ORIGINAL ARTICLE

Forecasting Mortality Rates with the Penalized Exponential Smoothing State Space Model

Yanlin Shi

Department of Actuarial Studies and Business Analytics, Macquarie University, Sydney, NSW 2019, Australia

ARTICLE HISTORY
Compiled October 23, 2020

ABSTRACT
It is well known that accurate forecasts of mortality rates are essential to various demographic research like population projection, and the pricing of insurance products such as pensions and annuities. Recent studies suggest that mortality rates of multivariate ages are usually not leading indicators in mortality forecasting. Therefore, multivariate stochastic mortality models including the classic Lee–Carter may not necessarily lead to more accurate forecasts, compared with sophisticated univariate counterparties like the exponential smoothing state space (ETS) model. Despite its improved forecasting accuracy, the original ETS model cannot ensure the age-coherence of forecast mortality rates. By introducing an effective penalty scheme, we propose a penalized ETS model to significantly overcome this problem, with discussions on related technical issues including the reduction of parameter dimensionality and the selection of tuning parameter. Empirical results based on mortality rates of the Australian males and females suggest that the proposed model consistently outperforms the Lee–Carter and original ETS models. Robust conclusions are drawn when various forecasting scenarios are considered. Long-term forecasting analyses up to 2050 comparing the three models are further performed. To illustrate its usefulness in practice, an application to price fixed-term annuities with the penalized ETS model is demonstrated.

KEYWORDS
Mortality forecasting; Exponential smoothing; Penalty scheme; Lee-Carter model

CONTACT Yanlin Shi Tel.: +61 2 9850 4750 Email: yanlin.shi@mq.edu.au

1. Introduction

Improvements in life expectancy have spurred serious concerns about mortality and longevity risks around the world over the past decades (see, for example, Beaumont (1981),Kamerud (1989),Wajiga and Adekola (1998),N. Li and Lee (2005) and H. Li and Lu (2017)). As argued in Giacometti, Bertocchi, Rachev, and Fabozzi (2012), longevity risk means that people are surviving longer than expected or observed deaths are lower than expected. Advances in medical science, technological improvements and lifestyle changes tend to increase the longevity risk. For demographic research, those risks significantly affect the accuracy of population projections. In demographic and actuarial practice, they affect the pricing of life insurance, pension and annuity products. For instance, overestimating the mortality rate may lead to underestimated future population counts and thus a worse-than-expected risk profile for the annuity providers.

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) develop 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.

However, the accuracy of forecasting with multivariate models like the LC is debatable. 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 analysis. Chatfield (1997) also suggests that simple univariate methods are often more robust to model misspecification 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, the sample cross-correlation function exhibits meaningful and statistically significant correlations. Regarding the mortality rates, a recent study by Feng and Shi (2018) argues that the cross-correlations of residuals obtained from univariate models on age groups are not significant. Consequently, multivariate rates may not be leading indicators in mortality forecasting.

In the content of univariate forecasting, the exponential smoothing state space (ETS) model systematically studied in Hyndman, Koehler, Snyder, and Grose (2002) is an appropriate choice to forecast mortality rates. First, ETS is designed to model and forecast non-stationary time series, which is the case of mortality rates since they consistently decline over time. Second, the ETS framework decomposes a time series into level and growth components, and models them separately. For mortality rates, such a framework is reasonable and analogous to the age and temporal factors considered in the LC model. When applied to the mortality rates, Feng and Shi (2018) demonstrate that the ETS model can consistently outperform a range of multivariate models including the LC, Functional Data Model (FDM) developed by Hyndman and Ullah (2007), the original VAR models and the sparse VAR model investigated in Davis, Zang, and Zheng (2016), as well as the univariate AR-ARCH model proposed by Giacometti et al. (2012).

Despite the outstanding performance of the ETS model, there is an important practical issue left unresolved. As pointed out by research such as H. Li and Lu (2017), forecasts of mortality rates should be age-coherent. That is, the rates of neighboring age groups should not diverge. The concept of age coherence is extended from the seminal work of N. Li and Lee (2005), who argue that the long-term mortality improvements among various populations should be convergent. Due to its biological and practical reasonableness, the (age) coherence feature is widely recognized and incorporated in many novel mortality models. For instance, with respect to the factor-type models (including the LC), Shang and Haberman (2020) discuss a grouped functional mortality model, which focuses on the sub-national populations. As a popular alternative, the coherent VAR-type model is firstly investigated in H. Li and Lu (2017) with technical ad-hoc constraints on the mortality structure. Those limitations are relaxed in recent studies of Feng, Shi, and Chang (2020) and Shi (2020), and a dynamic extension is proposed in Chang and Shi (2020). A discussion on the coherence, however, is absent for the ETS model. As evidenced in our empirical results, long-term forecasts with the ETS model can exhibit significant divergence across even neighboring age groups, which is then not coherent. This may lead to significant adverse biological and/or practical implications, such that the mortality rate of age 50 may be higher than that of age 60, and younger people may have to pay for higher life insurance premiums.

To overcome this problem, a penalizing scheme should be considered in the mortality modelling (see, for example, Renshaw and Haberman (2003), Delwarde, Denuit, and Eilers (2007) and Pitt, Li, and Lim (2018)). In this paper, we follow the approach of H. Li and Lu (2017) and proposes a penalized ETS model. More specifically, at the estimation step of the original ETS model, a penalized least square loss function is considered. The penalty scheme is realized via the sum of squared differences of forecast mortality improvements between neighboring ages. As suggested by H. Li and Lu (2017), there are at least two advantages of employing this penalty structure. First, unlike the original ETS model, mortality improvements are forced to be age-coherent (smoothed across neighboring ages). Second, such a penalty term utilizes the idea of spatial dependence among mortality rates of different age groups. Further, to increase the estimation efficiency, we adopt the Fourier flexible functional form to reduce the dimensionality of free parameters. The related tuning parameter selection issue is addressed via the cross-validation procedure discussed in Hyndman and Athanasopoulos (2018).

In order to illustrate the effectiveness of our proposed model, empirical evidence based on the Australian male and female mortality data sourced from Human Mortality Database (2019) is provided. Using the smoothed rates ranging from 1950 to 2016, we systematically compare the forecasting performance of the LC, original ETS and penalized ETS models. As measured by the root of mean squared error (RMSE), the penalized ETS model consistently beats the LC and ETS model at the 10-step-ahead forecasting horizon. Using replicates generated by the LC model with Normality-disturbances assumption, the superiority of the penalized ETS model is further evidenced via simulations. Robust results are also produced under various scenarios, including those that no dimensionality reduction is considered, longer sample period is used, 30-step-ahead horizon is studied, crude rates are employed, the Swedish data are examined, and a multivariate penalized ETS model is investigated. We then conduct long-term forecasting analyses. Forecast mortality rates and life expectancies up to 2050 produced by all the three models are compared. To illustrate its usefulness in practice, an application to price fixed-term annuities with the penalized ETS model is demonstrated. This analysis follows a similar design adopted in Fung, Peters, and Shevchenko (2015) and Shang and Haberman (2017).

The contributions of this paper are fourfold. First, the proposed ETS model significantly complements the study of Feng and Shi (2018). ETS model enjoys a great popularity in the forecasting practice and leads to superior results in the M3-competition (Makridakis & Hibon, 2000). By ensuring the age-coherence of forecast mortality rates, the penalized ETS model retains all the advantages and resolve the most outstanding issue of the original ETS model. Second, an effective strategy to reduce the dimensionality of the penalized ETS model is developed. Taking Australian males as an example, our approach reduces the number of unknown parameters by over 90% and still leads to robust forecasts as the unrestricted model. Third, the forecasting performance of the penalized ETS model is systemically studied with strong empirical evidence. Compared to the LC (original ETS) model, the out-of-sample RMSE generated by the penalized ETS model is more than 50% (35%) smaller for mortality rates of Australian males. Robust superiority of the proposed model is also evidenced across various scenarios. Finally, via an application to annuity pricing, we demonstrate the usefulness of the proposed model in practice. Hence, the penalized ETS model can provide more accurate and meaningful mortality forecasts for demographers, actuaries and other users interested in survival analysis.

