Advances in Stochastic Mortality Modelling
Robust Probabilistic Feature Extraction

Prof. Gareth W. Peters (FIOR, YAS-RSE)
Chair of Statistics for Risk and Insurance,
Department of Actuarial Mathematics and Statistics,
Heriot-Watt, Edinburgh, UK

“Stochastic Period and Cohort Effect State-Space Mortality Models
Incorporating Demographic Factors via Probabilistic Robust Principle
Components”
Available at SSRN: http://dx.doi.org/10.2139/ssrn.2977306

July 15, 2018
1. Mortality Modelling Context

2. State Space Stochastic Mortality Models

3. State-Space Hybrid Factor Models: Regression Extensions

4. Robust Probabilistic Feature Extraction Methods

5. Application to Mortality Modelling and Demographic Data

6. Appendix
1. Mortality Modelling Context
2. State Space Stochastic Mortality Models
3. State-Space Hybrid Factor Models: Regression Extensions
4. Robust Probabilistic Feature Extraction Methods
5. Application to Mortality Modelling and Demographic Data
6. Appendix
Mortality Modelling Context

- Ageing populations are a major challenge for many countries.
  - Fertility rates are declining while life expectancy is increasing.

- **Longevity risk**: the adverse financial outcome of people living longer than expected ⇒ possibility of outliving their retirement savings.

- Long term demographic risk has significant implications for societies and manifests as a systematic risk for pension plans and annuity providers.

- Policymakers rely on **mortality projection** to determine appropriate pension benefits and regulations regarding retirement.
World population by level of fertility over time (1950-2010)

On the x-axis you find the cumulative share of the world population. The countries are ordered along the x-axis descending by the total fertility rate of the country.

The interactive data visualization is available at OurWorldinData.org. There you find the raw data and more visualizations on this topic.

Licensed under CC-BY-SA by the author Max Roser.
Life Expectancy by Age in England and Wales, 1700-2013

Shown is the total life expectancy given that a person reached a certain age.

Data source: Life expectancy at birth. Data on life expectancy at age 1 and older from the Human Mortality Database (www.mortality.org).

The interactive data visualization is available at OurWorldInData.org. There you find the raw data and more visualizations on this topic. Licensed under CC-BY-SA by the author Max Roser.
Enhancing mortality models requires an understanding of common features of mortality behaviour [Cairns, Blake and Dowd, 2008]

- Mortality rates have fallen dramatically at all ages.
- Rate of decrease in mortality has varied over time and by age group.
- Absolute decreases have varied by age group.
- Aggregate mortality rates have significant volatility year on year.
Table of contents

1  Mortality Modelling Context

2  State Space Stochastic Mortality Models

3  State-Space Hybrid Factor Models: Regression Extensions

4  Robust Probabilistic Feature Extraction Methods

5  Application to Mortality Modelling and Demographic Data

6  Appendix
The uncertainty in future death rates can be divided into two components:

- **Unsystematic mortality risk.** Even if the true mortality rate is known, the number of deaths, will be random.
  - larger population ⇒ smaller unsystematic mortality risk (due to pooling of offsetting risks - diversification).

- **Systematic mortality risk.** This is the undiversifiable component of mortality risk that affects all individuals in the same way.
  - Forecasts of mortality rates in future years are uncertain.
[Lee and Carter, 92] proposed a stochastic mortality model (LC) to forecast the trend of age-specific mortality rates.

Several extensions to Lee-Carter model have been proposed, overview in [Fung et al. 2017].
• [Lee and Carter, 92] proposed a stochastic mortality model (LC) to forecast the trend of age-specific mortality rates.
• Several extensions to Lee-Carter model have been proposed, overview in [Fung et al. 2017].

*Survival probability is still consistently underestimated* especially in the last few decades ([IMF, 2012]).
[Lee and Carter, 92] proposed a stochastic mortality model (LC) to forecast the trend of age-specific mortality rates.

Several extensions to Lee-Carter model have been proposed, overview in [Fung et al. 2017].

*Survival probability is still consistently underestimated* especially in the last few decades ([IMF, 2012]).

This talk considers models aiming to help resolve this issue via

- **Stochastic State-Space Mortality Models** with **Period** and **Cohort** stochastic latent effects (LCC).
- + Extensions to State-Space Hybrid Regression Structures!

