LIVING to 100

SOCIETY OF ACTUARIES INTERNATIONAL SYMPOSIUM

2020 Symposium
Jan. 13–15
Lake Buena Vista, FL

3A – Aging Measurement and Mortality Modeling 1

New Developments in Mortality Modeling - Discussion

TOM EDWALDS, FSA, ACAS, MAAA

3A: Aging Measurement and Mortality Modeling 1

January 13, 2020

SOCIETY OF ACTUARIES®

Two Excellent Papers

- Congratulations to authors
  - Mathematically challenging
- Will discuss separately
- Longevity Risk
  - Aggregate vs individual

Huang – Affine Mortality Models

- Continuous annuity: $$\bar{a}_x = \int_0^{\omega} e^{-\mu_{x+t}-\delta t}\, dt$$

- Both μ and δ are complex stochastic processes
- Need to model both to capture risk

Arbitrage-Free Nelson-Siegel model

- Continuous time force of mortality model
- Closed form
  - Components for Level, Slope, and Curvature
  - But otherwise not very intuitive
- Consistent modeling of mortality & interest
  - Appropriate for aggregate longevity risk
- “Winner” in out of sample forecasting (SMAPE)

Cohort Mortality Model

- Advantages
  - Mortality patterns likely to be more related
  - More relevant for long term obligations
    - Annuities, pensions
- Disadvantages:
  - Complete cohorts were born a long time ago
    - Out of sample data was 1916 birth cohort
  - How relevant to current pensioner mortality?

Huynh – Multi-population Model

- Many insurance applications!
  - Male versus female mortality
  - Preferred versus standard mortality
    - And multi-level preferred
  - Company mortality versus industry mortality
  - Insured mortality versus population mortality
- Lengthscale parameter θ determines closeness

Coherent mortality improvement models

- See heat maps on slide 15
  - Color scales are different
- Would have been nice for 2015 VBT development

But Wait!

- Computationally expensive
  - Only used ages 70 – 84 to fit models
  - Suggested limit of 5 populations
- Covariance stationarity constrains shape
  - Suggested modeling age groups separately
  - Implicit Gompertz-Makeham assumption?

Congratulations Again to Authors!

SOCIETY OF ACTUARIES®

SOCIETY OF ACTUARIES®

The Application of Affine Processes in Cohort Mortality Risk Models

Zhiping Huang
with Michael Sherris, Andres Villegas, and Jonathan Ziveyi

School of Risk & Actuarial Studies, UNSW Business School, UNSW Sydney
ARC Centre of Excellence in Population Ageing Research (CEPAR)

2020 Living to 100 Symposium
Hilton Orlando Lake Buena Vista
Jan 13-15, 2020

cepar ARC CENTRE OF EXCELLENCE IN POPULATION AGEING RESEARCH

Introduction

- Mortality models have attracted research attention over recent years, particularly discrete time mortality models (Lee and Carter (1992), Cairns et al. (2006b), Cairns et al. (2009), Renshaw and Haberman (2006)).
    - focus on improvement trends,
    - impact of uncertainty or volatility of mortality, and
    - cohort effects.
- Continuous time affine cohort mortality models have attracted more recent research
    - single cohort models (Milevsky and Promislow (2001), Dahl and Møller (2006), Biffis (2005), Luciano et al. (2008), Schrager (2006), Cairns et al. (2006a), Blackburn and Sherris (2013))
    - multi-cohort models (Jevtic et al. (2013), Xu et al. (2019a), Chang and Sherris (2018), Huang et al. (2019)).

Why Continuous Time Affine Models?

- Analytical tractability - closed form survival curves for affine class,
- Consistency between mortality dynamics and functional form of the survival curve,
- Stability of parameter estimates,
- Use of mathematical finance methods for term structure and credit risk models familiar to financial market participants,
- Natural extensions to multi-factor models, capturing differing trends, volatility and correlations by age,
- Arbitrage-free formulation along with real world dynamics to allow calibration of prices of risk.

Coverage of this Presentation

- Introduce continuous-time affine cohort mortality models - closed-form expressions for survival curves, dynamics of mortality rates, AFNS mortality models with factors of level, slope and curvature for the mortality curve,
- Fitting with age-cohort data,
- Kalman filter and estimation of the models, highlighting how Poisson variation can be incorporated into the model estimation,
- Comparison of fits and cohort survival curve prediction using historical US mortality data.

Survival Curve - Continuous Time Affine Mortality Model

- Drawing on term structure of interest rate models - equivalence of average force of mortality rate to yield to maturity for zero coupon bond. Use of similar notation as in yield curve modelling.
- Survival probability $S(x, t, T)$ for single cohort aged x at time t for survival for a duration $(T - t)$ to age $x + (T - t)$, as an affine function of (latent) factors (3 factor case)

$$
\begin{aligned}
S(x, t, T) &= E[e^{-\int_t^T \mu^i(x,s)ds} | \mathcal{F}_t] \\
&= e^{-\bar{\mu}(t,T)(T-t)} \\
&= e^{B_1(t,T)X_1(t)+B_2(t,T)X_2(t)+B_3(t,T)X_3(t)+A(t,T)}, \qquad (1)
\end{aligned}
$$

- $B_j(t, T)$ are factor loadings (functional form derived from mortality dynamics for the latent factors, exponential terms) and $X_j(t)$ are the latent factors (stochastic parameters).

Mortality Rate - Continuous Time Affine Mortality Model

- Average mortality rate - age-period data or age-cohort data

$$\bar{\mu}(t, T) = -\frac{1}{T-t}\log\left[S(t, T)\right] = -\frac{B(t, T)'}{T-t}X_t - \frac{A(t, T)}{T-t}. \quad (2)$$

  where vector $B(t, T)$, the factor loadings, and $A(t, T)$ have explicit expressions (derivations similar to term structure models).

- Canonical form for these (Blackburn and Sherris, 2013), where δ_{jj} and σ_{jj} are parameters in the latent factor dynamics (estimated from historical data)

$$B_j(t, T) = -\frac{1 - e^{-\delta_{jj}(T-t)}}{\delta_{jj}}, \quad j = 1, 2, 3, \quad (3)$$

$$A(t, T) = \frac{1}{2}\sum_{j=1}^{3}\frac{\sigma_{jj}^2}{\delta_{jj}^3}\left[\frac{1}{2}\left(1 - e^{-2\delta_{jj}(T-t)}\right) - 2\left(1 - e^{-\delta_{jj}(T-t)}\right) + \delta_{jj}(T-t)\right]. \quad (4)$$

Mortality Rate - AFNS Mortality model

- Mortality model equivalent of the Nelson-Seigel term structure model in arbitrage-free dynamic implementation (Christensen et al., 2011)
- Independent AFNS mortality model - mortality rate curve has level, slope and curvature factors.

