als voorheen. In deze expressie zijn de eerder bepaalde waarden voor A_x , B_x en K_t weer ingevuld omdat die nodig zijn om $\mu_x(t)$ uit te rekenen. Opnieuw wordt genormaliseerd, door de som van elementen in κ_t over $t \in \bar{\mathcal{T}}$ en β_x over $x \in \mathcal{X}^o$ respectievelijk 1 en 0 te laten zijn.

5. De laatste stap maakt dan gebruik van de tijdreeks $\{(K_t, \kappa_t) \mid t \in \bar{\mathcal{T}}\}$ om de parameters a en θ en de matrix C te schatten die nodig zijn om toekomstige scenario's te genereren. Onder de gemaakte aanname dat de variabelen $E_t := (\epsilon_t, \delta_t)$ onderling onafhankelijk zijn en een bivariate normale verdeling met gemiddelde $(0, 0)$ en covariantiematrix C hebben, kiezen we de schatters voor a , θ en C zodanig dat de likelihood voor deze tijdreeks gemaximaliseerd wordt.

⁴Zie <http://statline.cbs.nl/StatWeb/publication/?DM=SLNL&PA=37530NED&D1=2,4&D2=0-100&D3=0&D4=60-62&HDR=T,G3&STB=G2,G1&VW=T> en <http://statline.cbs.nl/StatWeb/publication/?DM=SLNL&PA=71090ned&D1=0&D2=1-2&D3=1-100&D4=0&D5=0&D6=48,60,72,84,96&HDR=T,G3,G1,G5&STB=G2,G4&VW=T>

⁵Het protocol van de Human Mortality Database is te vinden op hun website, zie <http://www.mortality.org/Public/Docs/MethodsProtocol.pdf>. Er is gebruik gemaakt van versie 5.

