

RA **Economics and institutional change**

A continuous-time stochastic model for the mortality surface of multiple populations

Peter Jevtić
Luca Regis

Research Area

Economics and institutional change

A continuous-time stochastic model for the mortality surface of multiple populations

Peter Jevtić

Department of Mathematics and statistics, McMaster University, Canada

Luca Regis

IMT School for Advanced Studies Lucca

ISSN 2279-6894

IMT LUCCA EIC WORKING PAPER SERIES #03/2016

© IMT School for Advanced Studies Lucca

Piazza San Ponziano 6, 55100 Lucca

A CONTINUOUS-TIME STOCHASTIC MODEL FOR THE MORTALITY SURFACE OF MULTIPLE POPULATIONS

PETAR JEVTIĆ AND LUCA REGIS

ABSTRACT. We formulate, study and calibrate a continuous-time model for the joint evolution of the mortality surface of multiple populations. We model the mortality intensity by age and population as a mixture of stochastic latent factors, that can be either population-specific or common to all populations. These factors are described by affine time-(in)homogenous stochastic processes. Traditional, deterministic mortality laws can be extended to multi-population stochastic counterparts within our framework. We detail the calibration procedure when factors are Gaussian, using centralized data-fusion Kalman filter. We provide an application based on the mortality of UK males and females. Although parsimonious, the specification we calibrate provides a good fit of the observed mortality surface (ages 0-99) of both sexes between 1960 and 2013.

Keywords: multi-population mortality, mortality surface, continuous-time stochastic mortality, Kalman filter estimation, centralized data fusion.

JEL classification: C13, C38, G22, J11.

1. INTRODUCTION

The joint description of the mortality dynamics of different groups (populations) of individuals has recently attracted particular attention. Indeed, capturing the common trends in the evolution of mortality rates of different populations is relevant at many levels, from a broad perspective, regarding different national or regional populations, to more particular ones, regarding different cohorts or income or socio-economic groups within the portfolio of an insurer. From a financial perspective, the growing market for longevity derivatives poses challenges to insurers and reinsurers in terms of basis risk, originating from the differences in the dynamics of the mortality of insurer's policyholders and the standardized indexes underlying the mortality-linked contracts. The modelling of the joint dynamics of the portfolio and the reference population has obvious consequences on the pricing of products and on their hedging effectiveness evaluation (see Li and Hardy, 2011, for instance).

Date: July 8, 2016.

P. Jevtić: Department of Mathematics and Statistics, McMaster University, Canada
pjevtic@math.mcmaster.ca

L. Regis: AXES Research Unit, IMT School for Advanced Studies Lucca, Italy,
luca.regis@imtlucca.it.

The Authors would like to thank conference participants at the 19th IME Conference and SIMC 2015 conference as well as seminar participants at Waterloo University, Concordia University, HEC Lausanne and McMaster University for helpful comments. Luca Regis gratefully acknowledges financial support from the Crisis Lab project funded by the Italian Ministry of Education.

As a consequence, a growing literature in actuarial science is tackling the issue of developing multi-population models for the evolution of mortality rates. This paper contributes to this stream of literature, proposing a general framework for the modelling of multi-population mortality dynamics in continuous time. A similar aim has been pursued previously in a discrete-time set up, adapting the seminal Lee and Carter (1992) model. Li and Lee (2005) first extended the Lee-Carter model to multi-population mortality modelling approach, imposing a common trend to different populations. Other recent contributions explored the joint modelling of different national populations (Antonio et al., 2015) or sub-populations (Villegas and Haberman, 2014), either for hedging purposes (Coughlan et al., 2011), or for forecasting (Danesi et al., 2015)

We are the first, up to our knowledge, to develop a general multi-population model in a continuous-time setting. Continuous-time mortality models have gained increasing popularity in the last decade, starting from the paper by Milevsky and Promislow (2001), who first applied the continuous-time Cox modelling approach typical of credit risk (see Lando, 1998) to the survivorship of individuals. Subsequent papers developed this idea further (see Dahl, 2004 for instance), proposing appropriate models that fit the mortality pattern of single (Luciano and Vigna, 2005) or multiple cohorts (Jevtić et al., 2013, Blackburn and Sherris, 2013) within a population, and showing their convenience in pricing and hedging actuarial liabilities (Dahl and Möller, 2006, Luciano et al., 2015) under the usual martingale pricing approach. Focusing, as in Biffis (2005), on a particular class of processes, the affine one, which is analytically tractable and provides closed-form expressions for the survival probabilities in many instances, we propose a general model that can be used to describe the mortality dynamics of populations and sub-populations within them. We show that the framework can accommodate stochastic and multi-population extensions of some traditional models well-known to actuaries, i.e. the mortality laws. Such specifications, once properly selected depending on the database and calibrated, are suitable for pricing and hedging purposes and can be very parsimonious in terms of the number of parameters to be estimated. This is a particularly relevant feature, given the need for practical tools that can help fostering the market for longevity-linked securities.

Our general framework encompasses several models previously proposed in the literature. The uncertainty in the model is described by a vector of state processes, which represent the random changes in mortality patterns. This kind of approach has been used to model jointly the dynamics of multiple ages or generations belonging to the same population by Jevtić et al. (2013) and Blackburn and Sherris (2013). Our model extends that setup, by considering a set of populations and sub-populations within them, who are affected differently by the state processes. Moreover, each age responds to these changes in a specific way. In this sense, our paper is closely related to Schrager (2006), who proposed an affine model under general functional dependences between ages and their force of mortality, describing the dynamics of the whole mortality surface (all ages, in principle) of a population. The paper revisited some well-known mortality laws, extending them to a stochastic setting via the use of latent factors. We extend that contribution in several directions.

Firstly, we allow for the presence of many populations sharing the same latent factors, from which they are affected in different ways and to which they respond differently. Secondly, these populations may be divided in turn into sub-populations. Thirdly, our paper allows to account for age, period and cohort effects simultaneously in the mortality dynamics of multiple populations.

We describe a calibration procedure for our model, based on sequential Kalman filtering, when factors are described by Gaussian processes. To minimize computational instability, we apply a sequential Kalman filter algorithm as in Simon (2006). We provide an empirical application, showing how a rather simple specification of our model, related to the well-known Thiele law, fits the whole mortality surface (ages 0-100) of the populations of UK males and females. Despite being very parsimonious, this model provides a very good fit to the observed data.

The outline of the paper is the following. Section 2 presents our general framework and connects it to previous literature. Section 3 provides some particular specifications of our model, proposing stochastic extensions of some traditional actuarial laws for the force of mortality and an Age-Period-Cohort continuous-time model for the whole mortality surface of multiple populations. Section 4 presents an application of our model to the mortality of UK males and females, detailing the calibration procedure based on multi-dimensional Kalman Filter. Section 5 concludes. The Appendix contains the Proof of our main theorem, further details on the calibration procedure and on the properties of the multidimensional process we use in our application.

2. A GENERAL MULTI-POPULATION FRAMEWORK

In this section we present a model which, in a parsimonious framework, describes the joint mortality intensity dynamics of multiple populations. These populations are composed of individuals who are heterogeneous by age groups. The different populations can identify citizens from different countries, different sub-groups within a general country population (such as males and females, or employed vs. unemployed), different types of policyholders in the portfolio of an insurer.

2.1. The model. We develop our model following the continuous-time stochastic mortality approach initiated by Milevsky and Promislow (2001), in which the death process of homogeneous individuals is described by means of a Cox or doubly stochastic process.

To properly describe our setting, let us introduce a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, where \mathbb{P} denotes the real-world (historical) probability measure. The filtration $\{\mathcal{F}_t : 0 \leq t \leq T\}$ satisfies the usual properties of right-continuity and completeness. In addition, $\mathcal{F}_t = \mathcal{H}_t \vee \mathcal{G}_t$, i.e. the filtration includes all information from the historical life tables, contained in sub-filtration \mathcal{G}_t , and whether the individual is alive or not, whose information is contained in the sub-filtration \mathcal{H}_t .

In the most general setting, let us assume there exist J populations of interest, enumerated with an index $j \in \mathbf{J} = \{1, \dots, J\}$. Each population may contain K_j constituent sub-populations, enumerated by the index $k \in \{1, \dots, K_j\}$. The range of k depends on population j and may thus be different depending on the

population. With a pair (j, k) we uniquely identify population j and, within it, the sub-population k . Our object of interest is the modelling of the evolution of the mortality intensities of sets of ages in different populations. To this end, let us define as $\mu_{x+t}^{j,k}(t)$ the set of predictable processes that represent the mortality intensity at calendar time t of an age $x+t$ belonging to the cohort aged x at time t . Age x belongs to the set $\mathbf{X}^{j,k} = \{x_{\min}, \dots, x_{\max}\}$, which includes the (integer) ages in population j and sub-population k , ordered from the smallest, x_{\min} , to the biggest age x_{\max} , and to a certain population j and sub-population k . Such intensity drives the likelihood of death of the individual, being the intensity of the doubly stochastic process. The first jump time of such process determines the time of death of the individual.

The mortality intensity of members of population j depends on a population-specific state process vector, $\mathbf{Y}^j(t) := [Y_1^j(t), \dots, Y_N^j(t)]^T$, $j \in \mathbf{J}$.¹ Moreover, we consider that all J populations are affected by the same innovations, described by a vector \mathbf{Z} of orthogonal random noises. These innovations drive the randomness of the state processes of the different populations and, as a consequence, affect their mortality intensities. Figure 1 graphically illustrates our model.

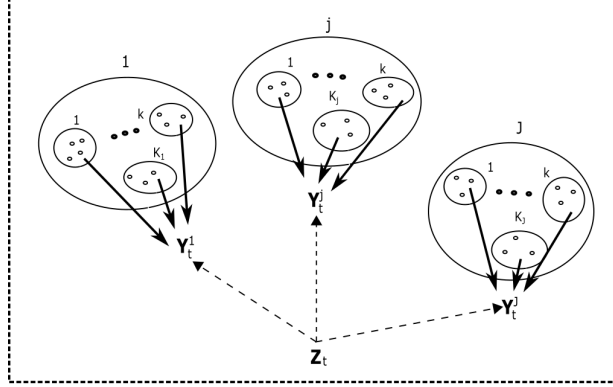


FIGURE 1. Conceptual view of the framework. The figure shows how the population-specific sources of uncertainty affect sub-populations within the J populations and how they are related to the vector \mathbf{Z}_t of common, orthogonal noises.

Given this setup, the mortality intensities of the different ages in population j and sub-population k respond differently to changes to factors \mathbf{Y}^j affecting population j . More formally, the mortality intensity at time t of an age $x+t$, belonging to population j and sub-population k , is described by a functional of the vector of state processes \mathbf{Y}^j

$$(1) \quad \mu_{x+t}^{j,k}(t; x) := R^{j,k}(\mathbf{Y}^j(t), x+t, t, x).$$

The functional $R^{j,k} : \mathbb{R}^3 \rightarrow \mathbb{R}$ is the response function of a life age $x+t$ at time t belonging to cohort x , population j and sub-population k to changes of underlying

¹In principle, we need not have all groups of populations to have state process vectors of same size, i.e. N might be indexed by j .

factors. It depends on both the age, the cohort and calendar time, thus representing an age-period-cohort response to underlying factors (see Figure 2).

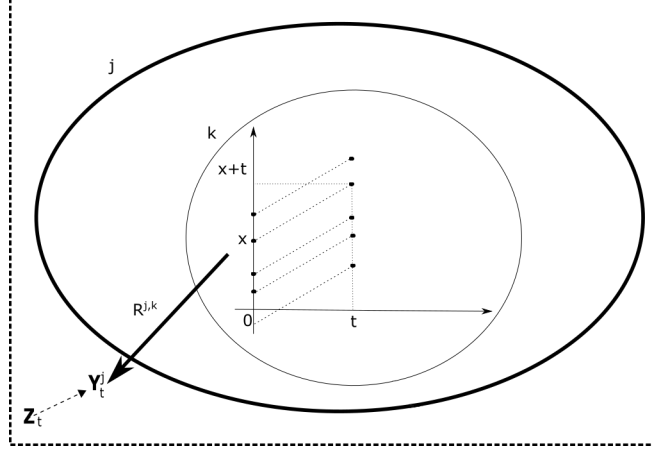


FIGURE 2. Conceptual view of the framework. The figure shows how the mortality dynamics of homogeneous individuals in the (j,k) -th population/sub-population, characterized by initial age x evolve in time t . The response function $R^{j,k}$ describes how the latent factors interact to drive the cohort-specific mortality intensities.