The rest of this paper is organized as follows. In Section 2, we review the specification and features of the ETS model. The penalized ETS model is described in Section 3. We conduct empirical studies with robustness checks, long-term forecast analyses and an application to annuity pricing in Section 4. Section 5 concludes the paper.

2. 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, Koehler, Ord, & Snyder, 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, Koehler, and Snyder (1997), Hyndman et al. (2002) and Hyndman, Koehler, Ord, and Snyder (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 specifications for each of the three components, there are thirty[1] distinct ETS methods (mainly by assuming an additive or multiplicative relationship for each component).[2] For details of those specifications, please refer to Section 2 of Hyndman and Khandakar (2008).

In the case of mortality rates, seasonality is not present. Additionally, only the additive trend and error specification are appropriate to model the logarithm structure. Therefore, only two out of those thirty methods are potentially applicable, and both are described below:

$$\begin{aligned} \ln m_{x,t} &= l_{x,t-1} + \phi_x b_{x,t-1} + \varepsilon_{x,t} \\ l_{x,t} &= l_{x,t-1} + \phi_x b_{x,t-1} + \alpha_x \varepsilon_{x,t} \\ b_{x,t} &= \phi_x b_{x,t-1} + \beta_x \varepsilon_{x,t} \end{aligned} \tag{1}$$

where $\ln m_{x,t}$ is the log mortality rates for age x at time t, $l_{x,t}$ and $b_{x,t}$ measure the level and growth of $\ln m_{x,t}$, respectively. α_x and β_x are their corresponding exponential smoothing parameters. Estimates of α_x and β_x are obtained by minimizing the sum of $\varepsilon_{x,t}^2$. However, due to its iterative structure, different from usual least-squares optimization, close-form solutions for the ETS model are not available.

Remark 1. With a sample size T, h-step-ahead forecast of $\ln m_{x,T}$ is $\ln \hat{m}_{x,T+h} = \hat{l}_{x,T} + \hat{b}_{x,t} \sum_{i=1}^{h} \hat{\phi}_x^i$, and $\hat{\phi}_x$ measures the dampened degree of $\hat{b}_{x,t}$. In general, when $\phi_x = 1$, Equation (1) reduces to the additive trend ETS model (also known as the Holt-Winters model). When $0 < \phi_x < 1$, it is a more general additive damped trend ETS model.

---

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

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

Remark 2. For mortality rates, however, our preliminary analysis shows that the $\hat{\phi}_x$ almost always reduces to 1 for all x, when a penalized scheme is imposed. Hence, we only consider the Holt-Winters model in this paper. This suggests that the growth of the log mortality rate is a constant in the long-run.

Remark 3. It is worth mentioning that when $\phi_x = 1$, the specification explained in Equation (1) is linked to the well-known ARIMA process. In particular, one can show that

$$\begin{aligned} \ln m_{x,t} &= l_{x,t-1} + b_{x,t-1} + \varepsilon_{x,t} \\ (1-L)l_{x,t} &= b_{x,t-1} + \alpha_x \varepsilon_{x,t} \\ (1-L)b_{x,t} &= \beta_x \varepsilon_{x,t} \end{aligned} \tag{2}$$

and thus rewrite $\ln m_{x,t}$ as

$$(1-L)^2 \ln m_{x,t} = (1 - \theta_{1,x}L - \theta_{2,x}L^2)\varepsilon_{x,t}$$

where $\theta_{1,x} = 2 - \alpha_x - \beta_x$ and $\theta_{2,x} = \alpha_x - 1$, if an infinite start-up is assumed (Hyndman et al., 2008). This may imply that $\ln m_{x,t}$ is non-stationary and a special I(2) process for each x, with the growth component described as a random walk without drift. Although different from the I(1) process usually assumed for mortality rates, it does not invalidate desirable statistical properties. To see this, the growth component is a random walk without drift and hence has 'stationary' (constant) mean. As for the mortality rates, the h-step-ahead forecast is $\ln \hat{m}_{x,T+h} = \hat{l}_{x,T} + h\hat{b}_{x,T}$, and the corresponding variance is $\hat{\sigma}^2_\varepsilon[1 + \sum_{i=1}^h(\hat{\alpha}_x + i\hat{\beta}_x)^2]$, where $\hat{\sigma}^2_\varepsilon$ is the variance of $\hat{\varepsilon}_{x,t}$ and $\sum_{i=1}^h(\hat{\alpha}_x + i\hat{\beta}_x)^2]$ captures the increasing uncertainty into the future with the growth of h. Hence, forecasts of $\ln m_{x,t}$ only depend on h (not t) and exponential smoothing parameters, with finite mean and variance for finite h. Also, when $0 < \alpha_x < 1$ and $0 < \beta_x < \alpha_x$, both level and growth components can be expressed as weighted averages of past values. More importantly, those conditions are sufficient for invertibility, stability and forecastability of $\ln m_{x,t}$ (Hyndman et al., 2008). In short, those properties ensure that observations in the distant past cannot have any effect on the forecasts.

As pointed out by Feng and Shi (2018), cross-correlation function cannot exhibit meaningful and statistically significant correlations for mortality rates among ages, after univariate models are fitted. On the other hand, the ETS model described above naturally fits the properties of the non-stationary $\ln m_{x,t}$. Compared with the famous Lee-Carter (LC) model, level and growth components of the ETS are directly analogous to the age and temporal factors of the LC model. When applied to the mortality data, Feng and Shi (2018) demonstrate that the univariate ETS model can outperform multivariate competitors for mortality forecasting. An important limitation for this approach, however, is that the forecasts of mortality rates for neighboring ages may diverge in the long run. 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, even if using smoothed mortality rates as inputs, there is no guarantee that the longitudinal forecasts will still have a smoothing structure. To overcome this potential issue, some penalization scheme needs to be imposed to the ETS model.

3. The penalized ETS model

When h is large (indicating long-run forecasts), $\ln \hat{m}_{x,T+h}$ will be dominated by $\hat{b}_{x,T+h} = h\hat{b}_{x,T}$, where T is the sample size. This is because that $\hat{l}_{x,T}$ is not changing with h and is therefore $o(h)$. Therefore, to ensure the age-coherence of forecast $\ln m_{x,T+h}$, it is sufficient to consider the impacts of $b_{x,T}$ among x only. This is consistent with N. Li and Lee (2005), in whose seminal paper, they argue that divergence of forecast mortality rates among ages is caused when the temporal effects are not converged, whereas the age effects are not relevant. Following the smoothing penalization scheme of H. Li and Lu (2017), a penalized ETS (PETS) model is constructed by minimizing[3]

$$\sum_{x=0}^{100}\sum_{t=1}^{T}\hat{\varepsilon}_{x,t}^2 + \lambda\sum_{x=0}^{99}(\hat{b}_{x+1,T} - \hat{b}_{x,T})^2 \tag{3}$$

where ages span from 0 to 100, and λ is the known non-negative tuning parameter. If $\lambda = 0$, this reduces to an unpenalized ETS model. The larger the λ is, the smoother the resulting forecasts will be. A formal discussion of the age-coherence property is presented below.

Definition 1. Age-coherence means that for the h-step-ahead forecasts, $|\ln \hat{m}_{i,T+h} - \ln \hat{m}_{j,T+h}| = O_p(1)$, for all $i, j \in \{1, 2, ..., N\}$. That is, when $h \to \infty$, $|\ln \hat{m}_{i,T+h} - \ln \hat{m}_{j,T+h}|$ will not diverge to infinity.

