

Article

A Parametric Factor Model of the Term Structure of Mortality

Niels Haldrup  and Carsten P. T. Rosenskjold *

Center for Research in Econometric Analysis of Time Series (CREATES), Department of Economics and Business Economics, Aarhus University, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark; nhaldrup@econ.au.dk

* Correspondence: cthillemann@econ.au.dk

Received: 11 June 2018; Accepted: 5 March 2019; Published: 11 March 2019



Abstract: The prototypical Lee–Carter mortality model is characterized by a single common time factor that loads differently across age groups. In this paper, we propose a parametric factor model for the term structure of mortality where multiple factors are designed to influence the age groups differently via parametric loading functions. We identify four different factors: a factor common for all age groups, factors for infant and adult mortality, and a factor for the “accident hump” that primarily affects mortality of relatively young adults and late teenagers. Since the factors are identified via restrictions on the loading functions, the factors are not designed to be orthogonal but can be dependent and can possibly cointegrate when the factors have unit roots. We suggest two estimation procedures similar to the estimation of the dynamic Nelson–Siegel term structure model. First, a two-step nonlinear least squares procedure based on cross-section regressions together with a separate model to estimate the dynamics of the factors. Second, we suggest a fully specified model estimated by maximum likelihood via the Kalman filter recursions after the model is put on state space form. We demonstrate the methodology for US and French mortality data. We find that the model provides a good fit of the relevant factors and, in a forecast comparison with a range of benchmark models, it is found that, especially for longer horizons, variants of the parametric factor model have excellent forecast performance.

Keywords: mortality forecasting; term structure of mortality; factor modelling; cointegration

JEL Classification: C1; C22; J10; J11; G22

1. Introduction

The [Lee and Carter \(1992\)](#) (LC) model has become a benchmark model when estimating and forecasting improvements in age-specific death rates and the calculation of life expectancy. The model is basically a one-factor model that allows for a single common time trend with age-specific loadings. The model has been extended in many different ways. For instance, [Booth et al. \(2002\)](#) and [Renshaw and Haberman \(2003\)](#) extend the model with a second common time trend with age-specific loadings. [Hyndman and Ullah \(2007\)](#) developed a functional data approach in which the data are smoothed across age prior to modelling using penalized regression splines and principal component analysis. We will refer to these models as nonparametric factor models. [De Jong and Tickle \(2006\)](#) use the state space framework to establish smoothness in the LC model using b-splines.

Typically, the estimation of factors is implemented nonparametrically via either singular value decomposition or principal component analysis. For models with multiple factors, these are identified via orthogonalization. Subsequently, the factors are modelled as individual time series models which can be used for forecast projections.

The LC model and its extensions are basically statistical models that summarize the variability of the measured age-specific death rates over time in a parsimonious way. No structure is imposed in the model specification. However, in the demographics literature on mortality laws, it is well established that age groups are exposed rather differently to death risk and it seems reasonable to believe that separate time factors may affect different age groups rather than assuming a single common factor as in the basic LC model.

Mortality laws for death rates observed at a given time have been suggested by amongst others [Gompertz \(1825\)](#); [Makeham \(1860\)](#); and [Heligman and Pollard \(1980\)](#); [Tabeau et al. \(2001\)](#) provide a review. These laws refer to separate mortality characteristics for different age groups such as infants, youths, adults, and the elderly. When accounting for the dynamic development of mortality over time, it seems natural to consider a factor model that accounts for these mortality laws. In this paper, we assume the presence of multiple factors and impose structure on the loadings via specific functional forms. The approach is similar to [McNown and Rogers \(1989\)](#). However, their model is both heavily over parametrized in terms of latent time-varying parameters and does not fully exploit the information contained in the factor dynamics of the model.

The idea is similar to e.g., the dynamic Nelson–Siegel model for the term structure of interest rates—see [Diebold and Li \(2006\)](#). Diebold and Li suggest a factor model with parametrized factor loadings which identify level, slope, and curvature of the yield curve, associated with the long, short, and medium term yields. In the context of the term structure of mortality, we define loading functions that identify the factors that drive infant, adult, and ‘accident hump’ (youth) mortality, respectively, plus a common factor uniformly affecting all age groups. We will generally refer to this class of factor models as parametric factor models (PFM). It follows from this approach that, opposed to traditional factor analysis, the factors to be extracted will not necessarily be independent. In fact, the factors may potentially cointegrate when these are found to have unit roots.

We consider estimation of the model parameters and the factors by cross-section regressions over age groups for each period of time. These estimations are conducted over a grid of tuning parameters that define the shape of the loading functions. Next, a least squares criterion is used to determine the desired tuning parameters and the corresponding factor elements. This approach is similar to the first step of the cross section projection procedure suggested in [Diebold and Li \(2006\)](#). After the factors have been extracted, the second step implies the estimation of a time series model for the factors. This can be done in different ways. For instance, univariate as well as multivariate models for the factors can be formulated and with the possibility of stationary as well as non-stationary factors that potentially cointegrate. The final time series specification of the factor dynamics is an empirical question and separate time series models are considered for this purpose.

We also consider a fully parametrized model specification formulated as a state space model. By use of the Kalman filter recursions, the model parameters and the factors can be estimated by full maximum likelihood. This approach is similar to that of [Diebold et al. \(2006\)](#) for the term structure of interest rates.

The proposed model for women and men is estimated using French and US data for the sample period 1950–2014. The estimated functional forms appear to be rather similar across countries with the duration of the accident hump being longer for men than for women. The shape of the four factors also generally appear similar across countries but with differences across genders. In terms of model fit compared with the LC model, it appears that our model is doing especially well for explaining the age-specific death rates around the age groups defining the ‘accident hump’.

We also evaluate the out of sample performance of the model where the predicted mortality rates are summarized in a loss function defined by the life expectancy. Specifically, we apply the model confidence set procedure of [Hansen et al. \(2011\)](#) to evaluate the relative forecast performance on the horizons of 1, 10, and 20 years ahead using a number of benchmark models. We find that, particularly for long horizon forecasts, our model tends to be in the set of best predicting models.

In Section 2, we briefly describe the LC model and provide a detailed description of the mortality data and set up a number of stylized facts of the mortality curve in Section 3. Section 4 introduces the PFM and its interpretation. Section 5 discusses estimation of the PFM and in Section 6 we present the empirical analysis, including the estimation results, the model fit, and the factor dynamics. Section 7 examines the relative out-of-sample forecast performance. Section 8 provides conclusions.

2. The Lee–Carter Model

The observed data of the analysis are the age specific death rates $m_{x,t}$ for age groups $x = 0, 1, \dots, N$ at year $t = t_0, \dots, T$, broadly defined as the number of deaths at age x in year t divided by the Exposure-to-Risk, which is the average population aged x in year t . The data used is obtained from the [Human Mortality Database \(2016\)](#).

The [Lee and Carter \(1992\)](#) (LC) model describes the (log) age-specific death rates by:

$$\ln m_{x,t} = \alpha_x + \beta_x \kappa_t + \varepsilon_{x,t}, \quad (1)$$

where α_x captures the average death rate for each age x . κ_t is a common time varying factor capturing the general trend in death rates over time t . β_x is the factor loading capturing the effect of the factor κ_t on each age group x and $\varepsilon_{x,t}$ is the age and time specific error term. The LC model is basically a one-factor model that allows a common time trend to have age-specific loadings with respect to the development of the age-specific log death rates. [Lee and Carter \(1992\)](#) obtain identification using the normalizations $\sum_{x=0}^N \beta_x = 1$ and $\sum_{t=t_0}^T \kappa_t = 0$. The constraints imply that α_x measures the age specific time-average of the log death rates, $\ln m_{x,t}$.¹ To estimate β_x and κ_t , the singular value decomposition is applied to the matrix $(A)_{xt} = \ln m_{xt} - \alpha_x$ for all x, t . [Lee and Carter \(1992\)](#) find that κ_t can be modelled as a random walk with drift, although they allow for other specifications as well. The LC model is designed to maximize the in-sample fit by fitting a general factor model structure to the death rates. Note that the LC model does not impose any particular structure on the age-specific graduation of mortality, which essentially is data driven. However, it imposes a rigid structure on the improvements of the age-specific death rates over time by requiring these to be proportional and governed by the single factor κ_t .

3. Stylized Facts of the Mortality Curve

A good mortality model should desirably account for both the age (cross section) dimension of mortality as well as its development over time, i.e., the time dimension. Here, we describe some stylized facts of the (log) death rates to be modelled. For illustration, we use data for France and USA available from the [Human Mortality Database \(2016\)](#).²

The age dimension: To illustrate the age dimension properties, which a good mortality model should be able to capture, we show the log mortality on 10 year intervals from 1950 to 2010 for men and women in Figure 1a–d for the US and France. The mortality curve shows a similar shape over the ages, but the level of mortality tends to decline over time; the shape is very similar across both genders and countries. The infant mortality is seen to decline rapidly during early childhood. In the late teens, the mortality rate experiences a rapid increase often termed the ‘accident hump’, which appears either as a distinct hump or as a flattening out of the death rates; see [Heligman and Pollard \(1980\)](#). After the accident hump, the mortality rates are gradually increasing with age (log-linearly). Thus, for a model to produce realistic results, these three facts should hold for each year. The three properties could also

¹ Following [Nielsen and Nielsen \(2014\)](#), the choice of restrictions is of no importance for the resulting forecasts. Other normalizations could be considered; however, this gives an intuitive interpretation of α_x .

² We restrict the data to 1950 and onwards as this removes outliers, and we avoid structural changes in the exposure; see [Lee and Miller \(2001\)](#) and [Booth et al. \(2002\)](#). To avoid uncertainty about the death rates, due to a few observations, we further restrict the ages to cover the ages 0 to 95 as is standard in the mortality forecasting literature.

be interpreted in terms of biological reasonableness as described by Cairns et al. (2006), which rules out patterns that are biologically unreasonable such as a decreasing mortality curve for the older as well as the crossing over of age-specific mortality rates.