Before proceeding, let us comment on the generality of the framework introduced above. The more comprehensive setting, which we described above, consists of J populations and K sub-populations. We call this the *JpKs* class. The K sub-populations within a group share the same driving factors, but differ in response functions. It is worth describing two distinct sub-cases. The first sub-case consists of only one population i.e. $j \in \{1 = J\}$ and K sub-populations. There would be only one (common) vector of state processes and K different response functions for the groups considered. This situation belongs to the *1 population K sub-populations (1pKs)* class. For instance, the problem of jointly modeling the mortality intensities of individuals belonging to different socio-economic groups in one population falls in this class. The second sub-case consists of J populations i.e. $j \in \{1, \dots, J\}$, each composed of only one sub-population. In this case, every different group has its own specific state process vector and response function. We refer to this setting as *J populations 1 sub-population (Jp1s)* class. The problem of jointly modeling the mortality of the total population of a set of countries is a good example of an application that falls in the *Jp1s* class.

Now, we are ready to specify further the function $R^{j,k}(\cdot, x + t, t, x)$, introducing a set of (population, sub-population)-specific functions $g_0^{j,k}(x + t, t, x)$ and

$$\mathbf{g}^{j,k}(x + t, t, x) := [g_1^{j,k}(x + t, t, x), \dots, g_N^{j,k}(x + t, t, x)]',$$

where $g_i^{j,k} : \mathbb{R}^3 \rightarrow \mathbb{R}, i \in \{0, 1, \dots, N\}$ are age $x + t$ time t and cohort x dependent functions for the j -th group and k -th population,

$$\begin{aligned} R^{j,k}(\mathbf{Y}^j(s); x + t, t, x) &:= \\ &= g_0^{j,k}(x + t, t, x) + \langle g_1^{j,k}(x + t, t, x), \dots, g_N^{j,k}(x + t, t, x) \rangle \cdot \mathbf{Y}^j(t) \\ &= g_0^{j,k}(x + t, t, x) + \mathbf{g}^{j,k}(x + t, t, x) \cdot \mathbf{Y}^j(t). \end{aligned}$$

Our objective is to develop expressions for the time t -survival probability of the ages in different groups. To enhance analytical tractability, we assume that the dynamics of the J state processes are described by continuous-time stochastic processes, belonging to the affine class characterized by its canonical form given in Filipović (2009) (pages 146-150). In this setting, the Markov process $\mathbf{Y}^j(t)$ drives the dynamics described by the following system of stochastic differential equations

$$d\mathbf{Y}^j(t) = \mathbf{A}^j(\boldsymbol{\theta}^j - \mathbf{Y}^j(t))dt + \boldsymbol{\Sigma}^j \sqrt{V_t^j} d\mathbf{W}^j(t),$$

where $\mathbf{A}^j = \text{diag}[a_1^j, \dots, a_N^j] \in \mathbb{R}_+^{N \times N}$ is a diagonal matrix, $\boldsymbol{\theta}^j \in \mathbb{R}^N$ is column vector, $\boldsymbol{\Sigma}^j = \text{diag}[\sigma_1^j, \dots, \sigma_N^j] \in \mathbb{R}_+^{N \times N}$ is a diagonal matrix and

$$V_t^j = V_t^j(\mathbf{Y}^j(t)) = \text{diag}[\gamma_1^j + \boldsymbol{\delta}_1^j \cdot \mathbf{Y}^j(t), \dots, \gamma_N^j + \boldsymbol{\delta}_N^j \cdot \mathbf{Y}^j(t)],$$

with $\boldsymbol{\delta}_i^j = [0, \dots, \underbrace{\delta_i^j}_{i\text{-th}}, \dots, 0]$ for $i \in \{1, \dots, N\}$. The stochastic process $\mathbf{W}^j(t)$ is a

vector of correlated Brownian Motions, and $\boldsymbol{\rho}^j$ denotes its instantaneous correlation matrix. Thus, using Cholesky's decomposition, the state process vector can be transformed as

$$d\mathbf{Y}^j(t) = \mathbf{A}^j(\boldsymbol{\theta}^j - \mathbf{Y}^j(t))dt + \boldsymbol{\Sigma}^j \sqrt{V_t^j} \mathbf{H}^j d\mathbf{Z}(t),$$

where \mathbf{H}^j is a population-specific matrix such that $\mathbf{H}^j(\mathbf{H}^j)' = \boldsymbol{\rho}^j$ and $\mathbf{Z}(t)$ is a standard Brownian Motion, which represents the baseline innovations that are driving the system.

2.2. Connections with related literature. The framework introduced is very general and relates to several existing models in the continuous-time stochastic mortality literature. As a very special case, single-cohort models such as for example proposed by Biffis (2005) and Luciano and Vigna, 2005 can be encompassed, when we consider 1 population, 1 sub-population and $\mathbf{X} = \{x\}$. The models considered in Schrager (2006) can also be recovered within this framework, as 1p1s models. The two-populations, single-cohort models in Dahl et al. (2008), Wong et al. (2014) and De Rosa et al. (2016) can also be seen as special cases of 1p2s models in our framework. Each sub-population contains again only one age group. Indeed, in these cases the state process vector \mathbf{Y} is two-dimensional and shared by the two populations, whose correlated mortality dynamics is described by means of two CIR-type processes involving this state process.

The 1pKs class includes the recently proposed models that jointly describe the mortality of a set of cohorts belonging to the same population. The multiple cohorts correspond to our sub-populations, whose mortality dynamics depends on n (in Jevtić et al., 2013) or 2 or 3 factors (Blackburn and Sherris, 2013) that are linked to a unique vector of state processes. The processes in these papers are of the

Ornstein-Uhlenbeck (either mean-reverting or not) type. The response functions have $g_0 = 0$ and $\mathbf{g} = \mathbf{1}$.

2.3. Survival Probabilities. Our affine framework, as introduced above, allows us to compute the expectation of functionals of the state processes in closed form. The survival probability ${}_{T-t}p_{x+t}^{j,k}(t, T)$, which represents the probability that at calendar time t an age $x + t$ belonging to population (j, k) will survive until time T , is indeed a functional of $\mu_{x+t}^{j,k}(t; x)$, since

$${}_{T-t}p_{x+t}^{j,k}(t, T) := \mathbb{E}^{\mathbb{P}}[\mathbb{1}_{\{\tau_x > T\}} \mid \mathcal{H}_t \cup \{\tau_x > t\}] = \mathbb{E}^{\mathbb{P}}[e^{-\int_t^T \mu_{x+s}^{j,k}(s; x) ds} \mid \mathcal{G}_t].$$

Given our definition in equation (1), the survival probability can be written as

$${}_{T-t}p_{x+t}^{j,k}(t, T) = \mathbb{E}^{\mathbb{P}}[e^{-\int_t^T R^{j,k}(\mathbf{Y}^j(s), x+s, s, x) ds} \mid \mathcal{G}_t].$$

We have now all the ingredients to present our main theorem.

Theorem. *Consider the above setting, then, the time- t survival probability of an individual, initially aged x , belonging to population group j and population k , up to time T is*

$$(2) \quad {}_{T-t}p_{x+t}^{j,k}(t) = e^{\alpha^{j,k}(x, t, T) + \beta^{j,k}(x, t, T) \cdot \mathbf{Y}^j(t)}.$$

In specific cases, $\alpha^{j,k}(x, t, T)$ and $\beta^{j,k}(x, t, T)$ solve the following differential equations:

i) in the Ornstein-Uhlenbeck (OU) case where $\gamma_i = 1$ and $\delta_i = 0$

$$\begin{aligned} \dot{\alpha}^{j,k} &= g_0^{j,k}(x+t, t, x) - \mathbf{A}^j \boldsymbol{\theta}^j \cdot \beta^{j,k} - \frac{1}{2} (\beta^{j,k})^T \boldsymbol{\Sigma}^j \boldsymbol{\rho}^j (\boldsymbol{\Sigma}^j)' \beta^{j,k}, \\ \dot{\beta}^{j,k} &= \mathbf{g}^{j,k}(x+t, t, x) + (\mathbf{A}^j)' \beta^{j,k}. \end{aligned}$$

ii) in the square-root (SR) process case where $\gamma_i = 0$ and $\delta_i = 1$, in the 1pKs case where $j \in \{1 = J\}$ and when $\boldsymbol{\rho}^1 = I_{N \times N}$ is an identity matrix²

$$\begin{aligned} \dot{\alpha}^{1,k} &= g_0^{1,k}(x+t, t, x) - \mathbf{A}^1 \boldsymbol{\theta}^1 \cdot \beta^{1,k}, \\ \dot{\beta}^{1,k} &= \mathbf{g}^{1,k}(x+t, t, x) + (\mathbf{A}^1)' \beta^{1,k} - \left[\frac{1}{2} (\beta_1^{1,k})^2 \sigma_1^2, \dots, \frac{1}{2} (\beta_N^{1,k})^2 \sigma_N^2 \right]', \end{aligned}$$

where $\alpha^{j,k}(x, T, T) = 0$ and $\beta^{j,k}(x, T, T) = \mathbf{0}$.

Proof. See the Appendix A. □

In particular cases, the systems of ODEs appearing in the theorem above have (semi) closed-form solutions. We will consider some of these cases, motivated by the actuarial tradition, in the next section.

²Special care must be taken when SR processes are considered as building blocks of multi-population models. Our particular choice of $\mathbf{Y}(t)$ and the restrictions we place in our theorem, are necessary to ensure that our model lies within the canonical form of affine processes, as in Filipović (2009).

3. EXAMPLES OF MULTI-POPULATION MORTALITY SURFACE MODELS

In this section, extending the single-population setting developed in Schrager (2006), we consider some specifications of the response functions $R^{j,k}$, that allow us to extend deterministic models well-known in the actuarial tradition to a stochastic and multi-population framework. We consider several specifications, that vary both in the number of driving factors and in the form of the response functions. We first consider three specifications, that extend to a stochastic and multi-population setting the Gompertz and Makeham traditional mortality laws. We then propose a fourth specification, that further disentangles the age, period and cohort effects, introducing what, up to our knowledge, is the first example of an age-period-cohort continuous-time model. This specification also encompasses a multi-population extension of the classic Thiele law. These examples need not be exhaustive: appropriate specifications, though similar in spirit, should be tailored to the data considered. In what follows, we consider parameters $c^{j,k} > 1$ and $b_i^{j,k} > 0$.

3.1. Multi-population Gompertz law. Consider the instantaneous mortality intensity for population j sub-population k defined as

$$\mu_{x+t}^{j,k}(t) := (c^{j,k})^{x+t} Y_1^j(t).$$

Motivated by the Gompertz law, this specification has one factor per population and one parameter per sub-population. For high $x+t$, i.e. for higher ages, more weight is given to the factor. This specification implies thus a pure age-based response to changes in the population-specific factor.

In our framework, in this case, we have

$$\begin{aligned} \mathbf{Y}^j &= [Y_1^j], \quad g_0^j(x+t, t, x) := 0 \quad \text{and} \\ \mathbf{g}^{j,k}(x+t, t, x) &:= [(c^{j,k})^{x+t}]'. \end{aligned}$$

3.2. First Makeham multi-population law. In this case the instantaneous mortality intensity for population j sub-population k is defined as

$$\mu_{x+t}^{j,k}(t) := Y_1^j(t) + (c^{j,k})^{x+t} Y_2^j(t).$$

Motivated by the Makeham law, this specification has two factors per population and one parameter per sub-population. The first factor represents a base-line mortality level, which changes in calendar time equally for every age in every population. More weight is instead given to the second factor the higher $x+t$, i.e. for high ages, as in the previous specification 3.1. Again, this response function implies an age-based response to the calendar-time changes in the factors.

In our framework, in this case, we have

$$\begin{aligned} \mathbf{Y}^j &= [Y_1^j, Y_2^j]^T, \quad g_0^{j,k}(x+t, t, x) := 0 \quad \text{and} \\ \mathbf{g}^{j,k}(x+t, t, x) &:= [1, (c^{j,k})^{x+t}]'. \end{aligned}$$

3.3. Second Makeham multi-population law. In this case, the instantaneous mortality intensity for population j and sub-population k is defined as

$$\mu_{x+t}^{j,k}(t) := Y_1^j(t) + (x+t)Y_2^j(t) + (c^{j,k})^{x+t}Y_3^j(t).$$

Motivated by the second Makeham law, this specification has three factors per population and one parameter per sub-population. On top of the two factors considered in the previous case 3.2, for which the same interpretation holds, in this specification an additional factor is present. The weight of such factor is the age of the individual at time t , implying a linear increase of the weight with age.

