Coherent mortality forecasting for small populations: an application to Swiss mortality data

Cheng Wan[1], Ljudmila Bertschi[2] and Yishan Yang[**]

March 2013

Abstract

Pension funds in Switzerland are exposed to longevity risk possibly to a greater extent than in many other developed economies. Swiss pensioners enjoy a very high standard of living, a well-developed social security system and generous medical insurance for the very oldest residents. For occupational pension plans, retirement and spouse pensions are guaranteed and should be paid lifelong, posing severe longevity risks. In the last five years we have begun to see the first steps to recognise the impact of longevity risk on the balance sheet in Switzerland. In 2010 the first generational mortality table for private pension funds was published (LPP/BVG 2010) – and this is the only actuarial basis for private pension funds. There is also a dearth of financial products to combat longevity risk, with a lack of buy-in and very limited variety of buy-out solutions available. The solutions that do exist frequently come at a very high price. To better understand the impact of longevity risk on the balance sheet, and to facilitate the development of capital market solutions for longevity hedging strategies, stochastic mortality models are indispensable.

To overcome these problems we have developed our own approach for modelling mortality rates for a small population (here Switzerland) with reference to developments in a larger population (global trends). The larger population is modelled with a Plat model and the spread between the larger population and the small population is modelled with a three factor Lee-Carter model. We allow the projections in the two populations to diverge in the short run while constraining them to share a similar pattern in the long run. Compared to existing methods, this method is more robust and allows more flexible age effects. The method is applied to data for Switzerland and 13 other countries (who are grouped as the large reference population) using data from Human Mortality Database. The outcomes and advantages of this approach are demonstrated with reference to the existing Swiss mortality tables (LPP/BVG 2010). Whilst applied to the case of Switzerland, the methodology is generalizable and can be flexibly adapted to other cases, such as addressing regional mortality rates within a large country etc.

Keywords: Longevity risk, Mortality forecasts, Pension fund liabilities, Switzerland, LPP 2010, Plat model, Lee-Carter model, SAINT model, Human Mortality Database.

JEL-Classification: C18, C22, C51, C11

[1] Cheng Wan, is a research associate in Towers Watson RAS, Youkeyuan Road 88, 430074, Wuhan, China. E-mail: cheng.wan@towerswatson.com

[2] Dr. Ljudmila Bertschi, SAV, Swiss certified pension actuary, is a Senior Benefits Consultant at Towers Watson AG, Talstrasse 62, 8021 Zurich, Switzerland. E-mail: ljudmila.bertschi@towerswatson.com

[**]Yishan Yang is a former associate in Towers Watson RAS, Xiangmei Road 1061, 518000, Shenzhen, China. E-mail: yishan.yang@whu.edu.cn

1. Introduction

In recent years, longevity risk has become a hot topic in developed countries. Many studies have shown that human beings are living much longer. The life expectancy increases at a rate ranging from 1 to 5 month every year according to different estimates during the last century. And there is no sign that the improvements in longevity are going to stop. Consider huge financial impacts of increasing life expectancy on the first and second pillars, it is no surprise that mortality modelling has attracted an increased attention.

There is a vast literature on mortality modelling, and we refer to Section 2.3 for a general review. However, most of these models focus on single-population modelling. For small populations like Switzerland, the development of mortality rates usually shows great variability and irregular patterns; therefore mortality projections generated by standard models might be less plausible.

In this paper, we propose a methodology to perform forecast for small populations with reference to a larger population. The reference population is modelled using the Plat model to generate the underlying trend. The spread or the difference between the small population and the reference trend is captured by an extended Lee-Carter model. Short-term deviations are allowed while long-term forecasts are ensured to be coherent. The projection also takes account of parameter uncertainty. The complete model has a simple structure and is easy to use in practice.

The paper is organized as follows. In Section 2 the background for Switzerland longevity risk is introduced and existing mortality modelling techniques are reviewed. In Section 3 a new mortality model for small populations is proposed. In Section 4 the model is applied to Switzerland mortality data with a pooled international population as reference and the forecasting results are compared with Switzerland official tables. Conclusion and discussions are provided in Section 5.

2. Background of our model and overview of relevant publications

2.1 Second pillar in Switzerland and the structure of Swiss labour force

Retirement and risk benefits in Switzerland are provided through a well-established system that incorporates state provision, employer provision and personal savings, known as the "three pillars" system. The benefits from the second pillar (BVG/LPP), together with first-pillar benefits, should enable recipients to maintain their previous standard of living according to the Swiss Constitution. Retirement and spouse pensions are guaranteed for life. All benefits in the second pillar are immediately and fully vested and the annuitisation rate is very high.

Switzerland has one of the highest immigration rates on the continent. Immigrants now make up 23% of the Swiss population (ca. 8 Mio in 2012) as the country traditionally used to be a destination for employment-seeking French, Germans, Italians, Spanish and Portuguese (as neighbour countries) and a highly skilled labour force from other distant countries. Although Switzerland is not a member of the EU, it has partially joined the liberalized European labour market. In June 2002, the Bilateral Agreement on the Free Movement of Persons between Switzerland and the EU Member States came into force. According to different forecasts prepared by the Swiss federal office for statistics (BFS/ESPOP) and Eurostat [1] Swiss population is going to grow in the next 50 years (in the period 2010-2060) by 15%-19% compared to the expected reduction by nearly 15% in Germany and a small average growth below 1% in EU-27 countries for the same period. Like the rest of Europe, Switzerland knows that further immigration will be needed to compensate for the aging population and to ensure economic growth in the future.

The labour force in Switzerland exceeded 4.5 Mio employees (2012). Due to the mandatory second pillar all these employees are insured in benefit plans set up by their employers which offer at least the mandatory minimum level (so called BVG/LPP-Minimum).

2.2 Swiss mortality tables and longevity risk in Switzerland

Switzerland is a small country. There is only one actuarial mortality table: LPP/BVG 2010, available for private pension funds[3]. Another table, VZ 2010, is used for Swiss public pension funds. Both tables are generational and deterministic, have a projection horizon up to 2030 (LPP/BVG 2010) and are not substantially different in relation to the level of future life expectancies. Insurance companies use their own mortality tables which are not disclosed and seem to be much more conservative than ones available for pension funds. There were no officially announced transactions with longevity swaps for Swiss pension funds.

Typical questions like "how a pension fund can assess the longevity risk it is exposed to" and/or "how it can protect itself from longevity risk and what the fair price is" cannot be addressed simply based on one generational table. Differentiation of mortality level based on the salary/ income scale is not possible, however it is known from the research that the life expectancy is positively correlated with the income level. Switzerland in international comparison has probably one of the highest levels of salaries and retirement pensions insured in the second pillar. A well-developed social security system, generous medical insurance and supply for very old Swiss residents including First Pillar support programs in financing nursing homes facilitate additionally a longer life expectancy of Swiss residents older than 80. The assessment of longevity risk requires a stochastic mortality model. The first publication on Swiss stochastic mortality modelling was prepared by Severine Gaille [2]. She indicated that life expectancies in LPP/BVG 2010 generational table appeared to be underestimated especially for women. Another attempt to calibrate other stochastic mortality models (M1-M7) for Switzerland was done by Abt [3].
Taking into account that in Switzerland the ratio of pension assets to GDP substantially exceeds 100% (it reached 150% before the financial crisis 2007-2009 – see Towers Watson Global Pension Asset Survey [4]) the system faces a major challenge in managing longevity risks. Swiss autonomous pension funds have accumulated substantial pensioner liability; in some mature pension funds it exceeds 50% and reaches 70%-80% of the total liability. The actual impact of longevity risk on pension liabilities varies according to interest rate levels and specific demographics of each individual pension plan.

The strains posed by the growing longevity risk and the historically lowest interest rates[4] force plan sponsors and Swiss pension funds to de-risk their benefit plans. Even if prospective benefits can be reduced by adjusting annuity conversion ratio at retirement, soft-closing defined benefit plans and so shifting investment risks to employees, pensions in payment are guaranteed and need proper funding. Indexation of retirement pensions in the second pillar is not mandatory and many pension funds try to index pensions in payment on an ad hoc basis, however public pension funds do it practically on a regular basis.

Most OECD countries have enjoyed large gains in life expectancy over the past decades, thanks to improvements in living conditions and progress in medical care. Compared to the OECD average level, Switzerland has a higher standard of living (a per capita GDP is larger than that of the big Western European economies) and a longer life expectancy. It is widely recognised in Switzerland that the longevity risk is a major threat to pension funds and their sponsors as it cannot be simply hedged like in case of interest rate and investment risks. Compared to the situation in the UK there is a dearth of financial products in Switzerland to combat longevity risk, with a lack of buy-in and a very limited variety of buy-out solutions. The solutions that do exist frequently come at a very high price and are unaffordable for pension funds and their sponsors.

For promoting products for longevity hedging on capital market as well as for prudent reserving for Swiss pension funds stochastic mortality models are indispensable.

---

[3] LPP/BVG tables have been published every 5 years starting from 2000 and are based on mortality statistics of 14 biggest pension funds in Switzerland which belong to the leading Swiss firms from different industry sectors (Swatch Group, ABB, Sulzer, Nestle, Migros, Coop, UBS, CS, Swiss Re, Alcan etc.); the biggest civil servant pension fund Publica in included. The next publication of the BVG table, BVG 2015, is expected to be done in 2015 and will be based on the last 5 years mortality data.

[4] Swiss 10-year-government bond yields reached in Dec 2012 0.45% - the lowest level worldwide and since the third quarter 2011 the Swiss inflation has been negative.

This study establishes a stochastic mortality model for Swiss pension funds which takes into account observes trends in longevity developments as well as the structure of the labour force in the Swiss market.

It was observed that Switzerland belongs to a group of countries with high mortality convergence [5] demonstrating steady growth of life expectancy without any slowdown[5]. We are going to use this finding for setting up the structure of our mortality model.