Theorem 1. *The PETS model estimated as by (3) has age-coherent out-of-sample forecasts when $T \to \infty$, given that h and T go to infinity at the same rate, $\lambda\sum_{x=0}^{99}(\hat{b}_{x+1,T} - \hat{b}_{x,T})^2$ is $O_p(T)$, and smoothing penalty λ goes large with T^3.*

Proof. Note that in (3), $\sum_{x=0}^{100}\sum_{t=1}^{T}\hat{\varepsilon}_{x,t}^2$ is $O_p(T)$. Thus, for the penalty terms to be effective, as in the given conditions, $\lambda\sum_{x=0}^{99}(\hat{b}_{x+1,T} - \hat{b}_{x,T})^2$ should also be $O_p(T)$.

[3]Different from H. Li and Lu (2017), we do not penalize α_x and β_x. One reason is that those parameters will be smoothed after applying the procedure described in Section 3.1. The other reason is that out-of-sample forecasts of $\ln m_{x,T+h}$ do not directly depend on them. In other words, smoothed α_x and β_x will not necessarily enforce the smoothness of $b_{x,T}$ across x.

Since λ goes large with T^3, it implies that $|\hat{b}_{x+1,T} - \hat{b}_{x,T}| = O_p(1/T)$. This can be inferred, for instance, from $\lambda \sum_{x=0}^{99} (\hat{b}_{x+1,T} - \hat{b}_{x,T})^2 = O_p(1/T^2)$.

Further, use Remark 3 and the assumption that h and T go to infinity at the same rate, it can be shown that

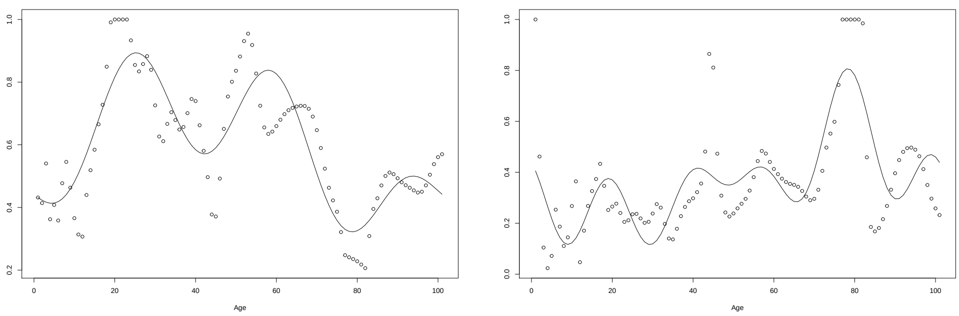
$$|\ln \hat{m}_{x+1,T+h} - \ln \hat{m}_{x,T+h}| = O_p(1) + h|\hat{b}_{x+1,T} - \hat{b}_{x,T}| = O_p(1).$$

This result can then be straightforwardly extended to any two ages within $0, 1, ..., 100$, since $\ln \hat{m}_{i+r,T+h} - \ln \hat{m}_{i,T+h} = \ln \hat{m}_{i+r,T+h} - \ln \hat{m}_{i+r-1,T+h} + \ln \hat{m}_{i+r-1,T+h} - \ln \hat{m}_{i+r-2,T+h} + ... + \ln \hat{m}_{i+1,T+h} - \ln \hat{m}_{i,T+h}$, for any $r \geq 1$, which completes the proof. $\square$

3.1. Reduction of dimensionality

Despite its theoretical soundness, the penalized ETS model described above is difficult to estimate. Since each age has two free parameters α_x and β_x, the total number of free parameters can be over two hundred. As no close-form solution is available, the estimation efficiency is further largely affected.

To overcome such an issue, we investigate the structure of α_x and β_x across ages to identify the possibility to reduce the dimensionality. We now focus on the smoothed log mortality rates of Australian male aged 0–100 over 1950–2006.[4] The fitted α_x and β_x ($\hat{\alpha}_x$ and $\hat{\beta}_x$) of the unpenalized ETS model are plotted in Figure 1 as scatter dots.



(a) α_x **(b)** β_x

Figure 1.: Estimated α_x and β_x for Australian male mortality data

Both $\hat{\alpha}_x$ and $\hat{\beta}_x$ demonstrate patterns across ages and similarities among neighboring groups. Thus, we adopt the Fourier flexible functional form and model

[4] Our full sample covers the range over 1950–2016. As will be discussed, we focus on a 10-year test sample (2007–2016) to evaluate the forecasting performance, and the rest (1950–2006) is employed as the training sample. Hence, the in-sample fitted parameters discussed in this section are based on the training sample over 1950–2006.

$\hat{\alpha}_x$ and $\hat{\beta}_x$ as follows

$$
\begin{aligned}
\hat{\alpha}_x &= \omega^{\alpha} + \sum_{i=1}^{n_\alpha}[\gamma_i^{\alpha} sin(\frac{2\pi i(x+1)}{101}) + \delta_i^{\alpha} cos(\frac{2\pi i(x+1)}{101})] \\
\hat{\beta}_x &= \omega^{\beta} + \sum_{i=1}^{n_\beta}[\gamma_i^{\beta} sin(\frac{2\pi i(x+1)}{101}) + \delta_i^{\beta} cos(\frac{2\pi i(x+1)}{101})]
\end{aligned}
\tag{4}
$$

where n_α and n_β are selected as the minimal integers that make the R^2 of the corresponding linear regression over 50%.[5] For the data examined here, n_α and n_β are 3 and 5, respectively. The fitted results are also demonstrated in Figure 1 as solid lines, which well represent the overall structures of α_x and β_x. Thus, instead of estimating α_x and β_x directly, given predetermined n_α and n_β, we can estimate ω^{α}, ω^{β}, γ_i^{α}, γ_i^{β}, δ_i^{α} and δ_i^{β} and use Equation (4) to obtain fitted α_x and β_x which then minimize Equation (3).[6] In this way, we have successfully reduced the number of free parameters by over 90% (from 202 to 18).

3.2. Selection of the tuning parameter

To select the tuning parameter λ, a natural solution is to employ the cross-validation (for example, see H. Li and Lu (2017)). However, due to the univariate time-series nature, related method such as leave-one-age-group-out is not applicable for the ETS framework. Hence, we employ the procedure discussed in Hyndman and Athanasopoulos (2018) to perform the cross-validation, which is also known as 'evaluation on a rolling forecasting origin.' The basic algorithm is explained below:

(1) Identify the first training set (e.g. $\ln m_{x,1}$,$\ln m_{x,2}$,...,$\ln m_{x,0.75T}$) out of the the entire sample;
(2) Given a value of λ, use the training set to fit the penalized ETS model and obtain the 1-step-ahead forecast $\ln \hat{m}_{x,0.75T+1}$;
(3) Extend the training set to include $\ln m_{x,0.75T+1}$ and refit the penalized ETS model to obtain the 1-step-ahead forecast $\ln \hat{m}_{x,0.75T+2}$;
(4) Repeat steps 2–3 until $\ln \hat{m}_{x,T}$ is generated; and

[5]The principle here is to balance the parsimony and accuracy in describing the structures of α_x and β_x. However, the optimal structures may change significantly when a penalty term is imposed. Thus, high-level precision in replicating the patterns of α_x and β_x from the unpenalized ETS model is not the focus. Therefore, we use 50% R^2 as a basic criterion to perform the selection. Robust results (available upon request) are produced when other choices including 30%, 60% and 90% are used. In one of the robustness check (can be found in the supplemental online material), we demonstrate that our selection has similar results to the model without dimensionality reduction. A systematic study on the optimal criteria for such a reduction remains for future research.