In our framework, in this case, we have

$$\begin{aligned} \mathbf{Y}^j &= [Y_1^j, Y_2^j, Y_3^j]^T, \quad g_0^{j,k}(x+t, t, x) := 0 \quad \text{and} \\ \mathbf{g}^{j,k}(x+t, t, x) &:= [1, x+t, (c^{j,k})^{x+t}]'. \end{aligned}$$

3.4. A three-factor age-period-cohort multi-population specification. In this rather general specification, the instantaneous mortality intensity for population j and sub-population k is defined as

$$\mu_{x+t}^{j,k}(t) := Y_1^j(t)e^{-b_1^{j,k}(x-b_2^{j,k}t)^{\frac{1}{m^{j,k}}}} + Y_2^j(t)e^{-b_3^{j,k}(x-b_4^{j,k}t-\eta^{j,k})^2} + Y_3^j(t)e^{+b_5^{j,k}(x+t)}.$$

This specification has three factors per population and eight parameters per sub-population. This specification is connected to the classic Thiele law. The three factors are related to infant mortality (first factor), young adults mortality (second factor) and old ages mortality (third factor), respectively. Indeed, the weight of the first factor declines with age, making it most relevant for young ages. The parameter $m^{j,k}$ allows to better capture the infant mortality spike observed at age 0. The weight attached to the second factor aims at reproducing the so-called “mortality hump”, observed for young adults around the population, sub-population specific age $\eta^{j,k}$. The weight imposed on the third factor increases with age, making it relevant for older ages. This specification disentangles calendar-time and initial age (cohort) effects in the responses of the first two factors. The response to the third factor, instead, is age-depedent. Overall, this specification constitutes then the first, up to our knowledge, age-period-cohort multi-population mortality model cast in continuous-time.

In our framework, in this case, we have

$$\begin{aligned} \mathbf{Y}^j &= [Y_1^j, Y_2^j, Y_3^j]^T, \quad g_0^{j,k}(x+t, t, x) := 0 \quad \text{and} \\ \mathbf{g}^{j,k}(x+t, t, x) &:= [e^{-b_1^{j,k}(x-b_2^{j,k}t)^{\frac{1}{m^{j,k}}}}, e^{-b_3^{j,k}(x-b_4^{j,k}t-\eta^{j,k})^2}, e^{+b_5^{j,k}(x+t)}]'. \end{aligned}$$

4. APPLICATION

In this section, we provide an application of a model that falls within our framework to the fit of the mortality surface of UK males and females aged 0-99 in the period ranging from 1960 to 2013. Although belonging to the same national population, the two sub-groups show differences in the mortality levels at different ages, and slightly different longevity improvement trends, making an analysis of their joint dynamics interesting.

We tackle this problem, in our framework, as belonging to the $1p2s$ class, even though we remark that it can also be considered as a $2p1s$ problem, as pictured in Figure 3.

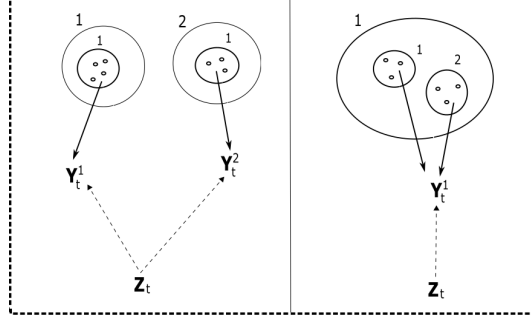


FIGURE 3. Two possible conceptualizations

In the first case (picture on the right), we consider males and females as being two sub-populations of the same population, driven by the same vector of state processes. Males and females are assumed to respond differently to changes in the same factors. In the second case (picture on the left), the two groups are considered as two different populations, with different driving factors (even though related to a unique “global” risk source) and different responses to them. In this Section, we present results from the first approach. We specify the model as a special case of the specification we presented in Section 3.4 and we consider Gaussian factors, whose dynamics follow a multidimensional OU process. We adopt a specification that has no period effect: $b_2^{j,k} = -1; b_4^{j,k} = -1$. This choice provides a very good fit of the whole mortality surface of the two genders and the widest possible age range (0-99), despite a remarkable parsimony and a reasonable computational effort.³ Due to the results in Theorem 2.3, it is possible and convenient to specify the model using a state-space linear representation and to proceed to its estimation using the Kalman Filter algorithm. The choice of the OU process allows us to estimate the parameters following a maximum likelihood procedure.

In the next sections, we first describe our calibration methodology, directing the reader to the Appendix for further details. Hence, we describe the state-space representation of our model and discuss its estimation using the implementation of the sequential Kalman filter algorithm. Then, we illustrate and discuss the results of our application.

4.1. Calibration Methodology. In this paper, we consider the case in which state processes are Gaussian, leaving the calibration of the SR process case for further research. In our OU case, we can directly apply the Kalman filtering algorithm,

³Obviously, the inclusion of calendar-time specific effects in the responses would help capturing some more features of the data. On the other side, time-inhomogeneity of the solutions to the Riccati equations associated to the model requires an extensive computational effort, resulting in a more time-consuming calibration procedure.

which returns an optimal linear estimate of the unobservable factors. We fit our model to one-year death probabilities, as in Schrager (2006).⁴

We apply the sequential Kalman filter algorithm, as described in the book by Simon (2006). This algorithm allows to treat the single elements of the observational vector as belonging to a sequence of individual observations, avoiding computational issues related to the inversion of large and sparse matrices. Moreover, we need to tackle the problem of how to filter the information coming from different sources, i.e. the data related to the different populations and sub-populations. This issue falls in the domain of data fusion and is a well-established problem in engineering (see for instance Mitchell (2007), Raol (2009) and Durrant-Whyte (2012)). As explained below, we consider the life tables of different groups as data collected by different channels of a unique sensor, adopting a “centralized data-fusion” approach. We now describe the two key steps of our calibration procedure, the representation of our model in a state-space linear form and the Kalman filter algorithm.

4.1.1. State-space Linear Representation. In order to apply the Kalman filtering algorithm to mortality data, that come at annual frequency, we discretize our problem and we represent it in state-space form. From now on, for simplicity, we consider integer ages and yearly-collected data, having in mind that this is the standard format of the mortality data collected in life tables. However, the case of non-integer ages and higher data frequency can be handled in this framework.

Given expression (2), the logarithms of survival probabilities are linear functions of the state process vector. Because survival probabilities take values between zero and one, we consider the opposite of this quantity, which is positive, as our measurement for each age. With a slight abuse of notation, we refer to the observation points as to $t \in \{1, \dots, T\}$, where $t = 1$ denotes the first year of observation, and denote with $\bar{q}^{j,k}(t)$ the elements of our observation vector for each couple (j, k) of populations and sub-populations:

$$\bar{q}^{j,k}(t) := -\ln(1 - \hat{q}^{j,k}(t)) = -\ln(\hat{p}^{j,k}(t)).$$

Indeed, given (2), we assume that each measurement comes from the following model

$$(3) \quad \bar{q}_{x+t}^{j,k}(t) = -\alpha^{j,k}(x, t, t+1) - \beta^{j,k}(x, t, t+1)' \mathbf{Y}^j(t) + \epsilon_x^{j,k}(t), \quad x \in \mathbf{X}^{j,k},$$

where $\epsilon_x^j(t)$ is the observation (or measurement) noise. We assume that this noise is Gaussian with mean zero and standard deviation proportional to the observation itself, i.e. $\text{Var}[\epsilon_x^{j,k}(t)] = [s^{j,k} \bar{q}_x^{j,k}(t)]^2$, $s^{j,k} \in \mathbb{R}$. This assumption is well justified by the fact that, due to the smaller cohort size, observations referring to old ages are usually noisier (in relative terms) than observations of younger ages. We further assume that measurement errors are uncorrelated across ages. The real parameter $s^{j,k}$ can in principle be different for every population and sub-population (to take into account, for instance, different data-collection methodologies) and belongs to the optimal parameter set.

⁴This choice allows us to exploit the information of the whole life table, but makes it more difficult to capture cohort effects in the data. An alternative approach, that we leave for further research, involves fitting death probabilities, or, equivalently, survival probabilities, with different maturities for the same cohort.

Given our model specification, we derive the expressions for $\alpha^{j,k}(x, t, t+1)$ and for the components of the vector $\beta^{j,k}(x, t, t+1)$. Hence,

$$\begin{aligned}\beta_3^{j,k}(x, t, t+1) &= \frac{e^{b_3^{j,k}(x+t)}}{b_3^{j,k} - a_3^j} \left(1 - e^{(b_3^{j,k} - a_3^j)}\right), \\ \dot{\beta}_1^{j,k} &= e^{-b_1^{j,k}(x+t)^{\frac{1}{m^{j,k}}}} + a_1^j \beta_1^{j,k}, \\ \dot{\beta}_2^{j,k} &= e^{-b_2^{j,k}(x+t-\eta^{j,k})^2} + a_2^j \beta_2^{j,k}, \\ \dot{\alpha}^{j,k} &= -\frac{1}{2}(\beta^{j,k})^T \Sigma^j \rho^j (\Sigma^j)^T \beta^{j,k}, \\ \alpha^{j,k}(x, T, T) &= 0 \text{ and } \beta^{j,k}(x, T, T) = \mathbf{0}.\end{aligned}$$

The only analytical solution concerns the third component of vector $\beta^{j,k}$, while numerical methods are required to obtain the other coefficients in our model (3) as solutions to the above ODEs.

For simplicity, we slightly transform our observations and define

$$\tilde{q}_{x+t}^{j,k}(t) := \tilde{q}_{x+t}^{j,k} + \alpha^{j,k}(x, t, t+1).$$

The crucial problem left to solve concerns how to treat the observations coming from the life tables of the different populations and sub-populations. We adopt a “centralized data fusion” approach (see Mitchell (2007), Raol (2009)). Technically speaking, it is akin to the problem of tracking a signal coming from a centralized multi-channel sensor(s). In practice, for all observation points, we group all the information coming from the different populations in a unique observational vector, as if they were coming from the same source. Still, heterogeneity in the observational noise relative to different groups is allowed for. Indeed, this approach is consistent with our modelling of the state vector \mathbf{Z} , that is assumed to be the same underlying source of uncertainty for all populations.

In a more compact form, our state-space linear representation reads

$$(4) \quad \tilde{\mathbf{Q}}(t) = -\mathbf{B}(t)\mathbf{Y}(t) + \boldsymbol{\epsilon}(t), \quad \boldsymbol{\epsilon}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{R}(t))$$

$$(5) \quad \mathbf{Y}(t+1) = \mathbf{D}\mathbf{Y}(t) + \boldsymbol{\eta}(t+1), \quad \boldsymbol{\eta}(t+1) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma}(t+1))$$

where $\tilde{\mathbf{Q}}(t)$ and $\boldsymbol{\epsilon}(t)$ are the vectors that collect all the (transformed) observations and (aggregated) noises respectively, for all the ages and all the populations and sub-populations. Analogously, $\mathbf{Y}(t)$ is the vector that collects the values of all the state processes, while $\boldsymbol{\eta}(t+1)$ collects their noises. We assume that the state process noises and measurement errors are inter-temporally and mutually uncorrelated, even across populations and sub-populations. The Appendix provides further details on how the matrices appearing in equations (4) and (5) are constructed, and derives the explicit expressions for the covariance matrix $\mathbf{\Gamma}(t)$ in the OU case. The next section focuses on the Kalman Filter estimation of the state-space model presented here.

4.1.2. Kalman Filter estimation. Given our multivariate Gaussian linear state space model, described by the equations (4) and (5), we apply the Kalman filter algorithm to find the estimates of the unobservable values of the state process vector $\mathbf{Y}(t)$,

$\mathbf{a}(t)$, and of its variance-covariance matrix conditional on the observations, $\mathbf{P}(t)$, i.e.

$$\begin{aligned}\mathbf{a}(t) &:= \mathbb{E}[\mathbf{Y}(t) \mid \mathcal{F}(t-1)] \quad \text{and} \\ \mathbf{P}(t) &:= \text{Var}[\mathbf{Y}(t) \mid \mathcal{F}(t-1)]\end{aligned}$$

where the filtration $\mathcal{F}(t-1) := \{\tilde{\mathbf{Q}}(t), \dots, \tilde{\mathbf{Q}}(t-1)\}$ contains all the information about past observations. The standard Kalman filter equations for our state space model are

$$\begin{aligned}\mathbf{v}(t) &:= \tilde{\mathbf{Q}}(t) - \mathbb{E}[\tilde{\mathbf{Q}}(t) \mid \mathcal{F}(t-1)] = \tilde{\mathbf{Q}}(t) + \mathbf{B}(t)\mathbf{a}(t) \\ \mathbf{F}(t) &:= \text{Var}[\tilde{\mathbf{Q}}(t) \mid \mathcal{F}(t-1)] = \text{Var}[\mathbf{v}(t)] = \mathbf{B}(t)\mathbf{P}(t)\mathbf{B}'(t) + \mathbf{R}(t) \\ \mathbf{K}(t) &:= \mathbf{P}(t)\mathbf{B}'(t) \\ \mathbf{a}(t+1) &= \mathbf{D}(t)(\mathbf{a}(t) + \mathbf{K}(t)\mathbf{F}^{-1}(t)\mathbf{v}(t)) \\ \mathbf{P}(t+1) &= \mathbf{D}(t)(\mathbf{P}(t) - \mathbf{K}(t)\mathbf{F}^{-1}(t)\mathbf{K}'(t))\mathbf{D}'(t) + \mathbf{\Gamma}(t+1)\end{aligned}$$

where $\mathbf{v}(t)$ is the one-step-ahead prediction error, $\mathbf{F}(t)$ is its variance matrix. The matrix $\mathbf{K}(t)$, commonly referred to as the Kalman gain, is the covariance matrix between the state process and the error, $\text{Cov}[\mathbf{Y}(t), \tilde{\mathbf{Q}}(t) \mid \mathcal{F}(t-1)]$. The last two equations recursively define the estimates of interest. The Kalman filter needs to be initialized, and we set the initial values of $\mathbf{a}(1)$ and $\mathbf{P}(1)$ to

$$\begin{aligned}\mathbf{a}(1) &= \mathbb{E}[\mathbf{Y}(1) \mid \mathcal{F}(0)] = \mathbb{E}[\mathbf{Y}(1)] = \mathbf{a} \\ \mathbf{P}(1) &= \text{Var}[\mathbf{Y}(1) \mid \mathcal{F}(0)] = \text{Var}[\mathbf{Y}(1)] = \mathbf{P}.\end{aligned}$$

where the elements of \mathbf{a} belong to the optimal parameter set and \mathbf{P} is the unconditional variance/covariance matrix of the initial state vector.