(see [Fung et al. 2017] and [Fung et al. 2018])
A state space model has two model components:

- a stochastic observation equation; and
- a stochastic latent Markov state process.
A state space model has two model components:

- a stochastic observation equation; and
- a stochastic latent Markov state process.

Key advantages of state space modelling approach:

- remove awkward identification specifications;
- computational efficiency and numerical robustness;
- accurate in-sample and out-of-sample forecasts;
- optimal statistical efficiency and unbiased estimation;
Period-Cohort effect state-space formulation

Observation equation: $y_{x,t} = \ln \hat{m}_{x,t}$, follow:

$$\ln \hat{m}_{x,t} = \alpha_x + \beta_x \kappa_t + \beta_x^\gamma \gamma_{t-x} + \epsilon_{x,t},$$

where $\epsilon_{x,t}$ is a regression noise term.

- $\alpha = \alpha_{x_1:x_p} := [\alpha_{x_1}, \ldots, \alpha_{x_p}]$ represents the age-profile of the log death rates
- $\beta = \beta_{x_1:x_p}$ measures the sensitivity of death rates for different age group to a change of period effect $\kappa_t$.
- $\beta^\gamma = \beta_{x_1:x_p}^\gamma$ measures the sensitivity of death rates for different age group to a change of cohort effect $\gamma_{t-x}$. 
Observation Process: in matrix form.

\[
\begin{pmatrix}
    y_{x_1, t} \\
    y_{x_2, t} \\
    \vdots \\
    y_{x_p, t}
\end{pmatrix}
= \begin{pmatrix}
    \alpha_{x_1} \\
    \alpha_{x_2} \\
    \vdots \\
    \alpha_{x_p}
\end{pmatrix}
+ \begin{pmatrix}
    \beta_{x_1} & \beta_{x_1}^\gamma & 0 & \cdots & 0 \\
    \beta_{x_2} & 0 & \beta_{x_2}^\gamma & \cdots & 0 \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    \beta_{x_p} & 0 & 0 & \cdots & \beta_{x_p}^\gamma
\end{pmatrix}
\begin{pmatrix}
    \kappa_t \\
    \gamma_{t, x_1} \\
    \gamma_{t, x_2} \\
    \vdots \\
    \gamma_{t, x_p}
\end{pmatrix}
+ \begin{pmatrix}
    \varepsilon_{x_1, t} \\
    \varepsilon_{x_2, t} \\
    \vdots \\
    \varepsilon_{x_p, t}
\end{pmatrix}.
\]

Here \((\kappa_t, \gamma_{t, x_1}, \ldots, \gamma_{t, x_p})^T\) is the \(p + 1\) dimensional latent state vector. \(\gamma_{t}^{x} := \gamma_{t-x}\)
Stochastic Mortality Modelling

State Equation in matrix form:

\[
\begin{pmatrix}
\kappa_t \\
\gamma_{t1} \\
\gamma_{t2} \\
\vdots \\
\gamma_{tp-1} \\
\gamma_{tp}
\end{pmatrix} =
\begin{pmatrix}
1 & 0 & 0 & \cdots & 0 & 0 \\
0 & \lambda_1 & \lambda_2 & \cdots & \lambda_{p-1} & \lambda_p \\
0 & 1 & 0 & \cdots & 0 & 0 \\
0 & 0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 1 & 0
\end{pmatrix} \begin{pmatrix}
\kappa_{t-1} \\
\gamma_{t-1} \\
\gamma_{t-1} \\
\vdots \\
\gamma_{t-1} \\
\gamma_{t-1}
\end{pmatrix} +
\begin{pmatrix}
\theta \\
\eta \\
\vdots \\
0 \\
0 \\
0
\end{pmatrix} +
\begin{pmatrix}
\omega_{t}^\kappa \\
\omega_{t}^\gamma \\
\vdots \\
0 \\
0 \\
0
\end{pmatrix}.
\]

Period effect \( \kappa_t \) is a random walk with drift process with \( \omega_t^\kappa \overset{iid}{\sim} \text{N}(0, \sigma_{\omega}^2) \) and Cohort effect \( \gamma_{t1} \) is a stationary AR(p) process with \( \omega_t^\gamma \overset{iid}{\sim} \text{N}(0, \sigma_{\gamma}^2) \).
Table of contents

1 Mortality Modelling Context

2 State Space Stochastic Mortality Models

3 State-Space Hybrid Factor Models: Regression Extensions

4 Robust Probabilistic Feature Extraction Methods

5 Application to Mortality Modelling and Demographic Data

6 Appendix
State-Space Hybrid Factor Models

**GOAL:** develop *stochastic mortality state-space hybrid factor models*.

- **Hybrid** := Stochastic Latent Factors \(+\) Observable Covariate Features

- **observable features** extracted from demographic data

- Feature extraction should aim for dimension reduction \(\Rightarrow\) model parsimony.