[6]The optimization can be performed with any usual numerical algorithms such as BFGS. In this paper, we adopted an effective and fast algorithm discussed in Ye (1987) to conduct the minimization, which is realized in the **Rsolnp** package of the statistical software **R**.

(5) Calculate the root of mean squared error (RMSE) as

$$\sqrt{\frac{1}{0.25T \times 101} \sum_{x=0}^{100} \sum_{h=1}^{0.25T} (\ln m_{x,0.75T+h} - \ln \hat{m}_{x,0.75T+h})^2}$$

λ is then chosen as that with the smallest RMSE. Even when no penalty is imposed, the neighboring $b_{x,T}$ may be already close to each other. Thus, it is expected that the value of λ can be quite large.

3.3. Overall fitting procedure

Combine the procedures of dimensionality reduction and tuning parameter selection, the overall fitting procedure of the penalized ETS model is explained below:

(1) Fit univariate ETS models for each age group individually to obtain $\hat{\alpha}_x$ and $\hat{\beta}_x$;
(2) Select n_α and n_β as described in Section 3.1;
(3) Given n_α and n_β, select the tuning parameter λ as described in Section 3.2; and
(4) Use the determined n_α, n_β and λ with Equation (4) to minimize Equation (3).

Forecasts of mortality rates can then be produced using the model as fitted above. Due to the penalized feature, the final obtained $\hat{\alpha}_x$ and $\hat{\beta}_x$ may not be asymptotically consistent. Thus, the non-parametric Bootstrap prediction intervals as described in Hyndman et al. (2008) can be employed to measure our forecast uncertainty.

4. Empirical application

In this paper, we use the Australian mortality data obtained from the Human Mortality Database (2019). Following Booth, Hyndman, Tickle, and De Jong (2006), we choose an opportune range of data starting from 1950 to 2016 in order to have a reliable and complete dataset. The smoothed male and female mortality rates[7] are studied separately, and the log rates are plotted in Figure 2 across all years. Consistent improvements over time are observed for both sexes.

To illustrate the powerfulness of our proposed model, we follow existing studies such as Feng and Shi (2018) and examine the 10-step-ahead forecasting[8], and the results of the LC, unpenalized ETS and penalized ETS models are compared. Therefore, our training sample ranges over 1950–2006, which is to be fitted by the three models individually, and the test sample of 2007–2016 is employed to calculate the out-of-sample RMSE as the criterion for comparison.

[7]The smoothed rates are produced using the weighted penalized regression splines with a monotonicity constraint. This is a standard method employed in the **demography** package of the statistical software **R**.

[8]In order to check the sensitivity of this chosen forecasting horizon, we also consider a long range of 30 steps. The results are robust and can be found in section 2 of the online supplementary materials.

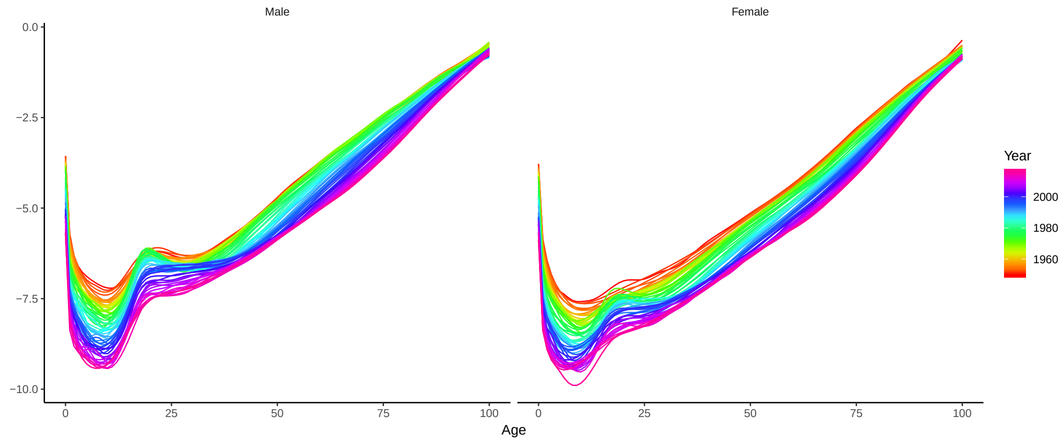
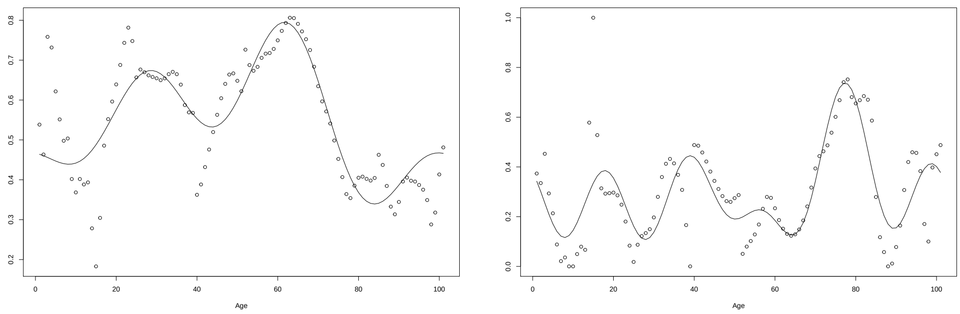


Figure 2.: Australian mortality data 1950–2016

4.1. Results of the tuning parameter selection

Before selecting the tuning parameter λ, we need to determine n_α and n_β as described in Section 3.1 to reduce dimensionality. As discussed previously, the results are 3 and 5, respectively, for Australian males. The estimated α_x and β_x from unpenalized ETS models for Australian females are plotted in Figure 3. According to R^2 of linear regressions, the selected n_α and n_β are also 3 and 5, respectively.[9] As demonstrated in Figure 3, the overall structures of both parameters are well captured, after the dimensionality reduction.



(a) α_x **(b)** β_x

Figure 3.: Estimated α_x and β_x for Australian female mortality data

According to Equation (3), the total loss function consists of sum of squared errors (SSE) and tuning parameter multiplying sum of squared differenced $b_{x,T}$. For the unpenalized ETS model, SSE for the Australian males is around 22.108, whereas $\sum_{x=0}^{99}(b_{x+1,T} - b_{x,T})^2$ is only 0.004, suggesting the similarities of neighboring $b_{x,T}$.

[9] Comparing Figures 1 and 3, there are some marginal differences in α_x and β_x among males and females. The main trends, however, are similar across sexes. This is consistent with our selected values of n_α and n_β.

However, such similarities are not adequate to ensure age-coherence, which is enforced by the penalty term described in Equation (3). Intuitively, if this term were as large as the SSE, λ needs to be at least 5000. Further, when the penalty is being considered in the optimization, SSE will be increased, while $\sum_{x=0}^{99}(b_{x+1,T} - b_{x,T})^2$ should reduce. Thus, it is expected that an optimal λ might be at the scale of 10^4 or more. We then perform a grid search of λ within the range of $(10^{-4}, 10^6)$. Using the procedure explained in Section 3.2 with 75% sample as the starting training set, the optimal λ chosen for male (female) is 111111.10 (101010.10), which is in-line with the assumption of Theorem 1 such that λ goes large with T^3 ($T = 57$ in this case).

4.2. Results of forecasting performance

To compare the forecasting performance across models, we follow H. Li and Lu (2017) and employ the RMSE. We consider both RMSEs over age groups and time horizons only and an overall measure as follows:

$$
\begin{aligned}
RMSE_x &= \sqrt{\frac{1}{10}\sum_{h=1}^{10}(\ln m_{x,T+h} - \ln \hat{m}_{x,T+h})^2} \\
RMSE_h &= \sqrt{\frac{1}{101}\sum_{x=0}^{100}(\ln m_{x,T+h} - \ln \hat{m}_{x,T+h})^2} \qquad (5) \\
RMSE_{all,h} &= \sqrt{\frac{1}{101 \times h}\sum_{i=1}^{h}\sum_{x=0}^{100}(\ln m_{x,T+i} - \ln \hat{m}_{x,T+i})^2}
\end{aligned}
$$

$RMSE_x$ ($RMSE_h$) is the RMSE averaged over all 10 forecasting steps (101 age groups) for age group x (time horizon h). $RMSE_{all,h}$ is the overall measure considering both dimensions up to step h. Relevant results for all three models are reported in Tables 1 and 2, as well as in Figures 4 and 5.