The above algorithm can easily run into numerical problems, due to the inversion of the highly dimensional and sparse matrix \mathbf{F} . In the previous section, we have assumed that the covariance matrix of the measurements, \mathbf{R} , is diagonal. We can therefore apply the sequential Kalman filter algorithm (see Simon (2006)), that allows us to treat each element in the observation vector $\tilde{q}^{j,k}(t)$ as a separate measurement⁵. The mathematical formulation of this algorithm is different, but equivalent to the standard Kalman Filter one. In the Appendix, we provide further details on the algorithm.

Given the above setup, we need to jointly estimate the unobservable values of our state processes in time and the model parameters, that include the coefficients of the state processes dynamics and the parameters that enter the response functions. Defining Θ as the vector containing all the (time-independent) parameters of our model, the log-likelihood function of our state-space model is

$$(6) \quad \log L(\Theta) = \text{const} - \frac{1}{2} \sum_{l=1}^T \sum_{i=1}^p \log F_i(t) + v_i^2(t) F_i^{-1}(t),$$

where p is the number of observations for each period, $F_i(t)$ and $v_i(t)$ denote the Kalman gain and the one-step ahead prediction error relative to the i -th observation

⁵Notice that this algorithm can be applied even when \mathbf{R} is non-diagonal, but constant.

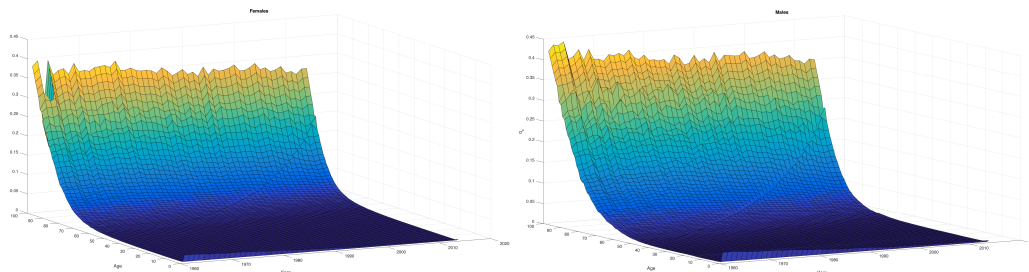


FIGURE 4. Observed mortality rates (q_x) of the UK females (left panel) and UK males (right panel) aged 0-99 from 1960 to 2013.

at time t , as defined in Appendix B. This function needs to be maximized to obtain the best fit of the model to the data. Numerical optimization procedures must be applied to find the optimal parameter set $\hat{\Theta}$, i.e.

$$\hat{\Theta} = \operatorname{argmax}_{\Theta} L(\Theta).$$

4.2. Estimation Results. We estimate the parameters of our model, following the above procedure. Our dataset is taken from the UK national population life table provided by the Human Mortality Database. The mortality surfaces of the two sub-populations of males and females are depicted in Figure 4. The figure allows to appreciate the most relevant phenomena concerning mortality evolution in the last 50 years, in particular the remarkable and constant decline in mortality rates over time at all ages, but especially at age 0 and the difference in the levels and patterns in the evolution of mortality rates between the two sub-populations considered.

To fit our model to the dataset, we perform the maximization of the log-likelihood through numerical methods, using a two-step procedure. First, we run a global Differential Evolution algorithm as a “burn-in”, and we use the parameters resulting from this procedure as the starting values for a local optimization algorithm. The model we fit requires estimation of 20 parameters only, while the number of observations is 10800. Table 1 collects the estimated parameters and some relevant statistics.

The three factors, that are related mostly to infant, young adults and elderly mortality respectively, all decline in calendar time (see Figure 4.2), as expected.⁶

The figure shows that the estimate of the factor related to infant mortality has experienced the largest drop, being almost ten times lower in 2013 than in 1960, while the other two factors dropped at around a half of the initial value.

The model specification we have selected implies that the two sub-populations are affected by factors have common dynamics, and thus common drift, diffusion

⁶The three factors, whose long-run average is set to zero, have initial values that are quite distant from this value, especially the first one. This is why we insert the initial values of the factors in our parameter space. The interpretation is that the ideal convergence to zero might be reached only in the very distant future. Alternatively, setting $\theta \neq 0$, one may try to fix or calibrate boundaries on the future improvement of human mortality.

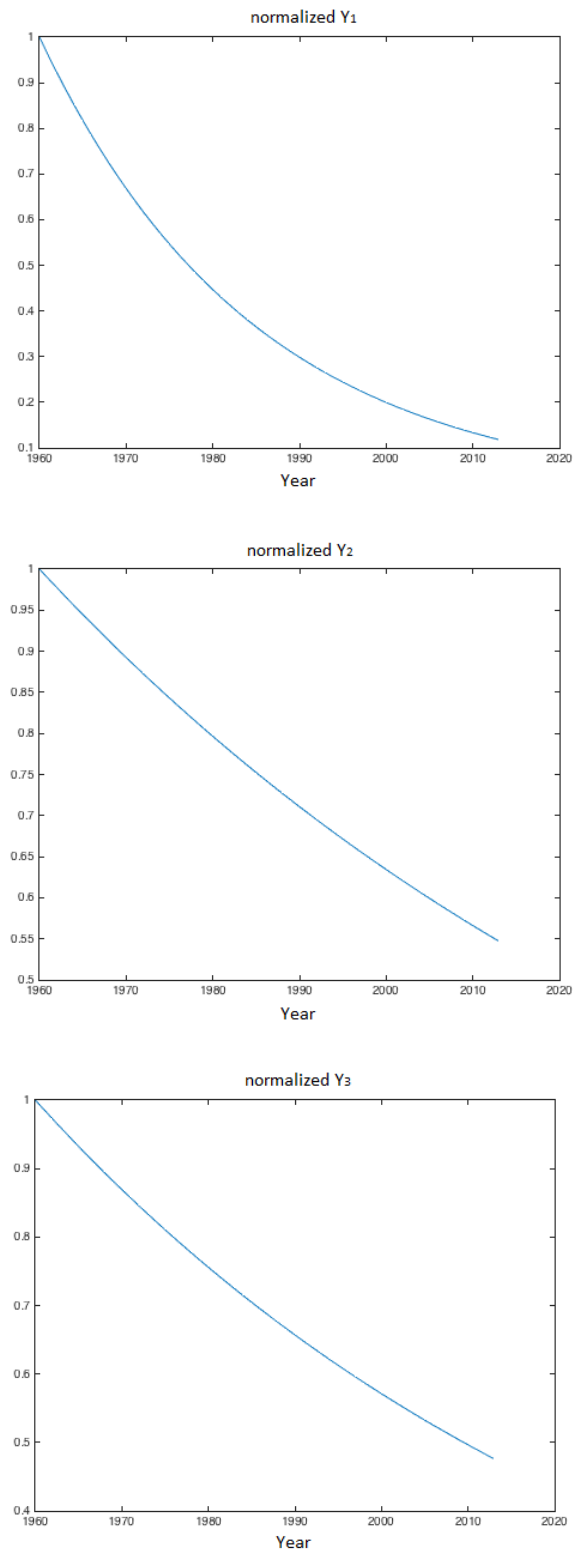


FIGURE 5. The figures plot the values of the three factors, Y_1 , Y_2 and Y_3 , normalized to the values at the first year.

Maximized log-Likelihood	69816.08	Number of parameters	20		
Optimal parameters					
Common		Males		Females	
Parameter	Value	Parameter	Value	Parameter	Value
a_1	0.0403	b_1^m	12.7960	b_1^f	13.0270
a_2	0.0114	b_2^m	0.0244	b_2^f	23.7818
a_3	0.0140	b_3^m	0.0999	b_3^f	0.0942
σ_1	2.71×10^{-4}	η^m	22.5661	η^f	18.9962
σ_2	1.74×10^{-14}	m^m	12.2580	m^f	12.0255
σ_3	4.01×10^{-19}				
ρ_1	0.0588				
ρ_2	-0.9566				
ρ_3	-0.8459				
s	0.1675				

TABLE 1. This table collects the optimal parameters for our model. The “common” column collects the parameters relative to the processes relative to the evolution of the factors and the measurement variance covariance matrix s . The other two columns collect the parameters relative to the response functions of the two sub-populations of males and females.

and correlation coefficients, that are reported in the left column. We remark that the second and the third factor, which affect young adult and old-age mortality respectively, show a very low level of variability. Indeed, almost deterministic trends have shaped the evolution of mortality at young and old-ages. A non-negligible part of the estimation error is captured by the age-specific measurement error, that we assumed to be proportional to the mortality rate. Comparing the parameters for the two sub-populations, we find that

- (1) the first factor affects very young females less (in level, see b_1^f vs. b_1^m) and its effect decays faster for them with age;
- (2) the second factor, that captures the mortality hump around age 20, is important in describing the mortality of males at that age, but helps describing the mortality of females only at a specific age (19);
- (3) the third factor weighs more for old man than women at the same age.

Despite its remarkable parsimoniousness, our tested three factor model is able to fit reasonably well the life tables of UK males and females from 1960 to 2013, and proves to be flexible enough to capture most of the observed features of mortality dynamics at all ages. This is evident when computing, for each age, the Mean Absolute Relative Estimation (MARE) Error in time, calculated as:

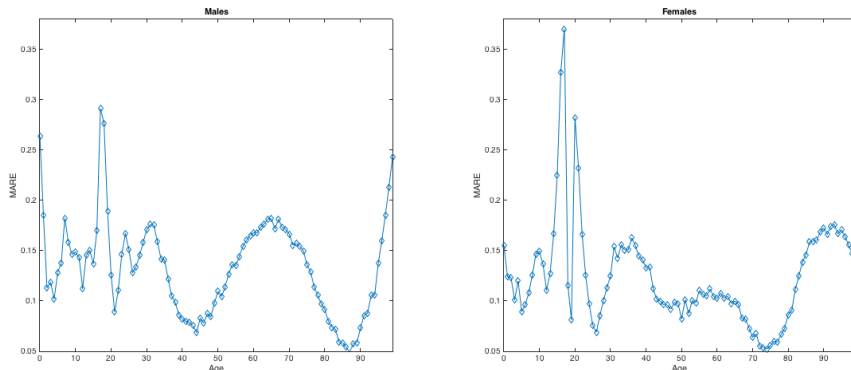


FIGURE 6. Mean Absolute Relative Errors by age for males (left panel) and females (right panel).

$$MARE(x) = \frac{1}{T} \sum_{t=1}^T \left| \frac{\hat{q}_x(t) - q_x(t)}{q_x(t)} \right|,$$

where $\hat{q}_x(t)$ denotes the fitted mortality rate. We report it, for the two sub-populations separately, of males and females, in Figure 4.2.

The model produces an average relative error of 12.90% for the 200 ages considered. Having set, for simplicity, the same specification for the two sub-populations, the model reproduces the dynamics at the different ages with a different level of accuracy for males and females. For instance, it does a better job in capturing the decline in infant mortality for females than for males, while it tends to fit more accurately young male adults (19-25) than female ones. Finally, the model produces lower relative estimation errors for female than for male old age mortality rates. These differences suggest that using different response functions across the two sub-populations might improve the results, at the price of introducing additional complexity in modelling. Overall, the MARE for males is 13.36%, larger than the one obtained for females, 12.44%. Analysis of the average relative error by calendar year (Figure 4.2) reveals that the fit accuracy is stable at around 12.5% for the whole period, except for the last five years of the sample, in which it exceeds 15%. It is likely that this feature of the data could be captured by the introduction of period-effects in the response functions, such as those we proposed in the specification described in Section 3.4.

