Journal of Population Research
https://doi.org/10.1007/s12546-018-9205-z

ORIGINAL RESEARCH

Forecasting mortality rates: multivariate or univariate models?

Lingbing Feng[1] · **Yanlin Shi**[2]

Abstract

It is well known that accurate forecasts of mortality rates are essential to various demographic research topics, such as population projections and the pricing of insurance products such as pensions and annuities. In this study, we argue that including the lagged rates of neighbouring ages cannot further improve mortality forecasting after allowing for autocorrelations. This is because the sample cross-correlation function cannot exhibit meaningful and statistically significant correlations. In other words, rates of neighbouring ages are usually not leading indicators in mortality forecasting. Therefore, multivariate stochastic mortality models like the classic Lee–Carter may not necessarily lead to more accurate forecasts, compared with sophisticated univariate models. Using Australian mortality data, simulation and empirical studies employing the Lee–Carter, Functional Data, Vector Autoregression, Autoregression-Autoregressive Conditional Heteroskedasticity and exponential smoothing (ETS) state space models are performed. Results suggest that ETS models consistently outperform the others in terms of forecasting accuracy. This conclusion holds for both female and male mortality data with different empirical features across various forecasting error measurements. Hence, ETS can be a widely useful tool to model and forecast mortality rates in actuarial practice.

Keywords Mortality rates · Multivariate model · Univariate model · Exponential smoothing · Lee–Carter model

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12546-018-9205-z) contains supplementary material, which is available to authorized users.

✉ Yanlin Shi
yanlin.shi@mq.edu.au

1 International Institute of Financial Studies, Jiangxi University of Finance and Economics, Nanchang, Jiangxi, China

2 Department of Actuarial Studies and Business Analytics, Macquarie University, Sydney, NSW 2109, Australia

Published online: 09 July 2018

Introduction

Improvements in life expectancy have spurred serious concerns about mortality and longevity risks around the world. As argued in Giacometti et al. (2012), longevity risk is the risk of people surviving longer than expected, or observed death rates being lower than expected. Advances in medical science, technological improvements and lifestyle changes tend to reduce the number of deaths. In demographic research, those risks significantly affect the accuracy of population projections. In actuarial practice, longevity risk affects the pricing of life insurance, pension and annuity products. Overestimating the mortality rate may lead to underestimated future population counts, and a high-risk profile for the annuity insurers.

To reduce such risks, the future of human survival has attracted considerable interest in the past few decades, and forecasting mortality has gained prominence in this context. Among the existing models, Lee and Carter (1992) developed their seminal work, which has been recognized as the most popular stochastic mortality model, namely the Lee–Carter (LC) model. Because of its enormous popularity, the LC model has become a benchmark for stochastic mortality modelling and forecasting. This model describes the variation of log central morality rate using age and time factors. As pointed out by Girosi and King (2007), the LC model can be viewed as a special type of a multivariate process in which the covariance matrix depends on the drift vector and the innovations are inter-temporally correlated. Based on the LC model, many extensions and modifications have been proposed. Among them, one of the most popular approaches is the Functional Data Model (FDM) developed by Hyndman and Ullah (2007) which extends the LC framework by employing the functional data paradigm that allows nonparametric smoothing, and by introducing the multiple principal components of time trends.

Apart from multivariate models like LC and FDM, univariate models can also be used to forecast mortality rates. Giacometti et al. (2012) propose an AR-ARCH stochastic mortality model, which relaxes the stringent homoscedastic assumption underlying the LC-type models. Using Italian mortality data, Giacometti et al. (2012) demonstrate that this univariate model can generate more accurate forecasts than the multivariate LC model. This surprising result draws our attention to a classic debate in forecasting: do the multivariate models necessarily beat their univariate counterparts?

Intuitively, multivariate models utilise more information than univariate models. Hence, forecasting with multivariate models is expected to be more accurate. However, Bell (1997) argues that a univariate simple random walk with drift model generates more accurate short-term forecasts than the more complicated approaches like curve fitting and principal components. Chatfield (1997) also suggests that simple univariate methods are often more robust to model mis-specification and to changes in the model than more complicated models. As pointed out by Du Preez and Witt (2003), forecasting with multivariate models is more accurate only when allowing for autocorrelations, as the sample cross-correlation function exhibits meaningful and statistically significant correlations. As mortality rates are very close to each other for neighbouring ages, those rates may contain almost identical information. On the other hand, the information of age groups far apart is more heterogeneous and may have much smaller cross-correlations. For instance, the mortality rates of ages 40 and 41 are quite close, so including the information of age 41 may not further benefit the filtration of age 40, after the lagged rates of age 40 are considered. Also, as mortality rates of ages 40 and 90 have quite different structures, using information of age 90 may not improve the forecasting of rates of age 40. Therefore, at first glance, it seems that the cross-sectional information of the mortality rates might not bring in additional useful information for the univariate rates, given that the autocorrelations are allowed. Thus, more sophisticated univariate models may lead to more accurate forecasts than the multivariate models.[1]

A potential problem of the AR-ARCH model proposed by Giacometti et al. (2012) is that the AR(1)-ARCH(1) specification is fixed for mortality rates of all ages. For one thing, the fixed ARCH(1) may not properly control for the potential heteroscedasticity of the mortality rates for all ages. For another, it is likely that the AR(1) specification cannot adequately fit the autocorrelations for all mortality rates.

This paper focuses on the time-dimension forecasting of mortality rates. We employ the univariate exponential smoothing (ETS) state space model proposed by Hyndman et al. (2002) and analyse its performance in forecasting mortality rates. Many popular univariate forecasting models (such as ARIMA) have the property that forecasts are weighted combinations of past observations, with recent observations given relatively more weights than older observations. Roughly speaking, those models can be classified as the ETS family. The word "exponential smoothing" reflects the fact that the weights decrease exponentially as the observations get older (Hyndman et al. 2008). ETS methods are originally classified by the taxonomy of Pegels (1969). They are later extended by Gardner (1985), modified by Hyndman et al. (2002) and Taylor (2003). Additionally, Ord et al. (1997) and Hyndman et al. (2002, 2005) have shown that all ETS methods (including non-linear ones) are optimal forecasts from innovations state space models. The most outstanding feature of an ETS model is that time series are decomposed into trend, seasonality and error components. Then, each of them can be modelled with a state-space form involving an exponential parameter. Hyndman and Khandakar (2008) summarise all possible specifications of the ETS models. In terms of forecasting performance, ETS models lead to superior results when they are applied to the 3003 series from the M3-competition (Makridakis and Hibon 2000), which are described in Hyndman et al. (2002).

To demonstrate the potential superiority of univariate models to multivariate ones, we follow the settings of Giacometti et al. (2012) and use Australian mortality data

[1] Intuitively, a sophisticated univariate model makes the best use of univariate filtration. However, if cross-sectional rates cannot bring in additional useful information but only noise, the quality of multivariate forecasting will be compromised.

from 1950 to 2011 including ages 40–91.[2] First, by employing the best univariate ARIMA model to fit the log mortality rates of each age 0–100 over time, we demonstrate that the cross-correlations of residuals obtained for ages 40–91 and those for all other ages are not significant. According to Du Preez and Witt (2003), this result indicates that multivariate models may not necessarily beat the univariate counterparts in terms of forecasting accuracy. In other words, multivariate rates are not leading indicators in the mortality forecasting. To verify this argument, we perform simulation studies based on the Australian mortality data using two distinct data generation processes (DGPs): AR(1) and VAR(1). We consider three multivariate (LC, FDM and VAR) and two univariate (ETS and AR-ARCH) models. VAR model is employed as a multivariate benchmark which has a structure comparable to ETS and AR-ARCH models. Our simulation results suggest that ETS consistently outperforms the others over age groups and across three different forecasting error measurements. Finally, we fit the five models to empirical datasets with various settings, including different age groups, time periods, forecasting horizons and smoothness. Among our empirical results, ETS still outperforms the others in terms of average forecasting error over age groups. Its variations of forecasting errors among age groups are also the smallest in almost all cases. Those observations hold consistently across the three error measurements for both females and males.

The contributions of this paper are twofold. First, we shed some light on the potential superiority of univariate models over the multivariate counterparts. Based on the study of Du Preez and Witt (2003), multivariate rates are not leading indicators in mortality forecasting, as the cross-correlations of residuals obtained from univariate models are not significant. This provides theoretical support for and hence complements related studies such as Bell (1997) who argues that forecasting with univariate models can be more accurate. Second, to the best of our knowledge, this study is the first to focus on the forecasting performance of ETS models on mortality data. Evidenced by both simulation and empirical results, ETS models outperform all the competing models across different error measurements. Hence, ETS models could be a widely useful tool to study mortality rates in actuarial practice

The rest of this paper is organised as follows. In "Functional data model (FDM)" section, we review the specifications of multivariate models used in this study, followed by introducing the employed univariate models in "Univariate models" section. Description of our dataset, analysis of cross-correlations of mortality rates and simulation studies are presented in "Analysis on the cross-correlations and simulation results" section. We conduct empirical studies with various distinct features in "Empirical results" section. "Concluding remarks and discussions" section concludes the paper.

[2] Age groups are selected to compare our results directly with those of Giacometti et al. (2012). A much wider range 0–100 is considered in "Australian mortality data: 0–100 years old, 10-step-ahead horizon, 1950–2011 and two-dimensionally smoothed rates" section. The complete results for such case are presented in the supplementary materials. Our conclusions derived in this paper consistently hold in that more comprehensive dataset.

