Prediction of mortality rates using a model with stochastic parameters

Chon Sern Tan and Ah Hin Pooi

Citation: AIP Conference Proceedings **1782**, 050016 (2016); doi: 10.1063/1.4966106

View online: http://dx.doi.org/10.1063/1.4966106

View Table of Contents: http://scitation.aip.org/content/aip/proceeding/aipcp/1782?ver=pdfcov

Published by the AIP Publishing

Prediction of Mortality Rates using a Model with Stochastic Parameters

Chon Sern Tan[1, a)] and Ah Hin Pooi[2, b)]

[1]*Department of Mathematical and Actuarial Sciences, Universiti Tunku Abdul Rahman Jalan Sungai Long, Bandar Sungai Long, Cheras, 43000 Kajang, Selangor, Malaysia.*

[2]*Centre for Actuarial Studies, Applied Finance and Statistics No. 5, Jalan Universiti, Bandar Sunway, 47500 Petaling Jaya, Selangor, Malaysia.*

[a)]tcsern@utar.edu.my
[b)]ahhinp@sunway.edu.my

Abstract. Prediction of future mortality rates is crucial to insurance companies because they face longevity risks while providing retirement benefits to a population whose life expectancy is increasing. In the past literature, a time series model based on multivariate power-normal distribution has been applied on mortality data from the United States for the years 1933 till 2000 to forecast the future mortality rates for the years 2001 till 2010. In this paper, a more dynamic approach based on the multivariate time series will be proposed where the model uses stochastic parameters that vary with time. The resulting prediction intervals obtained using the model with stochastic parameters perform better because apart from having good ability in covering the observed future mortality rates, they also tend to have distinctly shorter interval lengths.

INTRODUCTION

As individuals are living healthier and longer in this modern era with improving medical advancement, mortality rates are gradually decreasing which leads many developing countries globally to experience aging populations. Although this may be a positive change for individuals, increasing lifespan may render personal savings insufficient as financial support during retirement years. This longevity improvement is especially a concern for insurance companies which provide longevity-related insurance products such as life annuities. As initial assumptions of higher mortality rates may not take into account the current prolonged lifespan, insurance product premiums may potentially be underestimated and affect the solvency of insurers in the long term. Moreover, the long-run future mortality levels contain an uncertainty that arises with time and this presents a systematic mortality risk that is not diversifiable because it affects all individuals in the same manner [2]. Therefore, an appropriate and efficient stochastic mortality model is of significance for insurers to describe this uncertainty.

In 1992, Lee and Carter established a time series model [13] that makes long-term predictions of age-specific mortality rates. Firstly, the logs of central mortality rates are modeled as an addition of an age-specific constant and the product of a time-dependent index for the level of mortality and another age-specific constant. Secondly, the historical data is fitted with this model from which the age-specific constants are derived through singular value decomposition while the time-dependent index is modeled as a stochastic time series which then is used to predict

The 4th International Conference on Quantitative Sciences and Its Applications (ICOQSIA 2016)
AIP Conf. Proc. 1782, 050016-1–050016-7; doi: 10.1063/1.4966106
Published by AIP Publishing. 978-0-7354-1444-0/$30.00

the future index. Finally, these forecasted index and age-specific constants are used to predict the future age-specific mortality rates.

Building on the Lee-Carter method, similar models have been introduced to forecast mortality. For instance, Renshaw and Haberman proposed one of the first stochastic models [7] for population mortality that incorporates a cohort effect parameter to describe the observed mortality variations among individuals from different cohorts. This model adds on to the Lee-Carter model with a random cohort effect which is a function of the approximate year of birth. In addition to that, the Cairns-Blake-Dowd (CBD) two-factor model [1] includes a cohort effect parameter and two time-dependent components instead of only one unique component originally. Taking into account the different purposes and underlying shapes of mortality structures, a general class of flexible multivariate time series model was introduced and a particular model is selected where the best criterion is investigated by deciding the parameters to keep or remove. Chan, Li and Li [17] claimed that this CBD model has the most appropriate time-dependent model parameters as indexes to describe the longevity risk levels at different points of time. Furthermore, it is crucial to be able to incorporate uncertainty of model parameters into simulations to evaluate the impact of their estimation errors [2]. Having investigated this issue in [1] and [11] using the above two-factor model, the inclusion of parameter uncertainty is found to have a significant effect on the levels of uncertainty in forecasted mortality rates and expected future lifetimes, particularly in the long run.