[Toczydlowska and Peters, 2017] address important aspects of feature extraction:

1. **missing data** in time-series and panel (matrix) structured real demographic data;
2. **noisy observations and outliers** (in real data);
State-Space Hybrid Factor Models

Two fundamental approaches to develop Hybrid Factor Models:

1. **time varying factor with static loading coefficient**
   
   (classical distributed lag regressions such as ARDL models);

2. **static factor with time varying stochastic loading coefficients.**
   
   (state space models e.g. dynamic Nelson-Siegel yield curves).

- Option 2: suitable for **high dimensional** data, **time series / panel structured** but represented by relatively “short time series” lengths.

  - ⇒ particularly prevalent in demographic studies!
Consider the State-Space Hybrid Period-Cohort-Demographic Model

\[ y_t = \alpha + \tilde{B}_t \tilde{\varphi}_t + \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2 I_p), \]

\[ \tilde{\varphi}_t = \tilde{\Lambda} \tilde{\varphi}_{t-1} + \tilde{\Theta} + \tilde{\omega}_t, \quad \tilde{\omega}_t \overset{iid}{\sim} \mathcal{N}(0, \tilde{\Upsilon}) \]

where \( \tilde{\varphi}_t = (\varphi_t, \varrho_t) \) is a \((p + pk + 1) \times 1\) latent process vector of \( \varphi_t \) stochastic mortality factors (period-cohort) and \( \varrho_t \) dynamic factor loadings, with

\[ \tilde{\Theta} = \begin{pmatrix} \Theta_{(p+1) \times 1} \\ \Psi_{pk \times 1} \end{pmatrix} \]

\((p+pk+1) \times 1\)

a vector of drift parameters for state equations.
State-Space Hybrid Factor Models

Consider three models:

**Case 1:** Factors in Observation Equation Only;

**Case 2:** Factors in Period Effect State Equation Only;

**Case 3:** Factors in Cohort Effect State Equation Only.

\[
\tilde{B}_{t \times (p+pk+1)} = \begin{cases} 
( \mathbf{B}_{p \times (p+1)} | \tilde{F}_t ) & \text{for Case 1,} \\
( \mathbf{B}_{p \times (p+1)} | 0_{p \times pk} ) & \text{otherwise},
\end{cases}
\]

\[
\tilde{\Lambda}_{(p+pk+1) \times (p+pk+1)} = \begin{cases} 
\begin{pmatrix}
\Lambda_{(p+1) \times (p+1)} & 0_{(p+1) \times pk} \\
0_{pk \times (p+1)} & \Omega_{pk \times pk}
\end{pmatrix} & \text{for Case 1,} \\
\begin{pmatrix}
\Lambda_{(p+1) \times (p+1)} & \tilde{f}_t^T \\
0_{pk \times (p+1)} & \Omega_{pk \times pk}
\end{pmatrix} & \text{for Case 2,} \\
\begin{pmatrix}
\Lambda_{(p+1) \times (p+1)} & 0_{1 \times pk} \\
0_{pk \times (p+1)} & \Omega_{pk \times pk}
\end{pmatrix} & \text{for Case 3.}
\end{cases}
\]
Table of contents

1 Mortality Modelling Context

2 State Space Stochastic Mortality Models

3 State-Space Hybrid Factor Models: Regression Extensions

4 Robust Probabilistic Feature Extraction Methods

5 Application to Mortality Modelling and Demographic Data

6 Appendix
• Data $Y_t$ is observed (or partially observed) over periods $t \in \{1, \ldots, n\}$ and will be reduced to factors $\tilde{F}_t$

Example: $d$ countries demographic data and $p$ denotes the number of different demographic attributes observed
⇒ then $p \times d$ matrix of data in year $t$ is $Y_t$.