$$\begin{aligned} B^1(t,T) &= -(T-t), \quad B^2(t,T) = -\frac{1-e^{-\delta(T-t)}}{\delta}, \\ B^3(t,T) &= (T-t)e^{-\delta(T-t)} - \frac{1-e^{-\delta(T-t)}}{\delta}, \end{aligned} \tag{5}$$

$$\begin{aligned} \frac{A(t,T)}{T-t} &= \sigma_{11}^2 \frac{(T-t)}{6} + \sigma_{22}^2 \left[\frac{1}{2\delta^2} - \frac{1}{\delta^3} \frac{1-e^{-\delta(T-t)}}{T-t} + \frac{1}{4\delta^3} \frac{1-e^{-2\delta(T-t)}}{T-t} \right] + \\ & \sigma_{33}^2 \left[\frac{1}{2\delta^2} + \frac{1}{\delta^2} e^{-\delta(T-t)} - \frac{1}{4\delta}(T-t)e^{-2\delta(T-t)} - \frac{3}{4\delta^2} e^{-2\delta(T-t)} \right. \\ & \left. - \frac{2}{\delta^3} \frac{1-e^{-\delta(T-t)}}{T-t} + \frac{5}{8\delta^3} \frac{1-e^{-2\delta(T-t)}}{T-t} \right]. \end{aligned} \tag{6}$$

Affine Mortality Models - Q Pricing measure

- The price of a longevity zero-coupon bond on a specific cohort currently aged x is

$$\begin{aligned}\bar{P}_x(t,T) &= E^Q\left[e^{-\int_t^T (r(s)+\mu(x,s))ds}|\mathcal{F}(t)\right] \\ &= E^Q\left[e^{-\int_t^T r(s)ds}|\mathcal{G}(t)\right] E^Q\left[e^{-\int_t^T \mu(x,s)ds}|\mathcal{H}(t)\right] \\ &= P(t,T)S^Q(x,t,T), \end{aligned} \tag{7}$$

  where the dynamics of the mortality rates and the dynamics of the interest rates are independent.

- Dynamics of the mortality rate used to derive the (risk-neutral) survival probability $S^Q(x,t,T)$.
- Use the same methodology of term structure modelling applied to (pricing) survival probability.

Affine Mortality Models - Dynamics of Mortality Rates

- The affine dynamics of the latent factors X_t follow a system of stochastic differential equations (SDEs) under the risk-neutral measure Q (Duffie and Kan, 1996; Christensen et al., 2011):

$$dX_t = K^Q \left[\theta^Q - X_t\right] dt + \Sigma D\left(X_t, t\right) dW_t^Q, \tag{8}$$

  - where $K^Q \in \mathbb{R}^{n \times n}$ is the mean reversion matrix,
  - $\theta^Q \in \mathbb{R}^n$ is the long-term mean (usually zero in mortality models),
  - $\Sigma \in \mathbb{R}^{n \times n}$ is the volatility matrix,
  - $W_t^Q \in \mathbb{R}^n$ is a standard Brownian motion, and
  - $D\left(X_t, t\right)$ is a diagonal matrix with the ith diagonal entry as $\sqrt{\alpha^i\left(t\right) + \beta_1^i\left(t\right) x_t^1 + \ldots + \beta_n^i\left(t\right) x_t^n}$. α and β are bounded continuous functions.
- Continuous time equivalent of auto-regressive processes for factors.

Affine Mortality Models - ODEs for Factor Loadings

- Under these dynamics the (risk-neutral) survival probabilities for age x for survival from time t to time T are (see details in Blackburn and Sherris, 2013):

$$S(t, T) = \exp\left(B(t, T)' X_t + A(t, T)\right), \tag{9}$$

- Where $B(t, T)$ and $A(t, T)$ are the solutions to the following system of ordinary differential equations (ODEs):

$$\frac{dB(t, T)}{dt} = \rho_1 + \left(K^Q\right)' B(t, T), \tag{10}$$

$$\frac{dA(t, T)}{dt} = -B(t, T)' K^Q \theta^Q - \frac{1}{2} \sum_{j=1}^{3} \left(\Sigma' B(t, T) B(t, T)' \Sigma\right)_{j,j}, \tag{11}$$

with boundary conditions $B(T, T) = A(T, T) = 0$.

Affine Mortality Models - Gaussian and CIR Models

- Dynamics include Gaussian models (where there is a probability of negative mortality rates).
- These models
  - are readily estimated with (Gaussian) Kalman filter
  - are easily simulated
  - in practice, have very low probabilities of negative mortality rates.
- Dynamics also include square root process dynamics with potential to capture mortality heterogeneity (Cox-Ingersoll-Ross or CIR)
- These models
  - can capture mortality heterogeneity (gamma distributed mortality rates)
  - avoid probabilities of negative mortality rates
  - are more difficult to estimate with the Kalman filter (we use maximum quasi-likelihood).

Affine Mortality Models - Historical Mortality Rates

- To calibrate models to historical mortality rates (P measure) we need a link between the risk neutral dynamics and the historical dynamics - assumption for the price of risk.

- Assuming an essentially affine form for the risk premium (Duffee, 2002):

$$\Lambda_t = \begin{cases} \lambda^0 + \lambda^1 X_t, & \text{Gaussian processes;} \\ D\left(X_t, t\right)\lambda^0, & \text{the CIR model,} \end{cases} \tag{12}$$

  where $\Lambda_t \in \mathbb{R}^{n\times 1}$, $\lambda^0 \in \mathbb{R}^{n\times 1}$ and $\lambda^1 \in \mathbb{R}^{n\times n}$.

- The SDEs for factors under the measure P have the same (auto-regressive) form:

$$dX_t = \begin{cases} K^P\left[\theta^P - X_t\right]dt + \Sigma dW_t^P, & \text{Gaussian processes;} \\ K^P\left[\theta^P - X_t\right]dt + \Sigma D\left(X_t, t\right)dW_t^P, & \text{the CIR model.} \end{cases} \tag{13}$$

Affine Cohort Mortality Models - 3-factor Dynamics

The dynamics of the factors (Canonical and AFNS models) are (we estimate both P and Q measure dynamics):

- The independent Blackburn-Sherris model (Blackburn and Sherris, 2013)

$$\begin{pmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \end{pmatrix} = - \begin{pmatrix} \delta_{11} & 0 & 0 \\ 0 & \delta_{22} & 0 \\ 0 & 0 & \delta_{33} \end{pmatrix} \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} dt + \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix} \begin{pmatrix} dW_t^{1,Q} \\ dW_t^{2,Q} \\ dW_t^{3,Q} \end{pmatrix}. \tag{14}$$

- The independent AFNS model (Christensen et al., 2011). The dynamics of the factors under the Q-measure are given by:

$$\begin{pmatrix} dL_t \\ dS_t \\ dC_t \end{pmatrix} = - \begin{pmatrix} 0 & 0 & 0 \\ 0 & \delta & -\delta \\ 0 & 0 & \delta \end{pmatrix} \begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix} dt + \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix} \begin{pmatrix} dW_t^{1,Q} \\ dW_t^{2,Q} \\ dW_t^{3,Q} \end{pmatrix}. \tag{15}$$

- The dependent Blackburn-Sherris model

$$\begin{pmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \end{pmatrix} = - \begin{pmatrix} \delta_{11} & 0 & 0 \\ \delta_{21} & \delta_{22} & 0 \\ \delta_{31} & \delta_{32} & \delta_{33} \end{pmatrix} \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} dt + \begin{pmatrix} \sigma_{11} & 0 & 0 \\ \sigma_{21} & \sigma_{22} & 0 \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \begin{pmatrix} dW_t^{1,Q} \\ dW_t^{2,Q} \\ dW_t^{3,Q} \end{pmatrix}. \tag{16}$$