5. CONCLUSIONS

In this paper, we have described a general framework for multi-population mortality evolution, in the context of continuous-time, affine factor models. A set of latent, unobservable factors describes the evolution of mortality rates in time. Different populations, and within them, sub-populations, are assumed to respond differently to changes in these factors, following specific response functions, that

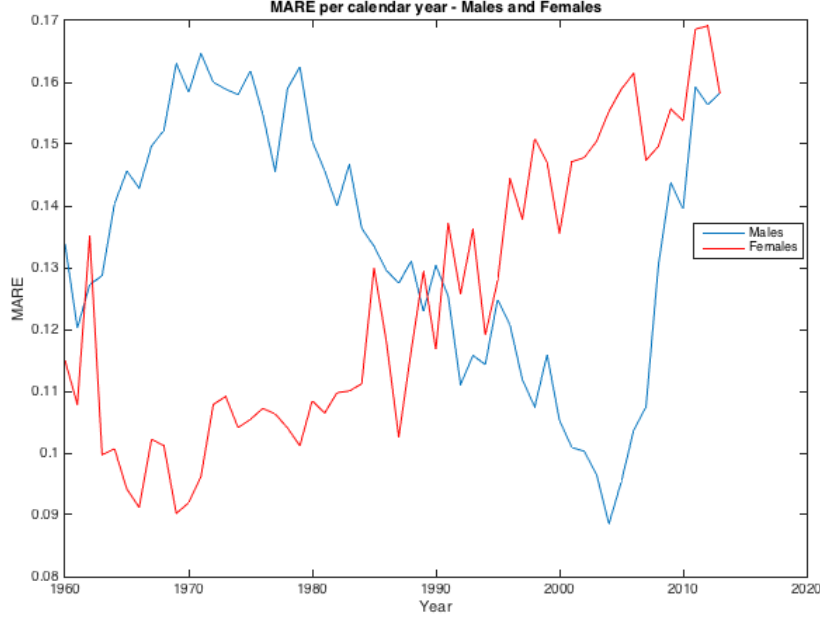


FIGURE 7. Mean Absolute Relative Errors by period for males (blue line) and females (red line).

can be themselves age and time-dependent. Our multi-population model allows to represent many different groups of homogeneous individuals, heterogeneous by age, such as males and females within different national populations or people with different socio-economic status belonging to the different states of a country. We showed how well-known deterministic mortality models can be extended to stochastic counterparts that fit into the framework we described. Our calibrated application exemplifies the flexibility and relevance of our approach. We fit the evolution of UK males and females mortality rates from 1960 to 2013 using a parsimonious model that relates to the classical Thiele law. Some relevant problems concerning the calibration procedure are discussed and tackled, in the context of Kalman filtering techniques applied to Gaussian factors. Extension of the calibration procedure to SR process, via quasi-Monte Carlo techniques, is in our research agenda, together with an exploration of the application of time-inhomogeneous specifications of our model. The models and the fitting procedure described in this paper can naturally serve as tools to be used in pricing and risk management applications. Nonetheless, we envisage the use of our framework to the stochastic forecast of multi-population mortality rates to be used in a wide variety of applications, such as the projections of public finance items.

REFERENCES

- Antonio, K., A. Bardoutsos, and W. Ouburg (2015). Bayesian poisson log-bilinear models for mortality projections with multiple populations. *European Actuarial Journal* 5(2), 245–281.
- Biffis, E. (2005). Affine processes for dynamic mortality and actuarial valuations. *Insurance: mathematics and economics* 37(3), 443–468.
- Blackburn, C. and M. Sherris (2013). Consistent dynamic affine mortality models for longevity risk applications. *Insurance: Mathematics and Economics* 53(1), 64–73.
- Coughlan, G. D., M. Khalaf-Allah, Y. Ye, S. Kumar, A. J. Cairns, D. Blake, and K. Dowd (2011). Longevity hedging 101: A framework for longevity basis risk analysis and hedge effectiveness. *North American Actuarial Journal* 15(2), 150–176.
- Dahl, M. (2004). Stochastic mortality in life insurance: market reserves and mortality-linked insurance contracts. *Insurance: mathematics and economics* 35(1), 113–136.
- Dahl, M., M. Melchior, and T. Möller (2008). On systematic mortality risk and risk-minimization with survivor swaps. *Scandinavian Actuarial Journal* (2-3), 114–146.
- Dahl, M. and T. Möller (2006). Valuation and hedging of life insurance liabilities with systematic mortality risk. *Insurance: mathematics and economics* 39(2), 193–217.
- Danesi, I. L., S. Haberman, and P. Millosovich (2015). Forecasting mortality in subpopulations using lee-carter type models: A comparison. *Insurance: Mathematics and Economics* 62, 151–161.
- De Rosa, C., E. Luciano, and L. Regis (2016). Basis risk in static versus dynamic longevity-risk hedging. *Scandinavian Actuarial Journal*, 1–23.
- Durrant-Whyte, H. F. (2012). *Integration, coordination and control of multi-sensor robot systems*, Volume 36. Springer Science & Business Media.
- Filipović, D. (2009). *Term-Structure Models: A Graduate Course* (1 ed.). Springer Finance. Springer-Verlag Berlin Heidelberg.
- Jevtić, P., E. Luciano, and E. Vigna (2013). Mortality surface by means of continuous time cohort models. *Insurance: Mathematics and Economics* 53(1), 122–133.
- Lando, D. (1998). On cox processes and credit risky securities. *Review of Derivatives research* 2(2-3), 99–120.
- Lee, R. D. and L. R. Carter (1992). Modeling and forecasting u. s. mortality. *Journal of the American Statistical Association* 87(419), pp. 659–671.
- Li, J. S.-H. and M. R. Hardy (2011). Measuring basis risk in longevity hedges. *North American Actuarial Journal* 15(2), 177–200.
- Li, N. and R. Lee (2005). Coherent mortality forecasts for a group of populations: An extension of the lee-carter method. *Demography* 42(3), 575–594.
- Luciano, E., L. Regis, and E. Vigna (2015). Single- and cross-generation natural hedging of longevity and financial risk. *The Journal of Risk and Insurance* (forthcoming).
- Luciano, E. and E. Vigna (2005). Non mean reverting affine processes for stochastic mortality. *ICER Applied Mathematics Working Paper No. 4-2005*.
- Milevsky, M. and D. Promislow (2001). Mortality derivatives and the option to annuitise. *Insurance: Mathematics and Economics* 29(3), 299–318.

- Mitchell, H. B. (2007). *Multi-sensor data fusion: an introduction*. Springer Science & Business Media.
- Raol, J. R. (2009). *Multi-Sensor Data Fusion with MATLAB®*. CRC Press.
- Schrager, D. (2006). Affine stochastic mortality. *Insurance: mathematics and economics* 38(1), 81–97.
- Simon, D. (2006). *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons.
- Villegas, A. M. and S. Haberman (2014). On the modeling and forecasting of socioeconomic mortality differentials: An application to deprivation and mortality in england. *North American Actuarial Journal* 18(1), 168–193.
- Wong, T. W., M. C. Chiu, and H. Y. Wong (2014). Time-consistent mean–variance hedging of longevity risk: Effect of cointegration. *Insurance: Mathematics and Economics* 56, 56–67.

6. APPENDIX

6.1. Appendix A - Proof of Theorem 2.3. To get the expressions for $\alpha^{j,k}(x, t, T)$ and $\beta^{j,k}(x, t, T)$, we apply the Feynman-Kac theorem. Let us first fix x, t, T and use the shortcut notation $\alpha^{j,k}$ and $\beta^{j,k}$ respectively. Our candidate solution is

$${}_{T-t}P_{x+t}^{j,k}(t) = e^{\alpha^{j,k}(x,t,T) + \beta^{j,k}(x,t,T) \cdot \mathbf{Y}^j(t)} := v^{j,k}.$$

Applying the Feynman-Kac theorem, we have

$$\begin{aligned} \frac{\partial v^{j,k}}{\partial t} + \mu^j(\mathbf{Y}_t^j, t) \cdot \nabla_{\mathbf{Y}^j} v^{j,k} + \frac{1}{2} \text{tr} \left\{ \sigma^j(\mathbf{Y}_t^j, t) \sigma^j(\mathbf{Y}_t^j, t)' \nabla_{\mathbf{Y}^j} \nabla_{\mathbf{Y}^j} v^{j,k} \right\} + \\ - R^{j,k}(\mathbf{Y}_t^j, x + t, t, x) v^{j,k} = 0, \end{aligned}$$

where we used the notation \mathbf{Y}_t^j to indicate $\mathbf{Y}^j(t)$. Since

$$\nabla_{\mathbf{Y}^j} v^{j,k} = v^{j,k} \beta^{j,k}, \quad \frac{\partial v^{j,k}}{\partial t} = v^{j,k} \left(\dot{\alpha}^{j,k} + (\mathbf{Y}_t^j)' \dot{\beta}^{j,k} \right) \quad \text{and} \quad \nabla_{\mathbf{Y}^j} \nabla_{\mathbf{Y}^j} v^{j,k} = \beta^{j,k} (\beta^{j,k})',$$

we have

$$\begin{aligned} v^{j,k} (\dot{\alpha}^{j,k} + (\mathbf{Y}_t^j)' \dot{\beta}^{j,k}) + \mathbf{A}^j (\boldsymbol{\theta}^j - \mathbf{Y}_t^j) \cdot (v^{j,k} \beta^{j,k}) \\ + \frac{1}{2} \text{tr} \left\{ \boldsymbol{\Sigma}^j \sqrt{V_t^j} \mathbf{H}^j (\mathbf{H}^j)' \left(\sqrt{V_t^j} \right)' (\boldsymbol{\Sigma}^j)' \nabla_{\mathbf{Y}^j} \nabla_{\mathbf{Y}^j} v^{j,k} \right\} \\ - [g_0^{j,k}(x + t, t, x) + (\mathbf{Y}_t^j)' \mathbf{g}^{j,k}(x + t, t, x)] v^{j,k} = 0. \end{aligned}$$

Collecting and rearranging, we get

$$\begin{aligned} v^{j,k} \left[(\dot{\alpha}^{j,k} + (\mathbf{Y}_t^j)' \dot{\beta}^{j,k}) + \mathbf{A}^j (\boldsymbol{\theta}^j - \mathbf{Y}_t^j) \cdot \beta^{j,k} \right. \\ \left. + \frac{1}{2} \text{tr} \left\{ \boldsymbol{\Sigma}^j \sqrt{V_t^j} \boldsymbol{\rho}^j \left(\sqrt{V_t^j} \right)' (\boldsymbol{\Sigma}^j)' \beta^{j,k} (\beta^{j,k})' \right\} \right. \\ \left. - g_0^{j,k}(x + t, t, x) - (\mathbf{Y}_t^j)' \mathbf{g}^{j,k}(x + t, t, x) \right] = 0. \end{aligned}$$

We use the properties of the trace of a matrix and obtain

$$\begin{aligned} v^{j,k} \left[\dot{\alpha}^{j,k} + (\mathbf{Y}_t^j)' \dot{\beta}^{j,k} + \mathbf{A}^j \boldsymbol{\theta}^j \cdot \beta^{j,k} - \mathbf{A}^j \mathbf{Y}_t^j \cdot \beta^{j,k} \right. \\ \left. + \frac{1}{2} \text{tr} \left\{ (\beta^{j,k})' \boldsymbol{\Sigma}^j \sqrt{V_t^j} \boldsymbol{\rho}^j \left(\sqrt{V_t^j} \right)' (\boldsymbol{\Sigma}^j)' \beta^{j,k} \right\} \right. \\ \left. - g_0^{j,k}(x + t, t, x) - (\mathbf{Y}_t^j)' \mathbf{g}^{j,k}(x + t, t, x) \right] = 0. \end{aligned}$$

Thus,

$$\begin{aligned} \dot{\alpha}^{j,k} + (\mathbf{Y}_t^j)' \dot{\beta}^{j,k} + \mathbf{A}^j \boldsymbol{\theta}^j \cdot \beta^{j,k} - (\mathbf{Y}_t^j)' (\mathbf{A}^j)' \beta^{j,k} \\ + \frac{1}{2} \text{tr} \left\{ (\beta^{j,k})' \boldsymbol{\Sigma}^j \sqrt{V_t^j} \boldsymbol{\rho}^j \left(\sqrt{V_t^j} \right)' (\boldsymbol{\Sigma}^j)' \beta^{j,k} \right\} \\ - g_0^{j,k}(x + t, t, x) - (\mathbf{Y}_t^j)' \mathbf{g}^{j,k}(x + t, t, x) = 0. \end{aligned}$$

At this point, we need to consider separately the case of the OU and SR processes.