The more recent extensions of the Lee-Carter method are as follows. Brockett et al [12] used the Lee-Carter model on the difference of log mortality rates as well as a Levy process and the Normal Inverse Gaussian distribution on the index of mortality, resulting in better performance than the original model. To retain the historical parametric structure and the intragroup error correlation structure, the predictions based on the Lee-Carter model are incorporated with a bootstrap procedure for dependent data in the Multiple Lee-Carter Panel Sieve (MLCPS) algorithm in [16]. This algorithm is applied to estimate the relationship between populations that have similar socioeconomic conditions, where the results showed that this method works well in the presence of the dependence structures. In [4], the issues of model over-parameterization and unjustifiable adding of terms in the model were reduced by using a toolkit of functions and expert judgment to establish a general procedure in constructing mortality models. These models can distinguish sequentially each significant demographic feature in the data which will then be given a parametric structural form, a relatively parsimonious model was consequently obtained with a good fit to the mortality data of United Kingdom (UK). Besides that, Cairns et al [3] proposed a general framework that describes the dynamics of mortality rates of two related populations simultaneously by modeling the difference in the stochastic factors between the two populations using a mean-reverting autoregressive process. The modeling of the stochastic factors is improved in [14] using a vector error correction model, which helps to remove the need to make assumption on which population is dominant.

In [6], Pooi et al applied a time series model based on multivariate power-normal (MPN) distribution to obtain prediction intervals for future age-specific mortality rates of the United States (US). The US mortality data from the years 1933 till 2000 was fitted with a MPN distribution, which then is used to obtain the prediction intervals for future death rates of the years 2001 till 2010. Subsequently, this methodology was also applied in cases where the data used was augmented to include female death rates, death rates of Canada and UK [8] as well as applied on incomplete US mortality training dataset that contains missing values [9]. In general, the resulting prediction intervals in [6], [8] and [9] were found to have good ability in covering the observed future death rates. Although the prediction interval lengths may become longer for long-run prediction in [9], the prediction intervals in [8] performed well as they tend to have shorter interval lengths.

In this paper, we incorporate into the multivariate time series with stochastic parameters that vary with time. Besides having good ability in covering the observed future death rates, the resulting prediction intervals obtained using the model with stochastic parameters also tend to have interval lengths that are distinctly shorter. The layout of this paper is as follows. In the next section, the multivariate time series given in [5] is briefly introduced and some results on prediction of deaths rates given in [6] and [8] are highlighted. The subsequent section provides the prediction results obtained using the model with stochastic parameters. Lastly, the final section concludes the paper.

A TIME SERIES MODEL BASED ON MULTIVARIATE POWER-NORMAL DISTRIBUTION

The multivariate time series model proposed by Pooi [5] applies the following power-normal distribution, which was presented by Yeo and Johnson in [10]. The random variable $\tilde{\varepsilon}$ is said to have a power-normal distribution with parameters λ^+ and λ^- if

$$\tilde{\varepsilon} = \psi(\lambda^+, \lambda^-, z) = \begin{cases} \dfrac{(z+1)^{\lambda^+} - 1}{\lambda^+} & if\,(z \geq 0, \lambda^+ \neq 0) \\ \log(z+1) & if\,(z \geq 0, \lambda^+ = 0) \\ \dfrac{-\left[(-z+1)^{\lambda^-} - 1\right]}{\lambda^-} & if\,(z < 0, \lambda^- \neq 0) \\ -\log(-z+1) & if\,(z < 0, \lambda^- = 0) \end{cases} \tag{1}$$

where z is standard normally distributed.

From the univariate power-normal distribution, a multivariate power-normal (MPN) distribution may be constructed for a vector $\boldsymbol{y}$ which consists of k correlated random variables. This vector $\boldsymbol{y}$ is said to have a k-dimensional power-normal distribution [5] with parameters $\boldsymbol{\mu}, \boldsymbol{H}, \lambda_i^+, \lambda_i^-, \sigma_i, 1 \leq i \leq k$ if $\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{H\varepsilon}$, where $\boldsymbol{\mu} = E(\boldsymbol{y})$, $\boldsymbol{H}$ is an orthogonal matrix that consists of the eigenvectors of the variance-covariance matrix of $\boldsymbol{y}$, $\varepsilon_1, \ldots, \varepsilon_k$ are uncorrelated,

$$\varepsilon_i = \sigma_i \left[\tilde{\varepsilon}_i - E(\tilde{\varepsilon}_i)\right] / \left\{\mathrm{var}(\tilde{\varepsilon}_i)\right\}^{1/2}, \tag{2}$$

$\sigma_i > 0$ is a constant, and $\tilde{\varepsilon}_i$ has a power-normal distribution with parameters λ_i^+ and λ_i^-.