Affine Cohort Mortality Models - 3-factor Dynamics

- The dependent AFNS model

$$\begin{pmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \end{pmatrix} = -\begin{pmatrix} \delta_{11} & 0 & 0 \\ \delta_{21} & \delta_{22} & 0 \\ \delta_{31} & \delta_{32} & \delta_{33} \end{pmatrix} \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} dt + \begin{pmatrix} \sigma_{11} & 0 & 0 \\ \sigma_{21} & \sigma_{22} & 0 \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \begin{pmatrix} dW_t^{1,Q} \\ dW_t^{2,Q} \\ dW_t^{3,Q} \end{pmatrix}. \tag{17}$$

- The CIR model

$$\begin{aligned} \begin{pmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \end{pmatrix} &= -\begin{pmatrix} \delta_{11} & 0 & 0 \\ 0 & \delta_{22} & 0 \\ 0 & 0 & \delta_{33} \end{pmatrix} \left[\begin{pmatrix} \theta_1^Q \\ \theta_2^Q \\ \theta_3^Q \end{pmatrix} - \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} \right] dt \\ &+ \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix} \begin{pmatrix} \sqrt{X_t^1} & 0 & 0 \\ 0 & \sqrt{X_t^2} & 0 \\ 0 & 0 & \sqrt{X_t^3} \end{pmatrix} \begin{pmatrix} dW_t^{1,Q} \\ dW_t^{2,Q} \\ dW_t^{3,Q} \end{pmatrix}. \end{aligned} \tag{18}$$

Model Estimation - Kalman Filter

- We model mortality rates but observe deaths - we need a measurement equation capturing the effects of Poisson variation and heterogeneity.
- Mortality rate curve changes stochastically through time, driven by latent factors with trend and uncertainty - we need a state transition equation for the dynamics.
- We then filter the values of latent factors from historical data - deriving means and covariances which are functions of the parameters in the dynamics.
- We can then construct the likelihood (Gaussian) in terms of means and covariance (a function of parameters to be estimated).
- Then numerically select the parameter set that maximises the likelihood using an iterative process.

Mortality Data - Estimating Mortality Models

- Mortality models are usually estimated with age-period historical data (life tables) - US data from 1933 to 2015 at ages from 50 to 100 is shown below.
- Cohort mortality rates are required in practice. Age-period models require forecasting of age-period curves and derivation of cohort mortality rates from the diagonal as the cohort ages.

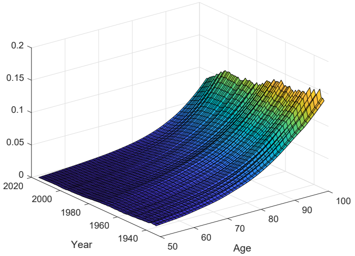


Figure 1: Average force of mortality of US Males Using Age-Period data, from 1933 to 2015

Mortality Data - Estimating Mortality Models

- Age-cohort data allows fitting of age-cohort curves directly but incomplete data for more recent cohorts - see US cohort data below.

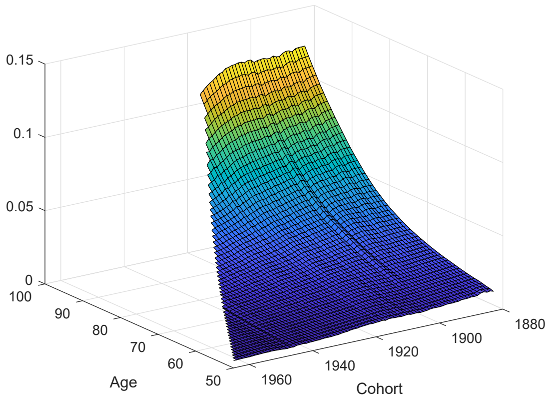


Figure 2: Average Force of Mortality for Males Born from 1883 to 1965

Mortality Data - Estimating Mortality Models

- Complete age-cohort data for cohorts born in earlier years. US complete cohort data below.

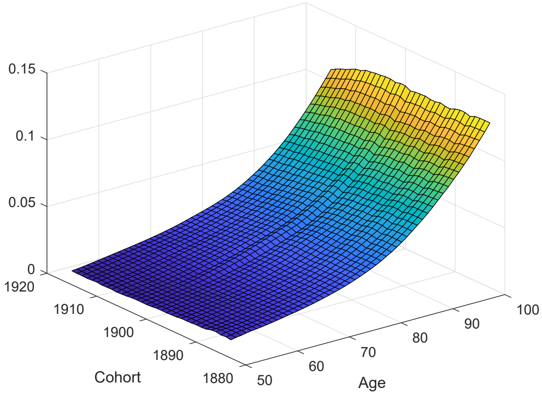


Figure 3: Average Force of Mortality for Males Born from 1883 to 1915

Mortality Data - Calibrating Affine Mortality Models

- US mortality age-cohort data from the Human Mortality Database (2017) (HMD) to calibrate and compare the mortality models.
- Mortality data of males from ages 50 to 100 for the cohorts born from 1883 to 1915.
- Historical survival probability, $S^i(x; t, T)$, and the historical average forces of mortality $\bar{\mu}^i(x; t, T)$ over the period $\tau = T - t$ for each cohort i aged x at time t from the data, using:

$$S^i(x; t, T) = \prod_{s=1}^{T-t} \left[1 - q^i(x + s - 1, t + s - 1)\right], \tag{19}$$

$$\bar{\mu}^i(x; t, T) = -\frac{1}{T - t} \log \left[S^i(x; t, T)\right], \tag{20}$$

where $q^i(x, t)$ is the one year death probability for an individual aged x at time t in cohort i.

Affine Cohort Mortality Models - Goodness of Fit

Table 1: Comparison of Affine Mortality Models

	The Blackburn-Sherris Model		The AFNS Model		The CIR Model
	Independent-Factor	Dependent-Factor	Independent-Factor	Dependent-Factor	
Log Likelihood	9896.419	9938.696	9665.801	9887.878	10045.70
RMSE	0.00250	7.601e-04	6.856e-04	9.160e-04	5.227e-04
No. of Parameters	12	18	10	13	18
AIC	-19570.837	-19643.392	-19113.602	-19551.757	-19857.40
BIC	-18968.292	-19008.277	-18521.914	-18943.783	-19222.29
Probability of Negative Mortality	0.02700	1.011e-32	1.722e-31	4.34e-14	-

- AFNS model fits historical age-cohort data well. Low negative mortality probabilities. CIR the best fit.

Canonical Age-Period Mortality Curve Factors

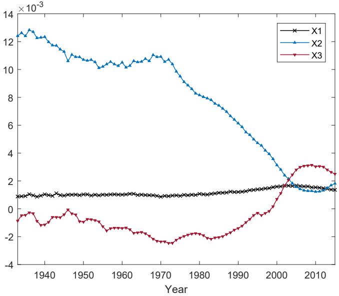
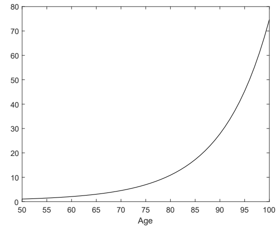


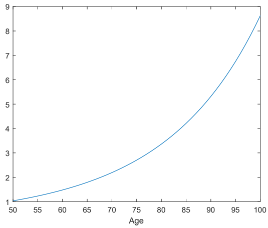
Figure 4: Factors in the Blackburn-Sherris Model with Age-Period Data

- Factor X_2 captures trend change around 1970's.