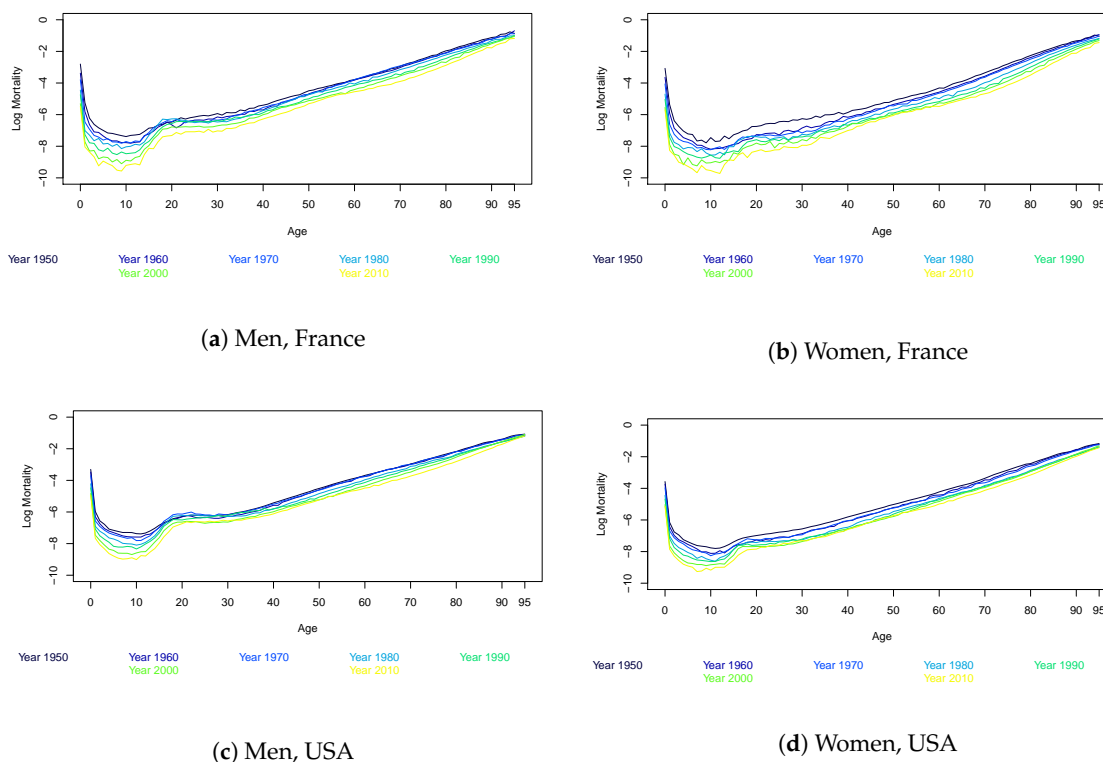


Figure 1. The log age-specific death rates for the years 1950, 1960, 1970, 1980, 1990, 2000, and 2010 for men and women in France and the USA.

The time dimension: When investigating the time dynamics in the development of the log age-specific death rates, the review paper by Wong-Fillipp and Haberman (2004, p. 56) notes that “There is a broad consensus across the resulting projections: (1) an approximately log-linear relationship between mortality rates and time, (2) decreasing improvements according to age”. The first point helps to explain the success of the LC model where the common time-varying factor is found to evolve almost linearly in most applications—see, e.g., Lee and Miller (2001) and Callot et al. (2016). The log-linear development of death rates over time is illustrated in Figure 2a–d. The second observation of decreasing improvements in mortality with respect to age can be described by the so called *compensation effect of mortality*—see, e.g., Gavrilov and Gavrilova (1979, 1991).³ In Figure 2a–d, this effect is seen by a slope of the log mortality-time plot that decreases with age.

Several studies find that a unit root is present in the individual age-specific death rates—see, for instance, Lazar and Denuit (2009). In addition, it is common that the time-factor of the LC model is modelled as a random walk with drift. Basically, this means that all death rates are governed by the same stochastic time trend component and hence for a system of $N + 1$ age groups, all death rates cointegrate pairwise and a total of N cointegrating relations exists among all age-specific death rates.

³ This is also called the Strehler-Mildvan correlation due to Strehler and Mildvan (1960).

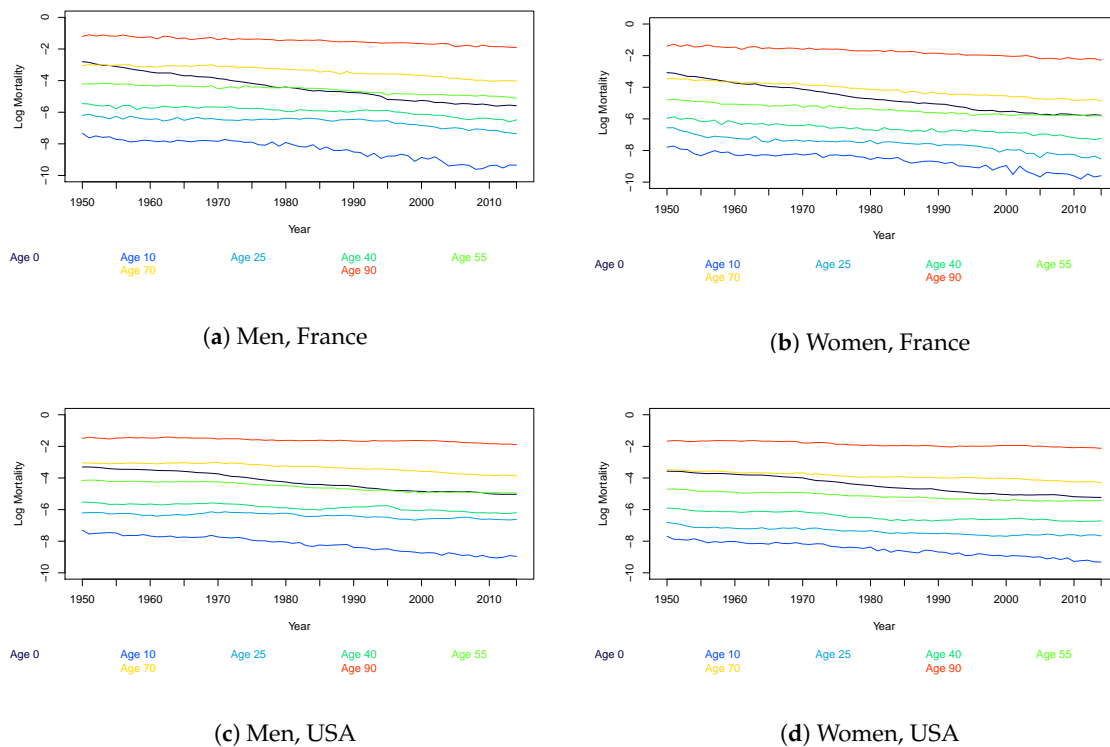


Figure 2. The log age-specific death rates for a range of ages for French and American men and women from 1950–2014.

To examine this feature of the LC model, we have conducted cointegration tests of all the pairwise combinations of log mortality across age groups. Note that, due to the dimension of the data, a full cointegration analysis cannot be conducted for the full data set. Figure 3a,d report a heatmap of the p -values from Johansen's trace test, Johansen (1991), of the null hypothesis of zero cointegrating relations against one cointegrating relation for all combinations of the log age-specific death rates. As seen, we cannot reject the null of no cointegration for most of the death rate pairs, especially for US data. It is apparent that most of the cointegrating relations found are between the adjacent ages, i.e., along the diagonal line. For both countries, we find clear rejection of no cointegration amongst the youngest ages, but not for newborns. For France, rejection of no cointegration amongst the oldest ages is found to a larger degree. Furthermore, it is found that around the accident hump and for the infants we cannot reject the null for relatively adjacent ages. Thus, overall the figures clearly show that the assumption of the LC model of N cointegrating relations is not consistent with the mortality data. We note that this is also consistent with Lazar and Denuit (2009) who found multiple stochastic trends when investigating cointegration across seven age groups of five-year age intervals. The stochastic trends driving mortality over time are generally different across the age groups. The model we are subsequently going to propose will not have the restrictive feature of the LC model, since different factors are constructed to affect separate age groups.

In summary, we observe seven stylized facts for the term structure of mortality that a good mortality model should be able to reproduce: (1) declining mortality for infants, (2) increasing mortality around the accident hump, (3) log-linearly increasing mortality with age for adults, (4) a log-linear relationship between the death rates and time, (5) the log age-specific death rates are integrated of order one around a linear trend, (6) decreasing improvements in mortality with age, and (7) multiple stochastic trends characterize the development of log mortality over time for the different age groups.

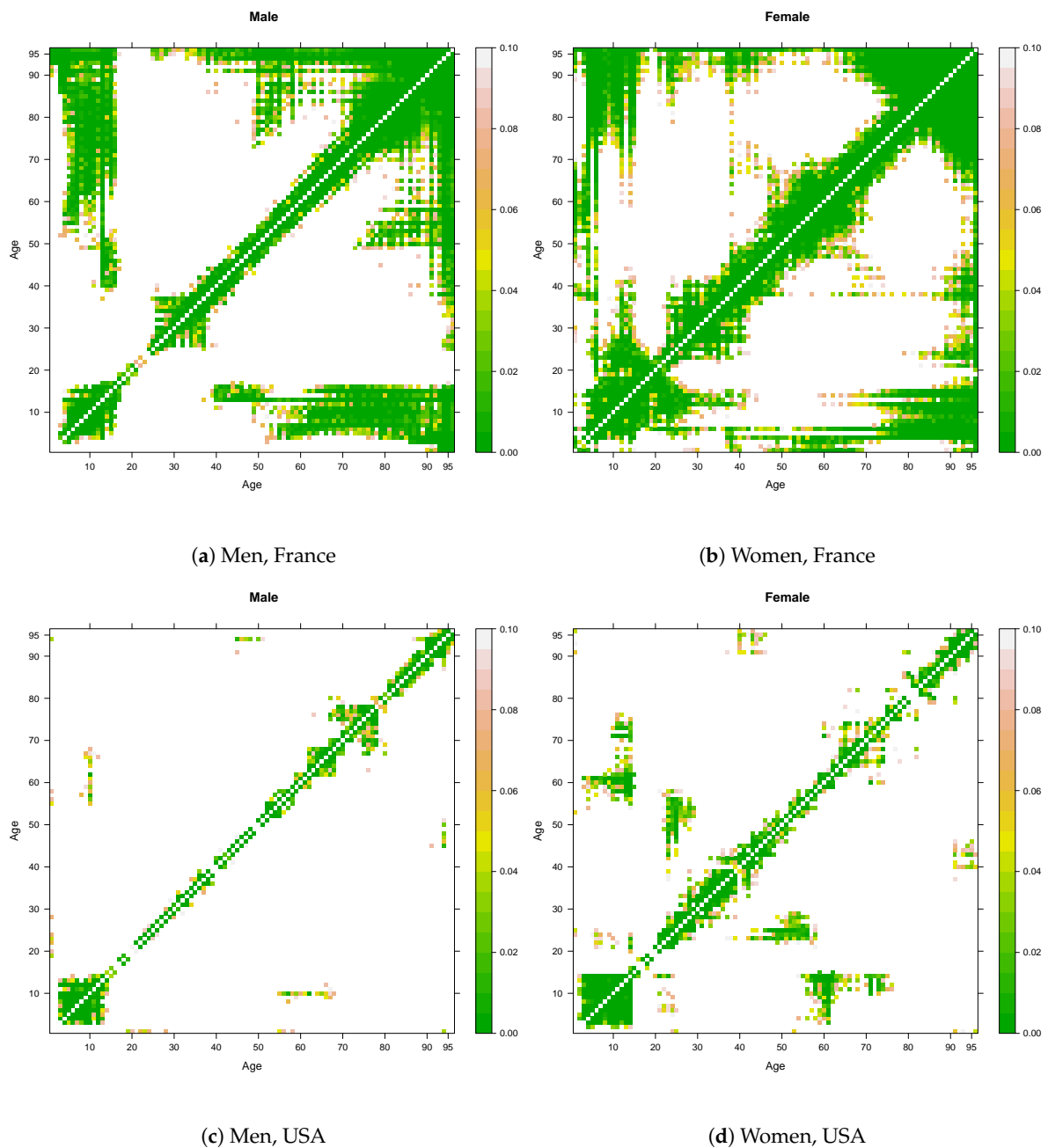


Figure 3. *p*-Values from Johansen’s Trace test are shown for all pairwise combinations of the (log) age-specific death rates for French and US men and women over the period 1950–2014. The test is performed with a restricted time trend in the cointegrating relation and 1 lag in the VAR specification. The *p*-Values are obtained via the gamma approximation following Doornik (1998, 1999) and shown for significance levels between 0 and 0.10.