2.3 Mortality modelling up-date

A comprehensive overview on mortality modelling is given in Pitacco, Denuit and Haberman (2009) [6], where general issues related to longevity risk as well as general guidance for building mortality models are discussed.
Lee and Carter (1992) [7] proposed a log-bilinear model (so called M1 model), which explains the log of death rates with a time-independent age component and a component with a time-varying stochastic factor multiplied by another age component describing the speed of mortality improvement with different years. The model is first fitted to the historical data and then the time factor is modelled as time series with an appropriate ARIMA model. The forecast is driven by projecting the time factors to future with the assumption that age component will remain the same. The Lee-Carter model is considered to be one of the most influential stochastic mortality models and have been widely used.

Since then, many stochastic mortality models have been proposed. Renshaw and Haberman (2006) [8] generalized the Lee-Carter model by introducing a cohort effect; Currie (2006) [9] used a simpler additive form of Age-Period-Cohort (APC) model with P-spline to ensure smoothness. Cairns, Blake and Dowd (2006) [10] (often referred as CBD) introduced a powerful alternative for mortality forecasting: modelling of mortality rates on the logit scale. Their model removes the independent age effect and assumes simpler parametric forms for age components in the rest age-period interactions. Cairns-Blake-Dowd model has several generalizations and improvements (hereafter CBD model family), such as allowing the cohort effect or adding a quadratic age component.

Plat (2009) [11] introduced a new APC model making use of the both Lee-Carter and CBD model family. Cairns et al. (2011a) [12] gave a quantitative comparison of several commonly used mortality models and Hunt and Blake (2013) [13] set out a general procedure (GP) for constructing mortality models. Hyndman and Ullah (2007) [14] used a functional data approach. For mortality modelling under Bayesian framework one is referred to Czado et al. (2005) [15] and Luoma et al. (2012) [16].

Despite advances in theoretical modelling approaches, there are only a few applied models used by pension funds for reserving purposes in Europe where the problem of population ageing is going to be acute. The ground for this is the fact that model projections could become less reliable if stochastic mortality models are fitted based on statistics from small populations, where substantial variability occurs due to limited exposures. Since mortality patterns are likely to be similar in closely related populations that have similar socioeconomic conditions (standard of living, medical care etc.) and close connections, modelling mortality for a group of populations offers a useful approach to gaining deeper insights into the nature of mortality improvements and so improving the forecast quality for small populations. Li and Lee (2005) [17] applied the Lee-Carter model to a group of populations, testing if populations can be modelled with common factors and compared the results from separate individual population forecasts. Biatat and Currie (2010) [18] introduced the idea of similarity for two populations using P-spline. Dowd et al. (2011) [19] proposed the gravity model for two related but different-sized populations with the coherency between them tuned by the gravity parameters. Jarner and Kryger (2011) [20] developed in 2006-2008 the SAINT model with application to Danish mortality data with reference to a pooled international data – practically the first applied stochastic mortality model successfully used for reserving purposes for the Danish APT pension fund. Cairns et al. (2011b) [21] introduced joint mortality modelling for two populations in a

[5] France, Italy, Spain, Portugal belong to this group as well. The life expectancy in this group is ca. 2 years longer than that in groups with the low mortality convergence [1].

Bayesian framework. Interesting results on mortality convergence in high-income countries are published in [22] by D'Albis at al. (2012).

Due to mortality convergence in countries with the similar living environment (climate, nutritional standards, hygienic and sanitary conditions etc.) and well-established three pillar systems we can win reliable information on trends in mortality improvements specific to countries with high level of living standards and well developed social security systems.

2.4 Reference populations

All data come from the Human Mortality Database (HMD) [23]. Following Li and Lee (2005) [17] approach, 13 low-mortality countries (excluding Canada and Italy due to non-available data as from 2008 at the time of writing, namely, Austria, Denmark, England and Wales, Finland, France, West Germany, Japan, The Netherlands, Norway, Spain, Sweden, Switzerland and USA) are selected and pooled as the international trend). Noticing that for age beyond 90 death rates are often too variable and suffer from data quality issue, we restrict ourselves to age ranging from 50 to 89 for model fitting purposes. For the period range we select data from 1971 to 2009. As regards the pooling method, we add death numbers or exposures across 13 countries at each age-period cell. Figure 1.1 shows both pooled international and Switzerland mortality data at age 85. It is obvious that in spite of the large variation exposed in the Swiss data, the general trends are closely related.

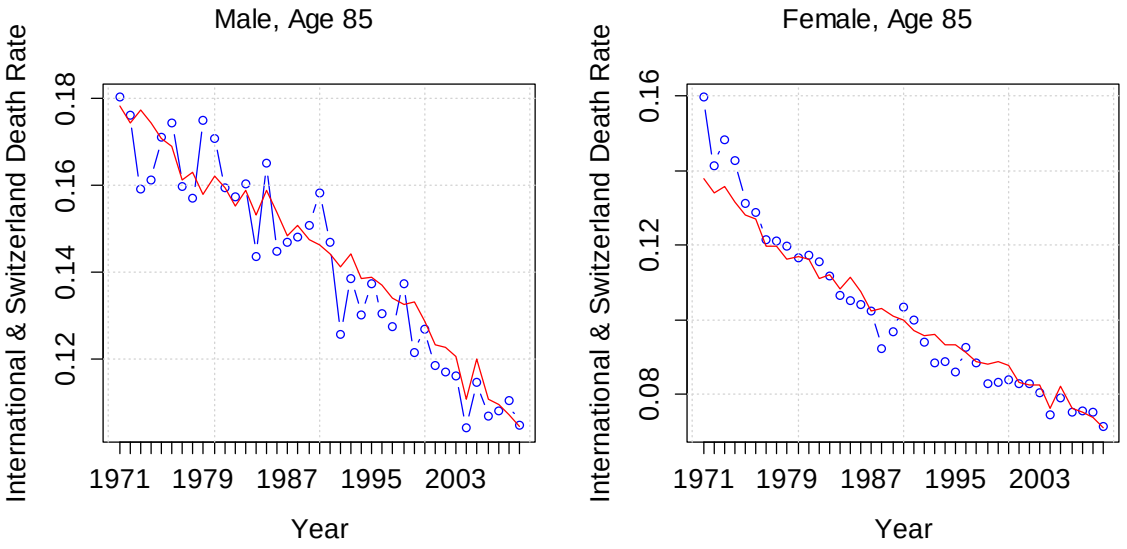


Figure 1.1: International death rates (red curve) and Switzerland death rates (blue points) plot at age 85. Left: Male population; Right: Female population.

We define $m(x,t)$ as the crude central death rate for age x in calendar year t, calculated by raw death number $d(x,t)$ divided by exposure to risk $e(x,t)$ from HMD data source:

$$m(x,t) = \frac{d(x,t)}{e(x,t)}.$$

Another mortality measure used is the mortality rate $q(x,t)$, which is the probability that an individual with age x in calendar year t will die between year t and $t+1$. Mortality rate and death rate are close to each other, in the following text, we adopt one of the common approximations for $q(x,t)$:

$$q(x,t) \approx \frac{1}{1+0.5m(x,t)}.$$

3 Model

3.1 Background of our model

For single population modelling, recall the Lee-Carter [7], CBD model family [10] and the Plat model [11]:

Lee-Carter type: $\log(m(x,t)) = \beta_x^{(1)} + \beta_x^{(2)}\kappa_t^{(2)} + \text{other terms}.$

CBD model family: $\text{logit}(q(x,t)) = \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \text{other terms}$

Plat model: $\log(m(x,t)) = \beta_x^{(1)} + \kappa_t^1 + \kappa_t^2(x-\bar{x}) + \text{other terms}$

Lee-Carter type models allow nonparametric form for age effect and have a separate age effect $\beta_x^{(1)}$. CBD type models assume simpler parametric form for age effect and use a separate period effect $\kappa_t^{(1)}$. Plat model provides a compromise between these two families. For multiple-population modelling, Li and Lee (2005**)** [17] first estimate a group of population for a common factor as a reference, then estimate the single population based on the estimated reference. During the process, Lee-Carter model is used for both stages. The SAINT model [20] shares a similar procedure, except that a frailty model is employed to estimate the mortality force $\mu(x,t)$ for the reference and a multivariate time series are used for Danish spread.

Li and Lee: $\log\big(m(x,\ t,\ i)\big) = a(x) + B(x)K(t) + a(x,\ i) + b(x,\ i)k(t,\ i).$

The SAINT model: $\log\big(\mu(x,\ t,\ i)\big) = \log\big(\mu(x,\ t)\big) + \sum_{j=1} r_j(x)k_j(t,\ i).$

For our case, we would like that all populations follow the same trend while spread is allowed for each specific population. Inspired by these models, we follow the common procedure shared by above approaches, which is to estimate a reference trend first, then to model the spread or difference between the individual population and the reference trend. Specifically, we use the model:

$$\log\big(m(x,\ t,\ i)\big) = \beta_x + \kappa_t^1 + \kappa_t^2(x-\bar{x}) + \gamma_{t-x} + \beta(x,\ i) + \sum_{j=1}^{m} b_j(x,\ i)k_j(t,\ i). \qquad 3.1$$

The first part $\beta_x + \kappa_t^1 + \kappa_t^2(x-\bar{x}) + \gamma_{t-x}$ is essentially the Plat model, which is employed to model the reference trend. The second part $\beta(x,\ i) + \sum_{j=1}^{m} b_j(x,\ i)k_j(t,\ i)$ is an extended Lee-Carter model for the spread between the reference trend and individual population i.

More specifically,

β_x is the age effect describing the general pattern of mortality by age for the reference trend.
κ_t^1 is the period effect describing the general pattern of mortality by period for the reference trend.
κ_t^2 is a period effect that allows mortality changes to vary for different ages for the reference trend,
γ_{t-x} is used to capture the cohort effects for the reference trend.

$\beta(x,\ i)$ is used to capture the general mortality difference by age between the reference trend and the individual population.
$\sum_{j=1}^{m} b_j(x,\ i)k_j(t,\ i)$ is used to capture all other unobserved differences between the reference trend and the individual population and m is a parameter indicating the number of age-period components needed for the individual population.