• We do not wish to utilise the raw demographic data $\tilde{F}_t \neq Y_t$:  
  \textit{in general it will produce a model with too many parameters}

• [Toczydlowska and Peters, 2017] considered stochastic projection methods of dimensionality reduction
⇒ \textbf{Probabilistic Principal Component Analysis (PPCA)} and \textbf{Robust} extensions.
Probabilistic Feature Extraction

**PCA by means of Factor Analysis:** with $n$ realisations of the $(p \times d')$-dimensional observed demographic data, vectorized into columns $Y$.

Consider linear decompositions:

$$Y_{n \times pd} = X_{n \times pd} W_{pd \times pd}^T + \epsilon_{n \times pd}.$$ 

*Factor analysis assumes diagonal covariance for $\epsilon_t$.***

**Stochastic Factor PCA:** differs from deterministic PCA as components $x_t$ and factor loading matrix $W$ account for correlation between elements of $y_t$ and only part of the variation:

$$\mathbb{E} y_t^T y_t = \mathbb{E} \left[ (x_t W^T + \epsilon_t)^T (x_t W^T + \epsilon_t) \right] = W \Lambda W^T + \Psi.$$ 

In standard PCA $x_t$ and $W$ account for the entire covariance.
Show $x_t$ and $W$ account for correlation!

**Example:** assume $x_t \sim \mathcal{N}(0, I_d)$ and $\epsilon_t \sim \mathcal{N}(0, \Psi)$ to obtain,

$$y_t | x_t, W, \Psi \sim \mathcal{N}(x_tW^T, \Psi),$$

$$\pi(y_t | W, \Psi) = \int_{\mathbb{R}^d} \pi(y_t, x_t | W, \Psi)dx_t = (2\pi)^{-\frac{d}{2}} |C|^{-1} \exp \left\{ -\frac{1}{2} y_tC^{-1}y_t^T \right\}$$

for $C = WW^T + \Psi$ where $|C|$ denotes the determinant of the matrix.

- Notice that since $\Psi$ is diagonal, the correlation structure between components $y_t$ is specified by the matrix $W$. 
Show $x_t$ and $W$ account for correlation cont.

Eigen decomposition of covariance $C = U_{d \times d}L_{d \times d}U^T$, for diagonal $L$ and orthonormal $U$, gives

$$0 = (C - L)U = \left(W^T W + \sigma^2 I_d - L\right)U = \left(WW^T - (L - \sigma^2 I_d)\right)U.$$

- Thus, the matrix $\Lambda = (L - \sigma^2 I_d)$ and $U$ are matrices of eigenvalues and corresponding eigenvectors of $WW^T$.
- Since $\lambda_i = l_i - \sigma^2 \geq 0$, the scalar $\sigma^2$ can be chosen as the smallest diagonal element of $\Lambda$.
- **Factor loadings are given by** $U\Lambda^{\frac{1}{2}}$.

Assuming the error term $\epsilon_t$ is homogeneous s.t. $\Psi = \sigma^2 I_d$, then estimating $W$ via PCA given $C = WW^T + \sigma^2 I_d$ is identifiable.
Probabilistic Feature Extraction

Feature Extraction via EM Algorithm Estimation!

Goal is to estimate:

- projection matrix $\mathbf{W}$,
- vector $\mathbf{\mu}$ and
- scalar $\sigma^2$

given marginal distribution of $\mathbf{Y}_t$

$$\mathbf{Y}_t | \Psi \sim \mathcal{N} \left( \mathbf{\mu}, \mathbf{WW}^T + \sigma^2 \mathbf{I}_d \right)$$

for the vector of static parameters $\Psi = [\mathbf{W}, \mathbf{\mu}, \sigma^2]$ of the model.

The EM algorithm uses logarithm of the complete data likelihood, i.e. the joint distribution of $\mathbf{Y}_{1:N}, \mathbf{X}_{1:N} | \Psi$ given by

$$\pi_{\mathbf{Y}_{1:N}, \mathbf{X}_{1:N} | \Psi} (\mathbf{y}_{1:N}, \mathbf{x}_{1:N}) = \prod_{t=1}^{N} \pi_{\mathbf{y}_t | \mathbf{x}_t, \Psi} (\mathbf{y}_t) \pi_{\mathbf{x}_t | \Psi} (\mathbf{x}_t)$$

$$= (2\pi)^{-N\frac{d+k}{2}} (\sigma^2)^{-N\frac{d}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^{N} (\mathbf{y}_t - \mathbf{\mu} - \mathbf{x}_t \mathbf{W}^T) (\mathbf{y}_t - \mathbf{\mu} - \mathbf{x}_t \mathbf{W}^T)^T - \frac{1}{2} \sum_{t=1}^{N} \mathbf{x}_t \mathbf{x}_t^T \right\}.$$
Feature Extraction via EM Algorithm Estimation!