Using the MPN distribution, a multivariate time series model may be obtained for a vector $\boldsymbol{x}(t)$ consisting of n_c observations recorded at time t. Letting Δt be a small increment of time after t, an $n_c(l+1)$-dimensional power-normal distribution is fitted on the vector $\boldsymbol{x}^{(1)} = \left[\boldsymbol{x}(t-(l-1)\Delta t), \ldots, \boldsymbol{x}(t-\Delta t), \boldsymbol{x}(t), \boldsymbol{x}(t+\Delta t)\right]$. The above $n_c(l+1)$-dimensional power-normal distribution may be used to obtain an n_c-dimensional conditional distribution of $\boldsymbol{x}(t+\Delta t)$ which will then specify a lag (l-1) multivariate time series model for the vector of n_c time-dependent correlated observations.

With the assumption that the multivariate time series is stationary, we may consider that the vector $\boldsymbol{x}^{(d)} = \left[\boldsymbol{x}(t+(d-l)\Delta t), \ldots, \boldsymbol{x}(t+(d-2)\Delta t), \boldsymbol{x}(t+(d-1)\Delta t), \boldsymbol{x}(t+d\Delta t)\right]$ for $d \geq 2$ has the same distribution as the vector $\boldsymbol{x}^{(1)}$. Thus, given the value of $\left[\boldsymbol{x}(t+(d'-l)\Delta t), \ldots, \boldsymbol{x}(t+(d'-2)\Delta t), \boldsymbol{x}(t+(d'-1)\Delta t)\right]$, a conditional distribution may be found for $\boldsymbol{x}(t+d'\Delta t)$ and next a value for $\boldsymbol{x}(t+d'\Delta t)$ may be generated for $d' = 2, 3, \ldots, d$. In this approach, a value of $\boldsymbol{x}(t+d\Delta t)$ may be generated when $d' = d$. This procedure of generating $\boldsymbol{x}(t+d\Delta t)$ may be repeated a large number of times and from these generated values of $\boldsymbol{x}(t+d\Delta t)$, a marginal distribution may be obtained for the j-th $(1 \leq j \leq n_c)$ component of $\boldsymbol{x}(t+d\Delta t)$. The prediction intervals whose end points are given by the $100(\alpha/2)$ and $100(1-\alpha/2)$ percentage points of the marginal distribution, may be used to predict the value of the j-th component of $\boldsymbol{x}(t+d\Delta t)$.

Constructing a lag-0 model for a vector of three age-specific total death rates in [6] and [8], the nominally 95% prediction intervals for the total death rates of the age group 60-64 were obtained, which are summarized in TABLE 1. In this table, L_d and U_d are respectively the lower and upper limits of the intervals for the death rate d years

after the year 2000. The column under the heading "US (Total)" contains the results based on the historical data for the US total (female and male combined) death rates. Similarly, "US (Female-Total)" indicates the historical data for the US female and total death rates, "Can-US" indicates the historical data for the Canadian female and male death rates together with the US female and total death rates, while "UK-Can-US" indicates the historical data for the UK as well as Canadian female and male death rates together with the US female and total death rates.