6.1.1. *OU Case.* Let $\gamma_i^{j,k} = 1$ and $\delta_i^{j,k} = \mathbf{0}$. Then $V_t = I_{N \times N}$ and, after rearranging the terms,

$$\begin{aligned} & \dot{\alpha}^{j,k} + \mathbf{A}^j \boldsymbol{\theta}^j \cdot \boldsymbol{\beta}^{j,k} - g_0^j(x+t, t, x) + \frac{1}{2} \text{tr} \left\{ (\boldsymbol{\beta}^{j,k})' \boldsymbol{\Sigma}^j \boldsymbol{\rho}^j (\boldsymbol{\Sigma}^j)' \boldsymbol{\beta}^{j,k} \right\} \\ & + (\mathbf{Y}_t^j)' \left[\dot{\boldsymbol{\beta}}^{j,k} - (\mathbf{A}^j)' \boldsymbol{\beta}^{j,k} - \mathbf{g}^{j,k}(x+t, t, x) \right] = 0. \end{aligned}$$

Since the trace of a scalar is a scalar, we have

$$\begin{aligned} & \dot{\alpha}^{j,k} + \mathbf{A}^j \boldsymbol{\theta}^j \cdot \boldsymbol{\beta}^{j,k} - g_0^j(x+t, t, x) + \frac{1}{2} (\boldsymbol{\beta}^{j,k})' \boldsymbol{\Sigma}^j \boldsymbol{\rho}^j (\boldsymbol{\Sigma}^j)' \boldsymbol{\beta}^{j,k} \\ & + (\mathbf{Y}_t^j)' \left[\dot{\boldsymbol{\beta}}^{j,k} - (\mathbf{A}^j)' \boldsymbol{\beta}^{j,k} - \mathbf{g}^{j,k}(x+t, t, x) \right] = 0. \end{aligned}$$

Now, by the “affine matching principle” (see Duffie and Kan, 1996) we have

$$\begin{aligned} \dot{\boldsymbol{\beta}}^{j,k} &= \mathbf{g}^{j,k}(x+t, t, x) + (\mathbf{A}^j)' \boldsymbol{\beta}^{j,k} \\ \dot{\alpha}^{j,k} &= g_0^j(x+t, t, x) - \mathbf{A}^j \boldsymbol{\theta}^j \cdot \boldsymbol{\beta}^{j,k} - \frac{1}{2} (\boldsymbol{\beta}^{j,k})' \boldsymbol{\Sigma}^j \boldsymbol{\rho}^j (\boldsymbol{\Sigma}^j)' \boldsymbol{\beta}^{j,k}. \end{aligned}$$

6.1.2. *SR process Case.* Let $\gamma_i = 0$ and $\delta_i = 1$. Then since $V_t = \text{diag}[Y_1^1(t), \dots, Y_N^1(t)] = \mathbf{Y}_t' I_{N \times N}$ and $\boldsymbol{\rho}^1 = I_{N \times N}$ we have

$$\begin{aligned} & \dot{\alpha}^{1,k} + \mathbf{A}^1 \boldsymbol{\theta}^1 \cdot \boldsymbol{\beta}^{1,k} - g_0^{1,k}(x+t, t, x) + \frac{1}{2} \text{tr} \left\{ (\boldsymbol{\beta}^{1,k})' \boldsymbol{\Sigma}^1 V_t (\boldsymbol{\Sigma}^1)' \boldsymbol{\beta}^{1,k} \right\} \\ & + (\mathbf{Y}_t^1)' \left(\dot{\boldsymbol{\beta}}^{1,k} - (\mathbf{A}^1)' \boldsymbol{\beta}^{1,k} - \mathbf{g}^{1,k}(x+t, t, x) \right) = 0. \end{aligned}$$

Realizing that the trace of a scalar is a scalar, we have

$$\begin{aligned} & \dot{\alpha}^{1,k} + \mathbf{A}^1 \boldsymbol{\theta}^1 \cdot \boldsymbol{\beta}^{1,k} - g_0^{1,k}(x+t, t, x) + \frac{1}{2} (\boldsymbol{\beta}^{1,k})' \boldsymbol{\Sigma}^1 V_t (\boldsymbol{\Sigma}^1)' \boldsymbol{\beta}^{1,k} \\ & + (\mathbf{Y}_t^1)' \left(\dot{\boldsymbol{\beta}}^{1,k} - (\mathbf{A}^1)' \boldsymbol{\beta}^{1,k} - \mathbf{g}^{1,k}(x+t, t, x) \right) = 0. \end{aligned}$$

which can be written as

$$\begin{aligned} & \dot{\alpha}^{1,k} + \mathbf{A}^1 \boldsymbol{\theta}^1 \cdot \boldsymbol{\beta}^{1,k} - g_0^{1,k}(x+t, t, x) + (\mathbf{Y}_t^1)' \left[\frac{1}{2} (\beta_1^{1,k})^2 \sigma_1^2, \dots, \frac{1}{2} (\beta_N^{1,k})^2 \sigma_N^2 \right]' \\ & + (\mathbf{Y}_t^1)' \left(\dot{\boldsymbol{\beta}}^{1,k} - (\mathbf{A}^1)' \boldsymbol{\beta}^{1,k} - \mathbf{g}^{1,k}(x+t, t, x) \right) = 0. \end{aligned}$$

Hence,

$$\begin{aligned} & \dot{\alpha}^{1,k} + \mathbf{A}^1 \boldsymbol{\theta}^1 \cdot \boldsymbol{\beta}^{1,k} - g_0^{1,k}(x+t, t, x) \\ & + (\mathbf{Y}_t^1)' \left(\dot{\boldsymbol{\beta}}^{1,k} - (\mathbf{A}^1)' \boldsymbol{\beta}^{1,k} - \mathbf{g}^{1,k}(x+t, t, x) + \left[\frac{1}{2} (\beta_1^{1,k})^2 \sigma_1^2, \dots, \frac{1}{2} (\beta_N^{1,k})^2 \sigma_N^2 \right]' \right) = 0. \end{aligned}$$

Now, applying the “affine matching principle” (see Duffie and Kan, 1996), we have

$$\begin{aligned} \dot{\alpha}^{1,k} &= g_0^{1,k}(x+t, t, x) - \mathbf{A}^1 \boldsymbol{\theta}^1 \cdot \boldsymbol{\beta}^{1,k} \\ \dot{\boldsymbol{\beta}}^{1,k} &= \mathbf{g}^{1,k}(x+t, t, x) + (\mathbf{A}^1)' \boldsymbol{\beta}^{1,k} - \left[\frac{1}{2} (\beta_1^{1,k})^2 \sigma_1^2, \dots, \frac{1}{2} (\beta_N^{1,k})^2 \sigma_N^2 \right]'. \end{aligned}$$

6.2. Appendix B - Further details on the state space representation and calibration procedure. In Section 4.1.1 we expressed the measurement process for population j sub-population k . In compact form, expression (3) can be written as

$$\bar{\mathbf{Q}}^{j,k}(t) = -\boldsymbol{\alpha}^{j,k}(t) - \mathbf{B}^{j,k}(t)\mathbf{Y}^j(t) + \boldsymbol{\epsilon}^{j,k}(t)$$

where

$$\begin{aligned}\bar{\mathbf{Q}}^{j,k} &:= [\bar{q}_{x_{\min}}^{j,k}(t), \dots, \bar{q}_{x_{\max}}^{j,k}(t)] \\ \boldsymbol{\alpha}^{j,k}(t) &:= [\alpha^{j,k}(x_{\min}^{j,k}, t, t+1), \dots, \alpha^{j,k}(x_{\max}^{j,k}, t, t+1)]' \\ \mathbf{B}^{j,k}(t) &:= [\beta^{j,k}(x_{\min}^{j,k}, t, t+1), \dots, \beta^{j,k}(x_{\max}^{j,k}, t, t+1)]' \\ \boldsymbol{\epsilon}^{j,k}(t) &:= [\epsilon_{x_{\min}}^{j,k}(t), \dots, \epsilon_{x_{\max}}^{j,k}(t)]' \text{ where } \boldsymbol{\epsilon}^{j,k}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{R}^{j,k}(t)) \text{ ,} \\ \mathbf{R}^{j,k}(t) &:= \text{diag}[(s\bar{q}_{x_{\min}}^{j,k}(t))^2, \dots, (s\bar{q}_{x_{\max}}^{j,k}(t))^2].\end{aligned}$$

We slightly transform the measurement process, obtaining

$$\tilde{\mathbf{Q}}^{j,k}(t) := \bar{\mathbf{Q}}^{j,k}(t) + \boldsymbol{\alpha}^{j,k}(t) = -\mathbf{B}^{j,k}(t)\mathbf{Y}^j(t) + \boldsymbol{\epsilon}^{j,k}(t).$$

The state-space process for population j has the following dynamics,

$$\mathbf{Y}^j(t+1) = e^{-\mathbf{A}^j}\mathbf{Y}^j(t) + (\mathbf{1} - e^{-\mathbf{A}^j})\boldsymbol{\theta} + \boldsymbol{\eta}^j(t+1), \quad j \in \mathbf{J}, \text{ where}$$

$$\boldsymbol{\eta}^j(t+1) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}^j(t+1)), \text{ and } \boldsymbol{\Gamma}^j(t+1) := \mathbb{E}\left[\left(\int_t^{t+1} e^{-\mathbf{A}^j(t+1-s)} \boldsymbol{\Sigma}^j \mathbf{H}^j d\mathbf{Z}_s\right)\right]^2.$$

From now on, to simplify notation and without significant loss of generality, we assume $\boldsymbol{\theta} = \mathbf{0}$.

The assumption that the state-space process noise $\boldsymbol{\eta}^j$ and measurement noise $\boldsymbol{\epsilon}^{j,k}$ are both inter-temporally and mutually uncorrelated formally implies that

$$\begin{aligned}\mathbb{E}[\boldsymbol{\epsilon}^{j,k}(m)\boldsymbol{\epsilon}^{j,k}(n)] &= \delta_{m,n}\mathbf{R}^{j,k}(m), \\ \mathbb{E}[\boldsymbol{\eta}^j(m)\boldsymbol{\eta}^j(n)] &= \delta_{m,n}\boldsymbol{\Gamma}^j(m) \text{ and} \\ \mathbb{E}[\boldsymbol{\eta}^j(m)\boldsymbol{\epsilon}^{j,k}(n)] &= \mathbf{0} \text{ where } \forall m, n \in \{0, \dots, T\}.\end{aligned}$$

For each population j , we gather the information coming from all sub-populations. There exists one state process only, common to all sub-populations, and mortality data from different sub-populations are akin to the signals received by different multi-channel sensors in a centralized architecture.⁷

$$\begin{aligned}\tilde{\mathbf{Q}}^j(t) &:= [\tilde{\mathbf{Q}}^{j,1}(t), \dots, \tilde{\mathbf{Q}}^{j,K_j}(t)]' \\ \mathbf{B}^j(t) &:= [\mathbf{B}^{j,1}(t), \dots, \mathbf{B}^{j,K_j}(t)]' \\ \boldsymbol{\epsilon}^j(t) &:= [\boldsymbol{\epsilon}^{j,1}(t), \dots, \boldsymbol{\epsilon}^{j,K_j}(t)]' \\ \mathbf{R}^j(t) &:= \mathbb{E}[\boldsymbol{\epsilon}^j(t)\boldsymbol{\epsilon}^j(t)'] = \text{blockdiag}[\mathbf{R}^{j,1}(t), \dots, \mathbf{R}^{j,K_j}(t)].\end{aligned}$$

⁷This problem is known in the radar (or satellite) target tracking domain as the single-target-multiple-sensors problem (see Mitchell (2007), Raol (2009)).