3.2 Parameter constraints

The model faces the problem of identifiability, which means that different parameterizations can lead to identical value for $\log\big(m(x,\ t,\ i)\big)$. Constraints for the parameters are therefore necessary for the model specification. We follow the approach by Cairns et al. (2009) [24] and the one used by Lee-Carter (1992) [7]:

$$\sum_{c=c_0}^{c_1} \gamma_c = 0 \tag{3.2}$$

$$\sum_{c=c_0}^{c_1} c\gamma_c = 0 \tag{3.3}$$

$$\sum_{t} k_j(t,\ i) = 0 \tag{3.4}$$

$$b_j(x,\ i) = \frac{b_j(x,\ i)}{\sum_x b_j(x,\ i)} \tag{3.5}$$

where c_0 and c_1 define the cohort range. The first two constraints ensure that cohort effect will not compensate a lack of age-period effect. It will fluctuate around 0 and has no trend so that on average it has no impact on mortality. The rest constraints are used to normalize the parameters in the spread model. We do not adjust the $k_j(t,\ i)$ as it was done by Lee and Carter (1992) [7] since Brouhns et al. (2002) [25] point out the re-estimation is not necessary.

3.3 Model estimate

The fitting procedure contains two stages. At the first stage we estimate the parameters in the model 3.1 while at the second stage we estimate the dynamic parameters for stochastic factors based on the estimated time series.

3.3.1 Model parameter estimation

In the first stage, we use a 2-step approach for estimation. In the first step we estimate parameters in the reference trend:

$$\log\big(m(x,\ t)\big) = \beta_x + \kappa_t^1 + \kappa_t^2(x - \bar{x}) + \gamma_{t-x}. \tag{3.6}$$

The procedure is done via maximum likelihood approach and the death numbers are assumed to follow a Poisson distribution, namely, for each age-period cell we have:

$$d(x,\ t) \sim Poisson\big(e(x,t)m(x,t)\big),$$

where $d(x,\ t)$ is the death number, and $e(x,\ t)$ is the exposure number. The log likelihood for 3.6 is given by:

$$\text{loglikelihood}(\phi) = \sum_{x,t} \Big(d(x,t) \log\big(e(x,t)m(x,t;\phi)\big) - e(x,t)m(x,t;\phi) \Big) + \text{constant},$$

where ϕ is the parameter set for the model. Brouhns et al. (2002) [25] introduce a commonly used fitting methodology via Newton-Raphson method. With the toolkit and codes provided by LifeMetrics[6] and Pensions Institute, it is convenient for us to fit the model with R[7]. In the second step we estimate parameters in the spread between the reference trend and individual population based on the estimated reference trend $\hat{m}(x,t)$ from the first step:

$$\log\big(m(x,\ t,\ i)\big) - \log(\hat{m}(x,t)) = \beta(x,\ i) + \sum_{j=1}^{m} b_j(x,\ i) k_j(t,\ i). \qquad 3.7$$

For this step singular value decomposition (SVD) is used to find the corresponding parameters. One of the applications by SVD is to find separable effects. Lee and Carter (1992) [7] uses this technique to estimate $\beta(x,\ i)$ and $k_j(t,\ i)$.

We estimate $\beta(x,\ i)$ by the average of the spread at each calendar year, then subtract the estimate from the left side of equation 3.7 and apply SVD technique:

$$\text{SVD}\{\log\big(m(x,\ t,\ i)\big) - \log\big(\hat{m}(x,t)\big) - \hat{\beta}(x,\ i)\} = UDV^T = \sum_{i=1}^{r} u_i d_i v_i,$$

where matrices U and V are left and right singular matrices with u_i, v_i the corresponding singular vectors. D is a diagonal matrix containing the singular values d_i in a descending order and r is the rank of the matrix.

Notice that Lee and Carter (1992) [7] only use the first components $u_1 d_1 v_1$ since in many situations it already explains a large portion of variation. Tuljapurkar et al. (2000) [26] also show that the first singular value has taken a very large portion compared to the rest. We have used three components in our case.

3.3.2 Dynamic parameter estimation

In the previous literatures, random walk with drift (RWD) is a common choice for κ_t^1 (Lee and Carter, 1992, CBD, 2006) and AR(1) process is often assumed for κ_t^2 (Plat 2009). ARIMA method is also used for a more objective reference.
Notice that, since we have two time factors, it is necessary to decide whether to use separate univariate processes or bivariate time series for them. The former approach is used here. For the other stochastic factors $k_j(t,\ i)$ in the spread modelling, we estimate the parameters under the structure of vector autoregression process (VAR).
Meanwhile, we assume no intercepts for each equation in the VAR model, which will ensure we get a coherent forecast in the long term. For robustness, we also use Bayesian Information Criterion (BIC) as a reference to test the appropriate structure for all the above dynamic processes. Notice that here we have made the assumption that the dynamic processes between the reference trend and individual spread are independent.

Finally, m needs to be determined (last term in equations 3.1 and 3.7). To find an appropriate value for it, we have plotted the fitting results together according to different values of m and visually selected its value. Like Cairns et al. (2009) [24], a residual plot is provided to examine the estimation results. The residuals

[6] See http://www.jpmorgan.com/pages/jpmorgan/investbk/solutions/lifemetrics/software

[7] R is a free and powerful language for statistical computing with a similar syntax to S language; it can be downloaded at http://www.r-project.org/.

are standardized and determined via raw death numbers minus the fitted death numbers, divided by the square root of the fitted death numbers:

$$Standardized\ residual(x,t,i) = \frac{d(x,t,i) - e(x,t,i)\hat{m}(x,\ t,\ i)}{\sqrt{e(x,t,i)\hat{m}(x,\ t,\ i)}},$$

We expect these residuals to be i.i.d. and follow a standard normal distribution approximately.

3.4 Model forecast

Mortality is influenced by stochastic factors. The forecast is made by projecting the stochastic factors into the future on top of deterministic trend. Longevity cannot be easily hedged since there are no large traded markets in longevity risk and the common approach by selling lot of policies will not make the risk diversifiable. Actuaries therefore pay more attention to the uncertainty levels where projections are needed. Model uncertainty and parameter uncertainty are the main sources, for example, see Cairns et al. (2006) [10]. Here we just consider the parameter uncertainty.

There are several methods to incorporate parameter uncertainty in the forecast. Among them, three choices are common in the literature. The first one is to model under a Bayesian framework, and one can examine the uncertainty by exploring the posterior distribution, see Cairns et al. (2006) [10]. This method is theoretically elegant but difficult to implement in practice. The second choice is by bootstrapping. Either via a residual bootstrap or a Poisson bootstrap, one re-sample the data and perform the whole fitting and projection process repeatedly, see Brouhns et al. (2005) [27] or Koissi et al. (2006) [28]. The last one is the most practical method to allow for parameter uncertainty. This method simulates new parameters from the asymptotical normal distributions of the maximum likelihood estimator. We follow the procedure by Li and Lee (2005) [17], which only considers the parameter uncertainty in the dynamic processes.

Specifically, for the reference trend projection, we suppose that we are using RWD for κ_t^1 and AR(1) without intercept for both κ_t^2 and γ_{t-x}; for the individual spread, as well as suppose that a VAR(1) process without intercepts is assumed for $k_j(t,\ i)$ with $m = 3$ in the last term in 3.1.

The reference trend forecast at year T looks like:

$$K_{T+1}^1 = \left(\hat{d} + \widehat{se}\ (\hat{d})\eta_{K_t^1}\right) + K_T^1 + \widehat{\sigma_{K_t^1}} z_{T+1}^1\,,$$

$$K_{T+1}^2 = \left(\widehat{\alpha_{K_t^2}} + \widehat{se}\ \left(\widehat{\alpha_{K_t^2}}\right)\eta_{K_t^2}\right) K_T^2 + \widehat{\sigma_{K_t^2}} z_{T+1}^2\,,$$

$$\gamma_{T+1} = \left(\widehat{\alpha_\gamma} + \widehat{se}\ \left(\widehat{\alpha_\gamma}\right)\eta_\gamma\right)\gamma_T + \widehat{\sigma_\gamma} z_{T+1}^3\,,$$

The individual spread forecast at year T is as follows:

$$k_1(T+1,i) = (\widehat{\alpha_{1i}} + \widehat{se}\ (\widehat{\alpha_{1i}})\eta_{1i})k_1(T,i) + \varepsilon_{1i}(T+1),$$

$$k_2(T+1,i) = (\widehat{\alpha_{2i}} + \widehat{se}\ (\widehat{\alpha_{2i}})\eta_{2i})k_2(T,i) + \varepsilon_{2i}(T+1),$$

$$k_3(T+1,i) = (\widehat{\alpha_{3i}} + \widehat{se}\ (\widehat{\alpha_{3i}})\eta_{3i})k_3(T,i) + \varepsilon_{3i}(T+1).$$

Where $\hat{d}$ is the estimated drift parameter in the RWD process; $\widehat{\alpha_{(.)}}$ are the estimated autoregression coefficients in the corresponding AR(1) or VAR(1) processes; $\widehat{\sigma_{(.)}}$ are the residual standard errors for each process. $\widehat{se}(.)$ is the estimated standard error of the parameter. z is drawn from the standard normal

distribution for each series in the reference trend and $\varepsilon_{,i}$ is drawn from the covariance structure of the series in the individual spread at every new forecasting year.

To account for parameter uncertainty in projections, for each simulation we use the following approach:

- Generate new parameters in the dynamic processes: for each parameter, draw $\eta_{(.)} \sim \text{normal}(0,1)$, multiply it with $\widehat{se}(.)$, then add it to the estimated parameter,
- Generate sample paths to further years with previously generated parameters.

4 Application to Swiss mortality data

4.1 Model calibration

We now apply our method to Switzerland mortality data for both males and females. Recall that we are following a two-step procedure to estimate parameters in Section 3. In the first step we fit the reference data with the Plat model [11]:

$$\log\big(m(x,\ t)\big) = \ \beta_x + \kappa_t^1 + \kappa_t^2(x - \bar{x}) + \gamma_{t-x}.$$

Figure 4.1, Figure 4.2, and Figure 4.3 show the estimated age, period and cohort effects. Table 4.1 gives the parameters in the dynamic proceeses at this step. For both genders, the average age effect β_x has a linear effect on log death rate; Period effect κ_t^1 has a downward trend while κ_t^2 exhibits a convex shape. The cohort effect γ_{t-x} is fluctuating around 0. The sudden hump for cohort born at 1919 due to the influenza epidemics is also observed. For the implied structure in the dynamic processes, random walk with drift is used for κ_t^1 and AR(1) without intercept is used for κ_t^2 and γ_{t-x} as in Plat (2009) [11]. Besides that, we only use cohort estimates up to 1945. The values beyond this year are considered inaccurate because the cohorts have not lived long enough to generate adequent data.