TABLE 1. The nominally 95% prediction intervals for future total death rates of the age group 60-64

d	L_d US (Total)	U_d US (Total)	Interval Length				Actual Observed Values
			US (Total)	US (Female-Total)	Can-US	UK-Can-US	
1	0.01120	0.01270	0.00150	0.00145	0.00150	0.00137	0.01205
2	0.01060	0.01300	0.00240	0.00146	0.00145	0.00124	0.01188
3	0.01020	0.01290	0.00270	0.00221	0.00211	0.00184	0.01177
4	0.00990	0.01300	0.00310	0.00250	0.00256	0.00230	0.01132
5	0.00954	0.01310	0.00356	0.00291	0.00289	0.00254	0.01126
6	0.00922	0.01310	0.00388	0.00350	0.00319	0.00293	0.01081
7	0.00894	0.01310	0.00416	0.00376	0.00315	0.00305	0.01063
8	0.00870	0.01320	0.00450	0.00460	0.00375	0.00315	0.01051
9	0.00850	0.01310	0.00460	0.00496	0.00366	0.00357	0.01026
10	0.00821	0.01330	0.00509	0.00580	0.00390	0.00380	0.01006

As a benchmark, the resulting prediction interval lengths from this paper will be compared with those from TABLE 1, where the interval lengths have been shown to be comparable to those obtained by Jarner et al [15] using the Lee-Carter method. Improvement in the model's predictive ability can be identified if the resulting prediction intervals have shorter interval lengths but yet can still cover the observed future values.

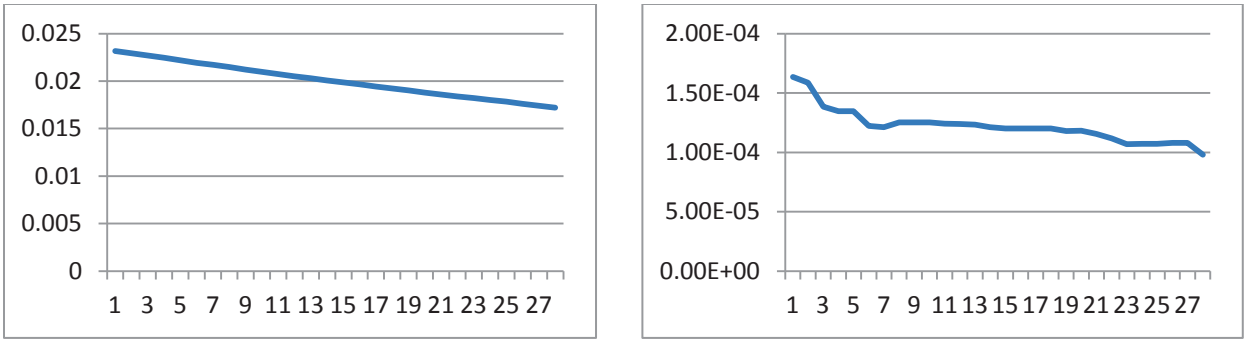
PREDICTION OF US MORTALITY RATES USING A MODEL WITH STOCHASTIC PARAMETERS

In this section, we will incorporate time-varying stochastic parameters into the methodology by Pooi et al [6]. In [6], [8] and [9], the training data consisted of mortality rates from the years 1933 till 2000 (68 years) while the testing data consisted of mortality rates from the years 2001 till 2010 (10 years). Intuitively, among the data in the full training dataset, the earlier death rates may be less informative than the recent death rates in reflecting the "future" mortality range in the testing dataset. Thus, from the available mortality data for 78 years, we may form a window using a number of consecutive years' data as training data and use the immediate future year's data as the testing data. By varying the set of consecutive years of data, we can form many such windows.

As an example, suppose the subset of age groups of interest is formed by the age groups 55-59, 60-64 and 65-69 while the year-t mortality rates are denoted by the vector $\boldsymbol{m}_t$. Using the data from 1933 till 1973, we form the first window which has 40 rows where each row contains the mortality rates $(\boldsymbol{m}_t, \boldsymbol{m}_{t+1})$ for the above three age groups in two consecutive years. Similarly, using the data from 1934 till 1974, we form the second window. By varying the set of consecutive years of data, we obtain 28 windows where the last window is formed using the data from 1960 till 2000.

A window contains the recent information on the mortality rates and would be suitable for predicting the mortality rate in the *next* year. To predict the mortality rates in the next d ($d \geq 2$) years, we may make use of the information contained in the current and several previous windows as explained below.

The data in the w-th window is fitted with a six-dimensional power-normal distribution. Let the parameters of the fitted distribution be denoted by $\mu_i^{(w)}, \sigma_i^{(w)}, H^{(w)}, \lambda_i^{+(w)}$ and $\lambda_i^{-(w)}$ where $1 \le i \le 6$ and $w = 1, 2, \ldots, 28$. For each $1 \le i \le 6$, we plot the values of μ_i and σ_i against w. In Fig. 1, we show two of these plots.