1. **Expectation step**: Expectation of the loglikelihood function of the join distribution of $\mathbf{Y}_{1:N}, \mathbf{X}_{1:N} | \psi$ for a fixed vector of static parameters $\psi^*$ with respect to the conditional distribution $\mathbf{X}_{1:N} | \mathbf{Y}_{1:N}, \psi$

   $$Q(\psi, \psi^*) = \mathbb{E}_{\mathbf{X}_{1:N} | \mathbf{Y}_{1:N}, \psi} \left[ \log \pi_{\mathbf{Y}_{1:N}, \mathbf{X}_{1:N} | \psi^*}(\mathbf{y}_{1:N}, \mathbf{x}_{1:N}) \right]$$

2. **Maximisation step**: Finding $W^*, \mu^*$ and $\sigma^2$ that maximize $Q(\psi | \psi^*)$

   $$\left( W^*, \mu^*, \sigma^* \right) = \text{argmax}_{W^* \in \mathbb{R}^{d \times k}, \mu^* \in \mathbb{R}^d, \sigma^2 > 0} Q(\psi, \psi^*)$$
Theorem

The E-step of the EM algorithm for Gaussian Probabilistic Principal Component Analysis given \( N \) realisations of the observation vector \( \mathbf{Y}_t \) denoted by \( \mathbf{y}_{1:N} = \{ \mathbf{y}_1, \ldots, \mathbf{y}_N \} \) is obtained in the closed form as follows

\[
Q(\Psi, \Psi^*) = \mathbb{E}_{\mathbf{x}_{1:N}|\mathbf{y}_{1:N}, \Psi} \left[ \log \pi_{\mathbf{y}_{1:N}, \mathbf{x}_{1:N}|\Psi^*}(\mathbf{y}_{1:N}, \mathbf{x}_{1:N}) \right] \\
= -\frac{N(d + k)}{2} \log 2\pi - \frac{Nd}{2} \log \sigma^* - \frac{1}{2} \sum_{t=1}^{N} \left\{ \frac{1}{\sigma^*} \text{Tr}\{\mathbf{y}_t^T \mathbf{y}_t\} \right\} \\
- \frac{2}{\sigma^*} \mathbf{y}_t \mu^*^T + \frac{1}{\sigma^*} \mu^* \mu^*^T - \frac{2}{\sigma^*} \text{Tr}\left\{ \mathbf{W}^* \mathbb{E}_{\mathbf{x}_t|\mathbf{y}_t, \Psi} \left[ \mathbf{X}_t^T \right] \mathbf{y}_t \right\} \\
+ \frac{2}{\sigma^*} \mathbb{E}_{\mathbf{x}_t|\mathbf{y}_t, \Psi} \left[ \mathbf{X}_t \right] \mathbf{W}^* \mu^*^T + \text{Tr}\left\{ \left( \frac{1}{\sigma^*} \mathbf{W}^* \mathbf{W}^* + \mathbb{I}_k \right) \mathbb{E}_{\mathbf{x}_t|\mathbf{y}_t, \Psi} \left[ \mathbf{X}_t^T \mathbf{X}_t \right] \right\}
\]

see details of expectations and proof in [Toczydlowska and Peters, 2017].
Theorem

The maximizers of the function $Q(\psi, \psi^*)$ are given by

$$
\mu^* = \bar{\mu}(y_{1:N}; \psi) \left( \mathbb{I}_d - WM^{-1}W^*T \right) + \mu WM^{-1}W^*T
$$

$$
W^* = \bar{C}_{\mu, \mu^*}(y_{1:N}; \psi, \psi^*)WM^{-1} \left( \sigma^2M^{-1} + M^{-1}W^T\bar{C}_\mu(y_{1:N}; \psi)WM^{-1} \right)^{-1}
$$

$$
\sigma^*^2 = \frac{1}{d} Tr \left\{ \bar{C}_{\mu^*}(y_{1:N}; \psi, \psi^*) - 2W^*M^{-1}W^T\bar{C}_{\mu, \mu^*}(y_{1:N}; \psi, \psi^*) 
+ W^* \left( \sigma^2M^{-1} + M^{-1}W^T\bar{C}_\mu(y_{1:N}; \psi)WM^{-1} \right) W^*T \right\}
$$

see details of components and proof in [Toczydlowska and Peters, 2017].
Probabilistic Feature Extraction

Probabilistic PCA with Missing Data:
Until now, we assumed the data did not contain any missing observations!