Figure 4 displays the $RMSE_x$. For Australian males, penalized ETS model produces smaller RMSE than LC for almost all age groups. Its performance is also better than that of the ETS model in most cases. Unlike LC (e.g. for ages 40–60) and ETS (e.g. for ages 0 and 20) models, penalized ETS does not produce any unusually large RMSE. As summarized in Table 1, the mean $RMSE_x$ across all age groups for penalized ETS model is only 40% of that for LC model. Compared to the ETS, adding the penalty scheme improves the forecasting performance by 30% on average. Q_1 and Q_3 measures further support that the penalized ETS model consistently outperforms LC and ETS counterparties. Standard deviation of $RMSE_x$ confirms that results of penalized ETS are much more narrowly spread than LC and ETS models. Those observations are largely robust with regard to those for the Australian female mortality rates, although the improvement of the penalized ETS model over the rest is smaller.

Table 1.: Summary of RMSE over age groups for the forecast Australian mortality data

Model	$RMSE_{all,10}$	Mean	*Std. Dev.*	Q_1	Q_3
Panel A: Male					
LC	0.1884	0.1625	0.0957	0.0794	0.2524
ETS	0.1217	0.1031	0.0649	0.0569	0.1243
PETS	**0.0788**	**0.0712**	**0.0339**	**0.0447**	**0.0973**
Panel B: Female					
LC	0.1383	0.1144	0.0781	**0.0369**	0.1846
ETS	0.1173	0.0952	0.0688	0.0448	0.1183
PETS	**0.1015**	**0.0824**	**0.0596**	0.0445	**0.0962**

Note: this table displays the RMSE over age groups for the 10-step-ahead forecasts of Australian male and female mortality rates. $RMSE_{all,10}$ is the overall RMSE across all ages and time horizons. Mean, *Std. Dev.*, Q_1 and Q_3 are the sample mean, standard deviation, first quartile and third quartile of the RMSEs over age groups, respectively. Bold numbers represent the smallest RMSEs among three models. LC, ETS and PETS stand for Lee-Carter, ETS and penalized ETS models, respectively.

Table 2.: RMSEs over time horizons of the forecast Australian mortality data

Steps	Male			Female		
	LC	ETS	PETS	LC	ETS	PETS
1	0.1368	0.0518	**0.0471**	0.0965	**0.0647**	0.0665
2	0.1546	0.0595	**0.0510**	0.0964	0.0662	**0.0601**
3	0.1632	0.0718	**0.0593**	0.1374	0.1038	**0.0940**
4	0.1981	0.1065	**0.0850**	0.1089	0.0790	**0.0610**
5	0.1642	0.0998	**0.0529**	0.1403	0.1097	**0.1047**
6	0.1653	0.1048	**0.0551**	0.1061	0.0817	**0.0801**
7	0.2139	0.1477	**0.0791**	0.1465	0.1343	**0.1101**
8	0.2178	0.1664	**0.0930**	0.1692	0.1580	**0.1360**
9	0.2241	0.1592	**0.1148**	0.1780	0.1693	**0.1462**
10	0.2205	0.1722	**0.1125**	0.1713	0.1468	**0.1140**

Note: this table displays the RMSE over time horizons of the forecast Australian male and female mortality rates for 10 steps. LC, ETS and PETS stand for Lee-Carter, ETS and penalized ETS models, respectively. Bold numbers represent the smallest RMSEs among three models.

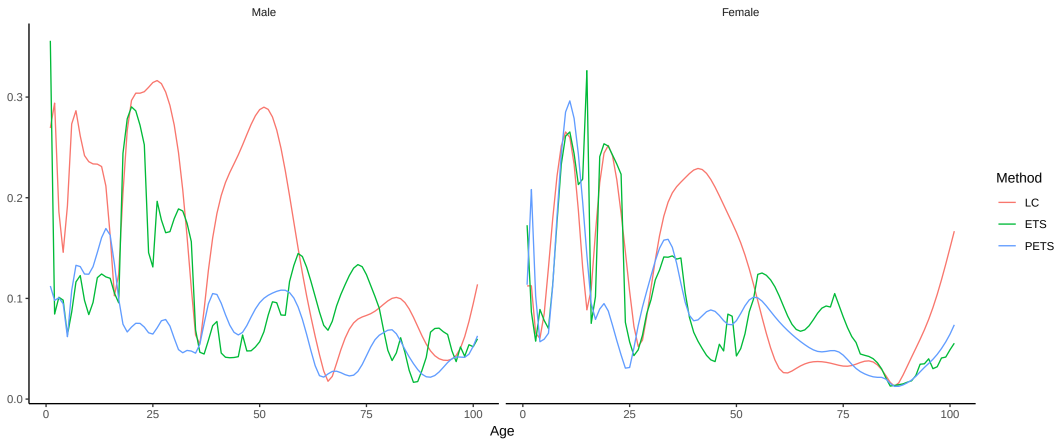


Figure 4.: RMSEs over age groups for Australian mortality data

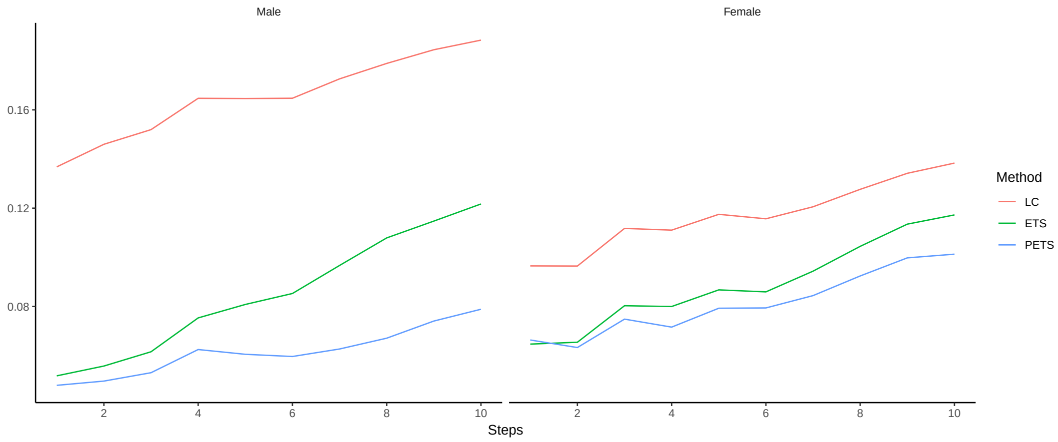


Figure 5.: RMSEs over forecasting steps for Australian mortality data

As indicated by $RMSE_{all,10}$, the overall performance of the penalized ETS model also beats the other two for both males and females.

Figure 5 plots the $RMSE_{all,h}$ for h ranging from 1 to 10. Distinct differences between penalized ETS and LC models can be observed for both males and females at all forecast horizons. Comparing with ETS model, $RMSE_{all,h}$ for the penalized ETS model is smaller in almost all scenarios. Further, with the growth of h, the increment in $RMSE_{all,h}$ is slower for the penalized ETS model, suggesting its better performance in the long run. Table 2 reports $RMSE_h$ at each of the 10 steps. Again, the penalized ETS model consistently beats the rest, except for the 1-step-ahead female forecast.