In the second step we fit an extended Lee-Carter model to the difference or the spread between the international reference trend and Switzerland mortalty data:

$$\log\big(m(x,\ t,\ i)\big) - \log(\widehat{m}(x,t)) = \beta(x,\ i) + \sum_{j=1}^{m} b_j(x,\ i)k_j(t,\ i).$$

Recall that for this step, we need to decide on m's value in the last term in the equation above.

In our case we find $m = 3$ is sufficient (See Figure 4.6 for comparison with other values of m). See also Table 4.2 for the largest six singular values and their weights, respectively. The corresponding residual plots with $m = 3$ are shown in Figure 4.7 For both males and females, the pattern is mixed well and no obvious clusters, bands or cohorts need to be further addressed.

Figure 4.4 shows the estimated average age effect $\beta\big(x,\ i\big)$ for the spread. A linear pattern is observed for males while for females the difference remains at low values like -0.25 until age 70. The general age effect for the spread gradually reduces to 0 at higher ages for both genders.

The three stochastic factors in the spread model are shown in Figure 4.5. The mean reverting pattern is well observed. We assume again no intercepts and test their appropriate structures in VAR(p). BIC criterion suggests a VAR(1) structure for both genders. However, we find that the results show that the coefficients significant at 0.1 level are only for one's own lag (male) and none is significant at 0.1 level in two equations (female), see Table 4.3 for details.

We then re-estimate the model with restrictions that all coefficients nonsignificant at 0.1 level equaling 0 for males, therefore each series is explained with its own lagged values. For females, we adopt the same structure

	κ_t^1: RWD	κ_t^2: AR(1), no intercept	γ_{t-x}: AR(1), no intercept
Male	-0.1960, 0.1256	0.9874, 0.0174	0.9611, 0.0327
Female	-0.0178, 0.0123	0.9598, 0.0476	0.9376, 0.0434

Table 4.1: Parameter estimates for stochastic factors (κ_t^1, κ_t^2 and γ_{t-x}) in the international trend.

Male	d_i	1.435	0.729	0.6446	0.6012	0.5476	0.4978
	$\sum_{i=1}^{j} d_i^2 / \sum_{i=1}^{r} d_i^2$	0.350	0.440	0.510	0.572	0.623	0.665
Female	$\mathrm{d_i}$	1.3999	0.8804	0.7666	0.7347	0.6885	0.6065
	$\sum_{i=1}^{j} d_i^2 / \sum_{i=1}^{r} d_i^2$	0.263	0.369	0.448	0.522	0.584	0.633

Table 4.2: Singular values for Switzerland spread. The largest 6 singular values and corresponding weights.

	Male			Female		
	$k_1(t,\ i)\_lag1$ s.e. (p-value)	$k_2(t,\ i)\_lag1$ s.e. (p-value)	$k_3(t,\ i)\_lag1$ s.e. (p-value)	$k_1(t,\ i)\_lag1$ s.e. (p-value)	$k_2(t,\ i)\_lag1$ s.e. (p-value)	$k_3(t,\ i)\_lag1$ s.e. (p-value)
$k_1(t,\ i)$	0.935 0.072 (5.16e-15)	1.112 2.052 (0.591)	-0.418 0.792, (0.601)	0.789 0.083, (3.26e-11)	0.812 1.064 (0.451)	-0.443 2.938 (0.881)

$k_2(t,\ i)$	0.0004 0.005 (0.939)	0.368 0.156 (0.024)	0.059 0.060 (0.332)	0.002 0.012 (0.869)	0.122 0.158 (0.444)	0.442 0.436 (0.318)
$k_3(t,\ i)$	0.007 0.011 (0.538)	-0.488 0.321 (0.137)	0.626 0.124 (1.38e-05)	0.001 0.005 (0.753)	-0.076 0.060 (0.217)	-0.006 0.166 (0.970)

Table 4.3: The VAR(1) fitting results (without restrictions on coefficients) for period effects $k_j(t,\ i)$ in the Switzerland spread.

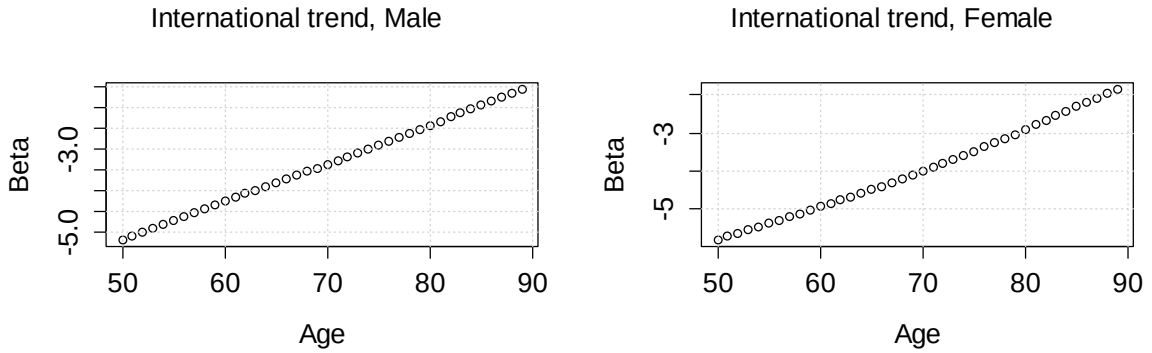


Figure 4.1: Estimated age effect (β_x) for the international reference trend. Left: Male population; Right: Female population.

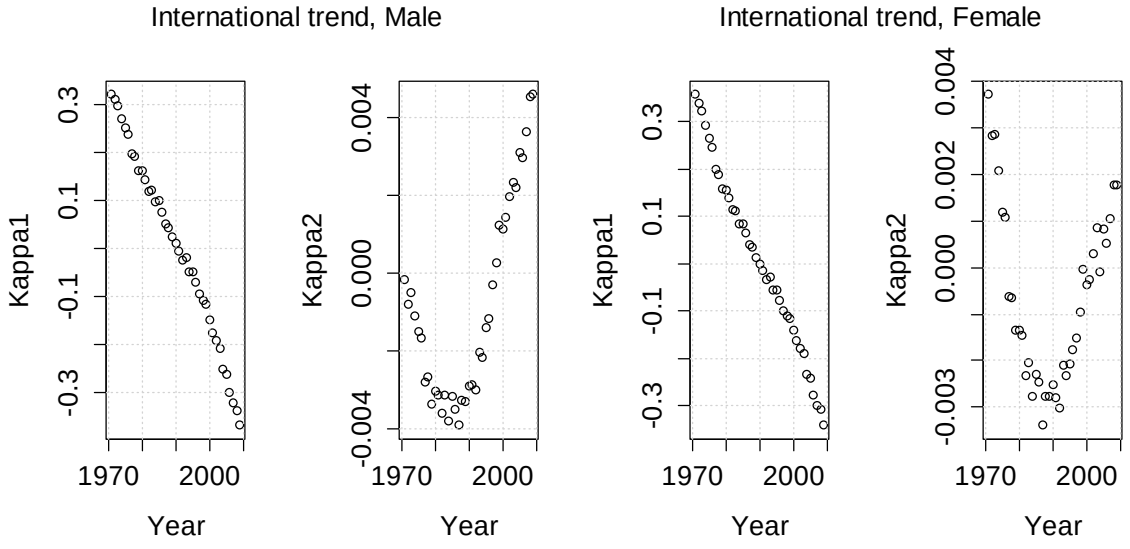


Figure 4.2: Estimated stochastic factors (κ_t^1 and κ_t^2) for the international reference trend. Left: Male population; Right: Female population.

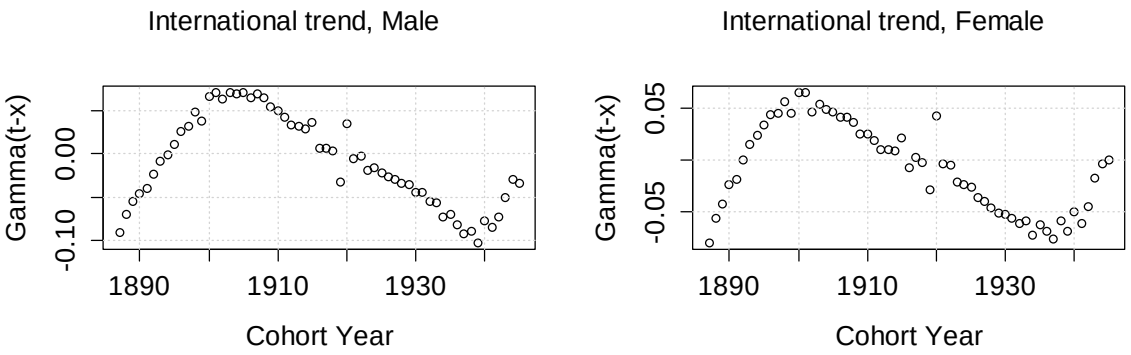


Figure 4.3: Estimated cohort factors γ_{t-x} for the international reference trend up to year 1945. Left: Male population; Right: Female population.

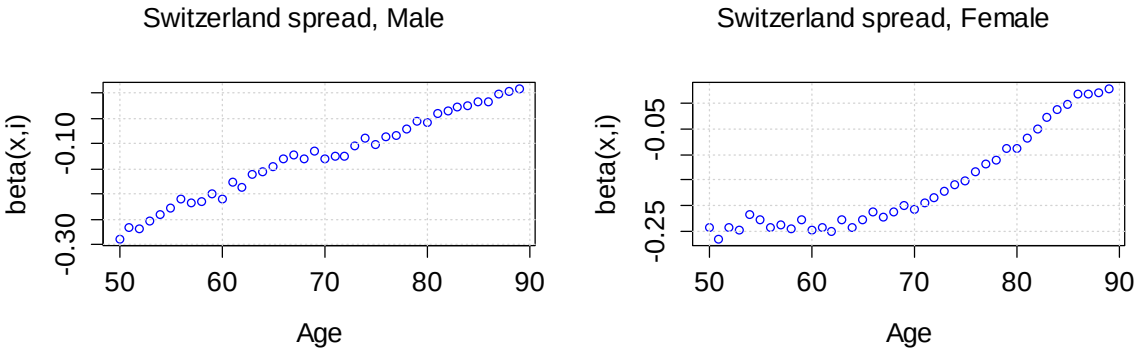


Figure 4.4: Estimated age effect $\beta(x,\ i)$ for the spread between Switzerland mortality and the international reference trend. Left: Male population; Right: Female population.