- Real demographic time series data have numerous types of missingness.
- \( \Rightarrow \) missingness is an important aspect to address in the feature extraction!


- Distributional Extensions: Student-t, Skewed and Grouped Student-t cases.
Define the indicator random variable $R_t$ which decides which entries of $Y_t$ are missing denoting them by 1, otherwise 0.

- Each observation consists of the pair $[Y^o_t, R_t]$ with distribution parameterized according to parameters $[\Psi, \Theta]$ respectively.

Likelihood is given by conditional probability $Y^o_t, R_t | \Psi, \Theta$:

$$
\pi_{Y^o_t, R_t | \Psi, \Theta} (y^o_t, r_t) = \int \pi_{Y^o_t, Y^m_t, R_t | \Psi, \Theta} (y^o_t, y^m_t, r_t) \, dy^m_t \\
= \int \pi_{R_t | Y_t, \Psi, \Theta} (r_t) \pi_{Y_t | \Psi, \Theta} (y_t) \, dy^m_t
$$

We assume for simplicity a pattern of missing data according to MAR - missing at random

- The assumptions imposes the indicator variable $R_t$ to be independent of of the value of missing data.
Then the vector $Y_t$ which is MAR satisfies

$$\pi_{R_t|Y_t,\psi}(r_t) = \pi_{R_t|Y^o_t,\psi}(r_t)$$

resulting in

$$\pi_{Y^o_t,R_t|\psi,\Theta}(y^o_t) = \pi_{R_t|Y^o_t,\Theta}(r_t) \int \pi_{Y_t|\psi}(y_t) \, dy^m_t = \pi_{R_t|Y^o_t,\Theta}(r_t) \pi_{Y^o_t|\psi}(y^o_t).$$

$\Rightarrow$ Under the MAR assumption, the estimation of $\psi$ via maximum likelihood of the joint distribution $Y^o_t, R_t|\psi, \Theta$ is equivalent to the maximisation of the likelihood of the marginal distribution $Y^o_t|\psi$. 
# Table of contents

1. Mortality Modelling Context
2. State Space Stochastic Mortality Models
3. State-Space Hybrid Factor Models: Regression Extensions
4. Robust Probabilistic Feature Extraction Methods
5. Application to Mortality Modelling and Demographic Data
6. Appendix
Demographic data that we extract “Observable” covariate regression Features from:

- Data from Human Mortality Database (http://www.mortality.org).

- We use four different data sets:
  - Birth counts;
  - Death counts;
  - Life tables: Life Expectancy at Birth and Death Rates.

- The time series vary in terms of data structure, the number of available observations and the missingness attributes of the records.
TYPES OF DATA:

- **One dimensional time series data per country per gender**
  (31 countries, M and F, gives 124 time series):
  - Birth counts and
  - Life expectancy at Birth.

- **Multivariate cross sectional time series data per country & gender**: age specific data for Death counts and Death Rates.

- **A single observation per country in time t describes**:
  - number of deaths of people with ages from 0 to 110+ (Death counts) or;
  - number of deaths for ages from 0 to 110+ scaled to the size of that population, per unit of time (Death Rates).
Model estimation performed by Forward-Backward Kalman Filter within Rao-Blackwellised Adaptive Gibbs Sampler (MCMC).

The state space models we considered in our studies were of type:

1. [LCC:] Lee-Carter model with the stochastic period + cohort effect.

2. [DFM-PC:] demographic factor model versions of Lee-Carter (Period-Cohort).

The factors are obtained by performing Probabilistic Principle Component Analaysis PPCA jointly on the set of data for all countries listed, excluding:

United Kingdom (response variable)
LC State Space Model - only a Period Effect $\kappa_t$ included.

**Figure**: In sample analysis residuals (left Female, right Male).
LCC State Space Model - with Period + Cohort Effects $\kappa_t, \gamma_{t-x}$ included.