4. The Parametric Factor Model for the Term Structure of Mortality

The model we propose assumes that mortality is driven by multiple factors and we impose structure on the factor loadings capturing the regularities discussed in the previous section.

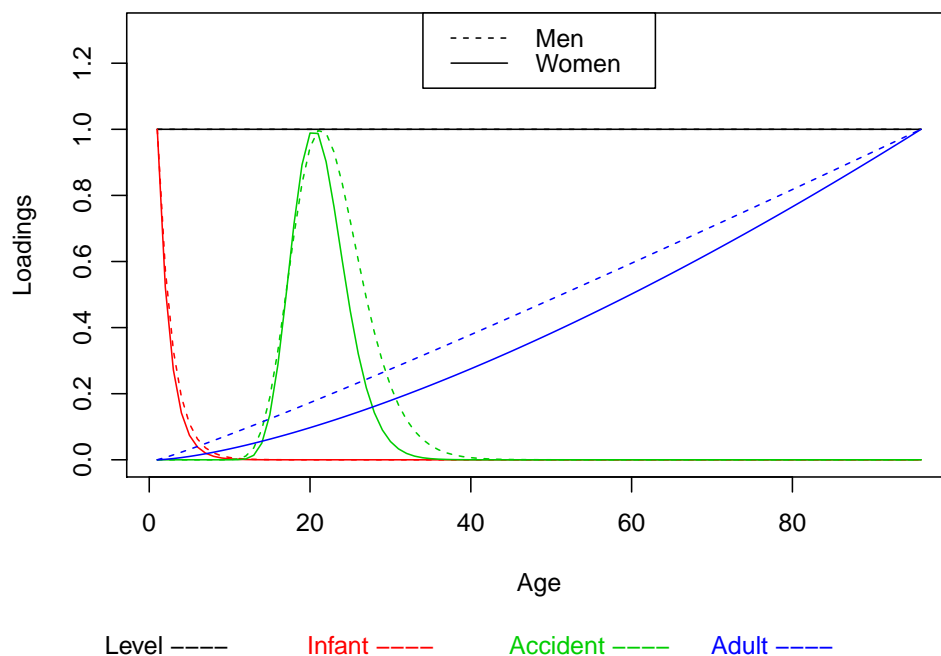
The PFM reads as follows:

$$\ln m_{x,t} = \kappa_{0,t} + \kappa_{1,t}e^{-\lambda_1 x} + \kappa_{2,t}e^{-\lambda_2(\ln(x)-\ln(k))^2} + \kappa_{3,t}\left(\frac{x}{N}\right)^{\lambda_3} + \varepsilon_{x,t}. \tag{2}$$

The model has four factors $\kappa_{i,t}$, $i = 0, 1, 2, 3$ with loading functions that are designed to capture distinct age groups. The shape of the loading functions are governed by the constant parameters $\lambda_1, \lambda_2, \lambda_3$ and k which are assumed positive. N is the maximum age used for the analysis, which is set to 95 due to data quality, as described in Section 3. $\kappa_{0,t}$ is a factor that is common to all age groups. The factor $\kappa_{1,t}$ captures child mortality, $\kappa_{2,t}$, the accident hump, and finally, $\kappa_{3,t}$ is a factor that tends to increase mortality with age. Note that the common factor has the constant loading one for all age groups. The loading for infant mortality declines rapidly with age. The loading for the accident hump is approximately bell-shaped around age k , which is estimated to be the age at which the accident hump equals one—see Figure 4a,b below. Finally, the loading for the adult factor grows almost linearly with age when λ_3 is close to one. The error term $\varepsilon_{x,t}$ is assumed to be IID normally distributed as $N(0, \sigma^2)$ for all ages and years.⁴ The loading functions estimated for France and the US are shown in Figure 4a,b; the estimation procedure will be discussed in the next section. Even though it may be claimed that the functional forms of the loading functions are arbitrary, they are designed such that the mortality laws and stylized facts, described in Section 3, are captured through the model specification.

The level and infant terms $\kappa_{0,t}$ and $\kappa_{1,t}e^{-\lambda_1 x}$, respectively, are used in many models using the age-specific graduation of mortality, see e.g., Siler (1979) and Rogers and Little (1994). The accident hump loading $e^{-\lambda_2(\ln(x) - \ln(k))^2}$ is taken from Heligman and Pollard (1980). The adult factor can be seen as a generalization of the Gompertz model, inspired by the Box and Cox (1964) power formulation. That is, the loading function captures the Gompertz specification if $\lambda_3 = 1$.

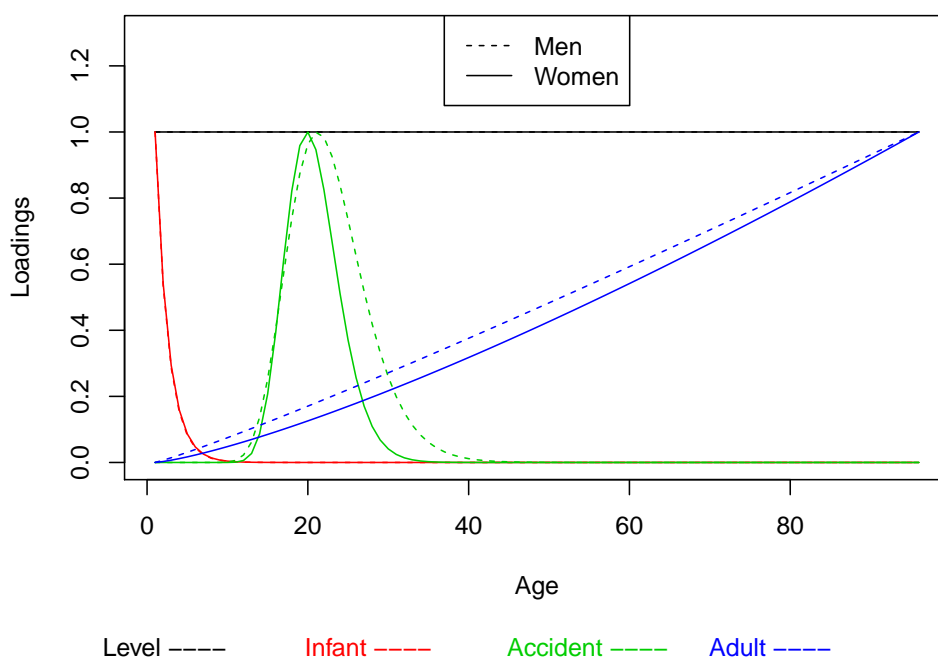
It is clear that the single factors $\kappa_{i,t}$ are only identified when λ_i is non-zero. If some λ_i is zero, it means that the associated factor is absent and can be left out from the analysis. Identification of factors when λ_i are non-zero is a result of imposing a particular functional form on the loadings. Hence, the identification issue of the Lee–Carter model is absent in the present model. See Nielsen and Nielsen (2014) about a general discussion of identification in mortality models.



(a) Loadings, France.

Figure 4. Cont.

⁴ This is similar to Lee and Carter (1992) who assumed a homoskedastic error term for the LC model. The i.i.d. homoscedasticity assumption is necessary for the analysis of the present paper, but the assumption may be critical in certain cases—see, e.g., Doz et al. (2011).



(b) Loadings, USA.

Figure 4. Plot of the estimated loading functions for the years 1950–2014 for men and women in France and USA. The loading functions correspond to the level, infant, accident hump, and adult age groups, respectively. The loadings are estimated following the two-step procedure described in Section 5.

5. Estimation Procedure for the Parametric Factor Model

We consider two estimation procedures for estimating the PFM, the two-step procedure of [Diebold and Li \(2006\)](#) and exact maximum likelihood estimation using the Kalman Filter recursions of the model written on state-space form similar to [Diebold et al. \(2006\)](#). Alternatively, one could use maximum likelihood estimation following [Brouhns et al. \(2002\)](#) assuming a Poisson distribution for the death counts.

5.1. The Two-Step Estimation Procedure

The two-step procedure considers first to estimate the model parameters and the factors of the model and, second, estimating a time series model of the extracted factors with the primary purpose of forecasting. Regarding the first step, [McNown and Rogers \(1989\)](#) propose to estimate the factors by nonlinear least squares for each point in time, hence giving a time series of the factors. This allows not only the factors but also the model (loading) parameters to be time-varying. [McNown and Rogers \(1992\)](#) fix the parameters of the model a priori and estimate the factors in a sequence of cross-section regressions. The latter procedure is also the one adopted by [Diebold and Li \(2006\)](#) when estimating the dynamic Nelson–Siegel model for the term structure of interest rates, where the different loadings refer to the level, slope, and curvature of the yield curve.

We suggest modifying [McNown and Rogers \(1992\)](#) and [Diebold and Li \(2006\)](#) by considering cross-sectional regressions at each time t for a fine grid of the model parameters and select the preferred model by minimizing the conditional sum of squares function. Hence, for a given set of loading parameters, the factors can simply be estimated by using ordinary least squares for each year. This can also be implemented by a nonlinear least squares optimization algorithm. Here, we use the limited memory BFGS procedure (“L-BFGS-B”) developed by [Byrd et al. \(1995\)](#) and implemented via the R package ‘Optim’ ([R Core Team 2015](#)). This step provides estimates of the four factors of the model. Note that, as opposed to traditional factor models, generally the estimated factors will not be orthogonal and in fact are most likely to be dependent. In the second step of the two-step procedure, time-series models are fitted to the factors. This step is only needed when the model is used

for predictions as we shall see in Section 7. These can be univariate time series models such as ARIMA model specifications, possibly with drifts or trends, or the factors can be modelled as stationary or nonstationary VAR models which potentially can allow for cointegration amongst the factors. It is an empirical question to properly select a time series model in the second step.

5.2. One-Step Estimation

The parametric factor model in Equation (2) can be formulated on state-space form and estimated by maximum likelihood by use of the Kalman Filter, see e.g., Durbin and Koopman (2012). This estimation procedure improves on the two-step estimation procedure by allowing joint estimation of the latent factors and their transition dynamics as well as the unknown parameters assuming Gaussian errors. Estimating the system jointly delivers the appropriate likelihood quantities, unlike the two-step approach, which ignores the uncertainty and estimation errors from the first step. We conjecture that standard inference results hold for the Gaussian Maximum Likelihood approach, although we do not provide a formal proof for this; see also Koopman et al. (2010).

The measurement equation of the state space model can be written as:

$$\ln \mathbf{m}_t = \Lambda \kappa_t + \varepsilon_t,$$

where

$$\ln \mathbf{m}_t = \begin{pmatrix} \ln m_{0,t} \\ \ln m_{1,t} \\ \vdots \\ \ln m_{N,t} \end{pmatrix}, \varepsilon_t = \begin{pmatrix} \varepsilon_{0,t} \\ \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{N,t} \end{pmatrix}, \kappa_t = \begin{pmatrix} \kappa_{0,t} \\ \kappa_{1,t} \\ \kappa_{2,t} \\ \kappa_{3,t} \end{pmatrix}$$

and

$$\Lambda = \begin{pmatrix} 1 & e^{-\lambda_1 \cdot 0} & e^{-\lambda_2(\ln(0) - \ln(k))^2} & \left(\frac{0}{N}\right)^{\lambda_3} \\ 1 & e^{-\lambda_1 \cdot 1} & e^{-\lambda_2(\ln(1) - \ln(k))^2} & \left(\frac{1}{N}\right)^{\lambda_3} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & e^{-\lambda_1 \cdot N} & e^{-\lambda_2(\ln(N) - \ln(k))^2} & \left(\frac{N}{N}\right)^{\lambda_3} \end{pmatrix}.$$

As in Section 4, the vector error term ε_t is assumed to be normally distributed as $N(0, I\sigma^2)$, where I is the identity matrix.