In order to understand this superiority, we compare the forecast mortality rates of all models at the tenth step. Figure 6 demonstrates the three forecast $\ln \hat{m}_{x,2016}$ sequences together with the actual data in 2016. Due to the lack of penalty scheme, ETS model displays incoherent patterns and divergences among ages for both sexes. Overall, the LC model (under-) over-forecasts the rates for age groups 20–30 (30–60). Relatively speaking, forecasts of the penalized ETS model are not diverged as those of ETS and better capture the variations among age groups than both the LC and ETS models. Thus, the results of the penalized ETS model are overall closer to the actual data.

Results of this section are robust when various scenarios are considered and a simulation study is performed. Details can be found in the supplemental online material.

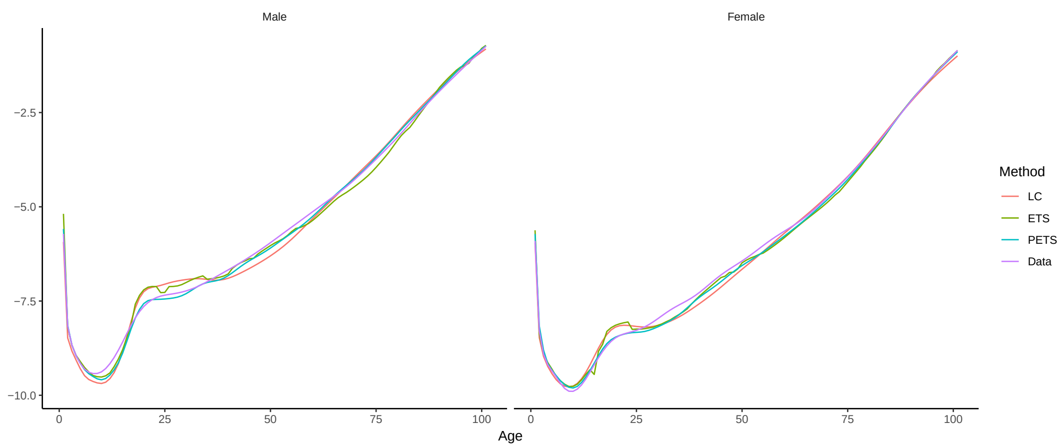


Figure 6.: Forecasts vs actual Australian mortality data in 2016

4.3. Long-term forecast analysis

We now examine the long-term forecasts of the three models. The complete datasets are employed (from 1950 to 2016), and the mortality rates are forecast to 2050 for both Australian males and females. For our penalized ETS model, n_α and

n_β are chosen as 3 and 5 (3 and 8), respectively, for males (females). The optimal tuning parameter is 80808.08 in both cases. Forecast mortality rates in 2050 and the life expectancies ranging 2001–2050 are plotted in Figures 7 and 8, respectively.

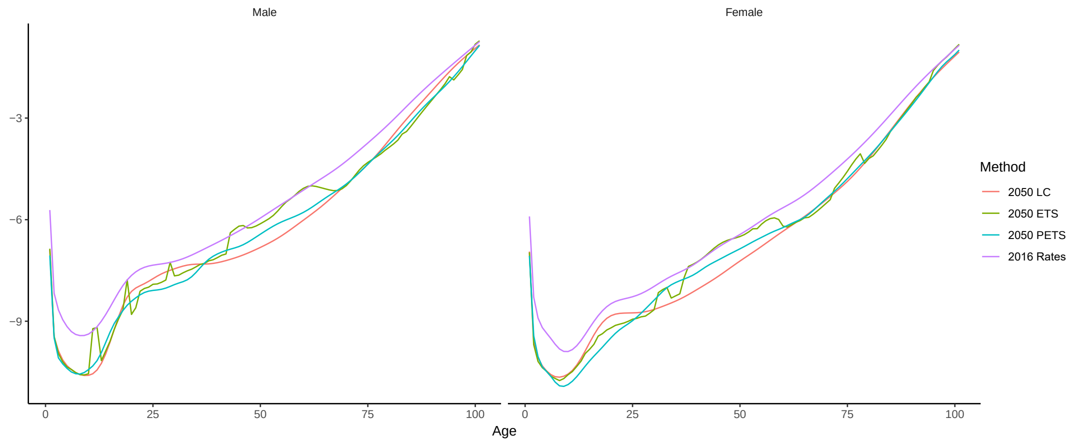


Figure 7.: Australian mortality rates in 2016 vs Forecast rates in 2050

Comparing with the actual mortality rates in 2016, forecast rates in 2050 suggest significant improvements in all cases. Due to the lack of penalty scheme, ETS model results in significant divergences among most age groups. In contrast, forecast rates of LC and penalized ETS models are much more smoothed across all ages. The differences between results of LC and penalized ETS models are largely consistent with our observations in Figure 6. Roughly speaking, LC tends to produce higher rates for ages 20–30 but lower rates for ages 30–60 than the penalized ETS model. Given the observed large $RMSE_x$ of the LC model at those ages in Figure 4, the corresponding forecasts of the penalized ETS model are potentially more reliable. The only difference between Figures 6 and 7 is that forecast 2050 rates of LC suggest smaller improvements for old male age groups (e.g. 80–95) than the penalized ETS model.

Among all information generated from mortality rates, life expectancies are widely studied in demographic research and investigated in actuarial practice. In Figure 8, we report the actual Australian life expectancies at birth from 2001 to 2016, together with mean/point forecasts (solid line) and 95% prediction intervals (dashed line) of forecasts up to 2050. The prediction intervals (PIs) of all models are calculated using the 2.5th and 97.5th percentiles of the Bootstrap samples with 5000 replicates. Despite the similarities of mean forecasts, the PIs of LC model are relatively wider than those of the other two models. This indicates the potential higher efficiency of our proposed model over the LC model for prediction uncertainty measure. For male life expectancies, results of LC and penalized ETS models are fairly close to each other, with a divergence starting at around 2025. Since then, the penalized ETS model indicates longer expectancies, with the final forecast as of 2050 being 86.36 years, and the result of LC is 85.75 years. Turning to the females, the differences between

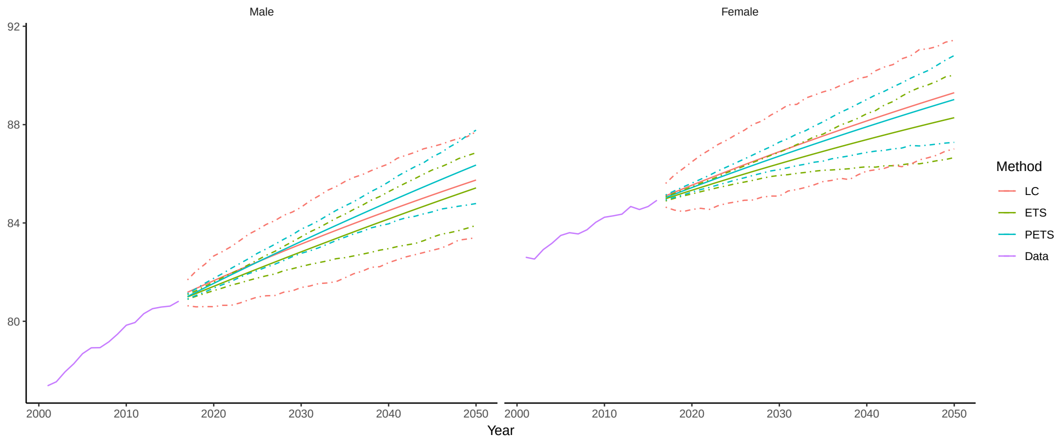


Figure 8.: Actual and forecast Australian life expectancies: 2001–2050

forecasts of LC and penalized ETS are even smaller. The final forecasts are 89.30 (LC) and 89.02 (penalized ETS) in 2050. ETS model produces the lowest life expectancies in all cases, although the widths of its PIs are as narrow as those of the penalized ETS model.