Figure: In sample analysis residuals (left Female, right Male).
Application

- [DFM-PC-B:] the mean of first principal component of Birth counts as a static parameter, age specific element of $\varrho_t$;
- [DFM-PC-D-r/s:] the first principal component of Death counts (which is age and country specific) as an exogenous factor, one element of $\varrho_t$ corresponds to a country specific subvector of the component;
- [DFM-PC-Mx-r/s:] the first principal component of Death Rates (which is age and country specific) as an exogenous factor, one element of $\varrho_t$ corresponds to a country specific subvector of the component.

r/s - is robust vs standard
• **Out-of-Sample Study:** Model calibration period is 1922 – 2002 ⇒ forecast performance analysis for 2003 – 2013

<table>
<thead>
<tr>
<th>Model</th>
<th>MSE</th>
<th>DIC</th>
<th>MSEP\textsubscript{MCMC}</th>
<th>MSEP\textsubscript{Kalman}</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCC</td>
<td>0.0097</td>
<td>-3627</td>
<td>0.1778</td>
<td>0.1774</td>
</tr>
<tr>
<td>DFM-PC-B</td>
<td>0.0072</td>
<td>-6500</td>
<td>0.0057</td>
<td>0.0062</td>
</tr>
<tr>
<td>DFM-PC-D-r</td>
<td>0.0182</td>
<td>-6380</td>
<td>0.0177</td>
<td>0.0251</td>
</tr>
<tr>
<td>DFM-PC-D-s</td>
<td>0.0065</td>
<td>-5996</td>
<td>0.0185</td>
<td>0.0156</td>
</tr>
<tr>
<td>DFM-PC-Mx-r</td>
<td>0.0081</td>
<td>-8225</td>
<td>0.0111</td>
<td>0.0129</td>
</tr>
<tr>
<td>DFM-PC-Mx-s</td>
<td>0.0174</td>
<td>-3951</td>
<td>0.0692</td>
<td>0.0285</td>
</tr>
</tbody>
</table>

• The results confirm that adding demographic features, as additional explanatory variables to the LCC model, improves both in-sample fit out-of-sample fit and therefore the predictability of log death rates.
Figure: 10-year out-of-sample forecasted log death (y axis) rates by age with corresponding prediction intervals.
Figure: 10-year out-of-sample forecasted log death (y axis) rates by age with corresponding prediction intervals.
Figure: 10-year out-of-sample forecasted log death (y axis) rates by age with corresponding prediction intervals.
Conclusions

- We explored how to construct a state space formulation of the stochastic mortality models for Period and Cohort factors.
- We explored how to extend to Hybrid Multi-Factor Stochastic State-Space Mortality models with Period-Cohort factors as well as demographic regressors.
- We briefly learnt about feature/covariate extraction methods to extract the demographic factors used in the extended HMF Stochastic State-Space Mortality models.
- Standard Lee-Carter Period-Cohort model consistently underestimates forecast log-death rates.
- Extended models proposed improve significantly the forecast performance of log-death rates.
Table of contents

1. Mortality Modelling Context
2. State Space Stochastic Mortality Models
3. State-Space Hybrid Factor Models: Regression Extensions
4. Robust Probabilistic Feature Extraction Methods
5. Application to Mortality Modelling and Demographic Data
6. Appendix
References

1. Fung M.C., Peters G.W., Shevchenko P.V.  
   *A unified approach to mortality modelling using state-space framework: characterisation, identification, estimation and forecasting.*  
   Available at SSRN: https://ssrn.com/abstract=2786559

2. Fung M.C. and Peters G.W. and Shevchenko P.V.  
   *A State-Space Estimation of the Lee-Carter Mortality Model and Implications for Annuity Pricing*  
   MODSIM Modelling and Simulation Society. 2015, July.  
   Available at SSRN: https://ssrn.com/abstract=2699624

3. Toczydlowska D., Peters G.W., Fung M.C. and Shevchenko P.V.  
   *Stochastic Period and Cohort Effect State-Space Mortality Models Incorporating Demographic Factors via Probabilistic Robust Principle Components*  
   Risks: Special Issue on “Aging Population Risks”.  
   Available at SSRN: https://ssrn.com/abstract=2977306

4. Fung M.C., Peters G.W. and Shevchenko P.V.  
   Available at SSRN: https://ssrn.com/abstract=2907868