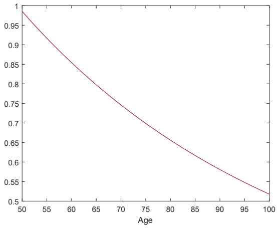
Canonical Age-Period Mortality Curve Factor Loadings



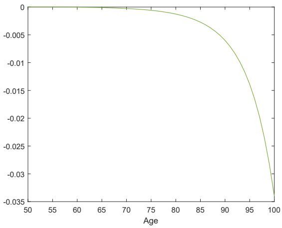
(a) $B_1 = -\frac{B^1(t,T)}{T-t}$



(b) $B_2 = -\frac{B^2(t,T)}{T-t}$



(c) $B_3 = -\frac{B^3(t,T)}{T-t}$



(d) $A = -\frac{A(t,T)}{T-t}$

AFNS Age-Cohort Mortality Curve Factors

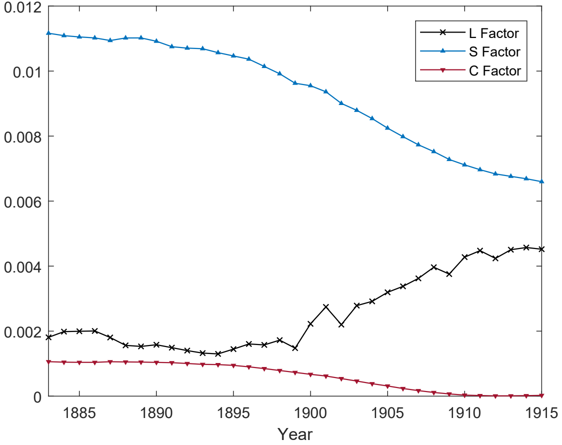
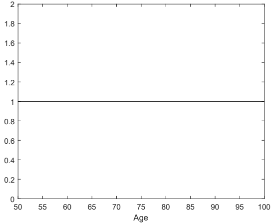
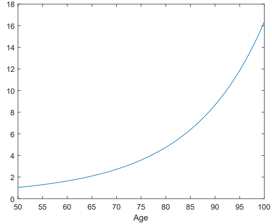


Figure 6: Factors in the Independent AFNS Model

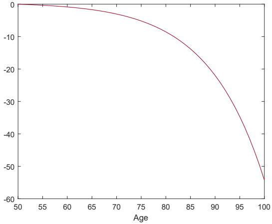
AFNS Age-Cohort Mortality Curve Factor Loadings



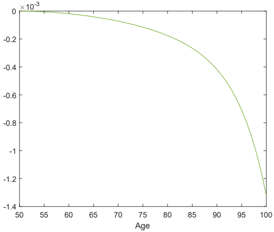
(a) $B_1 = -\frac{B^1(t,T)}{T-t}$



(b) $B_2 = -\frac{B^2(t,T)}{T-t}$

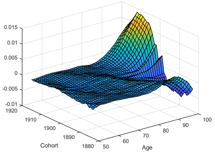


(c) $B_3 = -\frac{B^3(t,T)}{T-t}$

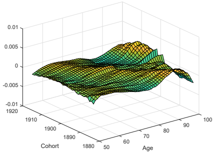


(d) $A = -\frac{A(t,T)}{T-t}$

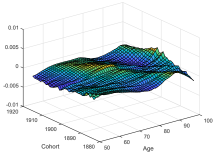
Affine Cohort Mortality Models - Residual Analysis



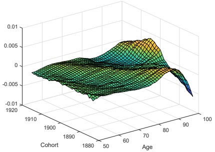
(a) The Independent Blackburn-Sherris Model



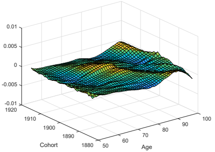
(b) The Dependent Blackburn-Sherris Model



(c) The Independent AFNS Model



(d) The Dependent AFNS Model



(e) The CIR Model

Figure 8: Residuals of Affine Mortality Models

Affine Cohort Mortality Models - MAPE

Mean Absolute Percentage Error (MAPE) for each age, across all cohorts

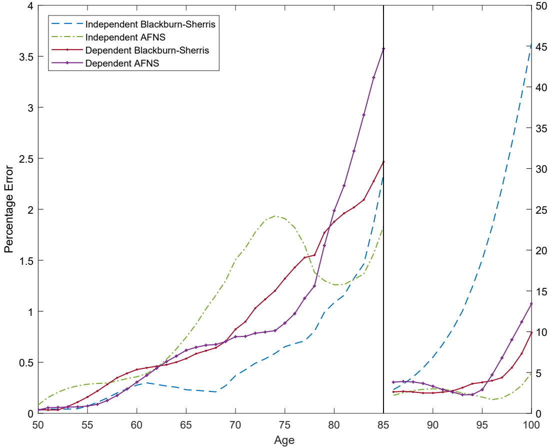


Figure 9: The Models with Gaussian Processes

Affine Cohort Mortality Models - MAPE

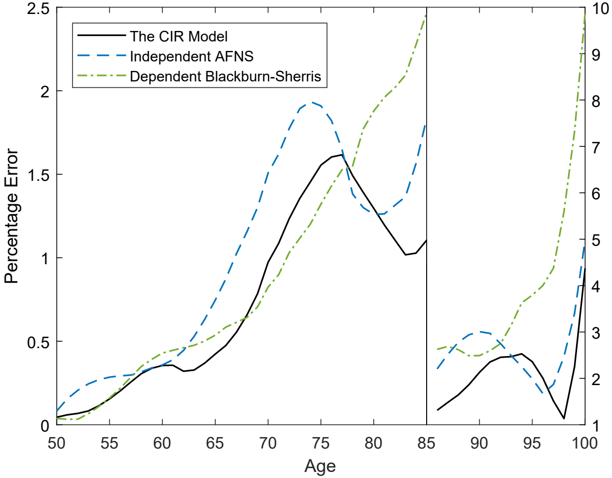


Figure 10: The CIR Model, the Dependent Blackburn-Sherris Model and the Independent AFNS Model

Forecasting Survival Curves

- Optimal forecasts, or best-estimate forecasts, are used to project average forces of mortality and survival probabilities for future cohorts.
- At time t, the one step ahead forecast of the average force of mortality is

$$\bar{\mu}(t+1, T+1) = -\frac{B(t,T)'}{T-t} E[X_{t+1}|X_t] - \frac{A(t,T)}{T-t}, \tag{21}$$

  where $B(t,T)$ and $A(t,T)$ depend on the model.
- The forcasts of survival probabilities are then:

$$S(t+1, T+1) = \exp\Big(B(t,T)' E[X_{t+1}|X_t] + A(t,T)\Big). \tag{22}$$

Forecasting Survival Curves

- The factor dynamics under measure P in the independent Blackburn-Sherris model and the 3-factor independent AFNS model are the same. The conditional expectation of state variables for these two models are as follows:

$$E\left[X_{t+1}^1 | X_t^1\right] = e^{-k_{11}^P} X_t^1, \quad E\left[X_{t+1}^2 | X_t^2\right] = e^{-k_{22}^P} X_t^2,$$
$$E\left[X_{t+1}^3 | X_t^3\right] = e^{-k_{33}^P} X_t^3. \tag{23}$$

- For the independent AFNS model, the conditional mean has the same structure but with $X_t = (L_t, S_t, C_t)$.