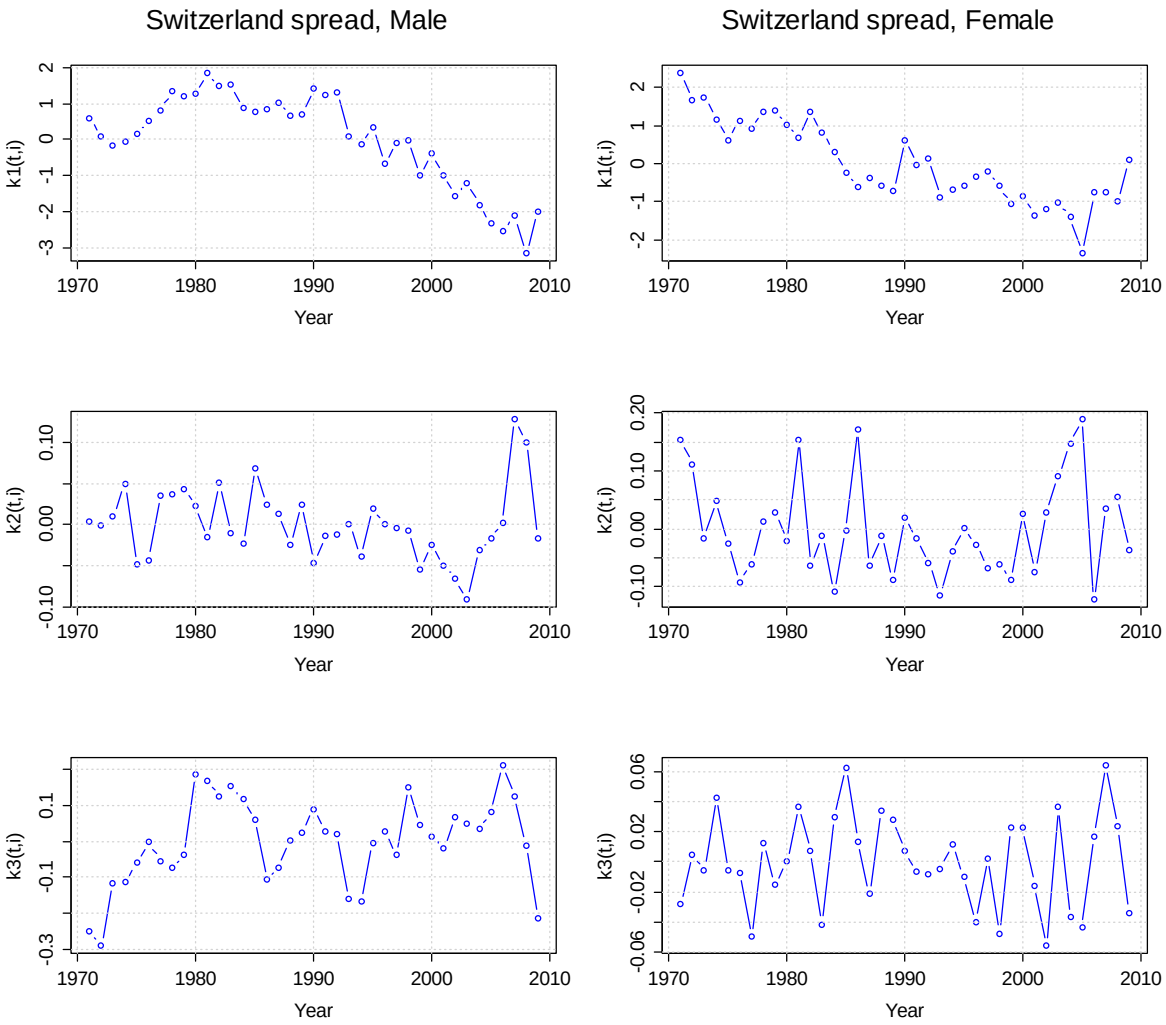


Figure 4.5: Estimated stochastic factors $k_j(t,\ i)$ in the spread between Switzerland mortality and the international reference trend. Left: Male population; Right: Female population.

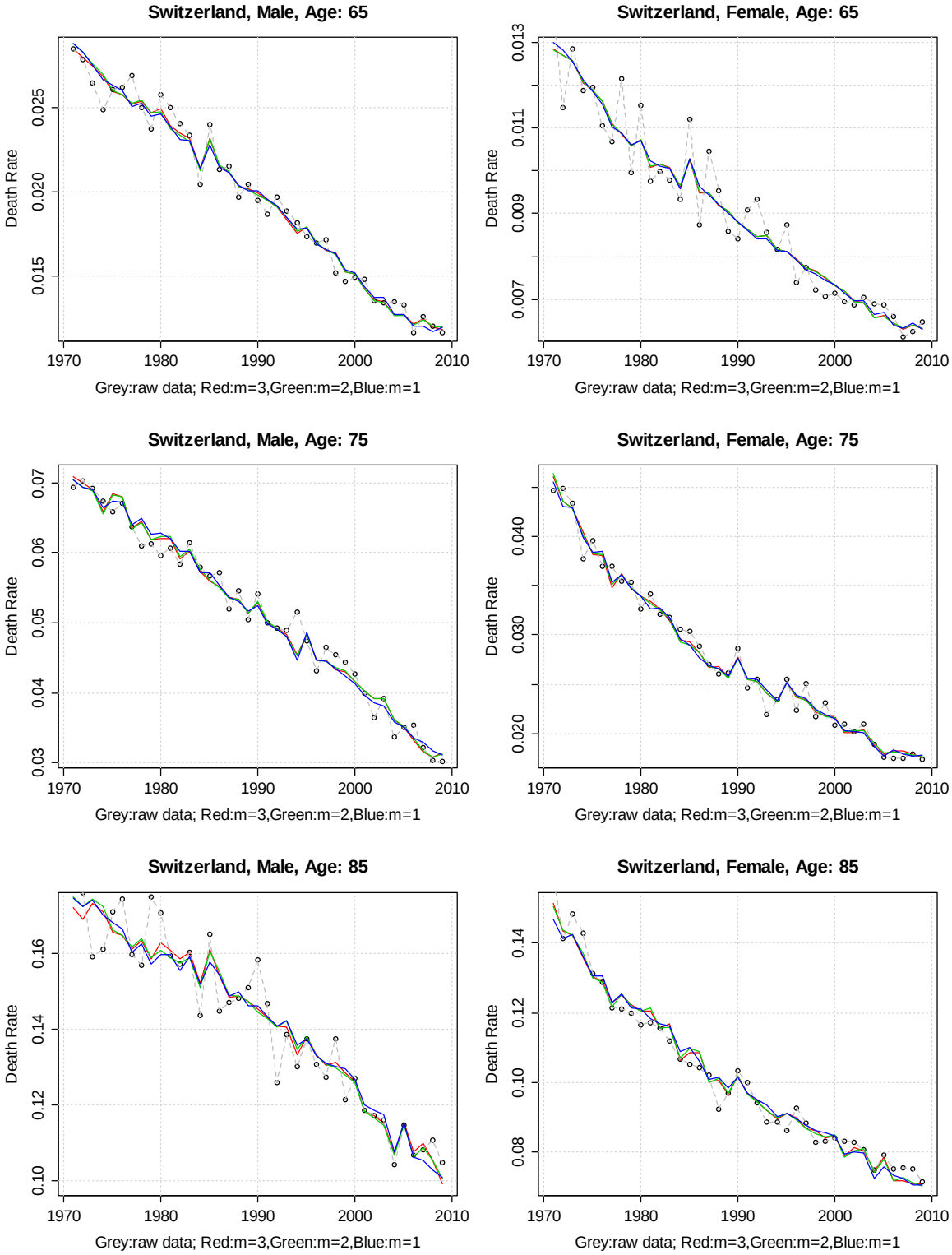


Figure 4.6: Switzerland death rate fitting results with different values of m at age 65, 75 and 85. Left: Male population; Right: Female population. Grey points and dashed line are used for raw data.

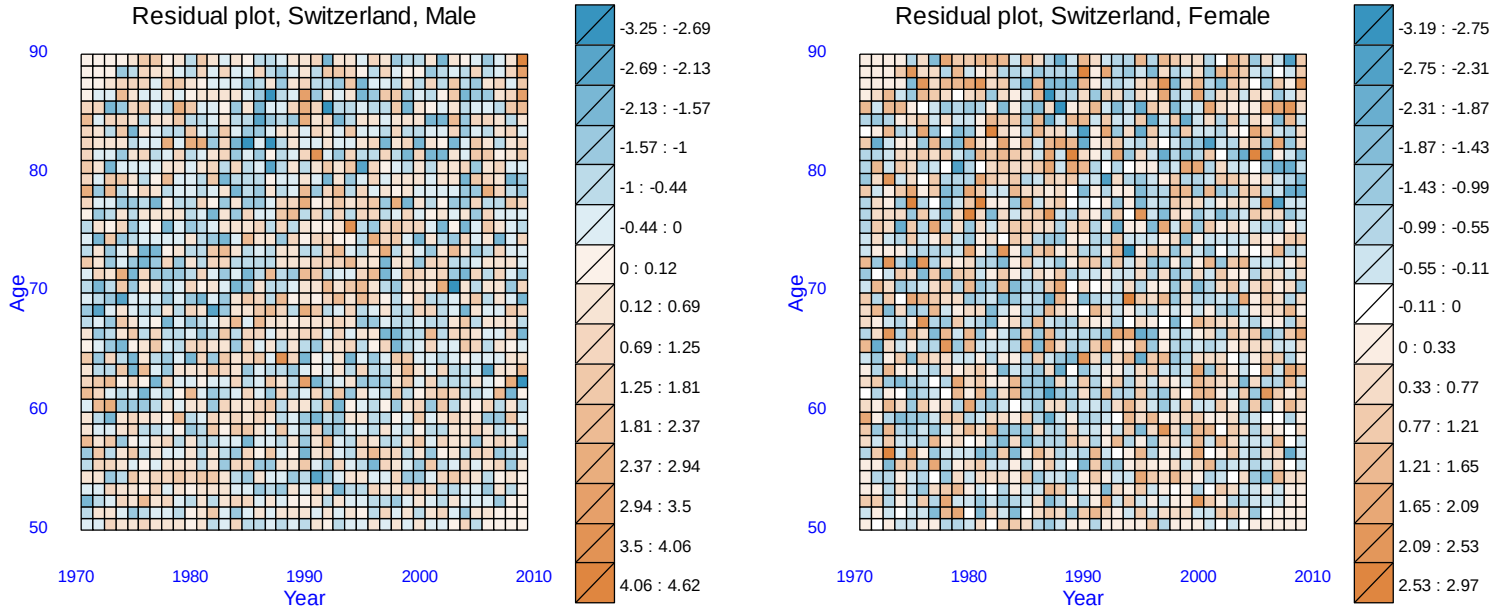


Figure 4.7: Residual plot for Switzerland mortality data using the proposed model with $m = 3$. Left: Male population; Right: Female population.