The transition equation of the state space model should be formulated to capture the dynamics of the factors. For instance, if we assume that the factors are governed by a VAR(1) process in first differences, the transition equation can be specified as:

$$\begin{pmatrix} \kappa_t \\ \Delta \kappa_{t+1} \end{pmatrix} = \begin{bmatrix} I_4 & I_4 \\ 0 & \Phi \end{bmatrix} \begin{pmatrix} \kappa_{t-1} \\ \Delta \kappa_t \end{pmatrix} + \begin{pmatrix} 0 \\ c \end{pmatrix} + \begin{pmatrix} 0 \\ v_t \end{pmatrix},$$

where v_t is multivariate normal distributed as $N(0, \Sigma)$ and c is a four-dimensional vector constant.

In the case where the factors cointegrate with r cointegrating relations, the transition dynamics can be written as:

$$\begin{pmatrix} \kappa_t \\ \Delta \kappa_{t+1} \end{pmatrix} = \begin{bmatrix} I_4 & I_4 \\ \alpha \gamma' & \alpha \gamma' \end{bmatrix} \begin{pmatrix} \kappa_{t-1} \\ \Delta \kappa_t \end{pmatrix} + \begin{pmatrix} 0 \\ c \end{pmatrix} + \begin{pmatrix} 0 \\ v_t \end{pmatrix},$$

where v_t is multivariate normal distributed as $N(0, \Sigma)$. Both α and γ are $4 \times r$ matrices. The second row gives the desired VECM specification for the transition dynamics:

$$\Delta \kappa_{t+1} = \alpha \gamma' \kappa_t + c + v_t. \tag{3}$$

Note that the constant c is treated as a state parameter vector within the Kalman filter. Estimation of the parameters $\psi = [\lambda_1, \lambda_2, \lambda_3, k, \sigma, \Sigma, \Phi \text{ (or } \alpha, \gamma), c]$ is achieved via numerical optimization of the prediction error decomposition of the likelihood function:

$$\mathcal{L}(\psi) = -\frac{NT}{2} \ln 2\pi - \frac{1}{2} \sum_{t=1}^T \ln |F_t| - \frac{1}{2} \sum_{t=1}^T v_t' F_t^{-1} v_t, \tag{4}$$

where v_t is the one step (innovation) prediction error of the measurements equation and F_t is the innovation covariance matrix of the measurement equation. The numerical optimization is performed via the low-memory BFGS procedure “L-BFGS-B” from [Byrd et al. \(1995\)](#) in the R package [Optim \(R Core Team 2015\)](#).

6. Empirical Analysis

6.1. Estimates Using the Two-Step Procedure

Figure 4a,b in Section 4 display the estimated shape of the loading functions for French and US men and women based on the two-step procedure. Table 1 reports the estimated shape parameters and their standard errors. Note that all parameters are significantly different from zero and hence the factors are identified. The estimated parameters are similar across countries. However, the loading functions for the adult curvature are more convex for women than for men. Similarly, the shape and location of the accident hump vary across genders with men suffering from the accident hump longer than for women.

Table 1. Estimated loading function parameters and standard errors from the first step in the two-step procedure for French and US men and women. The standard errors are calculated using the inverse Fisher information criterion.

| | | Men | | | | |
|----|----------|-------------|-------------|-------------|--------|------------|
| | | λ_1 | λ_2 | λ_3 | k | σ^2 |
| Fr | Estimate | 0.553 | 11.981 | 1.093 | 20.308 | 0.020 |
| | Std. Err | 0.013 | 0.024 | 0.004 | 0.002 | 0.018 |
| US | Estimate | 0.624 | 10.813 | 1.103 | 20.016 | 0.017 |
| | Std. Err | 0.013 | 0.020 | 0.003 | 0.002 | 0.018 |
| | | Women | | | | |
| Fr | Estimate | 0.649 | 18.092 | 1.453 | 19.492 | 0.023 |
| | Std. Err | 0.013 | 0.060 | 0.003 | 0.005 | 0.018 |
| US | Estimate | 0.607 | 19.029 | 1.295 | 18.675 | 0.013 |
| | Std. Err | 0.010 | 0.042 | 0.003 | 0.004 | 0.018 |

The estimated factors are shown in Figure 5a–d for France and Figure 6a–d for the US.

A number of insights follow from these figures. The factor governing the common mortality level decreases almost linearly and thus capturing a common decline in mortality across all age groups; this applies for both genders and countries. The infant factor for both men and women decline over the period showing that the infants have seen larger improvements in mortality reduction compared to the general level captured by the first factor. Moreover, it can be seen that the decline for the infant factor stagnates around 1995 for all populations considered. Hence, after 1995, the development in mortality for infants has generally followed the common rate.

The accident hump factor shows an increase in size from 1950 to about 1990 followed by stagnation for all but US men. Regarding the development of the adult factor, Figures 5d and 6d exhibit an upward slope over the sample period and hence reduce the mortality improvements for the relevant age group.

Thus, slower improvements in mortality with age are captured by the model, in line with the stylized facts previously reported.

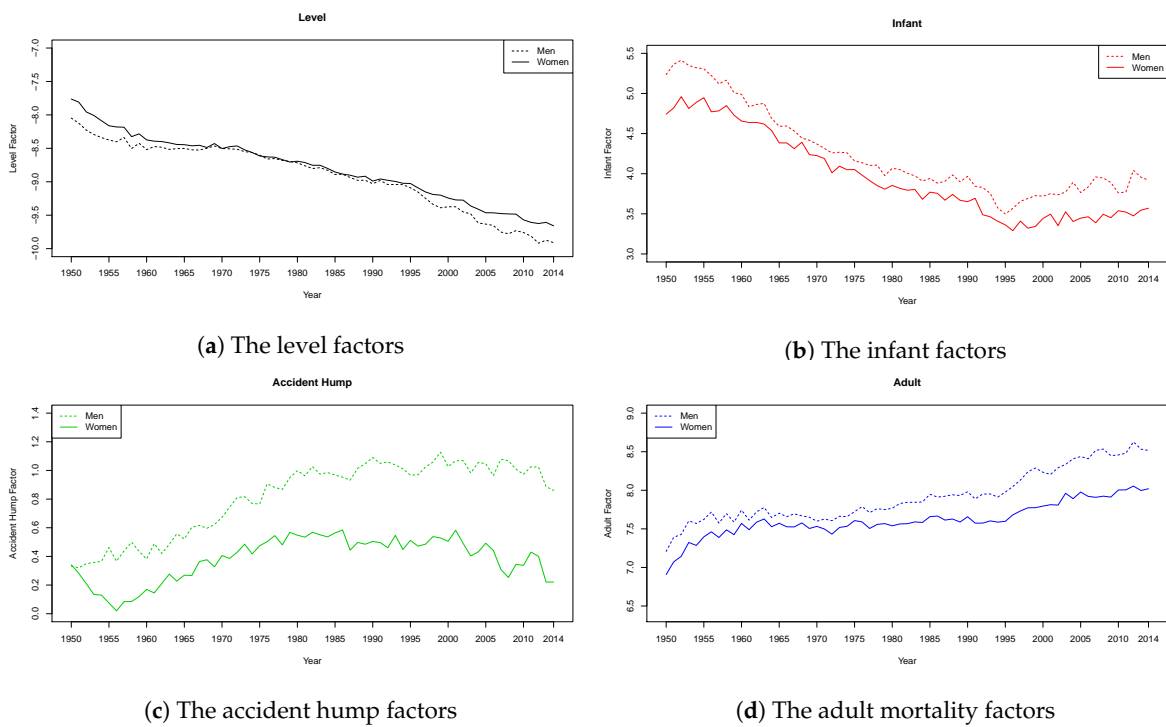


Figure 5. The factors $\kappa_{i,t}$, $i = 0, 1, 2, 3$ estimated by the two-step procedure for France using data from 1950–2014. The plots are showing the level factor, infant factor, accident hump factor, and adult factor for both genders, respectively.

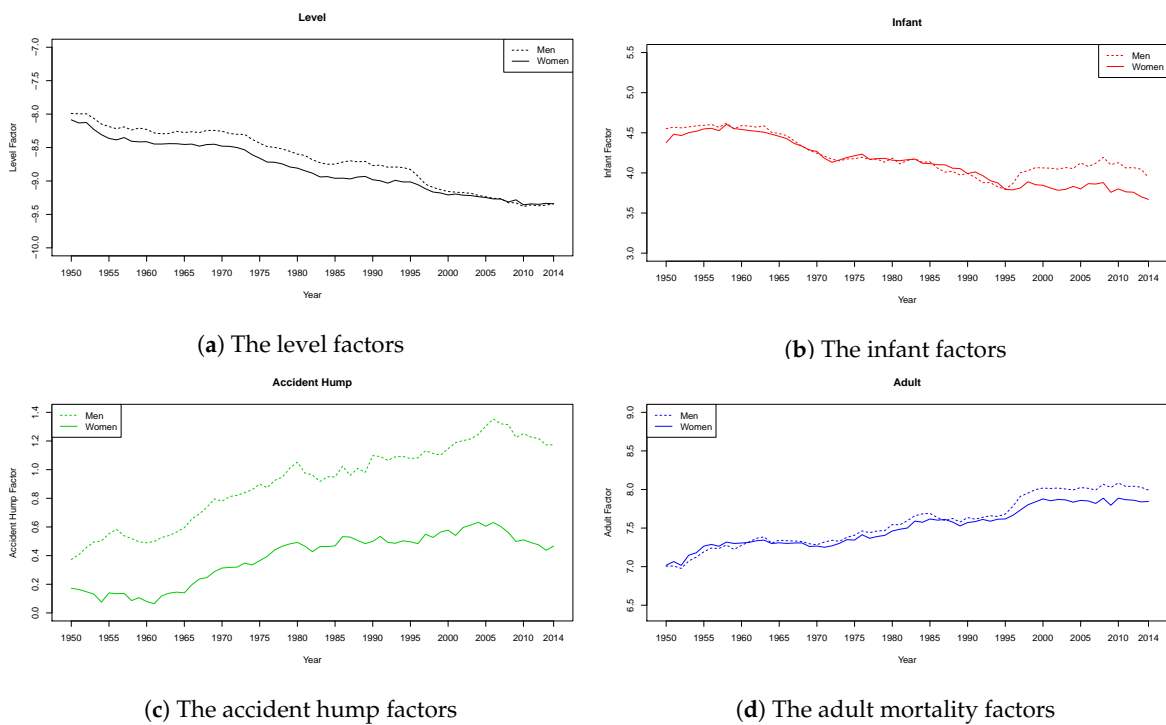


Figure 6. The factors $\kappa_{i,t}$, $i = 0, 1, 2, 3$ estimated by the two-step procedure for USA using data from 1950–2014. The plots are showing the level factor, infant factor, accident hump factor, and adult factor for both genders, respectively.