- The conditional mean of the CIR mortality model is:

$$E\left[X_{t+1}^1 | X_t^1\right] = e^{-k_{11}^P} X_t^1 + \theta_1^P \left(1 - e^{-k_{11}^P}\right),$$
$$E\left[X_{t+1}^2 | X_t^2\right] = e^{-k_{22}^P} X_t^2 + \theta_2^P \left(1 - e^{-k_{22}^P}\right), \tag{24}$$
$$E\left[X_{t+1}^3 | X_t^3\right] = e^{-k_{33}^P} X_t^3 + \theta_3^P \left(1 - e^{-k_{33}^P}\right).$$

Affine Cohort Mortality Models - Forecast RMSE

Table 2: RMSE by Comparing the Actual and Best-Estimate Survival Probabilities of the 1916 Cohort

	The Blackburn-Sherris Model		The AFNS Model		The CIR Model
	Independent	Dependent	Independent	Dependent	
RMSE	0.03197	0.00726	0.00668	0.00754	0.01835

- AFNS model performs well. CIR model has poorer forecasting performance. Forecast for a single cohort.

Affine Cohort Mortality Models - Forecast RMSE

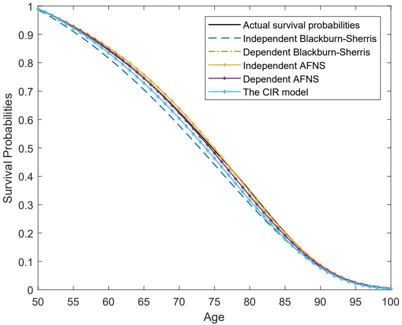


Figure 11: Actual and Best-Estimate Survival Probabilities of the 1916 Cohort

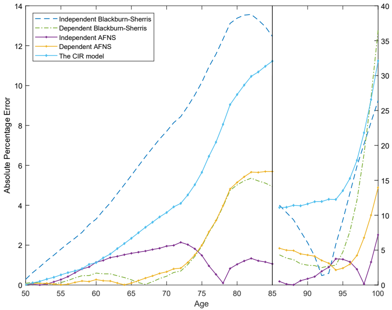


Figure 12: Absolute Percentage Errors between Actual and Best-Estimate Survival Probabilities

Wrap Up

- Introduced continuous-time mortality models including an AFNS cohort mortality model with interpretable latent stochastic factors for level, slope and curvature of the survival curve.
  - The model is based on factor loadings multiplied by (latent) factors, where the factors are equivalent to stochastic parameters and the factor loadings determine how the factors impact different ages.
- Outlined the dynamics of the mortality rates and the affine survival curves.
- Outlined the estimation of the models using the Gaussian Kalman filter.
- Outlined how the models can capture Poisson variation in the estimation. .

Some comments on the Models

- Empirical results show that the independent-factor AFNS cohort mortality model:
  - Is parsimonious, captures the variation in cohort mortality rates in US data, producing a better fit at older ages than the independent-factor Blackburn-Sherris model, and has good predictive performance.
  - Is easy to implement with closed-form expressions for survival probabilities, and as a Gaussian model is easy to estimate using the Kalman filter. Negative mortality rates have very low probability.
  - Has factors that fit historical data dynamics and have intuitive factor interpretation (Level, Slope, Curvature).
  - Multi-factor age-cohort models, and particularly the AFNS model, is well suited for financial and insurance applications - see for example Xu et al. (2019b).
- Work to be done: incorporating imcomplete cohorts inrto estimation, better capturing Poisson variation, age-dependence in trend and covariance, CIR model estimation and forecasting.

Acknowledgements

- Financial support from:
  - Society of Actuaries Center of Actuarial Excellence Research Grant 2017–2020: Longevity Risk: Actuarial and Predictive Models, Retirement Product Innovation, and Risk Management Strategies.
  - CEPAR Australian Research Council Centre of Excellence in Population Ageing Research project number CE170100005 for 2018–2024.

References I

Biffis, E. (2005). Affine processes for dynamic mortality and actuarial valuations. *Insurance: Mathematics and Economics*, 37(3):443–468.

Blackburn, C. and Sherris, M. (2013). Consistent dynamic affine mortality models for longevity risk applications. *Insurance: Mathematics and Economics*, 53:64–73.

Cairns, A. J., Blake, D., and Dowd, K. (2006a). Pricing death: Frameworks for the valuation and securitization of mortality risk. *ASTIN Bulletin*, 36(1):79 – 120.

Cairns, A. J., Blake, D., and Dowd, K. (2006b). A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance*, 73(4):687–718.

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., and Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal*, 13(1):1–35.

Chang, Y. and Sherris, M. (2018). Longevity Risk Management and the Development of a Value-Based Longevity Index. *Risks*, 6(1):1–20.

Christensen, J. H., Diebold, F. X., and Rudebusch, G. D. (2011). The affine arbitrage-free class of Nelson–Siegel term structure models. *Journal of Econometrics*, 164(1):4–20.

Dahl, M. and Møller, T. (2006). Valuation and hedging of life insurance liabilities with systematic mortality risk. *Insurance: mathematics and economics*, 39(2):193–217.

Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *The Journal of Finance*, 57(1):405–443.

Duffie, D. and Kan, R. (1996). A yield-factor model of interest rates. *Mathematical Finance*, 6:379–406.

Huang, Z., Sherris, M., Villegas, A., and Ziveyi, J. (2019). The application of affine processes in cohort mortality risk models. Working paper, CEPAR Working Paper.

References II

Human Mortality Database (2017). U.S.A. Life Tables. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at http://www.mortality.org/cgi-bin/hmd/country.php?cntr=USA&level=1, data downloaded on March 2018 and October 2018.

Jevtic, P., Luciano, E., and Vigna, E. (2013). Mortality surface by means of continuous time cohort models. *Insurance: Mathematics and Economics*, 53:122–133.

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, 87(419):659–671.

Luciano, E., Spreeuw, J., and Vigna, E. (2008). Modelling stochastic mortality for dependent lives. *Insurance: Mathematics and Economics*, 43(2):234–244.

Milevsky, M. A. and Promislow, S. D. (2001). Mortality derivatives and the option to annuitise. *Insurance: Mathematics and Economics*, 29(3):299–318.

Renshaw, A. and Haberman, S. (2006). A cohort-based extension to the LeeCarter model for mortality reduction factors. *Insurance: Mathematics and Economics*, 38(3):556 – 570.

Schrager, D. F. (2006). Affine stochastic mortality. *Insurance: Mathematics and Economics*, 38(1):81–97.

Xu, Y., Sherris, M., and Ziveyi, J. (2019a). Continuous time multi-cohort mortality modelling with affine processes. Working paper, UNSW Business School Research Paper - Forthcoming in Scandinavian Actuarial Journal.

Xu, Y., Sherris, M., and Ziveyi, J. (2019b). Market price of longevity risk for a multi-cohort mortality model with application to longevity bond option pricing. *Journal of Risk and Insurance https://doi.org/10.1111/jori.12273*.

Appendix

- Discretization of Continuous Model Dynamics
- Kalman Filter estimation

Model estimation - Measurement equation

- The measurement equation, based on the average force of mortality, for a given current age to different future survival ages is:

$$\bar{\mu}(t, T) = -\frac{B(t, T)^{'}}{T - t} X_t - \frac{A(t, T)}{T - t} + \varepsilon_t, \tag{25}$$

  where the measurement error ε_t is independently and identically distributed noise and X_t are the latent factors.