Multivariate Models

The Lee–Carter Model

The Lee–Carter (LC) model is proposed by combining a demographic model of mortality with time-series methods of forecasting. More specifically, a central mortality rate at age x in year $t(m_{x,t})$ is assumed to follow the below specification:

$$\ln m_{x,t} = a_x + b_x k_t + \varepsilon_{x,t} \tag{1}$$

where a_x is the average pattern of mortality by age across years, b_x is the relative speed of change at each age x, k_t is an index of the level of mortality at time t, $\varepsilon_{x,t}$ is the residual at age x and time t. As for their explanations, the random term $\varepsilon_{x,t}$ reflects a particular age-specific historical influence. Coefficients a_x are age-specific constants that describe the general shape of the age-mortality profile. Index k_t serves to capture the main temporal level of mortality. In terms of estimation, b_x and k_t are calculated by singular value decomposition (SVD) as suggested by Trefethen and Bau (1997).[3] In order to obtain a unique solution, parameters b_x and k_t should satisfy the constraints that b_x sum to 1 and k_t sum to 0. The second constraint implies that estimates of parameters a_x are given by the averages of $\ln m_{x,t}$ over time. Additionally, k_t is adjusted by refitting to total observed deaths. This adjustment gives greater weights to ages at which deaths are high, thereby partly counterbalancing the effect of using logarithm of rates in the LC model (Booth et al. 2006).

To forecast future mortality rates, Lee and Carter (1992) assume that a_x and b_x remain constant over time. The time factor k_t, on the other hand, is intrinsically viewed as a random walk with drift process as follows:

$$\hat{k}_t = \hat{k}_{t-1} + d + e_t \tag{2}$$

where d is the average annual change in $\hat{k}_t$, and e_t are independent and identically distributed Gaussian sequences with null mean. According to Giacometti et al. (2012), the expected h-step-ahead forecasted log mortality rate can be approximated by:

$$\ln \hat{m}_{x,t} = \hat{a}_x + \hat{b}_x + \left(\hat{k}_T + h \frac{\hat{k}_T - \hat{k}_1}{(T-1)} \right) \tag{3}$$

where T is the maximum of t.

[3] Estimates can also be obtained using maximum likelihood estimation, as suggested by Renshaw and Haberman (2006).

Functional Data Model (FDM)

Hyndman and Ullah (2007) adopt a functional data paradigm (Ramsay and Silverman 2005) to model log mortality rates, which extends the LC method in the following ways [concluded by Booth et al. (2006)]:

1. Mortality is assumed to be a smooth function of age that is observed with error.
2. More than one set of (k_t, b_x) components is used.[4]
3. More general time series methods than random walk with drift are used for forecasting the coefficients.
4. Robust estimation can be used to allow for unusual years due to wars or epidemics.
5. It does not adjust k_t.

The FDM approach can be expressed as follows.

$$\ln m_{x,t} = f(x,t) + \sigma_t(x)\varepsilon_{x,t} \tag{4}$$

where

$$f(x,t) = a(x) + \sum_{j=1}^{J} k_{t,j} b_j(x) + e_t(x)$$

Essentially, $f(x, t)$ is the underlying smooth function, whereas $\varepsilon_{x,t}$ is an independently and identically distributed normal random variable, and $\sigma_t(x)$ further allows noise to vary with x. In Hyndman and Ullah (2007), $f(x, t)$ is estimated via penalised regression splines (Wood 2003). Therefore, $\sigma_t(x)\varepsilon_{x,t}$ accounts for observational error that varies with age (the difference between observed rates and spline curves). As for the specification of $f(x, t)$, $a(x)$ is the average pattern of mortality by age across years, $b_j(x)$ is a "basis function" and $k_{t,j}$ is a time series coefficient. In terms of model fitting, $a(x)$ is estimated by employing penalised regression splines to each t and averaging the results. For each $j=1, \ldots, J$, the pairs ($k_{t,j}$, $b_j(x)$) are estimated using principal component decomposition. The error term $e_t(x)$ for $f(x, t)$ stands for the difference between spline curves and fitted curves from the model. Detailed steps of forecasting with FDM can be found in Section 4 of Hyndman and Ullah (2007).

In this paper, we adopt the settings of Hyndman and Ullah (2007) to forecast mortality rates. More specifically, we let J be 4 and use an ARIMA specification to forecast $k_{t,j}$ for each j in all cases.[5]

[4] Booth et al. (2002) and Renshaw and Haberman (2003) are the first to use more than one set of k_t and b_x to improve accuracy of the original LC model. In general, those studies allow for more flexible specifications of b_x and k_t (for instance, k_t can follow an ARIMA process).

[5] As noted by Booth et al. (2006), the forecasting results are relatively insensitive to the choice of J, given that J is large enough. We have also considered cases when $J=6$, 8 and 9. All of them lead to similar results.

VAR Model

VAR model is a multivariate extension of AR model. To control for possible non-stationarity of ln $m_{x,t}$ over time, we use its first-order difference in VAR model.[6] The following VAR(1) specification on $4lnm_{x,t}$ is employed as a benchmark to compare the performance of univariate and multivariate models.

$$\boldsymbol{\Delta}\mathbf{M}_t = \mathbf{A} + \mathbf{B}\boldsymbol{\Delta}\mathbf{M}_{t-1} + \boldsymbol{\varepsilon}_t \tag{5}$$

$\Delta\mathbf{M}_t = (\ldots, \ln m_{x-1,t} - \ln m_{x-1,t-1}, \ln m_{x,t} - \ln m_{x,t-1}, \ln m_{x+1,t} - \ln m_{x+1,t-1}, \ldots)'$ is a N by 1 vector of first-order differenced log mortality rates for selected N ages. $\mathbf{A}$ is a N by 1 vector measuring the average mortality rates at each age. $\mathbf{B}$ is a coefficient matrix (N by N) that measures the impacts of lag 1 mortality rates on current rates. $\boldsymbol{\varepsilon}_t$ is N by N error term at time t, which follows a multivariate Gaussian distribution with null mean.

To compare complete cross-sectional impacts, it is ideal to include as many ages as possible in the same VAR model. However, due to limited number of observations and potential non-stationarity of fitted models, not all interested ages can be put into one VAR in all cases. Hence, when it is not possible to contain all ages, we split them into a few groups (e.g. 10 ages in one group) and fit VAR(1) models individually for each group.

Univariate Models

AR-ARCH Model

Giacometti et al. (2012) propose a univariate AR(1)-ARCH(1) model to forecast mortality rates. As we are only interested in time-dimension forecasting, the following AR(1)-ARCH(1) specification is employed.

$$\begin{aligned} \ln m_{x,t} &= p(t) + \alpha_1 \ln m_{x,t-1} + \varepsilon_{x,t} \\ \sigma^2_{x,t} &= \beta_0 + \beta_1 \varepsilon^2_{x,t-1} \end{aligned} \tag{6}$$

where $p(t)$ is a polynomial function of t with degree n. $\varepsilon_{x,t}$ is the innovation of the time series process ln $m_{x,t}$ and $\varepsilon_{x,t} = z_t\sigma_{x,t}$, where z_t is an independently and identically-distributed Gaussian sequence with null mean and unit variance. $\sigma^2_{x,t}$ is then the conditional variance of $\varepsilon_{x,t}$.

In this paper, we follow Giacometti et al. (2012) and set n to be 2 in all cases to model time trends. AR(1)-ARCH(1) model is fitted individually for each of the interested ages

[6] We also consider using the original ln $m_{x,t}$ with polynomial trends at order 2, as employed by Giacometti et al. (2012). Results are consistent and available upon request.

Exponential Smoothing (ETS) State Space Model

Many popular forecasting models have the property that forecasts are weighted combinations of past observations, with recent observations given relatively more weights than older observations. Roughly speaking, those models can be classified as the exponential smoothing family. The word "exponential smoothing" reflects the fact that weights decrease exponentially as observations get older (Hyndman et al. 2008).

ETS methods are originally classified by the taxonomy of Pegels (1969). They are later extended by Gardner (1985), modified by Hyndman et al. (2002) and Taylor (2003). Additionally, Ord et al. (1997) and Hyndman et al. (2002, 2005) have shown that all exponential smoothing methods (including non-linear methods) are optimal forecasts from innovations state space models. By decomposing a time series into trend, seasonality and error, and considering different specification for each of the three components, there are thirty[7] distinct ETS methods (mainly by assuming an additive or multiplicative relationship for each component).[8] For details of those specifications, please refer to Section 2 of Hyndman and Khandakar (2008).

In the case of mortality rates (especially smoothed rates), seasonality is not present. Besides, as adopted in all models covered by this paper, the additive specification is widely used to model the logarithm of mortality rates. Therefore, by taking the additive trend and error terms without seasonality, we have the following ETS specification (also known as Holt's linear model) in state-space form:

$$\begin{aligned} \ln m_{x,t} &= l_{x,t-1} + b_{x,t-1} + \varepsilon_{x,t} \\ l_{x,t} &= l_{x,t-1} + b_{x,t-1} + \alpha\varepsilon_{x,t} \\ b_{x,t} &= b_{x,t-1} + \beta\varepsilon_{x,t} \end{aligned} \tag{7}$$

where $l_{x,t}$ and $b_{x,t}$ measure the level and growth of $\ln m_{x,t}$, respectively. α and β are their corresponding exponential smoothing parameters. Hence, h-step-ahead forecast of $\ln m_{x,t}$ is $\ln m^{\wedge}_{x,t} = l_t + b_t h$.

If an additive damped trend is considered, then Eq. (7) will be modified as:

$$\begin{aligned} \ln m_{x,t} &= l_{x,t-1} + \phi b_{x,t-1} + \varepsilon_{x,t} \\ l_{x,t} &= l_{x,t-1} + \phi b_{x,t-1} + \alpha\varepsilon_{x,t} \\ b_{x,t} &= \phi b_{x,t-1} + \beta\varepsilon_{x,t} \end{aligned} \tag{8}$$

7 Among them, eleven cases are unstable and not preferred in practice (Hyndman et al. 2008).

8 For the trend component, an additional "damped" type (Gardner Jr and McKenzie 1985) is also considered for additive and multiplicative cases. Altogether, there are five different scenarios for trend: none, additive (damped) and multiplicative (damped).

with h-step-ahead forecast $\ln m^{\wedge}_{x,t} = l_{x,t} + b_{x,t} \sum_{i=1}^{h} \phi^i$. It can be seen that growth of the predicted mortality rate is dampened by a factor of ϕ for each additional future time period.

Altogether, there are six stable ETS models that could be employed to forecast the mortality rate. Apart from the additive (damped) trend and error cases covered above, the other four models are additive (damped) trend and multiplicative error, and multiplicative (damped) trend and multiplicative error. We employ the automatic ETS algorithm proposed by Hyndman and Khandakar (2008) for forecasting. More specifically, we firstly fit all six models for the log mortality rates of each age and then choose the best one according to the small-sample-size corrected Akaike information criterion (AICc). Finally, the n-step-ahead forecasting will be generated based on the model chosen by AICc.