6.2. Cointegrating Analysis of the Factors

In order to use the estimated model for forecast projections, we need to examine the time series features of the estimated factors $\kappa_{i,t}, i = 0, 1, 2, 3$. By using a range of unit root tests, we find strong empirical support for the presence of unit roots, possibly with a drift, in all of the factors considered across both countries and gender. Given this observation, it is not surprising that the age-specific log death rates individually appear to have similar time series characteristics. The one-factor model of Lee and Carter (1992) also typically model the factor as a random walk with drift. From visual inspection of the factors in Figures 5 and 6, it is evident that the various factors tend to co-move across gender and thus the factors are likely to cointegrate. Accounting for cointegration amongst the factors will potentially lead to superior forecasts.

We have conducted cointegration analysis using the Johansen (1988) trace test for different subsets of factors. In Table 2, we report the test results for each country and for each gender using all four factors. In Table 3, we examine tests for each country using all eight factors for both men and women, and, finally, Table 4 displays the tests for men and women, respectively, by merging the factors across countries.

The results are rather different for the USA and France as can be seen from Table 2. For both genders, the US factors are found not to cointegrate and hence these factors are driven by four separate common stochastic trends. On the other hand, the factors for French men and women appear to cointegrate with two or three cointegrating vectors and thus the factors for each gender appear to be driven by a single or possibly two common stochastic trends. This finding is also in line with the heat maps reported in Figure 3 showing that, for France, the pairwise log mortality rates appear more cointegrated compared to the USA.

Table 2. Test for cointegration rank amongst factors for US and French men and women.

| USA | | | | |
|--------|------------|------------|------------|------------|
| | Men | | Women | |
| Rank | Trace-Test | p-Value | Trace-Test | p-Value |
| 0 | 52.240 | [0.323] | 47.939 | [0.512] |
| 1 | 27.748 | [0.641] | 30.689 | [0.468] |
| 2 | 13.926 | [0.668] | 15.243 | [0.562] |
| 3 | 1.1522 | [0.992] | 3.0677 | [0.858] |
| France | | | | |
| 0 | 106.790 | [0.000] ** | 108.730 | [0.000] ** |
| 1 | 56.044 | [0.001] ** | 47.599 | [0.014] * |
| 2 | 28.089 | [0.024] * | 24.893 | [0.064] |
| 3 | 3.642 | [[0.788] | 9.2236 | [0.171] |

Note: The Johansen trace test is calculated with a trend restricted to the cointegration space. The number of lags in the VAR is 1 for all cases. “***” and “**” signify significance at the 1% and 5% level, respectively.

In Table 3, the set of variables in the VAR model is expanded to include both men and women for each country. In this case, the eight factors for the US data are driven by four common stochastic trends. It is tempting to believe that the factors cointegrate across genders, however, a formal statistical test rejects this hypothesis. For the French data, the eight factors have six to seven cointegrating vectors and thus have one or two common stochastic trends. Again, a formal test rejects that the factors cointegrate pairwise across genders.

Finally, Table 4 shows that, when pooling the US and French data for men and women, respectively, both the male and female factors are likely to be driven by six factors and thus have two cointegrating relations. Hence, cross-country similarities exist across countries for both genders but only to a limited extent.

These findings demonstrate that different time series specifications should be considered when modelling the factors with the purpose of forecasting. For the US, it seems appropriate to specify a VAR in first differences with a vector of unrestricted constants to capture the drift of the single series. It could also be considered to base predictions on an expanded (cointegrated) VAR model including factors for both genders. For France, a cointegrated VAR with cointegration rank two or three seems appropriate. An expanded cointegrated VAR with eight factors and six to seven cointegrating vectors is also possible. When modelling the factors as univariate time series models, a random walk with drift specification is appropriate, but, since the cross dependence of factors is neglected in this case, it is likely that inferior forecasts will result.

Table 3. Test for cointegration rank amongst factors for men and women for USA and France.

| Rank | USA | | France | |
|------|---------------|-----------------|---------------|-----------------|
| | Men and Women | | Men and Women | |
| | Trace-Test | <i>p</i> -Value | Trace-Test | <i>p</i> -Value |
| 0 | 286.700 | [0.000] ** | 287.730 | [0.000] ** |
| 1 | 194.530 | [0.000] ** | 215.400 | [0.000] ** |
| 2 | 139.460 | [0.001] ** | 158.870 | [0.000] ** |
| 3 | 89.751 | [0.041] * | 116.050 | [0.000] ** |
| 4 | 52.267 | [0.321] | 76.354 | [0.002] ** |
| 5 | 34.795 | [0.257] | 47.410 | [0.015] * |
| 6 | 19.806 | [0.240] | 24.016 | [0.082] |
| 7 | 7.897 | [0.268] | 8.524 | [0.218] |

Note: The Johansen trace test is calculated with a trend restricted to the cointegration space. The number of lags in the VAR is 1 for all cases. “***” and “**” signify significance at the 1% and 5% level, respectively.

Table 4. Test for cointegration rank amongst factors for US and French men and women.

| Rank | Men | | Women | |
|------|----------------|-----------------|----------------|-----------------|
| | USA and France | | USA and France | |
| | Trace-Test | <i>p</i> -Value | Trace-Test | <i>p</i> -Value |
| 0 | 225.130 | [0.000] ** | 221.980 | [0.000] ** |
| 1 | 158.190 | [0.016] * | 152.790 | [0.036] * |
| 2 | 111.620 | [0.113] | 107.810 | [0.178] |
| 3 | 75.429 | [0.311] | 70.517 | [0.490] |
| 4 | 47.924 | [0.513] | 44.325 | [0.679] |
| 5 | 27.275 | [0.668] | 27.285 | [0.667] |
| 6 | 11.364 | [0.850] | 13.676 | [0.688] |
| 7 | 3.867 | [0.759] | 3.686 | [0.783] |

Note: The Johansen trace test is calculated with a trend restricted to the cointegration space. The number of lags in the VAR is 1 for all cases. “***” and “**” signify significance at the 1% and 5% level, respectively.

6.3. Estimates Using the One-Step Procedure

We now consider the one-step estimation of the model employing maximum likelihood estimation via the Kalman Filter recursions with the model specified on state space form. This method theoretically improves the efficiency as it avoids the issue from the two-step estimator of ignoring the estimation error from the first step in the second step. The estimation for US is based on the assumption of a VAR(1) in first differences for the transition dynamics and for France it is based on the cointegrated VAR model with two cointegrating relations. These specifications of the transient dynamics are chosen in accordance with the results reported in Section 6.2. Table 5 reports the estimated shape parameters and their standard errors. It is seen that the loading parameters and the variance are very similar to those obtained from the two-step procedure. Furthermore, the standard errors of the estimated

shape parameters are found to be very close to those of the two-step method (sometimes smaller). This indicates that small efficiency gains can be obtained by using the one-step procedure.

Table 5. Estimated loading function parameters and standard errors from the one-step procedure for French and US men and women. For US, the VAR(1) model in first difference is assumed for the transition dynamics and, for France, a VECM with two cointegrating relations is assumed. The standard errors are calculated using the inverse Fisher information criterion.

| | | Men | | | | |
|----|----------|-------------|-------------|-------------|--------|------------|
| | | λ_1 | λ_2 | λ_3 | k | σ^2 |
| Fr | Estimate | 0.556 | 12.151 | 1.094 | 20.324 | 0.021 |
| | Std. Err | 0.013 | 0.020 | 0.004 | 0.002 | 0.010 |
| US | Estimate | 0.624 | 10.809 | 1.103 | 20.014 | 0.018 |
| | Std. Err | 0.013 | 0.020 | 0.003 | 0.002 | 0.019 |

| | | Women | | | | |
|----|----------|-------------|-------------|-------------|--------|------------|
| | | λ_1 | λ_2 | λ_3 | k | σ^2 |
| Fr | Estimate | 0.648 | 18.711 | 1.453 | 19.450 | 0.024 |
| | Std. Err | 0.013 | 0.048 | 0.003 | 0.004 | 0.018 |
| US | Estimate | 0.608 | 18.930 | 1.295 | 18.703 | 0.013 |
| | Std. Err | 0.010 | 0.043 | 0.003 | 0.004 | 0.005 |

Figure 7 shows the estimated factors (or states) for the one-step state space estimation procedure for US based on the VAR(1) specification in differences. The factor estimates for the VECM specification for France are shown in Figure 8.

When comparing the estimated factors with those obtained in the first step of the two-step approach, the results appear similar. However, the factors from the one-step estimation show a smoother development because the one-step procedure directly accounts for the transition dynamics in the estimation.

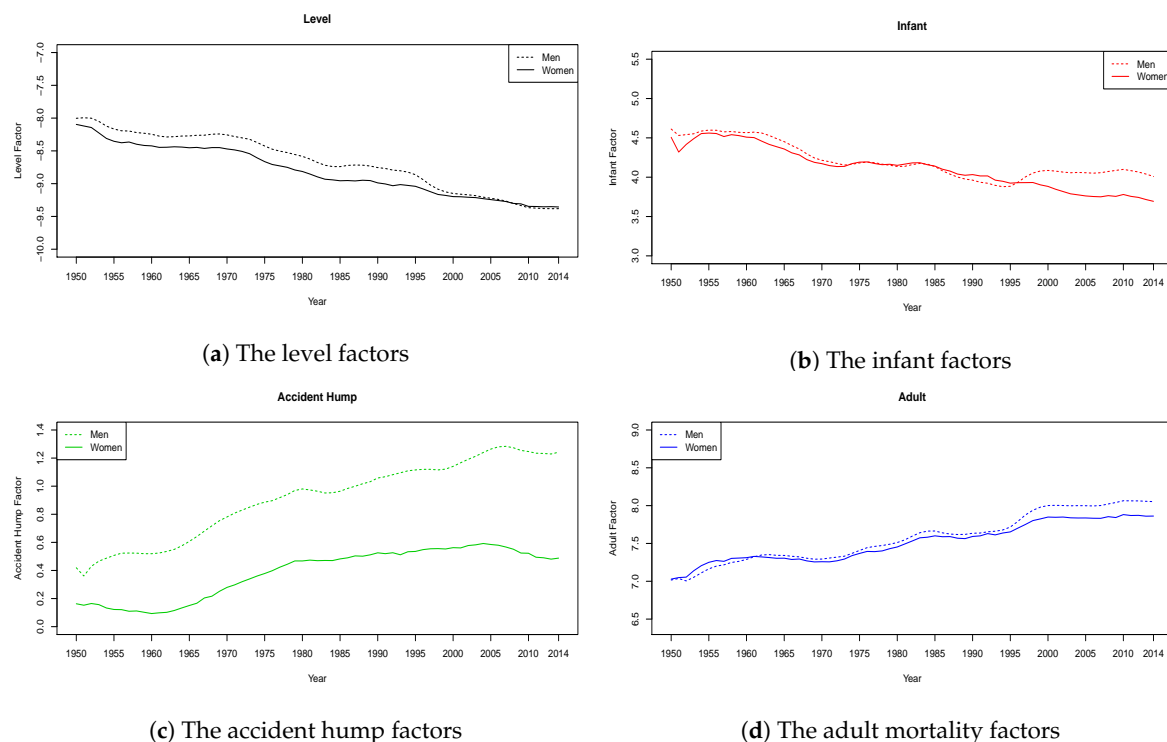


Figure 7. The factors $\kappa_{i,t}$, $i = 0, 1, 2, 3$ estimated by the one-step procedure for USA from 1950–2014 and assuming a VAR(1) model for the first difference of the factors, for both genders.