- We can write this as

$$y_t = BX_t + A + \varepsilon_t. \tag{26}$$

Model estimation - Measurement equation

- For a 3-factor affine mortality model, the measurement equation with N observed average forces of mortality for ages $x+1$ to $x+N$ is:

$$\begin{pmatrix} \bar{\mu}(t,t+1) \\ \bar{\mu}(t,t+2) \\ \vdots \\ \bar{\mu}(t,t+N) \end{pmatrix} = \begin{pmatrix} -B^1(t,t+1) & -B^2(t,t+1) & -B^3(t,t+1) \\ -\frac{B^1(t,t+2)}{2} & -\frac{B^2(t,t+2)}{2} & -\frac{B^3(t,t+2)}{2} \\ \vdots & \vdots & \vdots \\ -\frac{B^1(t,t+N)}{N} & -\frac{B^2(t,t+N)}{N} & -\frac{B^3(t,t+N)}{N} \end{pmatrix} \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} \tag{27}$$

$$+ \begin{pmatrix} -A(t,t+1) \\ -\frac{A(t,t+2)}{2} \\ \vdots \\ -\frac{A(t,t+N)}{N} \end{pmatrix} + \begin{pmatrix} \varepsilon_t(1) \\ \varepsilon_t(2) \\ \vdots \\ \varepsilon_t(N) \end{pmatrix}, \tag{28}$$

Kalman Filter - State Transition Equation

- The state transition equation is a discretized version of the SDE dynamics and is given by:

$$X_t = \exp\left(-K^P\right) X_{t-1} + \eta_t, \tag{29}$$

  where η_t is the transition error vector.

- The structure of stochastic error terms is assumed to be:

$$\begin{pmatrix} \eta_t \\ \varepsilon_t \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} R & 0 \\ 0 & H \end{pmatrix} \right], \tag{30}$$

  where both the matrix H and matrix R are diagonal, with R being the covariance matrix of the measurement error and H being the covariance matrix of the transition error.

Kalman Filter - Error Assumptions

- The error matrix R for the state transitions, derived from the dynamics of the latent factors, in discrete time form is

$$R = \int_{t-1}^{t} e^{-K^P(t-s)} \Sigma\Sigma' e^{-\left(K^P\right)'(t-s)} ds. \tag{31}$$

- Poisson variation is captured in the diagonal of the covariance matrix H, assumed to have exponential form (reflecting exponential increase in mortality rate) given by

$$H(t, T) = \frac{1}{T-t} \sum_{i=1}^{T-t} \left[r_c + r_1 e^{r_2 i} \right], \tag{32}$$

where the values of r_c, r_1 and r_2 are estimated as part of the optimal parameter set.

Kalman Filter - Forecasting step

- Denote the (average) mortality rates at time t by $Y_t = (y_1, \ldots, y_t)$ and the parameters by ψ.
- In the forecasting step we first update the state, X_{t-1}, and its mean square error, Σ_{t-1},

$$X_{t|t-1} = E\left[X_t | Y_{t-1}\right] = \Phi\left(\psi\right) X_{t-1}, \tag{33}$$

$$\Sigma_{t|t-1} = \Phi\left(\psi\right) \Sigma_{t-1} \Phi\left(\psi\right)^{'} + R\left(\psi\right), \tag{34}$$

where $\Phi = \exp\left(-K^P\right)$ and $R = \int_{t-1}^{t} e^{-K^P(t-s)} \Sigma\Sigma' e^{-\left(K^P\right)^{'}(t-s)} ds$.

Kalman Filter - Forecasting step

- We then use the historical mortality rate information at time t to update the forecasts to obtain:

$$X_t = E\left[X_t | Y_t\right] = X_{t|t-1} + \Sigma_{t|t-1} B\left(\psi\right)' F_t^{-1} \nu_t, \tag{35}$$

$$\Sigma_t = \Sigma_{t|t-1} - \Sigma_{t|t-1} B\left(\psi\right)' F_t^{-1} B\left(\psi\right) \Sigma_{t|t-1}, \tag{36}$$

where

$$\nu_t = y_t - E\left[y_t | Y_{t-1}\right] = y_t - A\left(\psi\right) - B\left(\psi\right) X_{t|t-1}, \tag{37}$$

$$F_t = cov\left(\nu_t\right) = B\left(\psi\right) \Sigma_{t|t-1} B\left(\psi\right)' + H\left(\psi\right). \tag{38}$$

Kalman Filter - Log-likelihood

- The log-likelihood function is then computed as:

$$\log L\left(y_1, \ldots, y_t; \psi\right) = \sum_{t=1}^{T}\left(-\frac{N}{2}\log\left(2\pi\right) - \frac{1}{2}\log\left|F_t\right| - \frac{1}{2}\nu_t^{'}F_t\nu_t\right), \tag{39}$$

  where N is the number of ages with observed average forces of mortality.

- The log-likelihood function is maximized with respect to ψ to obtain the optimal parameter set using an iterative process
- Start with initial values, use Kalman filter to determine likelihood of data, update parameter values and iterate until maximum of likelihood is derived.

Multi-Population Longevity Models: A Spatial Random Field Approach

Nhan Huynh

Department of Statistics & Applied Probability, UC Santa Barbara

Joint with Prof. Michael Ludkovski

2020 Living to 100 Symposium

January 13, 2020

Outline

Table of Contents

Multi-population Mortality Modeling

- Mortality changes over time for populations with similar geographic characteristics are correlated.
  - Countries in a region, states in a country, males/females in a country.

Multi-population Mortality Modeling

- Mortality changes over time for populations with similar geographic characteristics are correlated.
  - Countries in a region, states in a country, males/females in a country.

Aims:

- Examine the commonality in global longevity.
- Ensure the coherence in long-term projections across populations.
- Limited data $\rightarrow$ borrowing data from other populations.
- Enhance actuarial credibility from aggregation of multiple datasets.

Multi-population Mortality Modeling

- Mortality changes over time for populations with similar geographic characteristics are correlated.
  - Countries in a region, states in a country, males/females in a country.

Aims:

- Examine the commonality in global longevity.
- Ensure the coherence in long-term projections across populations.
- Limited data → borrowing data from other populations.
- Enhance actuarial credibility from aggregation of multiple datasets.

Challenges:

- Model complexity and computational challenges.

Multi-population Mortality Modeling

- Mortality changes over time for populations with similar geographic characteristics are correlated.
  - Countries in a region, states in a country, males/females in a country.

Aims:

- Examine the commonality in global longevity.
- Ensure the coherence in long-term projections across populations.
- Limited data $\rightarrow$ borrowing data from other populations.
- Enhance actuarial credibility from aggregation of multiple datasets.

Challenges:

- Model complexity and computational challenges.
- Data availability.
  - Used aggregated mortality datasets from *Human Mortality Database*.
  - Ages 50–84 and Years 1990–2016 from 10 European countries.

Lee & Li Method for Two-population Model

- Treats Age and Year as factors.
- Aggregated data $\rightarrow$ estimate global age and year factors $+$ country-specific age and year trend.
- Employs time-series method (e.g.: AR(1)) for mortality projection.