Analysis on the Cross-Correlations and Simulation Results

Intuitively, multivariate models utilise more information than the univariate counterparts. Hence, forecasting accuracy of multivariate models is expected to be higher. However, as pointed out by Du Preez and Witt (2003), this will only be the case when allowing for autocorrelations, i.e. the sample cross-correlation function exhibits meaningful and statistically significant correlations. In the case of mortality rates, we take the following simple univariate AR(1) model for $\ln m_{x,t}$ as an intuitive example.

$$\ln m_{x,t} = a_x + b_x \ln m_{x,t-1} + \varepsilon_{x,t} \tag{9}$$

Multivariate models will lead to more accurate forecasts than the above AR(1) only if cross-correlations of $\{\varepsilon_{x,t}\}$ and $\{\varepsilon_{x+j,t-k}\}$ obtained from Eq. (9) are significant for some integers $j \neq 0$ and $k > 0$.

Australian Mortality Data

In this paper, we use the Australian mortality data obtained from the Human Mortality Database (2016). Following Booth et al. (2006), we choose an opportune range of data starting from 1950 in order to have a reliable and complete dataset. Also, we follow Giacometti et al. (2012) and restrict the ages from 40 to 91 to avoid a hump around ages between 0 and 39. Finally, we adopt the two-dimensional smoothing proposed by Camarda (2012) to graduate the logarithms of mortality rates.

Following Du Preez and Witt (2003), the analysis of cross-correlations is performed as follows. First, we consider the best ARIMA model for each age $x \in (0, 100)$ and for both females and males. The p and q of the final ARIMA (p, d, q) model are selected to minimize the AICc, given that $p = 0, 1, \ldots, 5$, $d = 0, 1, 2$ and $q = 0, 1, \ldots, 5$. Hence, it is expected that the best ARMA model can utilise as much univariate information as possible and the resulting residuals are not autocorrelated any more. After all models are fitted, we then calculate the cross-correlations with up to 10 lags of residuals $\left\{\varepsilon_{x,t}\right\}$ and $\left\{\varepsilon_{y,t}\right\}$, where x and y range from 40 to 91 and from 0 to 100, respectively. The corresponding p-values (zero against non-zero correlation) are also generated. The maximum cross-correlations and the corresponding p-values for each of the selected ages are presented in Table 1.[9] Clearly, all maximum cross-correlations are considerably small. Except for females aged 45 and 50, all of them are less than 0.24 and are insignificant at 5% level. Therefore, lagged mortality rates of other ages have no additional useful information that is orthogonal to the autocorrelations but related to $\ln m_{x,t}$. In other words, the multivariate rates are not leading indicators in the mortality forecasting. Hence, the multivariate models will not necessarily lead to more accurate forecasts than the univariate models.[10]

To understand this result, we firstly consider the relationship between $\ln m_{x+j,t}$ and $|j|$. For small $|j|$, values of $\ln m_{x+j,t}$ are very close to each other for any t. Hence, those mortality rates contain almost identical information. Therefore, for neighbouring ages groups, the lagged rates $\ln m_{x+j,t}$ cannot bring in additional information, after the effects of univariate lags of $\ln m_{x,t}$ have been considered. On the other hand, for ages that are far apart from each other, their mortality structures are much more different from those of the neighbouring ages. Consequently, including mortality rates of distant ages may introduce more noise than useful information. Thus, although the corresponding correlations are the maximums, this result is more random in nature.

To sum up, after controlling for the autocorrelations, cross-correlations cannot bring additional information to the mortality rates of individual ages. Therefore, the multivariate models may not necessarily outperform their univariate counterparts in terms of forecasting accuracy. To augment this argument, we present simulation evidence using two different data generate processes (DGPs) in the next two subsections.[11]

Simulation Study: AR(1) DGP

We first consider a univariate setting of simulation using the AR(1) DGP. To avoid possible non-stationarity of $\ln m_{x,t}$, we firstly fit the following AR(1) model for each age from 40 to 91 for both females and males.

$$\Delta \ln m_{x,t} = a_x + b_x \Delta \ln m_{x,t-1} + \varepsilon_{x,t} \tag{10}$$

[9] Due to space limit, we only report the results of ages from 40 to 90 with a 5-year gap. Complete results are consistent and available upon request.

[10] Despite that the filtration of univariate model is a subset of the multivariate model, the additional cross-sectional information is mostly noise for the mortality rates of interested age group. Hence, the performance of multivariate models may be even worse than the univariate counterparts.

[11] A third DGP with LC model is also considered, which leads to consistent findings with those presented in the next subsections. Complete results can be found in the supplementary materials.

Table 1 Maximum cross-correlation between mortality rates of different ages

Variables	Ages										
	40	45	50	55	60	65	70	75	80	85	90
Panel A: Female											
Corr.	0.1646	0.2680	0.2735	0.1736	0.1806	0.1827	0.1617	0.2005	0.1813	0.1662	0.2393
P. val	(0.1950)	(0.0349)	(0.0313)	(0.1717)	(0.1551)	(0.1503)	(0.2029)	(0.1145)	(0.1534)	(0.1906)	(0.0595)
Panel B: Female											
Corr.	0.1224	0.1148	0.1191	0.1190	0.1163	0.1260	0.1343	0.1373	0.1401	0.1434	0.1350
P. val	(0.3350)	(0.3662)	(0.3483)	(0.3487)	(0.3600)	(0.3210)	(0.2903)	(0.2796)	(0.2701)	(0.2588)	(0.2878)

where $\Delta \ln m_{x,t} = \ln m_{x,t} - \ln m_{x,t-1}$. Then, we simulate T normally distributed sequences with null mean and variance being the sample variance of $\underset{x,t}{\hat{\epsilon}}$(fitted residuals). Hence, by using estimates of a_x and b_x from the fitted model, we can generate a series of simulated $\ln m_{x,t}$ with length T. Altogether, we simulate 300 replicates of $\ln m_{x,t}$.

In the second step, we divide each replicate into two parts:$t = 1, 2, \ldots, T-10$ and $t = T-9, T-8, \ldots, T$. All of the three multivariate and two univariate models are estimated using the first part, including LC, FDM, VAR, AR-ARCH and ETS. Then, we generate 10-step-ahead forecasts and compare them with the second part (true values) for each model. To measure the forecasting accuracy, we adopt three different statistics as considered in Hyndman and Koehler (2006): mean absolute forecast error (MAFE), mean absolute percentage error (MAPE) and mean absolute scaled error (MASE), which are just means of r_t, p_t and q_t, respectively. The formulae to calculate r_t, p_t and q_t are listed below.

$$\begin{aligned} r_t &= \ln m_{x,t} - \widehat{\ln m_{x,t}} \\ p_t &= \frac{100 r_t}{\ln m_{x,t}} \\ q_t &= \frac{r_t}{\frac{1}{T-1}\sum_{i=2}^{T} \left| \ln m_{x,i} - \ln m_{x,i-1} \right|} \end{aligned} \tag{11}$$

We summarise those measurements in Tables 2 and 3, and plot their means out of the 300 replicates for each model in Fig. 1. As the DGP is univariate AR(1), it is expected that the univariate models may outperform the multivariate models. In Table 2, for simulations of female mortality rates, ETS and VAR models yield much better results compared with others. All of the mean MAFE of ETS are below 0.01 and are the uniform minimums among five models for each age. Additionally, the standard deviations (SDs) of the MAFE of ETS are also quite small. The maximum SD is only 0.74, and almost all the SDs are the minimums among five models for each age. Hence, ETS leads to the smallest mean MAFE with smallest variation in almost all cases. Among the other four models, VAR outperforms the other three in most cases, and results of FDM are generally more accurate than LC and AR-ARCH. The above conclusions also consistently hold in cases of MAPE and MASE. For male mortality, ETS still outperforms the others in terms of both mean and variation of the forecasting errors. The difference is that the results of FDM and VAR are fairly close, and are almost uniformly better than those of LC and AR-ARCH.

Simulation Study: VAR(1) DGP

We now consider the multivariate DGP using the Australian mortality data. To demonstrate our arguments in "Australian mortality data" section that considering mortality rates of neighbouring ages may not improve the forecasting accuracy, we