We further assume that measurement noises of different populations and sub-populations are inter-temporally uncorrelated⁸, i.e.

$$\begin{aligned} \mathbb{E}[\epsilon^{j_1, k_1}(m) \epsilon^{j_2, k_2}(n)] &= \mathbf{0}, \quad \forall m, n \in \{0, \dots, T\} \text{ and} \\ \forall k_1, k_2 \in \{1, \dots, K_j\}, \quad k_1 &\neq k_2, \quad \forall j_1, j_2 \in \{1, \dots, J\}, \quad j_1 \neq j_2. \end{aligned}$$

Finally, as described in Section 4.1.1, to tackle our multi-population joint estimation problem, we need to combine all observations and observation models into a single composite group model, following a “centralized data-fusion approach”.⁹ This strategy allows us to treat all observations as if they were coming from a unique (albeit very big) multi-channel sensor, where each channel of the unique sensor receives the information from one population. In practice, we obtain the matrices in our equations (4) and (5) as follows:

$$\begin{aligned} \tilde{\mathbf{Q}}(t) &:= [\tilde{\mathbf{Q}}^1(t), \dots, \tilde{\mathbf{Q}}^J(t)]' \\ \mathbf{B}(t) &:= \text{blockdiag}[\mathbf{B}^1(t), \dots, \mathbf{B}^J(t)] \\ \epsilon(t) &:= [\epsilon^1(t), \dots, \epsilon^J(t)]' \\ \mathbf{R}(t) &:= \mathbb{E}[\epsilon(t) \epsilon'(t)] = \text{blockdiag}[\mathbf{R}^1(t), \dots, \mathbf{R}^J(t)], \\ \mathbf{Y}(t) &:= [\mathbf{Y}^1(t), \dots, \mathbf{Y}^J(t)]' \\ \mathbf{D} &:= \text{blockdiag}[e^{-\mathbf{A}^1}, \dots, e^{-\mathbf{A}^J}] \\ \mathbf{\Gamma}(t+1) &:= [\mathbf{\Gamma}^{j_1, j_2}(t+1)]_{j_1, j_2 \in J}, \text{ having} \\ \mathbf{\Gamma}^{j_1, j_2}(t+1) &:= \mathbb{E} \left[\int_t^{t+1} e^{-\mathbf{A}^{j_1}(t+1-s)} \mathbf{\Sigma}^{j_1} \mathbf{H}^{j_1} (\mathbf{H}^{j_2})' \mathbf{\Sigma}^{j_2} e^{-\mathbf{A}^{j_2}(t+1-s)} d\mathbf{Z}_s \right] \end{aligned}$$

and in particular $\mathbf{\Gamma}^{j,j}(t+1) = \mathbf{\Gamma}^j(t+1)$, $j \in J$.

Sequential KF algorithm

In the following, we provide further details on the sequential Kalman Filter algorithm we have used in estimating our model in Section 4. For this purpose, we will now see our matrices as

$$\begin{aligned} \tilde{\mathbf{Q}}(t) &= [\tilde{\mathbf{Q}}_1(t), \dots, \tilde{\mathbf{Q}}_p(t)]' \\ \epsilon(t) &= [\epsilon_1(t), \dots, \epsilon_p(t)]' \\ \mathbf{B}(t) &= [\mathbf{B}_1(t), \dots, \mathbf{B}_p(t)]' \\ \mathbf{R}(t) &= \text{diag}[\sigma_1^2(t), \dots, \sigma_p^2(t)]' \end{aligned}$$

where $p = \sum_{j,k} \# \mathbf{X}^{j,k}$ is the total number of ages considered, $\mathbf{Q}_i(t)$ is a scalar and $\mathbf{B}_i(t)$'s are row vectors. Now, since \mathbf{R} is a diagonal matrix, the equation that describes the model for each single observation of interest is

$$(7) \quad \tilde{\mathbf{Q}}_i(t) = -\mathbf{B}_i(t) \mathbf{Y}(t) + \epsilon_i(t).$$

The sequential Kalman filter algorithm runs then as follows:

⁸Notice that this assumption, although reasonable in many instances, can be rather strong if the same institution collects the measurements from all sub-populations.

⁹An alternative is track-to-track fusion, where bulk of computational effort is made in a decentralized manner (via many separate local Kalman filters) and then estimates are optimally merged at each point in time.

- (1) The filter is initialized using the unconditional mean and variance of the state-process vector,

$$\begin{aligned}\mathbf{a}(1) &= \mathbb{E}[\mathbf{Y}(1) \mid \mathcal{F}(0)] = \mathbb{E}[\mathbf{Y}(1)] = \mathbf{a} \\ \mathbf{P}(1) &= \text{Var}[\mathbf{Y}(1) \mid \mathcal{F}(0)] = \text{Var}[\mathbf{Y}(1)] = \mathbf{P},\end{aligned}$$

- (2) At each time step t , the state process vector and its variance covariance matrix are updated as in the standard Kalman Filter,

$$\begin{aligned}\mathbf{a}^-(t+1) &= \mathbf{D}(t)\mathbf{a}(t) \\ \mathbf{P}^-(t+1) &= \mathbf{D}(t)\mathbf{P}(t)\mathbf{D}'(t) + \mathbf{\Gamma}(t+1).\end{aligned}$$

- (3) At each time step t , the estimates are initialized

$$\begin{aligned}\mathbf{a}_0^+(t+1) &= \mathbf{a}^-(t) \\ \mathbf{P}_0^+(t+1) &= \mathbf{P}^-(t),\end{aligned}$$

and then updated at after each observation $i = 1, \dots, p$, as

$$\begin{aligned}\mathbf{a}_i^+(t) &= \mathbf{a}_{i-1}^+(t) + \mathbf{K}_i(t)v_i^+(t) \\ \mathbf{P}_i^+(t) &= \mathbf{P}_{i-1}^+(t) - \mathbf{K}_i(t)\mathbf{B}_i(t)\mathbf{P}_{i-1}^+(t),\end{aligned}$$

where

$$\begin{aligned}v_i^+(t) &= \tilde{Q}_i(t) + \mathbf{B}_i(t)\mathbf{a}_{i-1}^+(t) \\ \mathbf{K}_i(t) &= \frac{\mathbf{P}_{i-1}^+(t)\mathbf{B}_i'(t)}{\sigma_i^2(t)}.\end{aligned}$$

Note that $v_i(t)$ and $F_i(t)$ are scalars and $\mathbf{K}_i(t)$ is a column vector. Finally, we set

$$\begin{aligned}\mathbf{a}^+(t) &= \mathbf{a}_p^+(t); \\ \mathbf{P}^+(t) &= \mathbf{P}_p^+(t).\end{aligned}$$

For each $i = 1, \dots, p$, we further compute

$$\begin{aligned}F_i(t) &= \mathbf{B}_i(t)\mathbf{P}^+(t)\mathbf{B}_i'(t) + \sigma_i^2(t), \\ v_i(t) &= \tilde{Q}_i(t) + \mathbf{B}_i(t)\mathbf{a}^+(t).\end{aligned}$$

These last expressions for $F_i(t)$ and $v_i(t)$ define the quantities that enter the log-likelihood function (6), that is maximized to get the optimal parameter set.

6.3. Appendix C - Variance-Covariance matrix in the OU process case.

Let us consider the case in which the dynamics of the factors are described by OU processes, of the type

$$d\mathbf{Y}^j(t) = \mathbf{A}^j(\boldsymbol{\theta}^j - \mathbf{Y}^j(t))dt + \boldsymbol{\Sigma}^j d\mathbf{W}^j(t).$$

The above SDE can be written in integral form as

$$\mathbf{Y}^j(t) = e^{-\mathbf{A}^j t} \mathbf{Y}^j(0) + \int_0^t e^{-\mathbf{A}^j(t-s)} \boldsymbol{\theta}^j ds + \int_0^t e^{-\mathbf{A}^j(t-s)} \boldsymbol{\Sigma}^j \mathbf{H}^j d\mathbf{Z}(s).$$

This process is Gaussian. As a consequence, we can estimate our model for the death probabilities, that has a state-space linear representation, using a maximum likelihood approach. In this Appendix, we explicitly derive a general expression for the variance/covariance matrix $\mathbf{\Gamma}$ of the factors, when they are correlated. We define first

$$\mathbf{Y}(t) := \begin{bmatrix} \mathbf{Y}^1(t) \\ \vdots \\ \mathbf{Y}^J(t) \end{bmatrix}, \quad \mathbf{W}(t) := \begin{bmatrix} \mathbf{W}^1(t) \\ \vdots \\ \mathbf{W}^J(t) \end{bmatrix}, \quad \boldsymbol{\theta} := \begin{bmatrix} \boldsymbol{\theta}^1 \\ \vdots \\ \boldsymbol{\theta}^J \end{bmatrix},$$

$$\mathbf{A} := \begin{bmatrix} \mathbf{A}^1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^J & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}^J \end{bmatrix}, \quad \boldsymbol{\Sigma} := \begin{bmatrix} \boldsymbol{\Sigma}^1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}^J & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}^J \end{bmatrix}.$$

Then, the conditional mean and variance of the state process vector is

$$\begin{aligned} \mathbb{E}[\mathbf{Y}(t) \mid \mathbf{Y}(0)] &= e^{-\mathbf{A}t} \mathbf{Y}(0) + (\mathbf{1} - e^{-\mathbf{A}t}) \boldsymbol{\theta}, \\ \mathbf{\Gamma}(t) := \text{Var}[\mathbf{Y}(t) \mid \mathbf{Y}(0)] &= \mathbb{E} \left[\left(\int_0^t e^{-\mathbf{A}(t-s)} \boldsymbol{\Sigma} d\mathbf{W}(s) \right)^2 \right]. \end{aligned}$$

Defining

$$\mathbf{H}(t) = \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_J \end{bmatrix},$$

we have $\mathbf{W}(t) = \mathbf{H}\mathbf{Z}(t)$. Applying Itô isometry,

$$\mathbf{\Gamma}(t) = \int_0^t e^{-\mathbf{A}(t-s)} \boldsymbol{\Sigma} \mathbf{H} \mathbf{H}' \boldsymbol{\Sigma}' (e^{-\mathbf{A}(t-s)})' ds.$$

Since

$$\mathbf{H} \mathbf{H}' = \begin{bmatrix} \mathbf{H}^1 (\mathbf{H}^1)' & \mathbf{H}^1 (\mathbf{H}^2)' & \cdots & \mathbf{H}^1 (\mathbf{H}^J)' \\ \mathbf{H}^2 (\mathbf{H}^1)' & \mathbf{H}^2 (\mathbf{H}^2)' & \cdots & \mathbf{H}^2 (\mathbf{H}^J)' \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}^J (\mathbf{H}^1)' & \mathbf{H}^J (\mathbf{H}^2)' & \cdots & \mathbf{H}^J (\mathbf{H}^J)' \end{bmatrix},$$

then

$$\boldsymbol{\Sigma} \mathbf{H} \mathbf{H}' \boldsymbol{\Sigma}' = \begin{bmatrix} \boldsymbol{\Sigma}^1 \mathbf{H}^1 (\mathbf{H}^1)' (\boldsymbol{\Sigma}^1)' & \boldsymbol{\Sigma}^1 \mathbf{H}^1 (\mathbf{H}^2)' (\boldsymbol{\Sigma}^2)' & \cdots & \boldsymbol{\Sigma}^1 \mathbf{H}^1 (\mathbf{H}^J)' (\boldsymbol{\Sigma}^J)' \\ \boldsymbol{\Sigma}^2 \mathbf{H}^2 (\mathbf{H}^1)' (\boldsymbol{\Sigma}^1)' & \boldsymbol{\Sigma}^2 \mathbf{H}^2 (\mathbf{H}^2)' (\boldsymbol{\Sigma}^2)' & \cdots & \boldsymbol{\Sigma}^2 \mathbf{H}^2 (\mathbf{H}^J)' (\boldsymbol{\Sigma}^J)' \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}^J \mathbf{H}^J (\mathbf{H}^1)' (\boldsymbol{\Sigma}^1)' & \boldsymbol{\Sigma}^J \mathbf{H}^J (\mathbf{H}^2)' (\boldsymbol{\Sigma}^2)' & \cdots & \boldsymbol{\Sigma}^J \mathbf{H}^J (\mathbf{H}^J)' (\boldsymbol{\Sigma}^J)' \end{bmatrix}.$$

Because \mathbf{A} is a diagonal matrix, we have

$$e^{-\mathbf{A}(t-s)} = (e^{-\mathbf{A}(t-s)})'$$

As a consequence,

$$\begin{aligned} & e^{-\mathbf{A}(t-s)} \boldsymbol{\Sigma} \mathbf{H} \mathbf{H}' \boldsymbol{\Sigma}' (e^{-\mathbf{A}(t-s)})' = \\ &= \begin{bmatrix} e^{-\mathbf{A}^1(t-s)} \boldsymbol{\Sigma}^1 \mathbf{H}^1 (\mathbf{H}^1)' (\boldsymbol{\Sigma}^1)' e^{-\mathbf{A}^1(t-s)} & \cdots & e^{-\mathbf{A}^1(t-s)} \boldsymbol{\Sigma}^1 \mathbf{H}^1 (\mathbf{H}^J)' (\boldsymbol{\Sigma}^J)' e^{-\mathbf{A}^J(t-s)} \\ \vdots & \ddots & \vdots \\ e^{-\mathbf{A}^J(t-s)} \boldsymbol{\Sigma}^J \mathbf{H}^J (\mathbf{H}^1)' (\boldsymbol{\Sigma}^1)' e^{-\mathbf{A}^1(t-s)} & \cdots & e^{-\mathbf{A}^J(t-s)} \boldsymbol{\Sigma}^J \mathbf{H}^J (\mathbf{H}^J)' (\boldsymbol{\Sigma}^J)' e^{-\mathbf{A}^J(t-s)} \end{bmatrix}. \end{aligned}$$