Lee & Li Method for Two-population Model

- Treats Age and Year as factors.
- Aggregated data $\rightarrow$ estimate global age and year factors $+$ country-specific age and year trend.
- Employs time-series method (e.g.: AR(1)) for mortality projection.
  - (+) Simple and the parameters are easily interpretable.
  - (+) Stochastic forecasts with probabilistic prediction intervals.
  - (−) Smoothing via point estimators, no credible bands.
  - (−) Number of populations is limited to 2.
  - (−) Hundreds of parameters needed to be estimated.

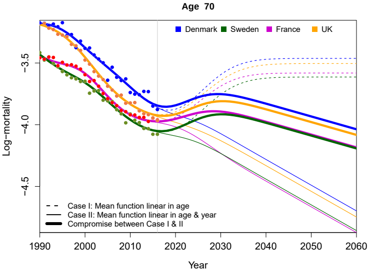
Main Contributions

- Employ spatial statistical framework of **Gaussian Process** (GP) regression as a machine learning method for multi-population modeling.

Main Contributions

Employ spatial statistical framework of **Gaussian Process** (GP) regression as a machine learning method for multi-population modeling.

Mortality forecast for Age 70 via Joint GP

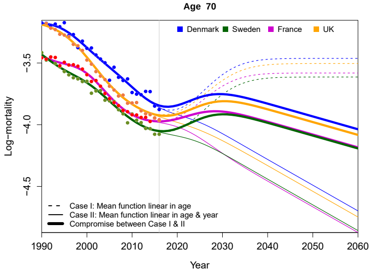


- Non-parametric → smoothed mortality surfaces over Age & Year dimensions.
- Bayesian approach → quantify predictive uncertainty and generate stochastic trajectories for predictions.

Main Contributions

Employ spatial statistical framework of **Gaussian Process** (GP) regression as a machine learning method for multi-population modeling.

Mortality forecast for Age 70 via Joint GP



- Non-parametric → smoothed mortality surfaces over Age & Year dimensions.
- Bayesian approach → quantify predictive uncertainty and generate stochastic trajectories for predictions.

- Captures the cross-population dependence for mortality from multiple populations.
- Number of parameters estimated is substantially smaller.
- Can handle more than 2 populations.

Table of Contents

Multi-population GP Regression

- Assume the output-input relationship:

$$y^n = f(x^n) + \epsilon^n$$

  - Input $= x^n =$ (Age, Year, Indicators for population) & Output $= y^n =$ log-mortality.
  - $f(.)$: true log-mortality surface as a random variable.
  - ϵ^n: errors from i.i.d. Gaussian with zero mean and constant variance.

Multi-population GP Regression

- Assume the output-input relationship:

$$y^n = f(x^n) + \epsilon^n$$

  - Input $= x^n =$ (Age, Year, Indicators for population) & Output $= y^n =$ log-mortality.
  - $f(.)$: true log-mortality surface as a random variable.
  - ϵ^n: errors from i.i.d. Gaussian with zero mean and constant variance.

- Specify prior distribution, then compute the posterior distribution after collecting data: $p(f|\mathcal{D}) \propto \; p(f).p(\mathbf{y}|\mathbf{x}, f) = \{prior\}.\{likelihood\}$.
  - Gaussian prior + Guassian likelihood → Gaussian posterior.

- A Gaussian process defines a prior over functions: any finite sample is a realization of a **Multivariate normal distribution**.
  - All properties specified via the mean and covariance function.

Covariance Kernel

- Covariance function describes the influences between data points.

  "Everything is related to everything else, but near things are more related than distant things" – Tobler's First Law of Geography.

Covariance Kernel

- Covariance function describes the influences between data points.

  "Everything is related to everything else, but near things are more related than distant things" – Tobler's First Law of Geography.

- Consider a squared-exponential kernel:

$$C(x^i, x^j) = \eta^2 \exp\left[-\frac{(x^i_{ag} - x^j_{ag})^2}{2\theta^2_{ag}} - \frac{(x^i_{yr} - x^j_{yr})^2}{2\theta^2_{yr}} \right] \prod_{\{l_1, l_2\}} \exp\left[-\theta_{l_1,l_2} \delta^{ij}_{l_1,l_2} \right]$$

where $\delta^{ij}_{l_1,l_2}$: an indicator whether ith & jth obs. from populations l_1 and l_2.

- Cross-population correlation is an exponential function of θ_{l_1,l_2}. Large value of $\theta_{l_1,l_2} \rightarrow$ low correlation between two populations.
- η^2: process variance, controls amplitude of f;
- θ_{ag} & θ_{yr}: characteristic lengthscales, determine the spatial smoothness in Age and Year dimensions.
- Kernel hyperparameters Θ are $\eta, \theta_{ag}, \theta_{yr}, \theta_{l_1,l_2}$, etc.

Fitting GP

- Fitting $\equiv$ learning the hyperparameters.
- **MLE**: optimization of the marginal likelihood function.
  - Used packages `DiceKriging` and `kergp` in R.

Fitting GP

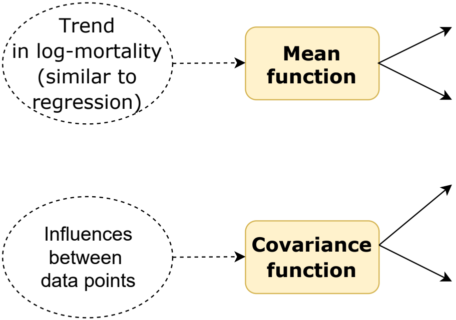
- Fitting $\equiv$ learning the hyperparameters.
- **MLE**: optimization of the marginal likelihood function.
  - Used packages `DiceKriging` and `kergp` in R.
- **Hierarchical approach**: specification of the priors for the hyperparameters (a fully Bayesian approach).
  - Quantify model risk - range of GP models consistent with the data via Bayesian framework.
  - Computationally more intensive but quantifies how well the spatial structure is being learned.
  - `Stan`: free, open-source software for Bayesian statistical inferences.

Table of Contents

Joint model for European Nations

- Incorporate more than 2 populations in the model.
  - ✎ Joint model for Male mortality from 4 European countries: {Denmark, France, Sweden, UK} ≡ {DNK, FRA, SWE, UK}.

    Fitted on Ages 70–84 and calendar Years 1990–2012:



10% increases in log-mortality for each additional unit in Age.

Sweden has lowest trend while Denmark has highest trend. Mean difference is 3%.

Common hyper-parameters that determine spatial smoothness in Age and Year

Cross-population correlation

	DNK	FRA	SWE
FRA	0.28		
SWE	0.23	0.60	
UK	0.91	0.61	0.30

Multi-population Models via Bayesian GP

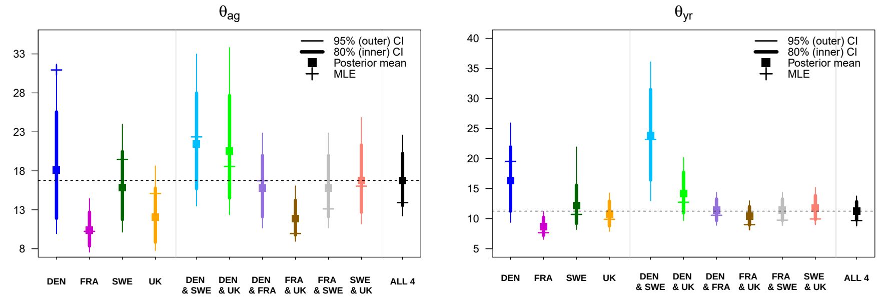
- Utilize aggregated data from multiple populations $\rightarrow$ provides tighter hyper-parameter posteriors (reduce the model risk).