Table 2 Female mortality with ARIMA DGP

Ages	ETS		VAR		LC		FDM		AR-ARCH	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Panel A: MAFE										
40	0.0044	0.0045	0.0053	0.0054	0.0298	0.0295	0.0139	0.0122	0.0157	0.0207
45	0.0045	0.0044	0.0059	0.0061	0.0297	0.0299	0.0179	0.0104	0.0157	0.0199
50	0.0035	0.0039	0.0049	0.0053	0.0231	0.0224	0.0120	0.0087	0.0123	0.0168
55	0.0027	0.0045	0.0029	0.0041	0.0115	0.0115	0.0152	0.0057	0.0083	0.0173
60	0.0055	0.0051	0.0067	0.0061	0.0086	0.0092	0.0123	0.0058	0.0172	0.0234
65	0.0077	0.0068	0.0110	0.0101	0.0222	0.0202	0.0166	0.0123	0.0274	0.0285
70	0.0082	0.0074	0.0117	0.0125	0.0276	0.0267	0.0159	0.0148	0.0307	0.0339
75	0.0066	0.0058	0.0105	0.0098	0.0163	0.0144	0.0108	0.0102	0.0246	0.0276
80	0.0084	0.0072	0.0143	0.0151	0.0290	0.0261	0.0128	0.0132	0.0324	0.0354
85	0.0030	0.0030	0.0038	0.0041	0.0063	0.0067	0.0075	0.0065	0.0108	0.0154
90	0.0013	0.0019	0.0014	0.0022	0.0084	0.0079	0.0044	0.0046	0.0049	0.0121
Panel B: MAPE										
40	0.0613	0.0637	0.0733	0.0760	0.4146	0.4066	0.1943	0.1704	0.2187	0.2896
45	0.0678	0.0675	0.0881	0.0922	0.4445	0.4424	0.2684	0.1568	0.2355	0.2962
50	0.0575	0.0665	0.0811	0.0881	0.3814	0.3679	0.1988	0.1437	0.2042	0.2793
55	0.0469	0.0778	0.0487	0.0700	0.1964	0.1960	0.2598	0.0981	0.1418	0.2944
60	0.0990	0.0927	0.1198	0.1112	0.1550	0.1659	0.2231	0.1070	0.3108	0.4264
65	0.1531	0.1364	0.2190	0.2018	0.4425	0.4078	0.3314	0.2507	0.5464	0.5794
70	0.1886	0.1734	0.2715	0.2942	0.6437	0.6386	0.3714	0.3513	0.7143	0.8018
75	0.1676	0.1489	0.2681	0.2510	0.4174	0.3718	0.2780	0.2701	0.6336	0.7198
80	0.2447	0.2147	0.4193	0.4493	0.8594	0.8007	0.3781	0.4023	0.9593	1.0824
85	0.1134	0.1132	0.1426	0.1545	0.2386	0.2561	0.2855	0.2514	0.4101	0.5942
90	0.0676	0.1057	0.0732	0.1201	0.4522	0.4231	0.2368	0.2477	0.2636	0.6600
Panel C: MASE										
40	0.2573	0.3114	0.3023	0.3532	1.5826	1.5266	0.7526	0.7065	0.8730	1.2406
45	0.2560	0.2783	0.3286	0.3690	1.8567	1.7956	1.1256	0.6149	0.9432	1.1672
50	0.2125	0.2645	0.2894	0.3283	1.6163	1.5321	0.8316	0.5683	0.8247	1.1291
55	0.1973	0.3443	0.2027	0.3127	0.8793	0.8610	1.1760	0.4414	0.6232	1.3212
60	0.4588	0.4568	0.5481	0.5269	0.7149	0.7908	1.0070	0.5165	1.4325	2.0581
65	0.7724	0.6995	1.1092	1.0335	2.0201	1.8405	1.4912	1.2033	2.5443	2.6757
70	0.6484	0.6098	0.9325	1.0136	2.0062	1.9741	1.1758	1.1778	2.2906	2.5774
75	0.4319	0.4092	0.6953	0.6685	1.0241	0.9473	0.6951	0.7635	1.5729	1.8927
80	0.4937	0.4477	0.8385	0.9410	1.6363	1.5992	0.7369	0.8663	1.8728	2.3242
85	0.1617	0.1862	0.2024	0.2328	0.3453	0.4338	0.4051	0.4298	0.5864	0.9706
90	0.0637	0.1098	0.0684	0.1163	0.4478	0.5005	0.2226	0.2356	0.2517	0.6657

construct a VAR(1) model for every five ages. However, we find that most of those VAR models are still not stationary after being first-differenced. Therefore, we difference $\ln m_{x,t}$ twice and fit the VAR(1) model as described in Eq. (5) for every five

Table 3 Male mortality with ARIMA DGP

Ages	ETS		VAR		LC		FDM		AR-ARCH	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Panel A: MAFE										
40	0.0028	0.0040	0.0033	0.0044	0.0265	0.0485	0.0067	0.0077	0.0091	0.0122
45	0.0044	0.0056	0.0052	0.0071	0.0484	0.0768	0.0093	0.0139	0.0142	0.0170
50	0.0044	0.0056	0.0049	0.0069	0.0496	0.0760	0.0090	0.0138	0.0141	0.0195
55	0.0052	0.0057	0.0063	0.0072	0.0178	0.0315	0.0069	0.0060	0.0186	0.0224
60	0.0082	0.0087	0.0116	0.0127	0.0200	0.0206	0.0106	0.0094	0.0320	0.0362
65	0.0071	0.0074	0.0099	0.0105	0.0198	0.0305	0.0102	0.0092	0.0293	0.0347
70	0.0101	0.0107	0.0153	0.0170	0.0281	0.0296	0.0127	0.0122	0.0404	0.0462
75	0.0113	0.0120	0.0154	0.0158	0.0371	0.0413	0.0140	0.0136	0.0441	0.0503
80	0.0091	0.0095	0.0146	0.0166	0.0310	0.0334	0.0115	0.0111	0.0353	0.0423
85	0.0081	0.0087	0.0133	0.0147	0.0243	0.0251	0.0103	0.0104	0.0311	0.0334
90	0.0054	0.0057	0.0065	0.0065	0.0168	0.0168	0.0070	0.0067	0.0216	0.0255
Panel B: MAPE										
40	0.0435	0.0620	0.0509	0.0683	0.4145	0.7629	0.1044	0.1194	0.1407	0.1901
45	0.0736	0.0943	0.0868	0.1167	0.8117	1.2571	0.1565	0.2365	0.2399	0.2881
50	0.0821	0.1048	0.0898	0.1238	0.9152	1.3722	0.1667	0.2578	0.2625	0.3593
55	0.1058	0.1155	0.1276	0.1426	0.3686	0.6758	0.1393	0.1225	0.3748	0.4486
60	0.1752	0.1887	0.2481	0.2726	0.4274	0.4430	0.2280	0.2125	0.6852	0.7861
65	0.1650	0.1748	0.2305	0.2500	0.4728	0.7879	0.2404	0.2244	0.6802	0.8053
70	0.2614	0.2886	0.3943	0.4449	0.7298	0.7785	0.3342	0.3479	1.0489	1.2241
75	0.3438	0.3860	0.4594	0.5017	1.1455	1.4220	0.4270	0.4635	1.3262	1.5553
80	0.3189	0.3518	0.5044	0.5793	1.1035	1.3299	0.4070	0.4336	1.2293	1.5331
85	0.3569	0.4062	0.5819	0.6997	1.0653	1.1457	0.4576	0.5130	1.3545	1.4940
90	0.3136	0.3430	0.3802	0.4131	0.9769	0.9999	0.4085	0.4315	1.2475	1.5465
Panel C: MASE										
40	0.2646	0.4162	0.3015	0.4272	2.2432	3.5977	0.6162	0.7887	0.8377	1.1697
45	0.4133	0.5519	0.4699	0.6082	5.1129	10.1088	0.8865	1.3343	1.3838	1.7563
50	0.4183	0.5448	0.4473	0.5788	5.9631	13.6905	0.8801	1.3043	1.3923	1.8979
55	0.5251	0.6138	0.6202	0.7102	2.3858	6.7948	0.7081	0.7129	1.8864	2.3103
60	0.8964	1.1605	1.2565	1.5972	2.1772	2.9976	1.1972	1.5689	3.5055	4.8196
65	0.8690	1.2141	1.2055	1.6039	2.8045	8.8585	1.2938	1.9040	3.3328	4.8146
70	1.0571	1.5059	1.5853	2.2781	2.7728	3.5731	1.4166	2.2458	4.0264	5.7319
75	1.1064	1.6428	1.3852	1.7838	3.6379	6.3164	1.4015	2.2835	4.0605	6.1015
80	0.8338	1.2345	1.2650	1.7878	2.8478	4.7934	1.0855	1.8291	3.0495	4.7505
85	0.6979	1.0193	1.1248	1.7203	1.9634	2.6540	0.9215	1.5484	2.5404	3.8339
90	0.4364	0.6146	0.5264	0.7209	1.3193	1.7640	0.5926	0.9632	1.6806	2.7041

ages (that is, 40–44, 45–49,…, 85–91). Then, we simulate T multi-normally distributed sequences with null mean and covariance being the sample covariance of $\hat{\varepsilon}_{x,t}$ (fitted residual matrix). Hence, using estimates of A and B from the fitted model, we generate the simulated $\ln m_{x,t}$ with length T for each of the five ages considered in

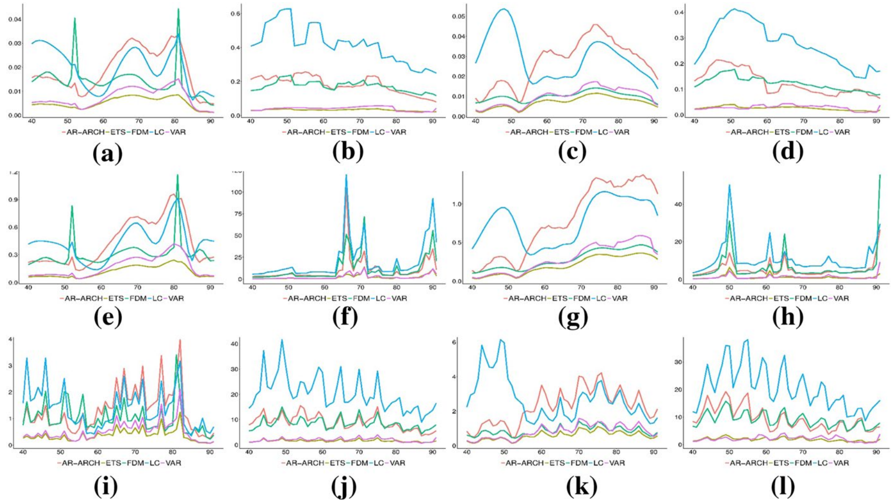

Fig. 1 Simulated Australian mortality rates. **a** MAFE: Female ARIMA, **b** MAFE: Female VAR, **c** MAFE: Male ARIMA, **d** MAFE: Male VAR, **e** MAPE: Female ARIMA, **f** MAPE: Female VAR, **g** MAPE: Male ARIMA, **h** MAPE: Male VAR, **i** MASE: Female ARIMA, **j** MASE: Female VAR, **k** MASE: Male ARIMA and **l** MASE: Male VAR

the VAR. We also simulate 300 such replicates for ages 40–91 and fit them by the five models. The results of MAFE, MAPE and MASE are summarised in Tables 4 and 5, and are plotted in Fig. 1.