4.2 Model projection

Generally, the forecast is following the procedure set up in Section 3. The parameter uncertainty is only allowed in forecasting the stochastic factors. We use 5000 simulations and forecast up to year 2038. Figure 4.8 plots the fitting results and fan charts of death rate $m(x,t)$ for the international reference trend (red circles) and Switzerland death rates (green squares) at age 65, 75 and 85 for both genders. Switzerland mortality shares a similar pattern from the international trend. The fluctuations in Switzerland raw data are somehow "smoothed". In the meantime, it is observed that Switzerland has a lower mortality at lower ages compared to the international trend, and this pattern is going to continue in the near future. For higher ages, we have found that the spread between the international trend and Switzerland mortality is small. Figure 4.9 gives the simulated percentiles (5%, 25%, 50%, 75%, 95%) for distribution of the mortality rate $q(x,t)$ at different ages for both genders.

4.2.1 Projection with LPP/BVG 2010 base table

We would like to apply our model projections to a given starting curve. To do this we need the improvement table, which is the mortality improvement at the next year for each given age. There are several ways to measure it, here we define it as:

$$Improvement(x,t) = 1 - \frac{q(x,t)}{q(x,t-1)},$$

Where $q(x,t)$ is the mortality rate. A positive improvement indicates a lower mortality level in the next year. For convenience of calculation, we also use the term improvement ratio, defined by:

$$Improvement\ ratio(x,t) = \frac{q(x,t)}{q(x,t-1)},$$

thus the mortality at the next year can be calculated by mortality rate at current year multiplied with the improvement ratio at the next year for each age.

LPP/BVG 2010 has higher ages compared with our data. The limiting age (the first age mortality rate reaches 1) is 107 for males and 113 for females. Therefore to apply our improvement tables, it is necessary to extrapolate them to higher ages.

One common approach is to taper the improvement down to 0 linearly at selected limiting age for each year, for example see CMIB Report no. 10 and no. 17 (1990, 1999) [30-31]. Here the selected limiting ages for male are set to be 107 for the first two projected years, then 108 for the next three years, and so on, depending on the last projected year. In our case the limiting age is 112 for males and 118 for females when projecting to 2038. We then apply 5000 extended improvement tables to LPP/BVG 2010 with the mortality curve in 2013 selected as our starting curve. The median of the improvement tables are illustrated in Figure 4.10. We can clearly see the effect by some cohorts like the jump between cohort 1939 and cohort 1940.

To compare the results, we use an aggregated mortality indicator such as period life expectancy, defined by:

$$Period\ life\ expectancy(x,t) = \big(1 - q(x,t)\big) + \big(1 - q(x,t)\big)\big(1 - q(x+1,t)\big) + \cdots + 0.5$$

Notice that LPP/BVG 2010 only forecasts to year 2030, so our forecasting range is selected as 2013 to 2030. The plots for life expectancy at age 65 and 85 are shown in Figure 4.11, where the red solid curve is calculated with LPP/BVG 2010 and blue dash curves are our simulated values at selected quantiles (5%, 25%, 50%, 75%, 95%). The "bend" in the median curve at age 85 in male population is due to a combination of a jump in estimated cohort effects and the taper used for improvement (the estimated effect of cohort 1939 is much lower than the cohort 1940, resulting that the improvement ratio is larger

than 1 at age 89 in year 2029. Tapering this improvement will cause the mortality rates in 2029 beyond age 89 higher than that in 2028 and lower the life expectancy.

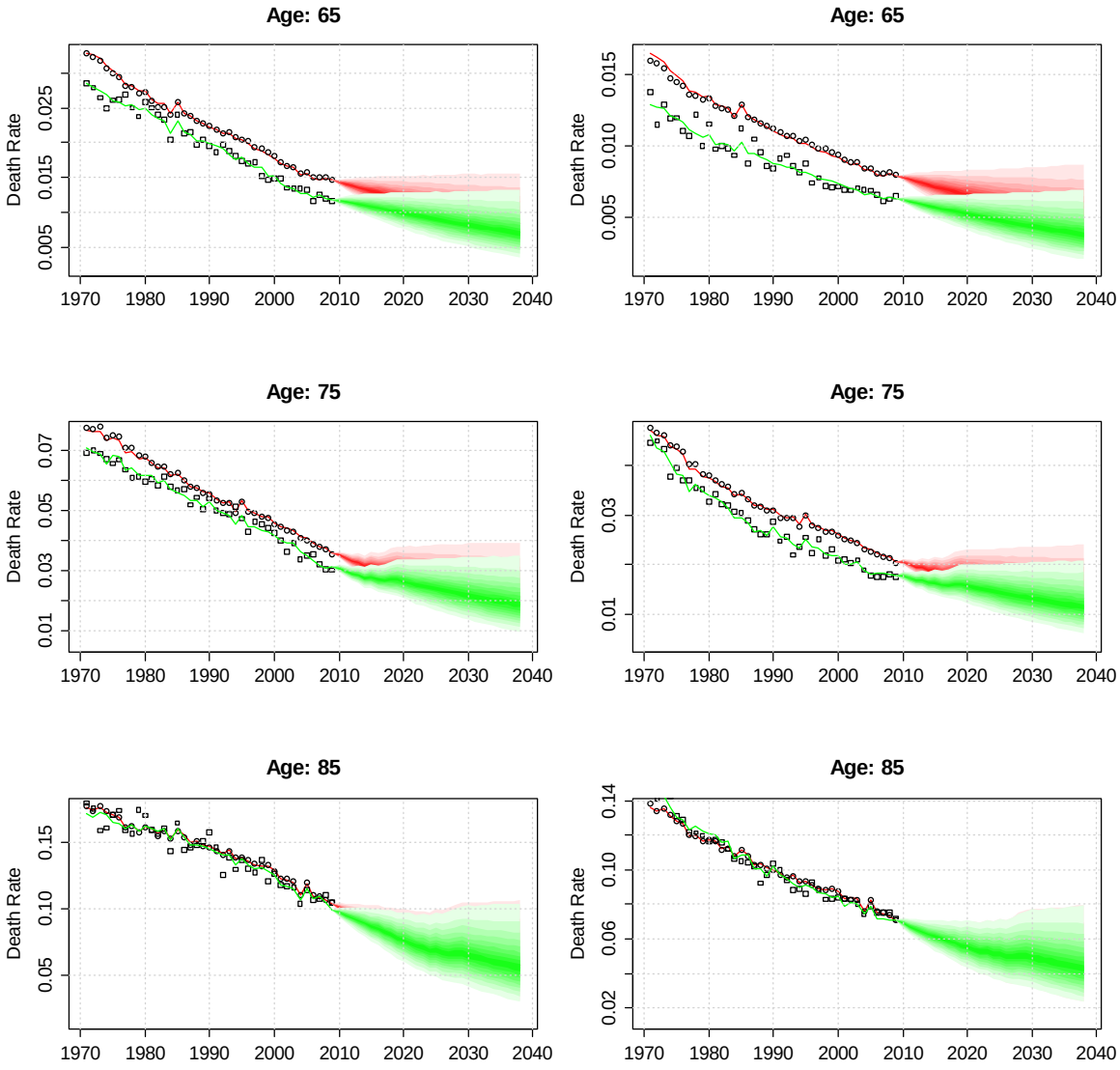


Figure 4.8: 5000 simulated death rate projections and raw data for Switzerland (green, square) and the international trend (red, circle) at age 65, 75 and 85. Left: Male population; Right: Female population.