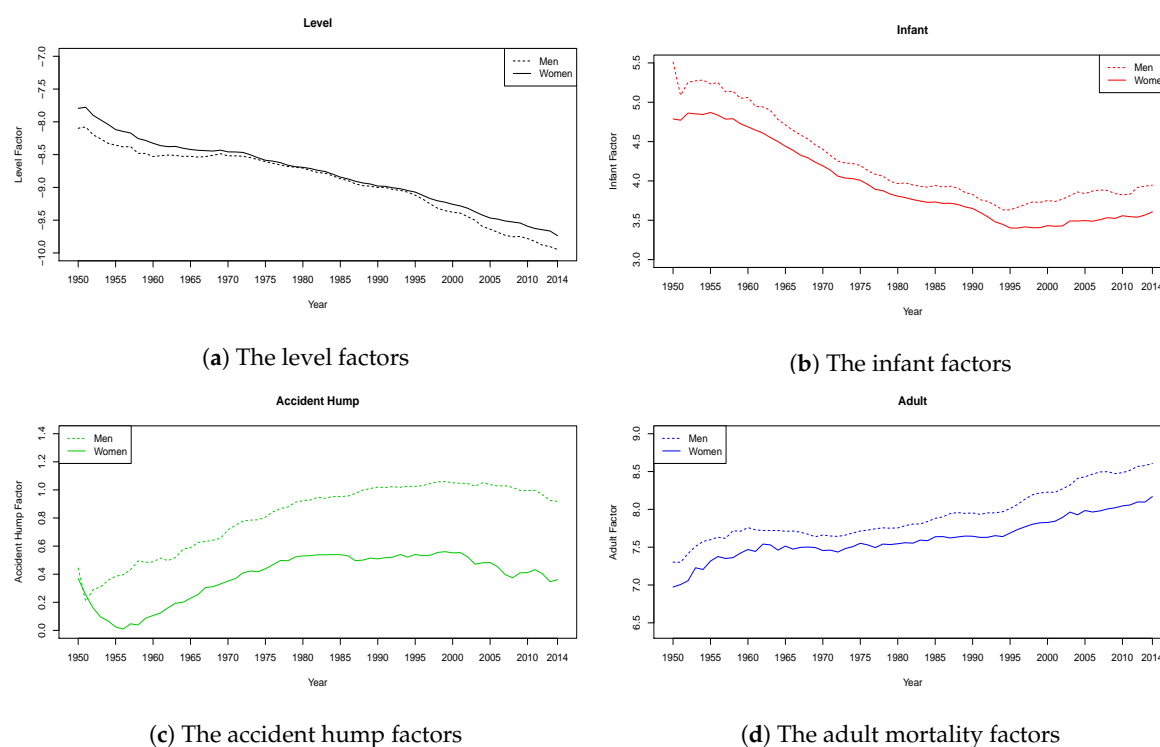


Figure 8. The factors $\kappa_{i,t}$, $i = 0, 1, 2, 3$ are estimated by the one-step procedure for France using data from 1950–2014 and assuming a VECM specification for the factors, for both genders.

6.4. Model Fit

We now compare the PFM with the LC model in terms of in-sample fit.

As the PFM does not include a constant for each age-specific death rate, we are interested in whether the model can capture the mean by relatively few parameters. As seen in Figure 9a–d, the model captures the mean well for all populations. Note that, by construction, α_x in the LC model is equal to the mean of the age specific log death rates, which corresponds to the data mean levels in the figures.

To further quantify the model fit, we calculate a pseudo R^2 for each age group by running a regression of the age-specific death rates on a constant and the fitted values.⁵ The pseudo R^2 's shown in Figure 10a–d display that both the LC and the PFM fit the observed data well. However, the PFM tends to perform better around the accident hump, where the LC model is found to have poor performance.

Next, we investigate how each of the factors contribute to the explanatory power of the model by calculating the partial correlation between the log mortality and a particular factor after adjusting for the influence of the fit obtained from the remaining factors. This adjustment is necessary because the factors are non-orthogonal. Figure 11a–d display the partial correlations in excess of a 65% threshold for all ages to identify where the different factors improve the fit.

It is seen that the infant mortality factor significantly improves the fit for infants as desired. The level factor substantially improves the performance for most ages, and the accident hump factor primarily affects the mortality in the years around the accident hump. Finally, the adult factor primarily improves the fit for the adult ages as desired, but its partial explanatory power is of a smaller magnitude compared with the other factors, mainly because the adult factor is highly correlated with the factor common to all age groups.

⁵ This corresponds to the partial correlation squared between the fitted and observed values.

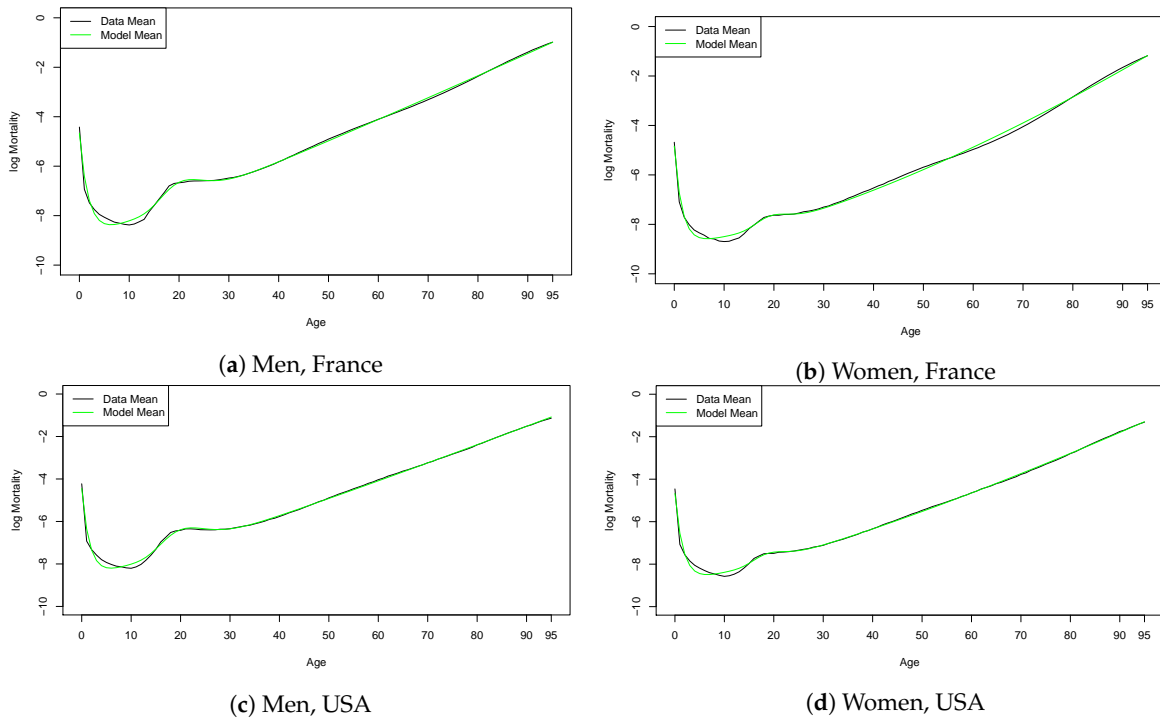


Figure 9. The mean of the data and the mean of the parametric factor model estimated using the two-step procedure for both men and women, for France and USA. The estimation period is 1950–2014.

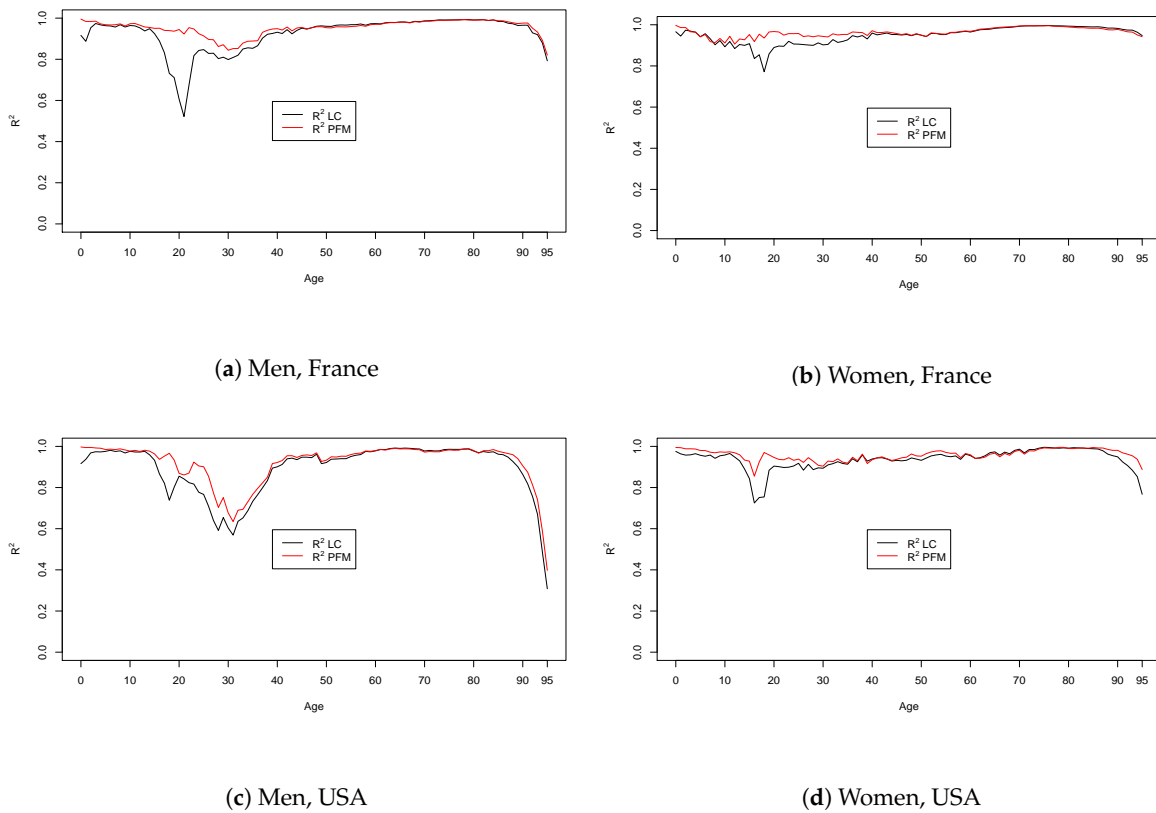


Figure 10. The pseudo R^2 (within) for the PFM and the LC model for all ages estimated using the two-step procedure. The R^2 is shown for both men and women in France and the USA, respectively.