Although the true DGP here includes adjacent 5-year age groups, we argue in "Australian mortality data" section that the cross-correlations cannot bring in additional information other than that of the autocorrelations. Hence, the multivariate models here may not necessarily outperform the univariate models, as the filtration of the former includes much more noise. We firstly consider the female mortality rates. In Table 4, it can be observed that the mean MAFE of ETS and VAR are still much smaller than those of the other models. Comparing the results of ETS and VAR, except for the ages 40, 80, 85 and 90, the mean MAFE of ETS are still the smallest, despite the true DGP being now multivariate. In terms of variations, ETS also leads to the smallest SD in most cases. For the complete ages 40–91, it is demonstrated in Fig. 1 that VAR only outperforms ETS for ages around 40 and after 80. LC is uniformly inferior to all the other models in all cases, while the performance of FDM and AR-ARCH are now quite close. The above conclusions also consistently hold in the cases of MAPE and MASE for female mortality rates. For males, ETS still outperforms all other models with respect to the mean and variation of forecasting error in most cases. The only difference is that VAR leads to the best forecasts for the ages 40–55, not just a few groups around 40 as in the female case. Our observations of the performance of LC, FDM and AR-ARCH for female mortality remain the same.

Table 4 Female mortality with VAR DGP

Ages	ETS		VAR		LC		FDM		AR-ARCH	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Panel A: MAFE										
40	0.0303	0.0232	0.0292	0.0265	0.4083	0.3289	0.1471	0.1182	0.2124	0.1808
45	0.0361	0.0270	0.0398	0.0404	0.5561	0.5439	0.1782	0.1427	0.2137	0.1963
50	0.0393	0.0322	0.0458	0.0593	0.6272	0.6190	0.2328	0.1651	0.2086	0.1832
55	0.0327	0.0244	0.0461	0.0604	0.4226	0.3707	0.1793	0.1500	0.2573	0.2191
60	0.0338	0.0262	0.0414	0.0534	0.4305	0.4535	0.1670	0.1222	0.1770	0.1587
65	0.0342	0.0279	0.0448	0.0474	0.4384	0.4569	0.1869	0.1484	0.1704	0.1494
70	0.0386	0.0298	0.0526	0.0519	0.4282	0.4373	0.1953	0.1566	0.1739	0.1452
75	0.0403	0.0305	0.0570	0.0576	0.4279	0.4524	0.1887	0.1483	0.2251	0.1890
80	0.0298	0.0223	0.0252	0.0203	0.3168	0.3246	0.1469	0.1107	0.1419	0.1311
85	0.0253	0.0183	0.0210	0.0180	0.2607	0.2317	0.1301	0.0972	0.1129	0.1051
90	0.0219	0.0177	0.0214	0.0172	0.2597	0.2766	0.1273	0.0943	0.0883	0.0771
Panel B: MAPE										
40	0.4380	0.3469	0.4318	0.4152	5.7058	4.3285	2.1435	1.7766	3.0587	2.6490
45	0.5917	0.5191	0.6468	0.8196	8.3588	7.9943	2.9047	2.7937	3.4267	3.6319
50	1.0081	2.9027	1.2753	3.0249	12.4704	18.3003	5.9857	19.4362	5.7554	18.1984
55	0.5712	0.4628	0.7792	0.9319	6.8896	5.3075	3.1536	2.8212	4.5442	4.5102
60	0.6920	0.5974	0.8522	1.0888	8.2769	8.0542	3.3690	2.8236	3.7375	3.9033
65	2.8979	9.0209	2.7901	7.4172	29.7051	82.7891	17.2235	52.8164	10.4984	32.7160
70	5.3524	17.7577	5.5446	15.0906	36.3656	97.6679	25.6943	96.7204	19.2777	53.5405
75	2.5682	12.8983	4.0047	24.9491	14.7634	51.4779	10.5008	56.1056	13.7769	59.5396
80	2.1230	18.6673	2.0145	18.5583	23.4260	232.2112	15.3847	176.4884	8.9103	58.7595
85	2.0297	8.3369	1.4362	3.9140	14.7472	25.5891	8.2598	15.6028	7.4799	20.7677
90	11.6743	74.6490	11.9187	82.3419	92.6180	341.0573	55.8027	236.1397	34.3750	179.5496

Table 4 (continued)

Ages	ETS		VAR		LC		FDM		AR-ARCH	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Panel C: MASE										
40	1.0909	1.0229	1.0482	1.1066	14.5438	11.2746	5.5596	4.8123	7.9025	7.2545
45	1.4492	1.4430	1.5345	1.7819	22.9719	21.4461	7.8140	6.5589	8.8385	9.6090
50	1.6735	1.5197	1.9911	2.9614	31.3522	34.3236	11.4158	9.3859	10.0834	9.8624
55	1.6279	1.3286	2.1641	2.4881	24.5231	22.7247	10.2662	10.1077	14.4357	14.1840
60	1.9619	1.7211	2.3492	3.4280	28.7396	36.6789	10.7373	9.0767	11.1290	11.6563
65	2.3996	2.2104	3.2997	4.3658	31.0074	36.7287	13.0430	11.7990	12.3623	12.6493
70	2.7621	2.5898	3.8648	6.1138	29.9135	32.9522	13.8816	13.8250	11.9067	10.9492
75	2.8169	2.6168	4.0337	4.2217	29.1662	33.9988	13.0962	12.2641	15.0703	14.7904
80	1.9043	1.6807	1.6703	1.5251	17.3765	17.1746	8.8259	7.3938	8.7071	8.5256
85	1.4458	1.3745	1.2493	1.4538	14.5855	16.6351	7.3155	6.6391	6.4849	8.3590
90	1.2133	1.1827	1.1551	1.0956	13.7218	16.8216	6.9044	6.3268	4.7032	4.8201

Table 5 Male mortality with VAR DGP

Ages	ETS		VAR		LC		FDM		AR-ARCH	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Panel A: MAFE										
40	0.0230	0.0166	0.0216	0.0292	0.1966	0.1825	0.1089	0.0839	0.1319	0.1031
45	0.0303	0.0224	0.0273	0.0725	0.3256	0.3260	0.1452	0.1180	0.2005	0.1627
50	0.0401	0.0328	0.0278	0.0353	0.4067	0.3774	0.1735	0.1386	0.2042	0.1589
55	0.0271	0.0190	0.0240	0.0378	0.3926	0.4104	0.1343	0.1094	0.1912	0.1716
60	0.0277	0.0194	0.0286	0.0335	0.2890	0.3177	0.1248	0.0941	0.0837	0.0834
65	0.0302	0.0227	0.0436	0.0815	0.3145	0.3881	0.1308	0.0993	0.1177	0.1032
70	0.0283	0.0225	0.0337	0.0383	0.2622	0.2467	0.1193	0.0954	0.1021	0.0881
75	0.225	0.0165	0.0316	0.0556	0.2419	0.2814	0.1048	0.0784	0.0842	0.0709
80	0.0179	0.0135	0.0282	0.0610	0.1956	0.2174	0.0971	0.0721	0.0896	0.0746
85	0.0163	0.0120	0.0100	0.0088	0.1466	0.1560	0.0945	0.0706	0.0998	0.0791
90	0.0126	0.0085	0.0098	0.0080	0.1684	0.1848	0.0775	0.0619	0.0718	0.0627
Panel B: MAPE										
40	0.4026	0.3094	0.3783	0.5270	3.4886	3.6293	1.8767	1.5535	2.2697	1.8046
45	0.6829	0.7343	0.6652	2.2043	7.3881	8.2963	3.3137	3.9790	4.2654	4.1690
50	6.3633	44.7434	4.6284	42.1925	50.2627	360.2154	31.0611	242.9370	13.9659	70.5328
55	0.6557	0.6820	0.5754	0.8361	8.5925	10.6227	3.0830	2.7783	4.7673	6.7683
60	1.2472	4.4636	0.9446	2.1249	10.3088	24.4941	4.4820	9.0050	2.9839	9.5304
65	1.1883	2.2774	1.9899	6.7311	14.6662	88.3993	24.0310	328.7608	13.9583	162.8270
70	0.8004	0.7889	1.0128	1.6271	6.8826	6.0720	3.3190	2.9144	2.9236	2.9549
75	0.7269	0.6505	1.1035	2.4316	7.1302	6.9166	3.3859	2.8457	2.8331	3.1096
80	0.6499	0.5586	1.1295	4.4541	6.5750	6.8337	3.4576	2.6738	3.3490	3.2619
85	0.6887	0.5327	0.4244	0.3872	5.9902	5.8462	3.9955	2.9944	4.4242	3.9971
90	1.7570	16.4915	0.7243	1.5816	20.1748	173.3910	8.1562	60.5850	15.3528	189.4241

Table 5 (continued)

Ages	ETS		VAR		LC		FDM		AR-ARCH	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Panel C: MASE										
40	1.2943	1.1655	1.1807	1.7008	11.9803	11.5781	6.7034	5.6354	8.3918	8.0209
45	1.9457	1.8354	1.5860	3.9047	22.6980	23.3487	10.5837	9.1774	14.5159	13.3923
50	3.0215	2.8510	2.0172	2.7132	35.8204	39.9992	14.5180	12.7576	17.2022	15.4238
55	2.4059	2.0436	2.0220	2.6809	38.0599	43.6684	12.6610	12.0483	18.7325	20.1563
60	2.7356	2.3199	2.7527	4.1975	29.1602	39.6415	12.4612	11.5171	8.3384	9.2039
65	3.1117	3.4304	4.2092	6.6233	32.3938	41.9949	13.4264	13.5232	11.6967	11.5040
70	2.8210	2.8171	3.1659	4.1454	25.6736	31.2671	11.3630	11.0293	9.4875	9.1977
75	2.0899	1.9468	3.3333	11.5961	22.7726	30.0711	9.6242	8.4381	7.5965	7.1636
80	1.6779	1.6470	2.8162	8.7558	16.3072	17.1143	8.7294	7.0902	8.2425	7.8102
85	1.4997	1.4144	0.9055	0.9029	13.1537	16.6490	8.3497	7.2016	9.2333	9.2672
90	1.1752	1.0830	0.9091	0.9263	14.6665	19.3245	6.8138	7.0350	6.2449	6.4550

In conclusion, Australian mortality data are used to demonstrate that cross-correlations of both female and male mortality rates are insignificant, after the univariate effects have been considered. It can be explained by the fact that mortality rates of neighbouring ages are quite similar, so that the cross-correlations contain information almost identical to the autocorrelations. Hence, in terms of forecasting accuracy, the multivariate mortality models do not necessarily outperform the univariate counterparts. Our simulation studies include two distinct DGPs: univariate AR(1) and VAR(1) including five adjacent age groups. In both studies, the ETS model outperforms VAR, LC, FDM and AR-ARCH in most cases. These results further support our explanation of the insignificant cross-correlations of mortality rates. In the next section, we will work on the empirical Australian mortality data and compare forecasting accuracy of the models.