4.4. Applications to fixed-term annuity pricing

An important application of long-term mortality forecast is for elderly population (approximately older than 65 years of age). As discussed in Section 4.3, due to the lack of the coherence, the ETS model will lead to questionable long-term results for this population group. For instance, from Figure 7, the forecast Australian male rate in 2050 of age 70 is higher than both the true rate in 2016 and the forecast rate of age 75. In life insurance practice, for example, this unreasonably suggests that younger policyholders will pay for higher premiums than older policyholders. Those questionable issues of long-term forecasts are completely resolved when the PETS model is employed, indicating its much more improved reliableness over the ETS model for actuarial practices.

To demonstrate practical applications of the PETS model, we follow Fung et al. (2015) and Shang and Haberman (2017) to consider fixed-term annuities in this paper. Such products have enjoyed growing popularity globally. Comparing to the lifetime annuities, fixed-term annuities pay a predetermined and guaranteed income of higher level. Deferred option for those products is usually available. In terms of the pricing, we adopt a cohort approach as employed by Fung et al. (2015) and Shang and Haberman (2017), and the maximal survival age is limited to 100-year-old. First, the

τ year survival probability of a person aged x at t = 0 is

$$ {}_{\tau}p_x = \prod_{j=1}^{\tau} {}_1p_{x+j-1} \tag{6} $$

where $\tau_1 p_{x+j-1} = e^{-m_{x+j-1}}$ and m_{x+j-1} can be obtained from mortality forecasts. This equation essentially assumes that the central mortality rates are constant throughout the one-year period. Thus, the price of an annuity with maturity T year, written for an x-year-old with benefit $1 per year and conditional on the survival is determined as

$$ a_x^T(m_{x,1:T}) = \sum_{\tau=1}^{T} B(0,\tau)E(I(T_x > \tau)|m_{x,1:T}) \tag{7} $$

where $I(\cdot)$ is the indication function, T_x is the survival time, and $B(0,\tau)$ is the τ-year bond price at a interest rate equal to the yield of the annuity. Given that $E(I(T_x > \tau)|m_{x,1:T}) = {}_{\tau}p_x(m_{x,1:T})$, the fixed-term annuity price is a function of the underlying yield and mortality rates. Therefore, for the purposes of pricing and risk management, it is critical to producing accurate forecasts of $m_{x,1:T}$ which appropriately capture the mortality experiences of policyholders.

To illustrate its application in practice, we employ up to 35-step-ahead forecasts (from 2017 to 2051) of the PETS model and price the fixed-term annuities as of 2016 in Table 3. The calculation is performed for both Australian males and females up to 30-year-maturity, starting from age 65. Following Fung et al. (2015), we examine four age groups and derive both mean and interval estimates. The PIs are produced with 5000 Bootstrap replicates, as constructed in Section 4.3. Also, we assume a constant interest rate of 3% throughout all maturities. Consistent with our previous observations, annuity prices for female are higher than the corresponding male counterparties, indicating lower exposure to mortality. In terms of PIs, the widths (measured in percentage of deviation from mean) of males and females are fairly close, although those of males are slightly wider. This is as expected, since mortality forecast of female is normally less uncertain than that of male. It is worth mentioning that the widths of our PIs are narrower than those stated in Fung et al. (2015) and Shang and Haberman (2017). Despite the difference in data coverage, this preliminarily indicates the potential efficiency of our proposed PETS model on measuring mortality forecasting uncertainty in practice. As argued in Fung et al. (2015), underpricing as small as by 0.1% can lead to dramatic shortfall in reserving with a large portfolio. Hence, accurately and efficiently measure the uncertainty of premium rate can significantly help insurers optimize their reserves to minimize the ruin probability.

Table 3.: Predicted fix-term annuity prices for Australian males and females

Age	Measure	$T = 5$	$T = 10$	$T = 15$	$T = 20$	$T = 25$	$T = 30$
Panel A: Male							
65	Mean	4.5192	8.3838	11.6717	14.4404	16.7291	18.5695
	LB	4.5168 (-0.05%)	8.3767 (-0.08%)	11.6546 (-0.15%)	14.4166 (-0.16%)	16.6417 (-0.52%)	18.3474 (-1.20%)
	UB	4.5216 (0.05%)	8.3911 (0.09%)	11.6877 (0.14%)	14.4622 (0.15%)	16.7970 (0.41%)	18.7515 (0.98%)
70	Mean	4.4819	8.2914	11.4936	14.1343	16.2510	17.8099
	LB	4.4765 (-0.12%)	8.2759 (-0.19%)	11.4724 (-0.19%)	14.0539 (-0.57%)	16.0376 (-1.31%)	17.3889 (-2.36%)
	UB	4.4872 (0.12%)	8.3061 (0.18%)	11.5153 (0.19%)	14.2006 (0.47%)	16.4280 (1.09%)	18.1782 (2.07%)
75	Mean	4.4127	8.1151	11.1601	13.5929	15.3858	NA
	LB	4.4030 (-0.22%)	8.0964 (-0.23%)	11.0892 (-0.64%)	13.3985 (-1.43%)	14.9955 (-2.54%)	NA
	UB	4.4218 (0.21%)	8.1325 (0.21%)	11.2210 (0.55%)	13.7534 (1.18%)	15.7293 (2.23%)	NA
80	Mean	4.2788	7.7881	10.5823	12.6436	NA	NA
	LB	4.2647 (-0.33%)	7.7315 (-0.73%)	10.4161 (-1.57%)	12.2869 (-2.82%)	NA	NA
	UB	4.2907 (0.28%)	7.8437 (0.71%)	10.7277 (1.37%)	12.9683 (2.57%)	NA	NA
Panel B: Female							
65	Mean	4.5405	8.4318	11.7533	14.5636	16.8979	18.7570
	LB	4.5392 (-0.03%)	8.4285 (-0.04%)	11.7433 (-0.09%)	14.5316 (-0.22%)	16.7692 (-0.76%)	18.5246 (-1.24%)
	UB	4.5418 (0.03%)	8.4352 (0.04%)	11.7623 (0.08%)	14.5898 (0.18%)	16.9829 (0.50%)	18.9378 (0.96%)
70	Mean	4.5164	8.3691	11.6247	14.3235	16.4699	18.0704
	LB	4.5142 (-0.05%)	8.3610 (-0.10%)	11.5985 (-0.23%)	14.2182 (-0.73%)	16.2631 (-1.26%)	17.7896 (-1.55%)
	UB	4.5186 (0.05%)	8.3760 (0.08%)	11.6488 (0.21%)	14.3991 (0.53%)	16.6407 (1.04%)	18.3163 (1.36%)
75	Mean	4.4682	8.2390	11.3581	13.8355	15.6814	NA
	LB	4.4632 (-0.11%)	8.2210 (-0.22%)	11.2744 (-0.74%)	13.6608 (-1.26%)	15.4042 (-1.77%)	NA
	UB	4.4735 (0.12%)	8.2570 (0.22%)	11.4227 (0.57%)	13.9913 (1.13%)	15.9160 (1.50%)	NA
80	Mean	4.3664	7.9700	10.8285	12.9571	NA	NA
	LB	4.3555 (-0.25%)	7.9167 (-0.67%)	10.6919 (-1.26%)	12.7123 (-1.89%)	NA	NA
	UB	4.3775 (0.25%)	8.0197 (0.62%)	10.9550 (1.17%)	13.1704 (1.65%)	NA	NA

Note: this table displays the forecast fix-term annuity price for Australian males and females as of 2016. The forecast mortality rates range from 2017 to 2051. LB and UB stand for the 2.5th and 97.5th percentiles of the Bootstrap prediction interval, respectively. Mean is the point/mean forecast price. T is the maturity term. Value in bracket is the percentage difference compared to the forecast mean annuity price. We only consider contracts with maturity so that age + maturity ≤ 100.