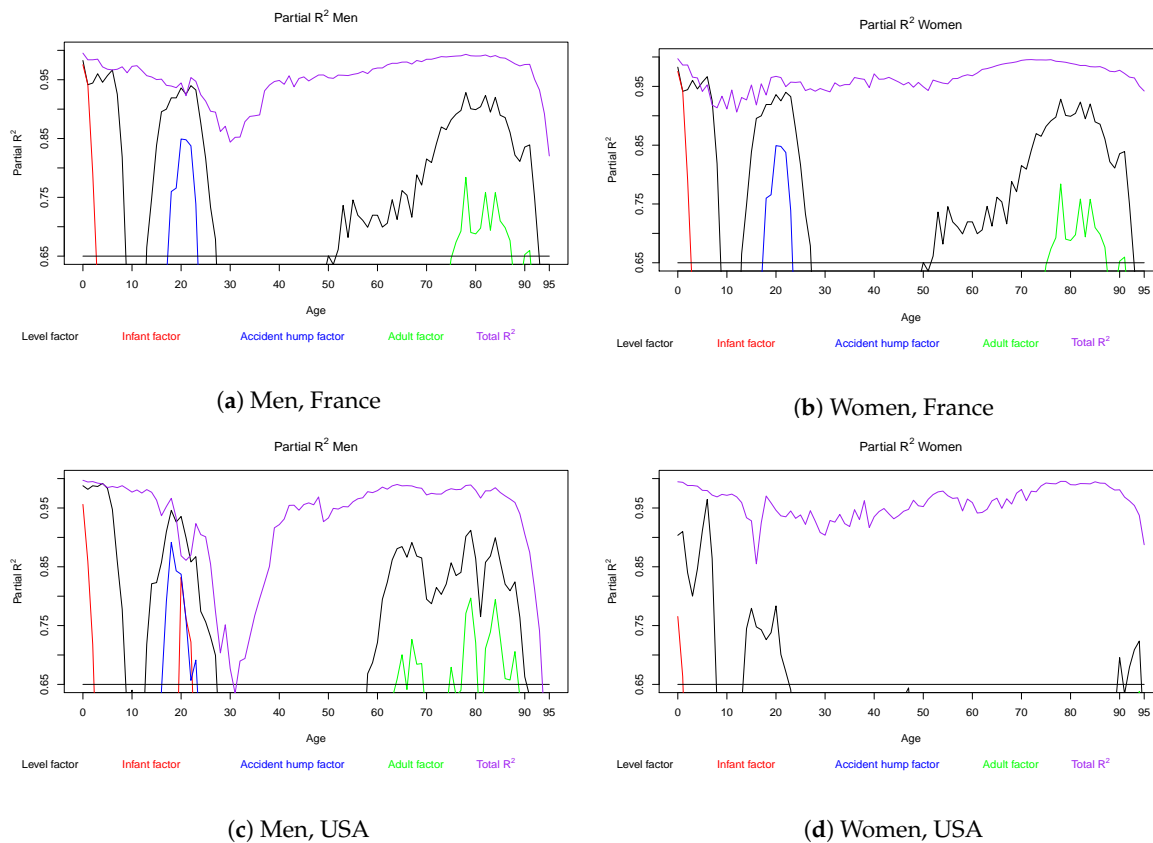


Figure 11. The partial R^2 for the infant, level, accident hump, and adult factor estimated using the two-step procedure for the years 1950–2014. The relative improvements from each of the factors are shown in excess of a 65% threshold. This is shown for both genders for France and the US, respectively.

7. Forecast Evaluation

In this section, we investigate the forecast performance of the PFM and compare with relevant benchmark models. For forecast evaluation and comparison, we use the Model Confidence Set (MCS) approach developed in [Hansen et al. \(2011\)](#).⁶

The MCS procedure is a test for predictive ability across a number of competing models, which sequentially removes the model that performs significantly worse than the remaining models left in the model confidence set. The procedure delineates the set of best performing models at a given confidence level among which we cannot say that any of the other models perform statistically better.

Hence, the MCS does not necessarily pick out a single best model but rather delineates a set of best models as the available information might not be able to discriminate between these models. The MCS procedure returns p -values, \hat{p}_i , for each model i considered, and, from this, the MCS can be determined. The MCS procedure returns a p -value of 1 to the best performing model.⁷

To reduce the dimension of the forecast evaluation, we calculate the (period) life expectancy at birth which aggregates the forecasted age-specific death rates into a single measure. The (period) life

⁶ The MCS approach is implemented via the Ox-package Mulcom 3.0 by [Hansen and Lunde \(2014\)](#) in Oxmetrics 7, see [Doornik \(2013\)](#).

⁷ For the case with only two models the forecast performance could be tested via the [Diebold and Mariano \(1995\)](#) test, which only allows for pairwise comparisons, whereas the MCS procedure allows for joint multiple model evaluation.

expectancy is calculated by using the standard assumption of a constant chance of death within each age interval as in [Brouhns et al. \(2002\)](#)⁸:

$$\hat{e}_0^\uparrow(t) = \frac{1 - \exp(-m_{0,t})}{m_{0,t}} + \sum_{k=1}^N \left(\prod_{j=0}^{k-1} \exp(-m_{j,t}) \right) \frac{1 - \exp(-m_{k,t})}{m_{k,t}}, \quad (5)$$

where $m_{j,t}$ signifies the age-specific death rates and $\hat{e}_0^\uparrow(t)$ is the (period) life expectancy at birth.

To show the robustness of the proposed model at producing reliable forecasts, we consider data for men and women for the USA and France in the forecast evaluation. The forecasts are constructed by recursively estimating each model from 1950 onwards until year $t = 1970, 1971, \dots$ and forecasting 1, 10, and 20 years ahead. This gives 43, 34, and 24 forecasts of the age-specific death rates for each model, respectively. The forecast performance is evaluated using the mean squared error of the life expectancy as the loss function. For implementation, we use the block bootstrap with a block length equal to the longest significant lag length from fitting an AR model and a confidence level of 5%—see [Hansen et al. \(2011\)](#) for details.

As benchmark models, we use (1) a random walk with drift (RWD) specification for each (log) age specific death rate, (2) the [Lee and Carter \(1992\)](#) model with a single factor, (3) and the functional data approach (FDA) of [Hyndman and Ullah \(2007\)](#). Based on the analysis in Sections 5 and 6, we consider two dynamic specifications of the factor structure, a VECM (with two cointegrating relations) and a VAR(1) in first differences of the factors. For comparisons, we use both specifications for each country and gender estimated by the two-step procedure. For the one-step procedure, we consider estimation assuming the VECM structure for France and the VAR(1) structure in first differences for the US as found to be appropriate in Section 6. Using the two-step procedure, we further compare a VAR(1) model in levels and univariate ARIMA models in the dynamic specification.⁹ Based on the finding of unit roots and trending behaviour for each of the factors, we decide to use a random walk with drift specification as ARIMA model specification.¹⁰ For the LC model, we use a random walk with drift specification for the single factor κ_t . The FDA model of [Hyndman and Ullah \(2007\)](#) can be considered an extension of the LC model by using K factors and smoothing across the death rates.¹¹ The results are reported in Table 6 for France and in Table 7 for the USA.

⁸ Note that we here use the period life expectancy (within year t), whereas the formula in [Brouhns et al. \(2002\)](#) computes the cohort life expectancy.

⁹ These specifications have often been used in studies applying graduation laws of mortality—see [Booth and Tickle \(2008\)](#); [McNown and Rogers \(1989, 1992\)](#).

¹⁰ In preliminary experiments, we also found this specification to give a better forecast performance compared with using other ARIMA models.

¹¹ The factors are estimated using weighted principal components in the R package Demography—see [Hyndman and Ullah \(2007\)](#) and [Hyndman et al. \(2014\)](#) for further details. All other models are estimated using own codes and the packages 'tsDyn', 'VARS', and 'Forecast' in R ([R Core Team 2015](#)) by [Pfaff \(2008\)](#); [Stigler \(2010\)](#) and [Hyndman \(2015\)](#).

Table 6. Forecasting life expectancy 1, 10, and 20 years ahead with mean-squared error criterion for US men and women evaluated using the Model Confidence Set. The mean squared error along with *p*-values for the estimated model confidence set for life expectancy. The models included in the set of best models are marked in boldface. The first five rows show the results for the parametric factor model assuming different specifications for the factor dynamics, whereas the last three rows show results for the benchmark models. The VAR1 in levels and ARIMA specifications are used for comparison.

| France | | Men | | | | | | Women | | | | | |
|--------|---------|--------|--------------|---------|--------------|---------|--------------|--------|--------------|---------|--------------|---------|--------------|
| | | 1 Year | | 10 Year | | 20 Year | | 1 Year | | 10 Year | | 20 Year | |
| | | MSE | Pval | MSE | Pval | MSE | Pval | MSE | Pval | MSE | Pval | MSE | Pval |
| PFM | VAR1 | 0.098 | 0.000 | 2.493 | 0.022 | 13.480 | 0.000 | 0.164 | 0.000 | 1.719 | 0.000 | 7.529 | 0.000 |
| | Arima | 0.103 | 0.001 | 0.967 | 1.000 | 4.147 | 1.000 | 0.071 | 0.002 | 0.144 | 0.233 | 0.370 | 0.920 |
| | ΔVAR1 | 0.095 | 0.006 | 0.985 | 0.931 | 4.611 | 0.146 | 0.086 | 0.001 | 0.157 | 0.013 | 0.444 | 0.415 |
| | VECM2 | 0.095 | 0.002 | 0.986 | 0.931 | 4.511 | 0.087 | 0.082 | 0.000 | 0.138 | 0.233 | 0.439 | 0.484 |
| | VECM2SS | 0.234 | 0.000 | 1.331 | 0.284 | 4.393 | 0.712 | 0.359 | 0.000 | 0.741 | 0.005 | 1.041 | 0.036 |
| RWD | | 0.032 | 0.611 | 1.135 | 0.284 | 5.439 | 0.000 | 0.032 | 1.000 | 0.103 | 1.000 | 0.367 | 1.000 |
| LC | | 0.099 | 0.000 | 1.479 | 0.001 | 6.367 | 0.000 | 0.119 | 0.001 | 0.229 | 0.050 | 0.460 | 0.329 |
| FDA | | 0.030 | 1.000 | 1.085 | 0.875 | 5.436 | 0.002 | 0.037 | 0.301 | 0.410 | 0.134 | 1.581 | 0.329 |

Table 7. Forecasting life expectancy 1, 10, and 20 years ahead with mean-squared error criterion for US men and women evaluated using the Model Confidence Set. Mean squared error along with *p*-values for the estimated model confidence set for life expectancy. The models included in the set of best models are marked in boldface. The first five rows show the results for the parametric factor model assuming different specifications for the factor dynamics, whereas the last three rows show results for the benchmark models. The VAR1 in levels and ARIMA specifications are used for comparison.

| USA | | Men | | | | | | Women | | | | | |
|-----|---------|--------|--------------|---------|--------------|---------|--------------|--------|--------------|---------|--------------|---------|--------------|
| | | 1 Year | | 10 Year | | 20 Year | | 1 Year | | 10 Year | | 20 Year | |
| | | MSE | Pval | MSE | Pval | MSE | Pval | MSE | Pval | MSE | Pval | MSE | Pval |
| PFM | VAR1 | 0.113 | 0.011 | 1.607 | 0.018 | 5.915 | 0.004 | 0.069 | 0.004 | 1.514 | 0.157 | 11.630 | 0.024 |
| | Arima | 0.112 | 0.013 | 0.787 | 1.000 | 2.493 | 1.000 | 0.054 | 0.011 | 0.519 | 0.225 | 1.378 | 0.017 |
| | ΔVAR1 | 0.110 | 0.010 | 0.897 | 0.140 | 3.094 | 0.038 | 0.057 | 0.011 | 0.562 | 0.005 | 1.369 | 0.002 |
| | VECM2 | 0.127 | 0.013 | 1.081 | 0.104 | 2.765 | 0.455 | 0.052 | 0.011 | 0.329 | 1.000 | 0.593 | 1.000 |
| | ΔVAR1SS | 0.117 | 0.013 | 1.244 | 0.024 | 3.662 | 0.006 | 0.084 | 0.001 | 0.575 | 0.124 | 1.316 | 0.024 |
| RWD | | 0.035 | 1.000 | 1.240 | 0.104 | 4.016 | 0.000 | 0.023 | 1.000 | 0.435 | 0.225 | 1.243 | 0.011 |
| LC | | 0.138 | 0.013 | 1.771 | 0.018 | 5.375 | 0.001 | 0.080 | 0.006 | 0.682 | 0.069 | 1.790 | 0.003 |
| FDA | | 0.044 | 0.154 | 1.577 | 0.104 | 4.824 | 0.006 | 0.026 | 0.032 | 0.495 | 0.225 | 1.441 | 0.024 |

France, men. For French men, the MCS using a 1-year forecast horizon includes the RWD and FDA specifications. However, when expanding the forecast horizon, the MCS now includes three variants of the PFM and, in fact, for a twenty-year forecast horizon, the MCS excludes the RWD and FDA specifications. It is interesting to observe that, in this forecast competition, the LC model is never included in the MCS. The same applies for the PFM model specification where the factors are modelled as a VAR(1) in levels. This is not surprising because all the factors were found to have unit roots.