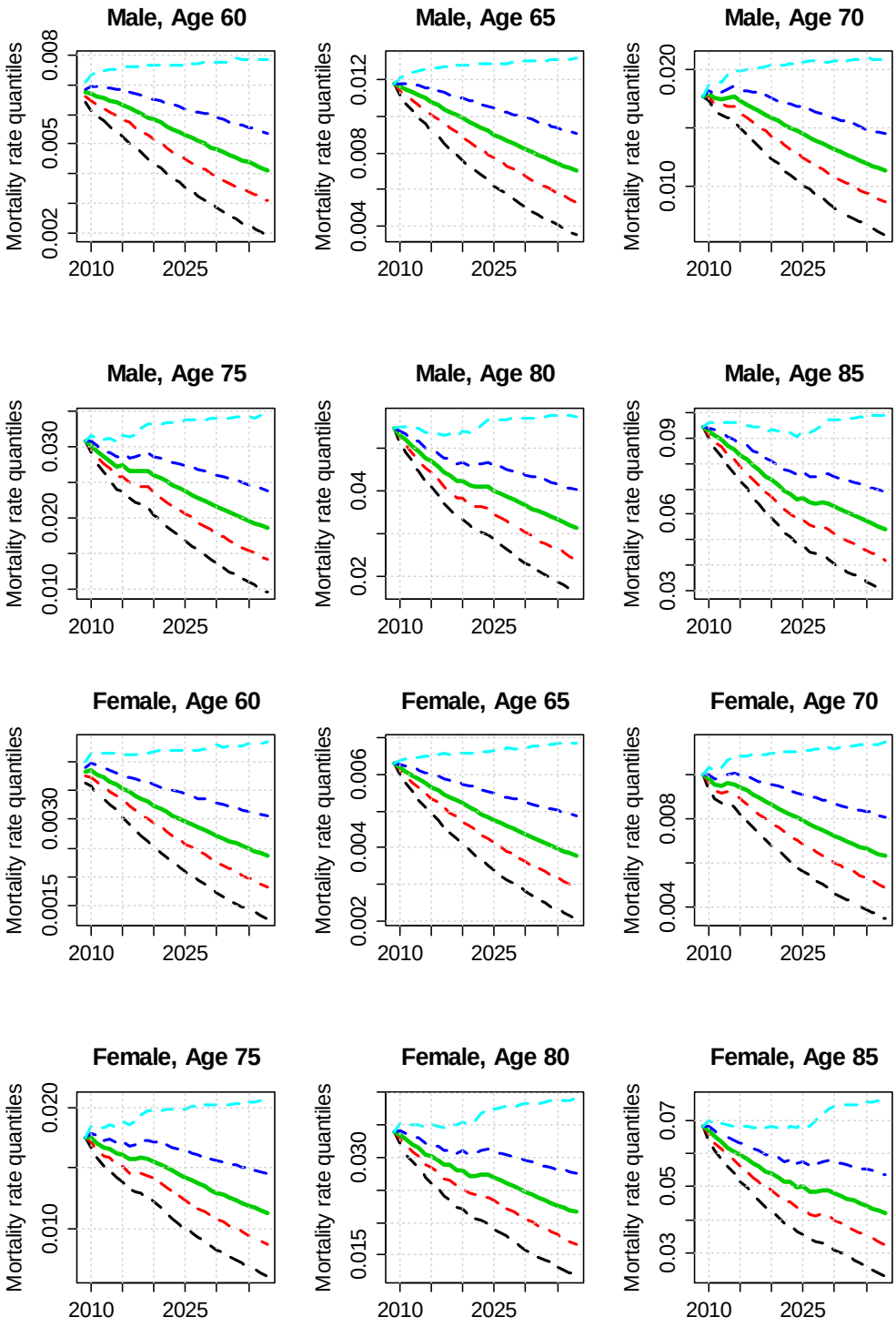
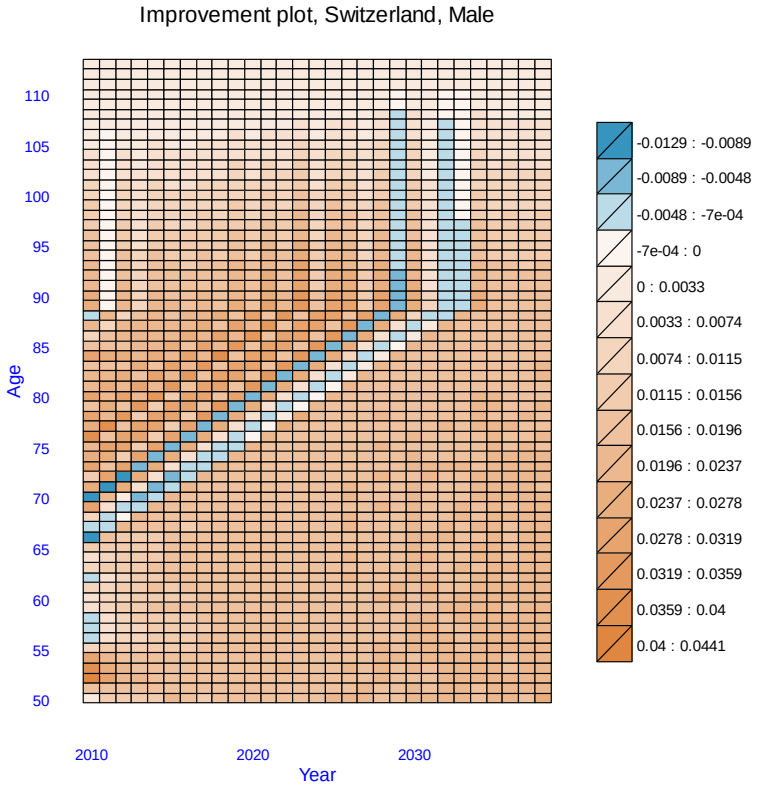


Figure 4.9: Quantiles (5%, 25%, 50%, 75%, 95%) of the simulated Switzerland mortality rate projections at age 60, 65, 70, 75, 80 and 85. Up: Male population; Down: Female population.



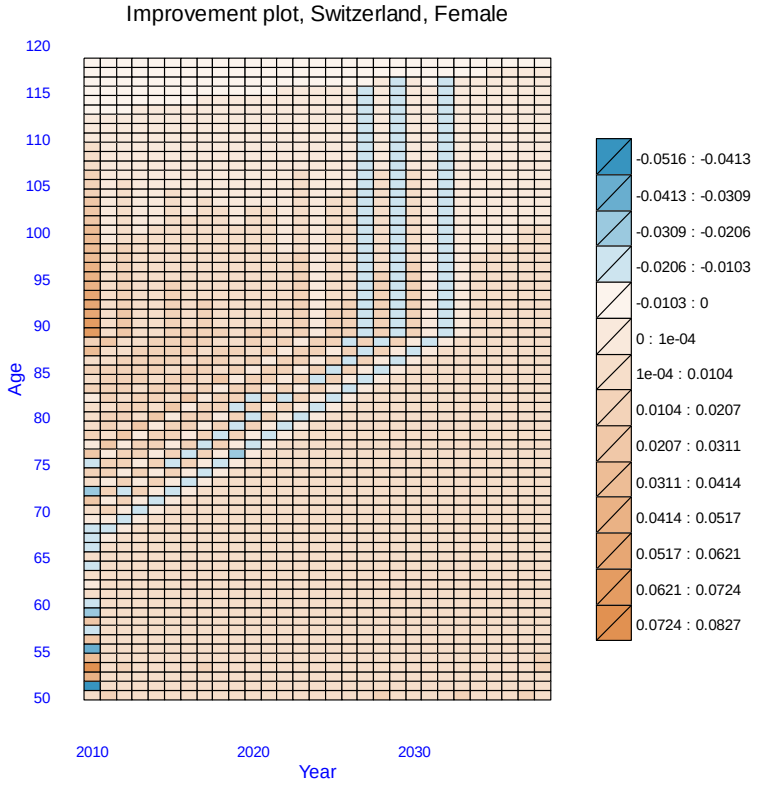


Figure 4.10: Median of the extended 5000 simulated improvement tables. Above: Male population; below: Female population.

4.2.2 Distribution of the Life expectancy (based on BVG 2010 mortality curve for calendar year 2013)

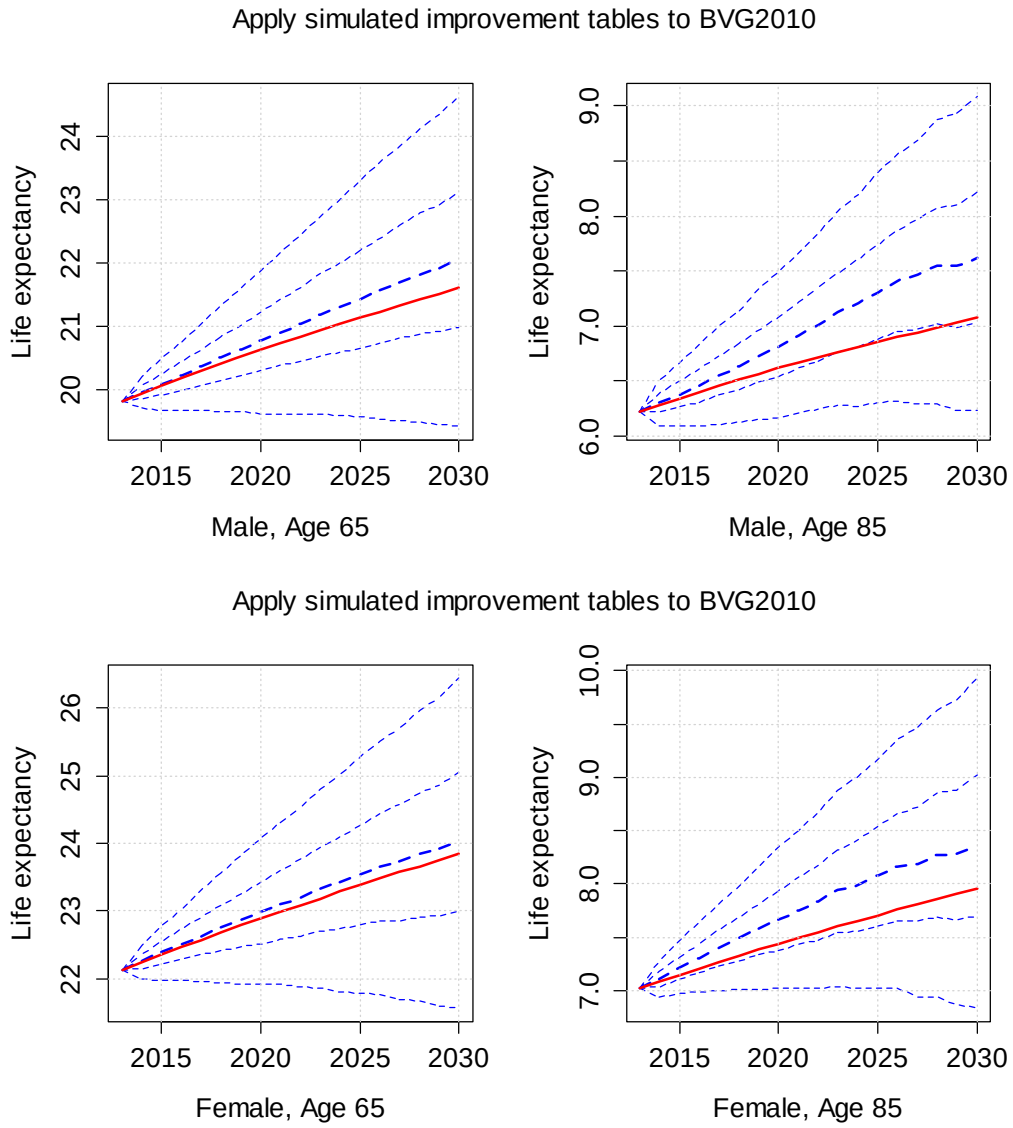


Figure 4.11: Period life expectancy comparison with LPP/BVG 2010 at age 65 and 85. Blue dashed curves: Percentiles (5%, 25%, 50%, 75%, 95%) for simulated life expectancy. Red solid curves: LPP/BVG 2010. Above: Male population; Below: Female population.