Empirical Results

In this section, we firstly analyse the same set of data described in "Australian mortality data" section and use MAFE, MAPE and MASE to compare the forecasting performance of the five models. We then consider various settings of time period, age groups, forecasting horizon and smoothness of rates to check the robustness of our conclusions.

Australian Mortality Data: 40–91 Years Old, 10-Step-Ahead Horizon, 1950–2011 and Two-dimensionally Smoothed Rates

We first consider the same dataset as introduced in "Australian mortality data" section. To compare forecasts, we divide the female and male mortality data into two parts: 1950–2001 and 2002–2011. The first part will be used to fit the five models and perform 10-step-ahead forecasting. The forecasted rates will then be compared with the second part (true data) to calculate the MAFE, MAPE and MASE. The results of selected ages are given in Table 4, and those of the complete ages are plotted in Fig. 2 and summarised in Panel A of Table 6.

In Table 7, it can be seen that ETS leads to the smallest MAFE for female mortality rates in most cases, where all of the errors are below 0.05. This can also be visually observed in Fig. 2a: except for a few age groups around 55, 75 and 90, MAFEs of ETS are almost uniformly smaller than those of the other models. In Panel A of Table 6, the average MAFE over the ages 40–91 of ETS is 0.0194 with a SD 0.0137. Hence, ETS generates the smallest average MAFE with the least variation over the interested ages. Compared to the other four models, VAR is the best in terms of both average and variation of MAFE. Results of LC and FDM are similar, while

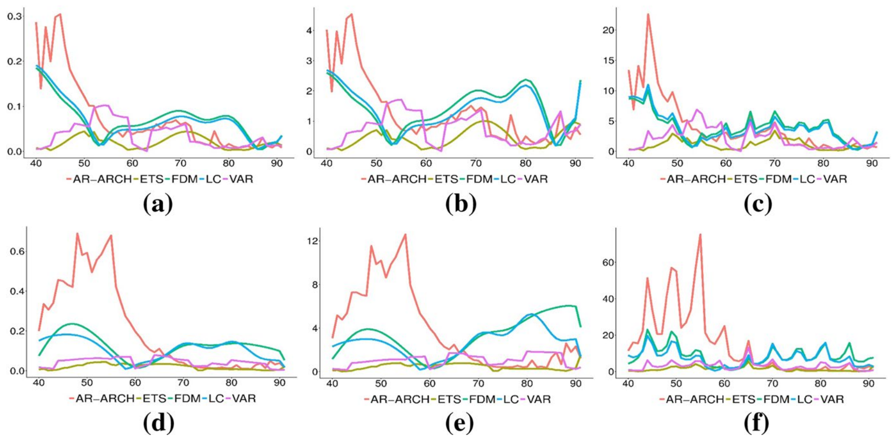

Fig. 2 Australian mortality rates: ages 40–91, 1950–2011, 10-year horizon and two- dimensionally smoothed rates. **a** Female: MAFE, **b** Female: MAPE, **c** Female: MASE, **d** Male: MAFE, **e** Male: MAPE and **f** Male: MASE

AR-ARCH is on average the least preferred model, due to its inferior performance for ages 40–55. The above conclusions also hold in the cases of MAPE and MASE.

Turning to male mortality, the advantage of ETS is even more distinct. In Fig. 2b, except for several age groups around 60, ETS uniformly outperforms the other models with respect to MAFE. Results of VAR model are not significantly different from those of ETS and considerably better than those of the rest. LC and FDM still lead to similar MAFEs, while AR-ARCH only outperforms them for ages above 65. Average MAFE over ages 40–91 of ETS is only 0.0193, less than half that of VAR (the second best model). Besides, ETS still has the smallest variation in MAFE. Despite its good performance for ages above 65, AR-ARCH has the largest errors for ages 40–65, which are much larger than those of the others (some MAFEs even exceed 0.6). Hence, AR-ARCH is still the least preferred model on average. The inferiority of AR-ARCH model may be due to its inflexibility of fixed order 1 for both AR and ARCH processes. Apparently, fixed specification does not fit all cases at the 10-step-ahead forecasting horizon.

The historical rates (1950–2001), actual rates and their forecasts from different models (2002–2011) for ages 40, 60, 80 and 90 are presented in Fig. 3. The shaded area stands for the 95% prediction interval produced by the ETS model. It can be seen that the forecasts of ETS are closest to the actual values in almost all cases. The performance of VAR is overall better than those of the remaining models.

Table 6 Australian mortality forecast summary

Models	Female						Male					
	MAFE		MAPE		MASE		MAFE		MAPE		MASE	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Panel A: 1950–2011, 40–91 ages and 10-year horizon												
ETS	0.0194	0.0137	0.4412	0.3130	1.0987	0.8013	0.0193	0.0137	0.4431	0.2948	1.3738	1.1698
VAR	0.0363	0.0294	0.7421	0.4841	2.0959	1.7329	0.0432	0.0217	1.0558	0.5236	3.1015	2.2794
LC	0.0671	0.0441	1.3831	0.6506	3.7706	2.4302	0.1065	0.0514	2.6836	1.2664	7.3902	4.4119
FDM	0.0698	0.0408	1.4618	0.6625	3.9374	2.2664	0.1219	0.0579	3.2501	1.7010	8.4893	5.1018
AR-ARCH	0.0745	0.0816	1.3153	1.1162	4.1742	4.6953	0.2032	0.2207	3.8273	3.6585	14.5120	18.0950
Panel B: 1950–2011, 0–100 ages and 10-year horizon												
ETS	0.0194	0.0137	0.4412	0.3130	1.0987	0.8013	0.0193	0.0137	0.4431	0.2948	1.3738	1.1698
VAR	0.0380	0.0286	0.8209	0.5492	2.2895	2.0410	0.0450	0.0193	1.1561	0.5015	3.9118	3.2243
LC	0.0700	0.0415	1.4454	0.6434	4.1150	2.7150	0.1089	0.0510	2.7557	1.2973	9.8119	8.6233
FDM	0.0526	0.0386	1.0568	0.6097	3.0683	2.4914	0.1066	0.0449	2.8156	1.3528	9.4542	7.9377
AR-ARCH	0.0745	0.0816	1.3153	1.1162	4.1742	4.6953	0.2032	0.2207	3.8273	3.6585	14.5120	18.0952
Panel C: 1921–2011, 40–91 ages and 10-year horizon												
ETS	0.0767	0.0640	1.6752	1.1993	4.3756	3.6267	0.0876	0.0456	2.2280	1.0740	6.1841	3.8839
VAR	0.2752	0.5025	5.0329	6.9820	15.0874	26.6381	0.5877	0.7940	18.2616	36.5062	40.6551	57.1606
LC	0.1039	0.0545	2.2469	0.9623	5.9678	3.2139	0.2589	0.1102	6.4550	2.4106	18.2906	9.2777
FDM	0.1410	0.0520	3.0504	0.7462	8.0889	3.1823	0.1180	0.0478	3.0683	1.1491	8.3144	4.1661
AR-ARCH	0.4176	0.1141	9.7882	3.2102	23.9086	7.2181	0.3077	0.1473	7.4285	2.1362	21.6171	11.9603
Panel D: 1950–2011, 40–91 ages and 20-year horizon												
ETS	0.0196	0.0151	0.4471	0.3418	1.4111	1.1422	0.0265	0.0185	0.6127	0.3482	2.2897	1.8707
VAR	0.0285	0.0164	0.6534	0.3570	2.0673	1.2860	0.0402	0.0304	0.8305	0.5342	3.3404	2.7484
LC	0.0991	0.0956	1.9532	1.3482	6.9751	6.4408	0.1635	0.1035	3.8447	1.6169	13.8581	9.3468
FDM	0.0481	0.0245	1.2837	1.4415	3.5810	2.3790	0.0829	0.0498	2.3070	1.8753	7.1655	4.9885

Table 6 (continued)

Models	Female						Male					
	MAFE		MAPE		MASE		MAFE		MAPE		MASE	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
AR-ARCH	0.0621	0.0558	1.2594	0.9767	4.4062	3.8730	0.0941	0.0534	2.4015	1.3601	8.0802	5.2076
Panel E: Crude rates, 1950–2011, 40–91 ages and 10-year horizon												
ETS	0.0650	0.0410	1.3846	0.6522	1.1062	0.8065	0.0710	0.0436	1.8458	1.3261	1.3110	0.8202
VAR	0.0904	0.0557	1.8897	0.8334	1.5303	0.9767	0.0936	0.0509	2.2863	1.1364	1.7163	0.9800
LC	0.0776	0.0488	1.5951	0.6802	1.2931	0.8192	0.1133	0.0644	2.7141	1.1390	2.0765	1.2387
FDM	0.0725	0.0320	1.5479	0.5158	1.2156	0.5800	0.0953	0.0624	2.1760	0.9483	1.7520	1.2172
AR-ARCH	0.1058	0.0832	1.9967	1.1108	1.7897	1.5011	0.1596	0.1389	3.1976	2.0978	2.9525	2.6431
Panel F: Dimensionally smoothed rates, 1950–011, 40–91 ages and 10-year horizon												
ETS	0.0520	0.0334	1.1122	0.5901	1.4509	1.0115	0.0588	0.0289	1.6211	1.1747	1.6254	0.8936
VAR	0.0712	0.0499	1.4231	0.6481	1.9728	1.4504	0.0740	0.0318	2.0612	1.3483	2.0574	0.9736
LC	0.0684	0.0477	1.4012	0.6371	1.9046	1.3617	0.1108	0.0600	2.6827	1.0447	3.0614	1.7481
FDM	0.0589	0.0250	1.2805	0.4315	1.6411	0.7567	0.0955	0.0751	2.0909	1.0646	2.6393	2.1382
AR-ARCH	0.1034	0.0929	1.9738	1.2374	2.8488	2.5176	0.1769	0.1738	3.4706	2.6884	4.8567	4.8410