Concluding,

$$\text{Var}[\mathbf{Y}(t) \mid \mathbf{Y}(0)] = \begin{bmatrix} \int_0^t e^{-\mathbf{A}^1(t-s)} \boldsymbol{\Sigma}^1 \mathbf{H}^1 (\mathbf{H}^1)' (\boldsymbol{\Sigma}^1)' e^{-\mathbf{A}^1(t-s)} ds & \dots & \int_0^t e^{-\mathbf{A}^1(t-s)} \boldsymbol{\Sigma}^1 \mathbf{H}^1 (\mathbf{H}^J)' (\boldsymbol{\Sigma}^J)' e^{-\mathbf{A}^J(t-s)} ds \\ \vdots & \ddots & \vdots \\ \int_0^t e^{-\mathbf{A}^J(t-s)} \boldsymbol{\Sigma}^J \mathbf{H}^J (\mathbf{H}^1)' (\boldsymbol{\Sigma}^1)' e^{-\mathbf{A}^1(t-s)} ds & \dots & \int_0^t e^{-\mathbf{A}^J(t-s)} \boldsymbol{\Sigma}^J \mathbf{H}^J (\mathbf{H}^J)' (\boldsymbol{\Sigma}^J)' e^{-\mathbf{A}^J(t-s)} ds \end{bmatrix}.$$

The variance/covariance matrix $\boldsymbol{\Gamma}$ is then a block matrix, whose blocks represent the variance/covariance matrices $\boldsymbol{\Gamma}^{j_1, j_2}$ of the pairs of vectors of processes driving the dynamics of populations j_1 and j_2 . In a more compact form, we have

$$\boldsymbol{\Gamma}(t) = \begin{bmatrix} \boldsymbol{\Gamma}^{1,1}(t) & \boldsymbol{\Gamma}^{1,2}(t) & \dots & \boldsymbol{\Gamma}^{1,J}(t) \\ \boldsymbol{\Gamma}^{2,1}(t) & \boldsymbol{\Gamma}^{2,2}(t) & \dots & \boldsymbol{\Gamma}^{2,J}(t) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Gamma}^{J,1}(t) & \boldsymbol{\Gamma}^{J,2}(t) & \dots & \boldsymbol{\Gamma}^{J,J}(t) \end{bmatrix}.$$

Let us now expand each block. Each element can be written as

$$\boldsymbol{\Gamma}^{j_1, j_2}(t) := \int_0^t e^{-\mathbf{A}^{j_1}(t-s)} \boldsymbol{\Sigma}^{j_1} \mathbf{H}^{j_1} (\mathbf{H}^{j_2})' (\boldsymbol{\Sigma}^{j_2})' e^{-\mathbf{A}^{j_2}(t-s)} ds.$$

We know that

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A}) \text{vec}(\mathbf{B}).$$

Then,

$$\text{vec} \left(e^{-\mathbf{A}^{j_1}(t-s)} \boldsymbol{\Sigma}^{j_1} \mathbf{H}^{j_1} (\mathbf{H}^{j_2})' (\boldsymbol{\Sigma}^{j_2})' e^{-\mathbf{A}^{j_2}(t-s)} \right) = \left(e^{-\mathbf{A}^{j_1}(t-s)} \otimes e^{-\mathbf{A}^{j_2}(t-s)} \right) \text{vec}(\boldsymbol{\Sigma}^{j_1} \mathbf{H}^{j_1} (\mathbf{H}^{j_2})' (\boldsymbol{\Sigma}^{j_2})').$$

We now use the identity

$$e^{\mathbf{A} \oplus \mathbf{B}} = e^{\mathbf{A}} \otimes e^{\mathbf{B}},$$

where

$$\mathbf{A}_{M \times M} \oplus \mathbf{B}_{N \times N} = \mathbf{A}_{M \times M} \otimes \mathbf{I}_{N \times N} + \mathbf{I}_{M \times M} \otimes \mathbf{B}_{N \times N}.$$

We have that

$$\text{vec} \left(e^{-\mathbf{A}^{j_1}(t-s)} \boldsymbol{\Sigma}^{j_1} \mathbf{H}^{j_1} (\mathbf{H}^{j_2})' (\boldsymbol{\Sigma}^{j_2})' e^{-\mathbf{A}^{j_2}(t-s)} \right) = \text{vec} \left(e^{-\mathbf{A}^{j_1} \oplus \mathbf{A}^{j_2}(t-s)} \right) \text{vec}(\boldsymbol{\Sigma}^{j_1} \mathbf{H}^{j_1} (\mathbf{H}^{j_2})' (\boldsymbol{\Sigma}^{j_2})'),$$

and, finally,

$$\begin{aligned} \text{vec}(\boldsymbol{\Gamma}^{j_1, j_2}(t)) &= \left(\int_0^t e^{-\mathbf{A}^{j_1} \oplus \mathbf{A}^{j_2}(t-s)} ds \right) \text{vec}(\boldsymbol{\Sigma}^{j_1} \mathbf{H}^{j_1} (\mathbf{H}^{j_2})' (\boldsymbol{\Sigma}^{j_2})') \\ &= (\mathbf{A}^{j_1} \oplus \mathbf{A}^{j_2})^{-1} e^{-\mathbf{A}^{j_1} \oplus \mathbf{A}^{j_2}(t-s)} \Big|_0^t \text{vec}(\boldsymbol{\Sigma}^{j_1} \mathbf{H}^{j_1} (\mathbf{H}^{j_2})' (\boldsymbol{\Sigma}^{j_2})') \\ &= (\mathbf{A}^{j_1} \oplus \mathbf{A}^{j_2})^{-1} \left(\mathbf{I} - e^{-\mathbf{A}^{j_1} \oplus \mathbf{A}^{j_2} t} \right) \text{vec}(\boldsymbol{\Sigma}^{j_1} \mathbf{H}^{j_1} (\mathbf{H}^{j_2})' (\boldsymbol{\Sigma}^{j_2})'). \end{aligned}$$

6.3.1. Three-factor models. In particular, let us focus on the 3-dimensional case, as in our specification described in Section 3.4, that we used in our application. In such a case,

$$\mathbf{H}^k = \begin{bmatrix} 1 & 0 & 0 \\ \rho_{12}^k & \sqrt{1 - (\rho_{12}^k)^2} & 0 \\ \rho_{13}^k & \frac{\rho_{23}^k - \rho_{13}^k \rho_{12}^k}{\sqrt{1 - (\rho_{12}^k)^2}} & \sqrt{\frac{1 + 2\rho_{12}^k \rho_{13}^k \rho_{23}^k - [(\rho_{12}^k)^2 + (\rho_{13}^k)^2 + (\rho_{23}^k)^2]}{1 - (\rho_{12}^k)^2}} \end{bmatrix}.$$

for $k \in \{1, \dots, J\}$, while

$$\mathbf{\Sigma}^k = \begin{bmatrix} \sigma_1^k & 0 & 0 \\ 0 & \sigma_2^k & 0 \\ 0 & 0 & \sigma_3^k \end{bmatrix},$$

hence

$$\mathbf{\Sigma}^k \mathbf{H}^k = \begin{bmatrix} \sigma_1^k & 0 & 0 \\ \sigma_2^k \rho_{12}^k & \sigma_2^k \sqrt{1 - (\rho_{12}^k)^2} & 0 \\ \sigma_3^k \rho_{13}^k & \sigma_3^k \frac{\rho_{23}^k - \rho_{13}^k \rho_{12}^k}{\sqrt{1 - (\rho_{12}^k)^2}} & \sigma_3^k \sqrt{\frac{1 + 2\rho_{12}^k \rho_{13}^k \rho_{23}^k - [(\rho_{12}^k)^2 + (\rho_{13}^k)^2 + (\rho_{23}^k)^2]}{1 - (\rho_{12}^k)^2}} \end{bmatrix}.$$

Now

$$\begin{aligned} \mathbf{\Sigma}^{j_1} \mathbf{H}^{j_1} (\mathbf{H}^{j_2})' (\mathbf{\Sigma}^{j_2})' &= \\ & \begin{bmatrix} \sigma_1^{j_1} & 0 & 0 \\ \sigma_2^{j_1} \rho_{12}^{j_1} & \sigma_2^{j_1} \sqrt{1 - (\rho_{12}^{j_1})^2} & 0 \\ \sigma_3^{j_1} \rho_{13}^{j_1} & \sigma_3^{j_1} \frac{\rho_{23}^{j_1} - \rho_{13}^{j_1} \rho_{12}^{j_1}}{\sqrt{1 - (\rho_{12}^{j_1})^2}} & \sigma_3^{j_1} \sqrt{\frac{1 + 2\rho_{12}^{j_1} \rho_{13}^{j_1} \rho_{23}^{j_1} - [(\rho_{12}^{j_1})^2 + (\rho_{13}^{j_1})^2 + (\rho_{23}^{j_1})^2]}{1 - (\rho_{12}^{j_1})^2}} \end{bmatrix} \times \\ & \begin{bmatrix} \sigma_1^{j_2} & \sigma_2^{j_2} \rho_{12}^{j_2} & \sigma_3^{j_2} \rho_{13}^{j_2} \\ 0 & \sigma_2^{j_2} \sqrt{1 - (\rho_{12}^{j_2})^2} & \sigma_3^{j_2} \frac{\rho_{23}^{j_2} - \rho_{13}^{j_2} \rho_{12}^{j_2}}{\sqrt{1 - (\rho_{12}^{j_2})^2}} \\ 0 & 0 & \sigma_3^{j_2} \sqrt{\frac{1 + 2\rho_{12}^{j_2} \rho_{13}^{j_2} \rho_{23}^{j_2} - [(\rho_{12}^{j_2})^2 + (\rho_{13}^{j_2})^2 + (\rho_{23}^{j_2})^2]}{1 - (\rho_{12}^{j_2})^2}} \end{bmatrix} = \\ & \begin{bmatrix} \sigma_1^{j_1} \sigma_1^{j_2} & \sigma_1^{j_1} \sigma_2^{j_2} \rho_{12}^{j_2} & \sigma_1^{j_1} \sigma_3^{j_2} \rho_{13}^{j_2} \\ \sigma_2^{j_1} \rho_{12}^{j_1} \sigma_1^{j_2} & \sigma_2^{j_1} \sigma_2^{j_2} \left(\rho_{12}^{j_1} \rho_{12}^{j_2} + \lambda^{j_1} \lambda^{j_2} \right) & \sigma_2^{j_1} \sigma_3^{j_2} \left(\rho_{12}^{j_1} \rho_{13}^{j_2} + \lambda^{j_1} \gamma^{j_2} \right) \\ \sigma_3^{j_1} \sigma_1^{j_2} \rho_{13}^{j_1} & \sigma_3^{j_1} \sigma_2^{j_2} \left(\rho_{13}^{j_1} \rho_{12}^{j_2} + \gamma^{j_1} \lambda^{j_2} \right) & \sigma_3^{j_1} \sigma_3^{j_2} \left(\rho_{13}^{j_1} \rho_{13}^{j_2} + \gamma^{j_1} \gamma^{j_2} + \psi^{j_1} \psi^{j_2} \right) \end{bmatrix}, \end{aligned}$$

where

$$\begin{aligned} \lambda^j &= \sqrt{1 - (\rho_{12}^j)^2}, \gamma^j = \frac{\rho_{23}^j - \rho_{13}^j \rho_{12}^j}{\sqrt{1 - (\rho_{12}^j)^2}}, \\ \psi^j &= \sqrt{\frac{1 + 2\rho_{12}^j \rho_{13}^j \rho_{23}^j - [(\rho_{12}^j)^2 + (\rho_{13}^j)^2 + (\rho_{23}^j)^2]}{1 - (\rho_{12}^j)^2}}, \quad j \in \{1, \dots, J\} \end{aligned}$$

Hence, the i_1, i_2 -th element of matrix, where $i_1, i_2 \in \{1, \dots, J\}$, is

$$[\mathbf{\Gamma}^{j_1, j_2}]_{i_1, i_2} = \frac{\sigma_i^{j_1} \sigma_j^{j_2}}{a_i^{j_1} + a_j^{j_2}} \left(1 - e^{-(a_i^{j_1} + a_j^{j_2})t} \right) [\mathbf{H}^{j_1} (\mathbf{H}^{j_2})']_{i_1, i_2}$$

and unconditional variance-covariance matrix, obtained by letting $t \rightarrow \infty$, leads to

$$(8) \quad [\mathbf{\Gamma}^{j_1, j_2}]_{i_1, i_2}^\infty = \frac{\sigma_i^{j_1} \sigma_j^{j_2}}{a_i^{j_1} + a_j^{j_2}} [\mathbf{H}^{j_1} (\mathbf{H}^{j_2})']_{i_1, i_2}.$$



2016 © IMT School for Advanced Studies, Lucca
Piazza San ponziano 6, 5100 Lucca, Italy
www.imtlucca.it