The life expectancy at 65 is a very important parameter in the first and second pillars in Switzerland as it is the official retirement age. The median of our model is consistent in pattern with the LPP/BVG 2010 forecasts up to 2030, but the level is higher. Its 95th percentile shows the magnitude of uncertainty and indicates that in 2030 the life expectancy at 65 can exceed its median by ca. 3 years.
The curve LPP/BVG 2010 life expectancy at 85 is very close to the 25th percentile of the life expectancy at 85 calculated with our model. It shows that LPP/BVG 2010 appears to underestimate the trend in mortality improvements for this age group. This is probably due to the fact that LPP/BVG 2010 is based on the mortality statistics of 14 big autonomous pension funds (ca. 2 mio exposed members) during the last 5-7 years and its reference population is small compared to ours. The uncertainty intrinsic to the longevity risks is well pronounced in 95th percentile of the life expectancy at 85 – it is expected to amount to ca. 2 years in 2030 compared to LPP/BVG 2010.

5 Conclusion and Discussion

In this paper we introduce a new stochastic mortality model for small populations and apply it to Switzerland mortality data (HMD) for estimating the trend risk for the whole Swiss population in relation to the big reference population. In the second step we applied mortality improvements calculated in our model to the LPP/BVG 2010 generational mortality curve for the year 2013; so addressing the difference in mortality rates between the overall population and the pension fund-specific population (the level risk). The results are compared with current official mortality tables used by pension funds in Switzerland (LPP/BVG 2010). The proposed model shows that the official table likely underestimates longevity risk especially for members older than 80 and the mortality improvement could be larger than anticipated in the LPP/BVG 2010. Pension funds might need to reserve more and pay closer attention to better assessing of the uncertainly intrinsic to the nature of longevity risks.

The cohort effect is found to have a large impact in our model, especially when extending the improvement to the limiting age, more advanced taper is needed to address problem caused by large jump in cohort effects. It is worth exploring in future work whether the cohort effect is indeed transitory or is likely to emerge as a trend. Transitory factor could be largely ignored while trend development will significantly affect pension liabilities.

Apart from modelling, one should pay attention to the selection of the reference population. Here we simply pooled the 13 countries' data together, however, we have noticed that population of Japan and USA is obviously quite large compared to the West European countries. Weighted approach is a natural generalization and could be considered as a counterpart as well as including or excluding countries from the reference population for gaining deeper insights into the nature of mortality convergence. Besides that, building a "world frontier" of the lowest mortality could also be an interesting reference.

It is worth paying closer attention to the extrapolation models to get stochastic mortality models for practical purposes for private pension funds. Further, it should be kept in mind that these models do not deal with the assumptions for sudden jump of mortality, for example, jump due to medical advance or jump due to unexpected epidemics. The only information taken into account is from the historical data.

At this stage of our project, we are more interested in long-term impact of longevity risks and their trends as actuarial provisions can be adjusted based on the mortality trends in the reference populations every year. It should be kept in mind that longevity risks emerge over time. On a short-term basis, the impact of longevity risk on a pension plan is never going to be very high. However for a typical pension fund, over the long term horizon the longevity risk impact could be of a similar magnitude as interest rate risk and equity risk.

The model we developed is not limited to the demographic situation of Switzerland. With some modifications it is possible to apply our model to other countries or regions. For example, populations in a group in the model definition could be actually a population of a big country split into different provinces or population groups based on specific criteria, or even a group of countries in a region. This will help address for example a region specific level of mortality in a big country.

6 Acknowledgment

We are indebted to Jonathan Garner, Towers Watson UK, and to Gaobo Pang, Towers Watson US, for their feedback during the project phase.

Reference

1. Szenarien zur Bevölkerungsentwicklung der Schweiz 2010-2060, Bundesamt für Statistik BFS, Neuchatel, 2010, Nr. 201-1000, URL: http://www.bfs.admin.ch

2. Gaille, S. (2011) "Forecasting Mortality: When Academia Meets Practice." *European Actuarial Journal* 2(1): 49-76.

3. Abt, B. "Longevity Risk in Switzerland. A stochastic modelling approach", 2012, Master Thesis, HEC Lausanne.

4. Towers Watson Global Pension Asset Survey, URL: http://www.newunionism.net/library/working%20life/Watson%20Wyatt%20-%20Global%20Pension%20Assets%20Study%20-%202007.pdf

5. Die Zukunft der Langlebigkeit in der Schweiz, Bundesamt für Statistik BFS, Neuchatel, 2009, Nr. 1043-0901, URL: http:/www.bfs.admin.ch

6. Pitacco, E., Denuit, M., Haberman, S. and Olivieri, A. (2009) *Modelling Longevity Dynamics for Pensions and Annuity Business*. Oxford University Press, New York.

7. Lee, R.D. and Carter, L.R. (1992) "Modeling and Forecasting U.S. Mortality." *Journal of the American Statistical Association* 8: 659-651.

8. Renshaw, A.E. and Haberman, S. (2006) "A Cohort-based Extension to the Lee-Carter Model for Mortality Reduction Factors." *Insurance: Mathematics and Economics* 38: 556-570.

9. Currie, I.D. (2006) "Smoothing and Forecasting Mortality Rates with P-splines." Talk at the Institute of Actuaries, June 2006. URL: http://www.ma.hw.ac.uk/~iain/research/talks.html

10. Cairns, A.J.G., Blake, D. and Dowd, K. (2006) "A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration." *Journal of Risk & Insurance* 73(4): 687-718.

11. Plat, R. (2009) "Stochastic Portfolio Specific Mortality and Qualification of Mortality Basis Risk." *Insurance: Mathematics and Economics* 45: 123-132.

12. Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D. and Khalaf-Allah, M. (2011a) "Mortality Density Forecasts: An Analysis of Six Stochastic Mortality Models." *Insurance: Mathematics and Economics* 48(3): 355-367.

13. Hunt, A. and Blake, D. (2013) "A General Procedure for Constructing Mortality Models", (Working Paper No. 1301). Pensions Institute, URL: http://pensions-institute.org/workingpapers/wp1301.pdf

14. Hyndman, R. and Shahidullah, M. (2007) "Robust Forecasting of Mortality and Fertility Rates: A Functional Data Approach." *Computational Statistics & Data Analysis* 51(10): 4942-4956.

15. Czado, C., Delwarde, A. and Denuit, M. (2005) "Bayesian Poisson Log Bi-linear Mortality Projections." *Insurance: Mathematics and Economics* 36: 260-284.

16. Luoma, A., Puustelli, A. and Koskinen, L. (2012) "A Bayesian Smoothing Spline Method for Mortality Modelling." *Annals of Actuarial Science* 6(2): 284-306.

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

18. Biatat V.D. and Currie, I.D. (2010) "Joint Models for Classification and Comparison of Mortality in Different Countries." *Proceedings of the 25th International Workshop of Statistical Modelling, Glasgow*, 89-94.

19. Dowd, K., Blake, D., Cairns, A.J.G., Coughlan, G.D., and Khalaf-Allah, M. (2011b) "A gravity model of mortality rates for two related populations". *North American Actuarial Journal*, 15: 334-356

20. Jarner, S. F. and Kryger, E. M. (2011) "Modelling Adult Mortality in Small Populations: The SAINT Model." *Astin Bulletin* 41(2): 377-418.

21. Cairns, A.J.G, Blake, D., Dowd, K., Coughlan, G.D., and Khalaf-Allah, M. (2011) "Bayesian Stochastic Mortality Modelling for Two Populations." *Astin Bulletin* 41(1): 29-59.

22. D'Albis, H., Esso, L.J. and Arolas, H.P. "Mortality Convergence Across High-Income Countries: An Econometric Approach", Documents de Travail du Centre d'Economie de la Sorbonne - 2012.76*Human Mortality Database*. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany), Data download at November 2012. URL: http://www.mortality.org

23. Cairns, A.J.G., Blake, D., Coughlan, K., Epstein, G.D., Ong, A. and Balevich, I. (2009) "A Quantitative Comparison of Stochastic Mortality Models Using Data from England & Wales and the United States." *North American Actuarial Journal* 13: 1-35.

24. Brouhns, N., Denuit, M. and Vermunt, J. (2002) "A Poisson Log-bilinear Regression Approach to the Construction of Projected Lifetables." *Insurance: Mathematics and Economics* 31(3): 373-393.

25. Tuljapurkar, S., Li, N. and Boe, C. (2000) "A Universal Pattern of Mortality Decline in the G7 Countries." *Nature* 405: 789-792.

26. Brouhns, N., Denuit, M. and Keilegom, I. (2005) "Bootstrapping the Poisson Log-bilinear Model for Mortality Projection." *Scandinavian Actuarial Journal* 3: 212-224.

27. Koissi, M., Shapiro, A. and Hognas, G. (2006) "Evaluating and Extending the Lee–Carter Model for Mortality Forecasting: Bootstrap Confidence Interval." *Insurance: Mathematics and Economics* 38(1): 1-20.

28. Renshaw, A.E. and Haberman, S. (2008) "On Simulation-based Approaches to Risk Measurement in Mortality with Specific Reference to Poisson Lee–Carter Modelling." *Insurance: Mathematics and Economics* 42(2): 797-816.

29. CMIB (1990). Reports no. 10. Continuous Mortality Investigation Bureau, Institute of Actuaries and Faculty of Actuaries.

30. CMIB (1999). Reports no. 17. Continuous Mortality Investigation Bureau, Institute of Actuaries and Faculty of Actuaries.

31. Booth, H., Maindonald, J., and Smith, L. (2002). "Applying Lee-Carter under Conditions of Variable Mortality Decline." *Population Studies* 56(3): 325-336.