Table 7 Australian mortality forecast result: ages 40–91, 1950–2011, 10-step-ahead horizon and two-dimensionally smoothed rates

Ages	MAFE					MAPE					MASE				
	ETS	VAR	LC	FDM	AR-ARCH	ETS	VAR	LC	FDM	AR-ARCH	ETS	VAR	LC	FDM	AR-ARCH
Panel A: Female															
40	0.0077	0.0050	0.1912	0.1850	0.2864	0.1075	0.0708	2.6887	2.6018	4.0260	0.3578	0.2358	8.9463	8.6569	13.3964
45	0.0137	0.0428	0.1332	0.1203	0.3043	0.2034	0.6369	1.9859	1.7927	4.5360	0.6609	2.0275	7.6236	6.8411	16.9189
50	0.0444	0.0572	0.0750	0.0642	0.1217	0.7074	0.9107	1.1965	1.0234	1.9403	2.3993	2.8534	4.5039	3.8257	7.2821
55	0.0285	0.1015	0.0277	0.0380	0.0431	0.4813	1.7123	0.4710	0.6453	0.7280	1.8021	6.1799	2.2698	2.9718	2.7859
60	0.0046	0.0170	0.0480	0.0561	0.0379	0.0847	0.3114	0.8829	1.0319	0.6962	0.3774	1.3513	3.9589	4.6225	3.1079
65	0.0160	0.0472	0.0577	0.0728	0.0604	0.3181	0.9395	1.1581	1.4606	1.2052	1.5808	4.8335	5.2063	6.5506	5.8824
70	0.0424	0.0570	0.0779	0.0900	0.0595	0.9422	1.2670	1.7458	2.0168	1.3203	3.4030	4.5687	5.7203	6.6133	4.7770
75	0.0308	0.0154	0.0640	0.0702	0.0168	0.7757	0.3884	1.6263	1.7847	0.4249	2.2000	1.0782	4.3600	4.7663	1.1879
80	0.0026	0.0080	0.0723	0.0785	0.0165	0.0775	0.2424	2.1859	2.3760	0.4985	0.1665	0.5017	4.6696	5.0514	1.0823
85	0.0075	0.0196	0.0126	0.0181	0.0062	0.2868	0.7485	0.4792	0.6904	0.2362	0.4610	1.2050	0.7953	1.1150	0.3914
90	0.0186	0.0133	0.0214	0.0201	0.0152	0.9575	0.6815	1.1109	1.0430	0.7932	1.1138	0.7990	1.2095	1.1503	0.8075
Panel B: Male															
40	0.0065	0.0167	0.1506	0.0754	0.1999	0.0987	0.2549	2.3039	1.1516	3.0542	0.3534	0.9006	8.8559	4.3202	11.4169
45	0.0132	0.0511	0.1818	0.2235	0.4504	0.2131	0.8230	2.9358	3.6051	7.2645	0.8461	3.0403	15.2364	17.2495	33.8933
50	0.0373	0.0614	0.1624	0.2094	0.5921	0.6412	1.0539	2.7976	3.6041	10.1905	2.8585	4.4349	15.8374	18.9453	54.9736
55	0.0311	0.0595	0.0775	0.1069	0.6790	0.5737	1.0991	1.4375	1.9781	12.5849	3.0436	5.6541	8.5956	11.1281	75.2280
60	0.0159	0.0114	0.0274	0.0165	0.1974	0.3194	0.2285	0.5529	0.3345	3.9937	2.0128	1.4169	3.5732	1.8755	25.0130
65	0.0340	0.0763	0.0615	0.0571	0.1103	0.7597	1.6993	1.3821	1.2948	2.4734	5.4707	13.2232	8.9610	7.2872	16.9177
70	0.0270	0.0543	0.1357	0.1255	0.0403	0.6744	1.3590	3.4330	3.1891	1.0055	3.3139	6.6630	15.3466	13.7669	5.0089
75	0.0118	0.0240	0.1139	0.1282	0.0143	0.3406	0.6944	3.3274	3.7682	0.4138	1.2590	2.4864	11.3347	12.3715	1.4666
80	0.0074	0.0535	0.1465	0.1367	0.0126	0.2563	1.8630	5.1714	4.8427	0.4429	0.7085	5.0686	13.0458	11.9994	1.1622
85	0.0053	0.0400	0.0773	0.1281	0.0287	0.2355	1.7791	3.4643	5.7794	1.2991	0.4707	3.4674	6.3597	10.1755	2.2311
90	0.0056	0.0040	0.0496	0.0988	0.0384	0.3393	0.2358	2.9708	5.9771	2.3171	0.4195	0.3243	3.7388	7.1506	2.7977

Nevertheless, all actual values fall in the 95% prediction interval of ETS, indicating its satisfactory forecasting accuracy.

Australian Mortality Data: 0–100 Years Old, 10-Step-Ahead Horizon, 1950–2011 and Two-dimensionally Smoothed Rates[12],[13]

We now include the ages 0–100 in the models, but still focus on the forecasted mortality rates for the ages 40–91. Note that changing the age dimension will only affect the results of multivariate models. Including more ages will enrich the filtration of multivariate models, but may not necessarily improve forecasts. As discussed in "Analysis on the cross-correlations and simulation results" section, this may be due to the fact that the cross-correlations are mostly insignificant, so that including more ages can only bring in more noise. The new results of selected ages are presented and plotted in Table 1 and Fig. 1 of the supplementary materials and summarised in Panel B of Table 6.

Comparing the new figure and Fig. 2, the difference between MAFEs of VAR model is quite small for both female and male mortality rates. The only distinct change is that FDM now outperforms LC in most cases for both female and male data. In Panels A and B of Table 6, it can be seen that the average MAFEs of LC and VAR increase to some extent, after including more ages. Taking the female mortality for instance, the mean MAFE of LC (VAR) increases from 0.0671 (0.0363) to 0.0700 (0.0380). Only the average MAFE of FDM decreases, from 0.0698 to 0.0526. Also, the SDs of MAFE for female data all decline to some degree. However, these results are not consistent among all cases. For example, the averages and variations of MASE of the male mortality increase greatly for all multivariate models. More specifically, the mean (SD) of MASE of FDM grows from 8.4893 (5.1018) to 9.4542 (7.9377). Despite the changes of multivariate models, ETS still leads to the least MAFE, MAPE and MASE in most cases of female and male data.

Australian Mortality Data: 40–91 Years Old, 10-Step-Ahead Horizon, 1921–2011 and Two-dimensionally Smoothed Rates[14]

We now consider the complete available data back to 1921. As discussed in Booth et al. (2006), the Australian mortality structures before and after 1950 are considerably different. Hence, including the data before 1950 will bring in more noise. Therefore, the results of this section can be used as a robustness check against irrelevant

[12] We present only the fitted results for ages 40–91 in this paper for comparison purpose. Complete results for ages 0–100 can be found in the supplementary materials. In all cases, our results lead to consistent conclusions.

[13] From this subsection onward, all empirical results are used as robustness checks from different perspectives. As suggested by one of the anonymous referees, to avoid overloaded exhibits, we have presented all the relevant tables and figures in the supplementary materials.

[14] As noted by Booth et al. (2002), there are structural changes in Australian mortality around 1970 and the resulting Lee–Carter k_t has a kink in the trend. Therefore, we have checked the robustness for data ranging from 1971 to 2011 instead. Complete results are presented in the supplementary materials.

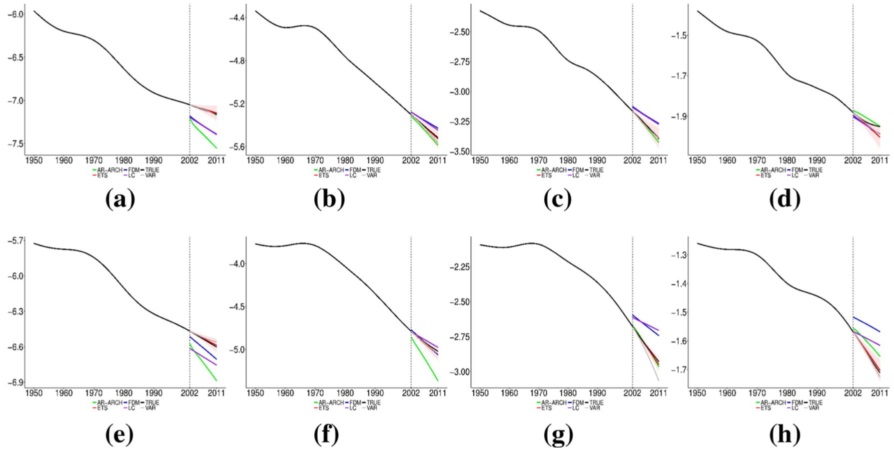

Fig. 3 Forecasts of the Australian mortality rates: ages 40, 60, 80 and 90, 1950–2011, 10-year horizon and two-dimensionally smoothed rates. **a** Female: 40, **b** Female: 60, **c** Female: 80, **d** Female: 90, **e** Male: 40, **f** Male: 60, **g** Male: 80 and **h** Male: 90

information. The new results are presented and plotted in Table 2 and Fig. 2 of the supplementary materials and summarised in Panel C of Table 6.

As expected, the errors presented in the new table are uniformly larger than those of the Table 7, showing that including noise will lead to more errors in all models. In the new figure, ETS still outperforms the other models in most cases. Also, VAR model leads to much larger errors for the ages around 40 (60 and 85) for female (male) mortality data. AR-ARCH model has the largest errors for most ages of female mortality rates, but outperforms LC for male rates in general. In Panel C of Table 6, ETS is still the best model regarding its average forecast errors over ages, after more noise is introduced.

Australian Mortality Data: 40–91 Years Old, 20-Step-Ahead Horizon, 1950–2011 and Two-dimensionally Smoothed Rates

We now extend forecasting horizon from 10 to 20-step-ahead. The new results are presented and plotted in Table 3 and Fig. 3 of the supplementary materials and summarised in Panel D of Table 6.