USA, men. The pattern observed for French men generally applies for US men as well. However, for the 20-year horizon, only two of the PFM models are included in the MCS.

France, women. For French women and a forecasting horizon of one year, the results are rather similar to those of French men and in particular the RWD specification and the FDA model are the ones included in the MCS. For a 10-year horizon, the MCS also includes a single PFM specification and, for a 20-year horizon, only the PFM with a VAR(1) in levels is not included in the MCS.

USA, women. For US women, the RWD model is always in the MCS. For a 10-year horizon, the results are similar to French women and, for a 20-year horizon, the MCS is slightly smaller

than for French women and includes in particular the two PFM specifications the FDA and the RWD specifications.

In summary, the PFM class of models appears to perform especially well for longer forecast horizons and in most cases performs better than the LC model. An explanation for this result could be the structural features of the PFM class of models compared to the LC model. For longer horizons, the structural restrictions on the loadings account for different factors affecting the separate age groups. The structure implied by the PFM specification ensures a realistic shape of the mortality curve, which cannot be captured by a single factor LC model. Another conclusion is that, in situations where competing models are performing well, especially for longer horizons, the different PFM models also perform well. On the other hand, in situations where competing models are not performing so well, the PFM models are included in the MCS as seen especially for men.

8. Conclusions

We have suggested a multi-factor model for the term structure of mortality. The factors are identified after restrictions on the loading functions in such a way that different age groups and their factor dynamics can be addressed separately. Thus, rather than having a single factor governing all age groups as for the LC model, different factors (or trends) play a role in the way that mortality across age groups develop. In particular, we consider separate factors driving infant mortality, the accident hump mortality, mortality for the elderly in addition to a common factor affecting all age groups. We have suggested two estimation methods that are similar to estimation of term structure models considered in other contexts. In an application, we apply the methodology to mortality data for the US and France for each gender. The models are shown to provide a good fit and, for certain age groups, provides a much better fit compared to the LC model. In a forecast comparison across a range of competing models, the new class of models that we consider in the paper are shown to perform well, especially over longer forecast horizons.

Author Contributions: The authors made equal contributions.

Funding: Danish National Research Foundation: DNRF78.

Acknowledgments: The authors appreciate financial support from CREATES, Center for Research in Econometric Analysis of Time Series, funded by the Danish National Research Foundation (Grant No. DNRF78). We acknowledge helpful comments and suggestions from Siem Jan Koopman and from seminar and conference participants at CREATES lunch seminars, at Max-Planck Odense Center on the Biodemography of Aging, at the annual conference of the International Association for Applied Econometrics in Sapporo, Japan, 2017, and at Dansk Økonometrisk Selskabs annual meeting in 2016.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Booth, Heather, John Maindonald, and Len Smith. 2002. Applying Lee–Carter under conditions of variable mortality decline. *Population Studies* 56: 325–36. [[CrossRef](#)] [[PubMed](#)]
- Booth, Heather, and Leonie Tickle. 2008. Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science* 3: 3–43. [[CrossRef](#)]
- Box, G. E., and D. R. Cox. 1964. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)* 26: 211–52. [[CrossRef](#)]
- Brouhns, Natacha, Michel Denuit, and Jeroen K. Vermunt. 2002. A poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics* 31: 373–93. [[CrossRef](#)]
- Byrd, Richard H., Peihuang Lu, Jorge Nocedal, and Ciyu Zhu. 1995. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing* 16: 1190–208. [[CrossRef](#)]
- Cairns, Andrew J., David Blake, and Kevin Dowd. 2006. Pricing death: Frameworks for the valuation and securitization of mortality risk. *Astin Bulletin: The Journal of the International Actuarial Association* 36: 79–120. [[CrossRef](#)]

- Callot, Laurent, Niels Haldrup, and Malene Kallestrup-Lamb. 2016. Deterministic and stochastic trends in the Lee–Carter mortality model. *Applied Economics Letters* 23: 486–93. [CrossRef]
- De Jong, Piet, and Leonie Tickle. 2006. Extending Lee–Carter mortality forecasting. *Mathematical Population Studies* 13: 1–18. [CrossRef]
- Diebold, Francis, and Robert Mariano. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13: 253–63.
- Diebold, Francis X., and Canlin Li. 2006. Forecasting the term structure of government bond yields. *Journal of Econometrics* 130: 337–64. [CrossRef]
- Diebold, Francis X., Glenn D. Rudebusch, and S. B. Aruoba. 2006. The macroeconomy and the yield curve: A dynamic latent factor approach. *Journal of Econometrics* 131: 309–38. [CrossRef]
- Doornik, Jurgen A. 1998. Approximations to the asymptotic distributions of cointegration tests. *Journal of Economic Surveys* 12: 573–593. [CrossRef]
- Doornik, Jurgen A. 1999. Erratum [approximations to the asymptotic distribution of cointegration tests]. *Journal of Economic Surveys* 13.
- Doornik, Jurgen A. 2013. An Object-Oriented Matrix Programming Language Ox 6. Available online: <https://ora.ox.ac.uk/objects/uuid:242f0a19-0665-4d9a-b863-774a35ce98c7> (accessed on 10 March 2019).
- Doz, Catherine, Domenico Giannone, and Lucrezia Reichlin. 2011. A two-step estimator for large approximate dynamic factor models based on kalman filtering. *Journal of Econometrics* 164: 188–205. [CrossRef]
- Durbin, James, and Siem Jan Koopman. 2012. *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press, vol. 38.
- Gavrilov, L. A., and N. S. Gavrilova. 1979. Determination of species length of life. *Doklady Akademii nauk SSSR: Biological Sciences* 246: 905–8.
- Gavrilov, L. A., and N. S. Gavrilova. 1991. *The Biology of Life Span: A Quantitative Approach*. New York and Chur: Harwood Academic Publisher.
- Gompertz, B. 1825. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London* 115: 513–83. [CrossRef]
- Hansen, Peter, and Asger Lunde. 2014. Mulcom 3.00. Econometric Toolkit for Multiple Comparisons.
- Hansen, Peter R., Asger Lunde, and James M. Nason. 2011. The model confidence set. *Econometrica* 79: 453–97. [CrossRef]
- Heligman, L., and J. H. Pollard. 1980. The age pattern of mortality. *Journal of the Institute of Actuaries* 107: 49–80. [CrossRef]
- Human Mortality Database. 2016. Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available online: www.mortality.org or www.humanmortality.de (accessed on 29 June 2016).
- Hyndman, Rob J. 2015. ‘forecast’: Forecasting Functions for Time Series And Linear Models. R package version 6.2. Available online: <http://github.com/robjhyndman/forecast> (accessed on 13 November 2016).
- Hyndman, Rob J., Heather Booth, Leonie Tickle, and John Maindonald. 2014. Demography: Forecasting Mortality, Fertility, Migration and Population Data. R package version 1.18. Available online: <http://CRAN.R-project.org/package=demography> (accessed on 13 November 2016).
- Hyndman, Rob J., and Shahid Ullah. 2007. Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis* 51: 4942–56.
- Johansen, Soren. 1988. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12: 231–54. [CrossRef]
- Johansen, Soren. 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica: Journal of the Econometric Society* 59: 1551–80. [CrossRef]
- Koopman, Siem Jan, Max I. Mallee, and Michel Van der Wel. 2010. Analyzing the term structure of interest rates using the dynamic nelson-siegel model with time-varying parameters. *Journal of Business & Economic Statistics* 28: 329–43.
- Lazar, Dorina, and Michel M. Denuit. 2009. A multivariate time series approach to projected life tables. *Applied Stochastic Models in Business and Industry* 25: 806–23. [CrossRef]
- Lee, Ronald D., and Lawrence R. Carter. 1992. Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association* 87: 659–71. [CrossRef]

- Lee, Ronald D., and Timothy Miller. 2001. Evaluating the performance of the Lee–Carter method for forecasting mortality. *Demography* 38: 537–49. [CrossRef] [PubMed]
- Makeham, William Matthew. 1860. On the law of mortality and the construction of annuity tables. *The Assurance Magazine, and Journal of the Institute of Actuaries* 8: 301–10. [CrossRef]
- McNown, Robert, and Andrei Rogers. 1989. Forecasting mortality: A parameterized time series approach. *Demography* 26: 645–60. [CrossRef] [PubMed]
- McNown, Robert, and Andrei Rogers. 1992. Forecasting cause-specific mortality using time series methods. *International Journal of Forecasting* 8: 413–32. [CrossRef]
- Nielsen, Bent, and Jens P. Nielsen. 2014. Identification and forecasting in mortality models. *The Scientific World Journal* 2014: 1–24. [CrossRef] [PubMed]
- Pfaff, Bernhard. 2008. VAR, SVAR and SVEC models: Implementation within R package vars. *Journal of Statistical Software* 27: 1–32. Available online: <http://www.jstatsoft.org/v27/i04/> (accessed on 30 January 2017). [CrossRef]
- R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online: <http://www.R-project.org/> (accessed on 13 November 2016).
- Renshaw, Arthur E., and Steven Haberman. 2003. Lee–Carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics* 33: 255–72. [CrossRef]
- Rogers, Andrei, and Jani S. Little. 1994. Parameterizing age patterns of demographic rates with the multiexponential model schedule. *Mathematical Population Studies* 4: 175–95. [CrossRef] [PubMed]
- Siler, William. 1979. A competing-risk model for animal mortality. *Ecology* 60: 750–57. [CrossRef]
- Stigler, Matthieu. 2010. Threshold Cointegration: Overview and Implementation in R. R package version 0.7-2. Available online: <http://stat.ethz.ch/CRAN/web/packages/tsDyn/vignettes/ThCointOverview.pdf> (accessed on 30 January 2017).
- Strehler, Bernard L., and Albert S. Mildvan. 1960. General theory of mortality and aging. *Science* 132: 14–21. [CrossRef] [PubMed]
- Tabeau, Ewa, Anneke van den Berg Jeths, and Christopher Heathcote. 2001. *Forecasting Mortality in Developed Countries: Insights From a Statistical, Demographic and Epidemiological Perspective*. Dordrecht: Kluwer Academic Publishers.
- Wong-Fupuy, Carlos, and Steven Haberman. 2004. Projecting mortality trends: Recent developments in the United Kingdom and the United States. *North American Actuarial Journal* 8: 56–83. [CrossRef]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).