Multi-population Models via Bayesian GP

- Utilize aggregated data from multiple populations $\rightarrow$ provides tighter hyper-parameter posteriors (reduce the model risk).
- More populations added into the model, the closer at discovering 'common' representation of mortality pattern.

Multi-population Models via Bayesian GP

- Utilize aggregated data from multiple populations → provides tighter hyper-parameter posteriors (reduce the model risk).
- More populations added into the model, the closer at discovering 'common' representation of mortality pattern.



Length-scale in Age dimension Length-scale in Year dimension

Prediction Performance in Multi-population Models

- Out-of-sample predictions via multi-pop. model are more accurate.
    - Symmetric mean absolute percentage error (SMAPE) to assess model performance.

Prediction Performance in Multi-population Models

- Out-of-sample predictions via multi-pop. model are more accurate.
  - Symmetric mean absolute percentage error (SMAPE) to assess model performance.
- Multi-population model provides tighter prediction intervals (more certain about the future).
  - Posterior marginal variance is smaller in multi-population models.

Prediction Performance in Multi-population Models

- Out-of-sample predictions via multi-pop. model are more accurate.
  - Symmetric mean absolute percentage error (SMAPE) to assess model performance.
- Multi-population model provides tighter prediction intervals (more certain about the future).
  - Posterior marginal variance is smaller in multi-population models.

Multi-population model for two Nordic countries: Denmark & Sweden (Male populations, Ages 70–84 and Years 1990-2012).

Prediction accuracy for Ages 70–84

SMAPE	2013 (1-yr out)		2016 (4-yr out)	
	Single-pop	Multi-pop	Single-pop	Multi-pop
Denmark	1.58	**1.52**	1.26	**1.22**
Sweden	1.05	**0.82**	2.53	**0.83**

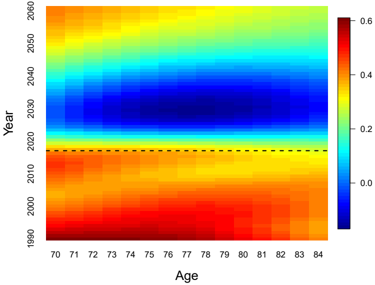
Posterior S.D. for Sweden

$s_{SWE}(.)$	2013 (1-yr out)		2016 (4-yr out)	
	Single-pop	Multi-pop	Single-pop	Multi-pop
Age 75	0.0300	**0.0292**	0.0399	**0.0317**
Age 85	0.0330	**0.0311**	0.0461	**0.0351**

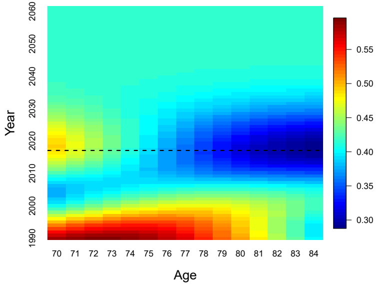
Limitation of Single-population Models

- Often generate divergent long-term forecasts that are inconsistent with historical patterns.
  - Implausible difference in mean forecasts, or excessively fast changes in relative mortality.

Log-mortality difference in mean forecast between Danish Males and Females



(a) Single-population models

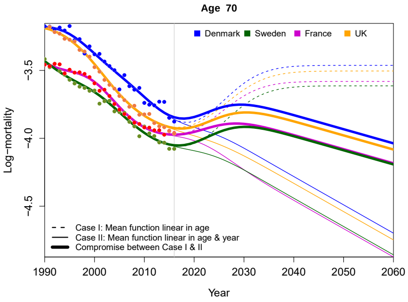


(b) Two-population model

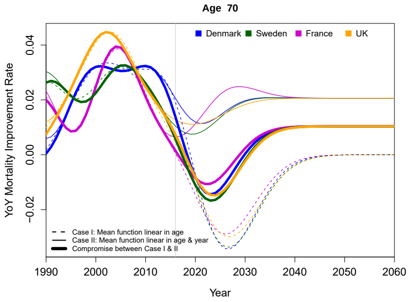
Coherence Forecasts

- Forecasts via multi-pop. models maintain historical characteristics observed in the data into the future.
- Use the mean function to generate long-term forecasts and enforce desired coherence.
  - Mortality across populations move in unison → strong coherence.

Long-term mortality forecasting in 1990–2060



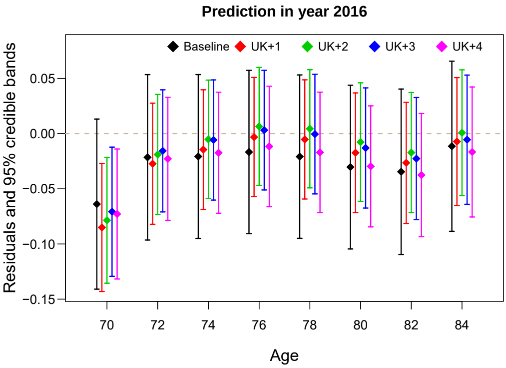
(a) Log-mortality



(b) Improvement factors

Borrowing latest data

- Opportunities to borrow latest information from other populations to improve prediction about the recent domestic mortality.
  - In HMD, data from different countries arrives non-synchronously.
  - 2016 (one-year out) forecast in UK:
    Baseline ≡ single model for UK up to 2015.
    Multi-pop'n models with UK up to 2015 & other populations up to 2016.



Borrowing latest data (Cont.)

- Improvement in recent mortality prediction via multi-population models depend on the cross-population correlation.
  - SMAPE: difference between forecast and observed log-mortality.
  - Continuous Ranked Probability Score (CRPS): difference between forecast and empirical cumulative distribution function of the observation.

Correlation vs. prediction performance in one-year out cross validation

2-pop'n model with UK	Correlation	Improvement in SMAPE	Improvement in CRPS
Switzerland	0.89	−22.0%	−1.7%
Czech	0.84	−12.1%	−3.2%
Sweden	0.54	−8.0%	−1.6%
Hungary	0.34	−1.7%	1.6%
Latvia	0.03	44.4%	6.9%

Table of Contents

Next steps

- Cluster methods to identify groups of populations with similar characteristics in multi-population models.
- Implementation on 50-state US Mortality Database.
- Modeling cause-of-death mortality.
- Investigate other kernel families to explore spatial covariance structure.
- Computational speed-up to handle larger data sets.


THANK
YOU

References

Williams, C. K. and Rasmussen, C. E. 2006.
Gaussian processes for machine learning, the MIT Press.

Roustant, O., Ginsbourger, D., Deville, Y., et al. 2012.
Dicekriging, Diceoptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization.
Journal of Statistical Software, 51(1):1–55.

M. Ludkovski, J. Risk, H. Zail. 2018.
Gaussian Process Models for Mortality Rates and Improvement Factors.
ASTIN Bulletin 48(3), 1307–1347.
github.com/jimmyrisk/GPmortalityNotebook

N. Huynh, M. Ludkovski, H. Zail
Multipopulation Longevity Analysis with Gaussian Process Models.
2020 Living to 100 Symposium, Accepted