5. Concluding remarks and discussions

This paper proposes a penalized exponential smoothing state space (PETS) model to forecast mortality rates. Three key conclusions can be drawn from our study. First, the new model extensively complements the original ETS model and ensures the age coherence in the long run. Second, PETS model outperforms the famous Lee-Carter and the ETS counterparties, in terms of mortality forecasting accuracy. Thirdly, the improved long-term forecasting reliability makes the PETS model much more desirable than ETS for actuarial practices, such as the fixed-term annuity pricing.

The proposed PETS model has many managerial implications in practice. For instance, the model may assist life insurers in the profitability analysis, which heavily rely on accurate forecasts. One example is that life insurers need to hold a significant amount of reserves, which is usually determined by a conservative risk metric, such as the 99% value-at-risk produced by the adopted mortality model. Since those reserves can only earn a minimum (risk-free) return, over-reserving (i.e. the value-at-risk is too large) will lead to lower investment returns and thus negatively affect the profitability of the life insurance products. As discussed in Section 4.3, the interval forecasts of PETS are more efficient than those of LC. Adopting the PETS model could therefore improve the profitability of life insurers.

Another application is related to the risk mitigation analysis. Because that longevity and mortality risks are natural offsets of each other, life insurers may consider offering life annuities in addition to life insurance products. Such a viability investigation will require portfolio-level scenario analyses, under which both life annuities and insurances are offered simultaneously. In this case, one can use the forecast rates of the PETS model to simulate a wide range of forecast rates with various certainties. A before-and-after comparison can then be conducted to analyze the effectiveness in mortality risk mitigation, if the life annuities are to be offered.

Overall, the merits of our research are briefly summarized as follows. First, the major research question, i.e. the concept of age coherence, is mathematically defined and is therefore not vague. Second, comprehensive analyses are conducted to support our arguments. Those include the RMSE of point forecasts at both the age and temporal dimensions, a long-term analysis with point and interval forecasts, a carefully designed simulation study and robustness checks of five different factors. Finally, we provide a practical application on the annuity pricing with the proposed model to demonstrate its real-life usefulness. The major limitation of this research is its empirical coverage. Due to the space constraints, the current study mainly focuses on the Australian mortality data with one robustness check of the Swedish data.

In addition to exploring more empirical data, there are three potential future directions to extend this study. First, existing literature has extensively discusses the temporal forecasts of mortality rates, but little has been explored for the age-dimension forecasts. For its univariate feature, the PETS model can be naturally employed to forecast mortality rates of very old ages with time periods fixed. A relevant example can be found in Giacometti et al. (2012), who use the AR-ARCH model to fit the Italian data of ages 40–91 and forecast those of ages 92–94 over 1960–2006. This is particularly important when the mortality rates of the very old ages are scare, volatile and thus unreliable to use directly. Second, the uncertainty measure may be improved if an appropriate modeling strategy is employed. One potential approach is to employ the two-step framework as in H. Li and Lu (2017), where the contemporaneous dependency of age-specific residuals is captured in the fitted variance-covariance matrix. Another pathway is to capture the temporal heteroskedasticity using ARCH or GARCH models as in Giacometti et al. (2012). Those parametric methods may produce more efficient prediction intervals, compared to those generated by the Bootstrap method as in this paper. Finally, some technical details influencing the estimation of the PETS model are worth further investigation. For instance, in Section 3.1, we adopt the Fourier flexible functional form to reduce the parametric dimensionality of the PETS model. Alternatively, many spline-based parametric models, such as the natural spline, p-spline and super smoother (Friedman & Silverman, 1989), may be employed. Also, a relevant criterion (to replace the 50% R^2 used in this paper) is worth examining to balance the trade-off of parsimony and accuracy and to help select an optimal dimensionality strategy.

Supplemental online material

Please see supplemental online material for the discussions and results of the simulation study and robustness check.

Acknowledgments

We are grateful to the Macquarie University for their support. The author would also like to thank Rob Hyndman, Jackie Li, Han Li, James Raymer and Chong It Tan for their helpful comments and suggestions. We particularly thank the Editor, Associate Editor and two anonymous referees for providing valuable and insightful comments on earlier drafts. The usual disclaimer applies.

References

Beaumont, C. D. (1981). The analysis of mortality and other acturial statistics. *Journal of the Operational Research Society*, *32*(2), 159–159.

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

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

Chang, L., & Shi, Y. (2020). Dynamic modelling and coherent forecasting of mortality rates: A time-varying coefficient spatial-temporal autoregressive approach. *Scandinavian Actuarial Journal*, 1–21.

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.

Delwarde, A., Denuit, M., & Eilers, P. (2007). Smoothing the lee–carter and poisson log-bilinear models for mortality forecasting: a penalized log-likelihood approach. *Statistical modelling*, *7*(1), 29–48.

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.

Feng, L., & Shi, Y. (2018). Forecasting mortality rates: Multivariate or univariate models? *Journal of Population Research*, *35*(3), 289–318.

Feng, L., Shi, Y., & Chang, L. (2020). Forecasting mortality with a hyperbolic spatial temporal VAR model. *International Journal of Forecasting*.

Friedman, J. H., & Silverman, B. W. (1989). Flexible parsimonious smoothing and additive modeling. *Technometrics*, *31*(1), 3–21.

Fung, M. C., Peters, G. W., & Shevchenko, P. V. (2015). A state-space estimation of the Lee-Carter mortality model and implications for annuity pricing. *arXiv preprint arXiv:1508.00322*.

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

Gardner Jr, E. S., & 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.

Human Mortality Database. (2019). *University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany).* Retrieved 2019-03-20, from `http://www.mortality.org`

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice.* OTexts.

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., 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. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: The state space approach.* Springer Science & Business Media.

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. J., & Ullah, S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics and Data Analysis*, *51*(10), 4942–4956.

Kamerud, D. B. (1989). Mortality risk and life expectancy. *Journal of the Operational Research Society*, *40*(2), 199–200.

Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting US mortality. *Journal of the American statistical association*, *87*(419), 659–671.

Li, H., & Lu, Y. (2017). Coherent forecasting of mortality rates: A sparse vector-autoregression approach. *ASTIN Bulletin: The Journal of the IAA*, *47*(2), 563–600.

Li, N., & Lee, R. (2005). Coherent mortality forecasts for a group of populations: an extension of the Lee-Carter method. *Demography*, *42*(3), 575-594.

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.

Pitt, D., Li, J., & Lim, T. K. (2018). Smoothing poisson common factor model for projecting mortality jointly for both sexes. *ASTIN Bulletin: The Journal of the IAA*, *48*(2), 509–541.

Renshaw, A. E., & Haberman, S. (2003). Lee–carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics*, *33*(2), 255–272.

Shang, H. L., & Haberman, S. (2017). Grouped multivariate and functional time series forecasting: An application to annuity pricing. *Insurance: Mathematics and Economics*, *75*, 166–179.

Shang, H. L., & Haberman, S. (2020). Forecasting multiple functional time series in a group structure: An application to mortality. *ASTIN Bulletin: The Journal of the IAA*.

Shi, Y. (2020). Forecasting mortality rates with the adaptive spatial temporal autoregressive model. *Journal of Forecasting*.

Taylor, J. W. (2003). Exponential smoothing with a damped multiplicative trend. *International Journal of Forecasting*, *19*(4), 715–725.

Wajiga, G., & Adekola, O. (1998). Life-expectancy in a nonhomogeneous population. *Journal of the Operational Research Society*, *49*(9), 1011–1012.

Ye, Y. (1987). *Interior algorithms for linear, quadratic, and linearly constrained non-linear programming* (Unpublished doctoral dissertation). Department of ESS, Stanford University.