Comparing Fig. 2a and the corresponding new figure, general shapes of ETS, VAR and LC curves do not change much, after extending the forecasting horizon from 10 to 20-step-ahead. Results of FDM and AR-ARCH change to a larger degree, and it seems that some extreme errors have largely decreased (e.g. errors of ages 40–50). In the male case, the shape of the FDM curve remains roughly the same, while that of the LC curve shows larger errors for the ages 40–50. Overall, although ETS is not consistently the best model, it has a structure that is comparatively more stable than the others, and its results are very close to the best ones over ages. This conclusion holds across both female and male data and the three error measurements. In Panel D of Table 6, ETS has the smallest averages and variations among MAFE, MAPE and MASE for both female and male data.

Australian Mortality Data: 40–91 Years Old, 10-Step-Ahead Horizon, 1950–2011 and Crude Rates

We now consider the crude rates instead of the two-dimensionally smoothed data.[15] The results can be viewed as the robustness check given that random sampling error exists. The new results are presented and plotted in Table 4 and Fig. 4 of the supplementary materials and summarised in Panel E of Table 6.

In the new figures, all curves have unstable structures, and it is unclear which one is generally lower. Despite that, the AR-ARCH model generates much larger errors for both females and males aged 40–60. VAR model also yields larger errors in most cases of females aged 45–60. In Panel E of Table 6, ETS and FDM lead to similar results for female data. More specifically, ETS has the lowest average error, while FDM has the smallest variation. For male data, ETS outperforms the others for both average and variation of errors.

Australian Mortality Data: 40–91 Years Old, 10-Step-Ahead Horizon, 1950–2011 and One-dimensionally Smoothed Rates

Finally, we consider one-dimensionally smoothed mortality rates, where the random sampling error has been removed among age groups.[16] The results can be viewed as the robustness of different smoothing approaches. New results are presented and plotted in Table 5 and Fig. 5 of the supplementary materials and summarised in Panel F of Table 6.

In the new figures, the curves now have much more stable structures, compared with those of the crude rates. It can be seen that although ETS is still not consistently the best model among ages, its curve is overall relatively more stable than those of the others. For female data, VAR and AR-ARCH lead to comparatively larger errors for some ages of 40–60. As to male data, AR-ARCH still generates much larger errors in most cases of ages 40–65. In Panel F of Table 6, we have similar observations as stated in "Australian mortality data: 40–91 years old, 10-step-ahead horizon, 1950–2011 and crude rates" section. ETS leads to the smallest average errors for female data, although its variations are slightly greater than those of the FDM. In the case of male data, ETS generates smallest averages and variations across the three error measurements.

To sum up, we use the Australian mortality data as an empirical example to compare the forecasting accuracy of five models of interest. Among different settings, including age groups, time periods, forecasting horizons and smoothness, ETS

[15] In order to fit the un-smoothed time pattern, here we only consider the ETS models with damped trends.

[16] As the time pattern is still not smoothed, we only consider the ETS models with damped trends here.

consistently outperforms the other models in terms of average error over ages. Its forecasting variation is also the smallest in most cases. Those conclusions hold for both female and male mortality data across three error measurements.

Concluding Remarks and Discussions

Mortality and longevity risks affect accuracy in population projection and pricing of insurance products like life insurance, pensions and annuities. A precise forecast of mortality rates is therefore invaluable for demographic research and actuarial practice. Over the past few decades, most of the existing stochastic mortality models have been developed based on the seminal work of Lee and Carter (1992). Such models can be viewed as a special multivariate process in which the covariance matrix depends on the drift vector and the innovations are intertemporally correlated (Girosi and King, 2007). However, for a general time series, only when the cross-correlations are meaningful and statistically significant can multivariate models outperform univariate models in terms of forecasting accuracy (Du Preez and Witt 2003). However, as mortality rates of neighbouring ages are almost identical, their cross-correlations may be insignificant, after controlling for the effects of autocorrelations.[17]

In this study, we work on the Australian mortality data from 1950 to 2011 including ages 40–91. First, we employ the univariate AR(1) model to fit log mortality rates of each age over time. Results show that the vast majority of cross-correlations are not significant. Hence, this is preliminary evidence that rates of neighbouring ages are usually not leading indicators in mortality forecasting. Therefore, multivariate mortality models may not necessarily lead to more accurate forecasts than the univariate counterparts. To verify this argument, we perform simulation studies with two distinct data generating processes. Among the five models including ETS, VAR, LC, FDM and AR-ARCH, simulation results show that ETS consistently outperforms the others. It holds for both female and male data over age groups and across three different forecasting error measurements. Finally, we fit the five models to empirical datasets with various settings, including different age groups, time periods, forecasting horizons and smoothness. As in the simulation results, ETS still consistently outperforms the others in terms of the average forecasting error over age groups in all cases.

One potential problem of the univariate models like ETS is that the forecasts for neighbouring ages may diverge. For example, it is possible that the 50-step-ahead forecast of the mortality rate of age 71 is higher than that of age 72. In other words, there is no guarantee that the longitudinal forecasts still have a smoothing structure,

[17] Technically, when cross-correlations are insignificant, a multivariate model like VAR should reduce to its univariate form (AR). However, excessive number of parameters in VAR may lead to infeasible estimation. Sparse VAR (sVAR) proposed by Davis et al. (2016) is an effective solution in this regard. Using a LASSO-type penalty, irrelevant parameters are forced to be 0 in this model. We consider the sVAR(1) model in the supplementary materials. Although its results are generally better than those of VAR, ETS is still overall the most preferred model considering all investigated ages.

which is desirable in demographic and actuarial practice. One simple but useful solution to such a problem is to treat those "raw" forecasts as the crude rates and apply a common one-dimension smoothing method. For instance, one can employ the weighted penalised regression splines with a monotonicity constraint as used by Hyndman and Ullah (2007) to smooth the "raw" forecasts generated by ETS. Therefore, the possible non-smoothing structure of the longitudinal forecasts will be removed.

Our results demonstrate the enormous effectiveness of the univariate models and especially the ETS models in mortality forecasting. Future research can be conducted to apply or extend ETS models to study other features of mortality rates. For example, ETS models can be employed to forecast mortality rates in the age-dimension with time periods fixed. Also, general ETS models can be extended to consider the potential heteroscedasticity of mortality rates. Additionally, one can extend our research to study the mortality data of various countries. For instance, ETS may be employed to fit mortality rates of multiple developing and developed countries, and the difference among forecasts can be compared and investigated.

Acknowledgements We are grateful to the Macquarie University and Jiangxi University of Finance and Economics for their support. The authors would also like to thank Rob Hyndman, James Raymer and Hanlin Shang for their helpful comments and suggestions. We are also grateful for the valuable advice provided by the Editor (Santosh Jatrana) and two anonymous referees. The usual disclaimer applies.

References

Bell, W. R. (1997). Comparing and assessing time series methods for forecasting age specific fertility and mortality rates. *Journal of Official Statistics, 13,* 279.

Booth, H., Hyndman, R., Tickle, L., De Jong, P., et al. (2006). Lee–Carter mortality forecasting: A multi-country comparison of variants and extensions. *Demographic Research, 15,* 289–310.

Booth, H., Maindonald, J., & Smith, L. (2002). Applying Lee–Carter under conditions of variable mortality decline. *Population Studies, 56*(3), 325–336.

Camarda, C. G. (2012). Mortalitysmooth: An R package for smoothing poisson counts with P-splines. *Journal of Statistical Software, 50*(1), 1–24.

Chatfield, C. (1997). Forecasting in the 1990s. *Journal of the Royal Statistical Society: Series D (The Statistician), 46*(4), 461–473.

Davis, R. A., Zang, P., & Zheng, T. (2016). Sparse vector autoregressive modeling. *Journal of Computational and Graphical Statistics, 25*(4), 1077–1096.

Du Preez, J., & Witt, S. F. (2003). Univariate versus multivariate time series forecasting: An application to international tourism demand. *International Journal of Forecasting, 19*(3), 435–451.

Gardner, E. S. (1985). Exponential smoothing: The state of the art. *Journal of Forecasting, 4*(1), 1–28.

Gardner, E. S., Jr., & McKenzie, E. (1985). Forecasting trends in time series. *Management Science, 31*(10), 1237–1246.

Giacometti, R., Bertocchi, M., Rachev, S. T., & Fabozzi, F. J. (2012). A comparison of the Lee–Carter model and AR-ARCH model for forecasting mortality rates. *Insurance: Mathematics and Economics, 50*(1), 85–93.

Girosi, F., & King, G. (2007). Understanding the Lee-Carter mortality forecasting method 1. Technical report, Rand Corporation, Santa Monica, CA.

Human Mortality Database. (2016). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany), http://www.mortality.org. Accessed 1 Feb 2016.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software, 27*(3), 1–22.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22*(4), 679–688.
Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2005). Prediction intervals for exponential smoothing using two new classes of state space models. *Journal of Forecasting, 24*(1), 17–37.
Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: The state space approach*. Berlin: Springer.
Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting, 18*(3), 439–454.
Hyndman, R., & Ullah, S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis, 51*(10), 4942–4956.
Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting US mortality. *Journal of the American statistical association, 87*(419), 659–671.
Makridakis, S., & Hibon, M. (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting, 16*(4), 451–476.
Ord, J. K., Koehler, A., & Snyder, R. D. (1997). Estimation and prediction for a class of dynamic nonlinear statistical models. *Journal of the American Statistical Association, 92*(440), 1621–1629.
Pegels, C. (1969). Exponential forecasting: Some new variations. *Management Science, 15*(5), 311–315.
Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis*. Berlin: Springer.
Renshaw, A. E., & Haberman, S. (2003). Lee–carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics, 33*(2), 255–272.
Renshaw, A. E., & Haberman, S. (2006). A cohort-based extension to the Lee–Carter model for mortality reduction factors. *Insurance: Mathematics and Economics, 38*(3), 556–570.
Taylor, J. W. (2003). Exponential smoothing with a damped multiplicative trend. *International Journal of Forecasting, 19*(4), 715–725.
Trefethen, L. N., & Bau, D. (1997). *Numerical linear algebra* (Vol. 50). Siam.
Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 65*(1), 95–114.