(a) (b)

FIGURE 1. Trend of mean and standard deviation of the total death rates for the age group 60-64

Generally, it is observed that both μ_i and σ_i experience a downward trend over the 28 windows. This suggests that these parameters are varying with time. Therefore, instead of assuming that the μ_i and σ_i calculated from the data in the 28-th window will remain the same for the years 2001 till 2010, we will model the trend of μ_i and σ_i appropriately and make forecasts of the future parameters.

Firstly, due to the deterministic trend of $\mu_i^{(w)}$ above, the means of the mortality rates from these 28 windows are modeled as a quadratic equation $\mu_i^{(w)} = a_i w^2 + b_i w + c_i$ where w represents the window number. The mean $\mu_i^{(w+d)}$ for the next d-th ($1 \le d \le 10$) window is then extrapolated from this quadratic equation.

Secondly, due to the stochastic nature of $\sigma_i^{(w)}$ when w varies above, we model the vector of standard deviations as a lag-0 multivariate time series via the conditional distribution of $\left(\sigma_1^{(w+1)}\ \sigma_2^{(w+1)} \ldots \sigma_6^{(w+1)}\right)$ which is derived from the fitted 12-dimensional power-normal distribution for $\left(\sigma_1^{(w)} \ldots \sigma_6^{(w)} \sigma_1^{(w+1)} \ldots \sigma_6^{(w+1)}\right)$. Using this conditional distribution, the value for $\left(\sigma_1^{(w+1)}\ \sigma_2^{(w+1)} \ldots \sigma_6^{(w+1)}\right)$ is generated iteratively from the value of $\left(\sigma_1^{(w)} \ldots \sigma_6^{(w)}\right)$ where $w = 28, 29, \ldots, 37$.

As the available values for the other parameters given by $H^{(w)}, \lambda_i^{+(w)}$ and $\lambda_i^{-(w)}$ may not be enough to model these remaining parameters as a multivariate time series, we assume that their values remain the same for $w = 28, 29, \ldots, 37$.

Now, for $d = 1$, by using the values of $\mu_i^{(w+d)}, \sigma_i^{(w+d)}, H^{(28)}, \lambda_i^{+(28)}, \lambda_i^{-(28)}, 1 \le i \le 6$, we generate a set of values for $\boldsymbol{m}_{28+d}$. This procedure is repeated for $d = 2, 3, \ldots, 10$ to produce iteratively a set of values for $\boldsymbol{m}_{28+d}$. In this way, we can generate a set of values for $\boldsymbol{m}_{38}$ which represents the mortality rates in 2010. Using the parameters calculated from this set of values, we next generate a large number of $\boldsymbol{m}_{38}$ and use these generated values to form a nominally 95% prediction interval for the mortality rates in 2010 for each of the three age groups. TABLE 2 displays the nominally 95% prediction intervals based on this constructed lag-0 model for the total death rates of the age group 60-64.

TABLE 2. The nominally 95% prediction intervals using a model with stochastic parameters for future total death rates of the age group 60-64

d	L_d - US (Stochastic)	U_d - US (Stochastic)	Observed Future Death Rate, O_d	Prediction Interval Length				
				US (Stochastic)	US (Total)	US (Female-Total)	Can-US	UK-Can-US
1	0.01137	0.01236	0.01205	0.00099	0.00150	0.00145	0.00150	0.00137
2	0.01081	0.01217	0.01188	0.00136	0.00240	0.00146	0.00145	0.00124
3	0.01041	0.01218	0.01177	0.00177	0.00270	0.00221	0.00211	0.00184
4	0.01010	0.01203	0.01132	0.00193	0.00310	0.00250	0.00256	0.00230
5	0.00981	0.01195	0.01126	0.00214	0.00356	0.00291	0.00289	0.00254
6	0.00967	0.01193	0.01081	0.00226	0.00388	0.00350	0.00319	0.00293
7	0.00935	0.01186	0.01063	0.00250	0.00416	0.00376	0.00315	0.00305
8	0.00913	0.01175	0.01051	0.00262	0.00450	0.00460	0.00375	0.00315
9	0.00901	0.01172	0.01026	0.00270	0.00460	0.00496	0.00366	0.00357
10	0.00872	0.01162	0.01006	0.00290	0.00509	0.00580	0.00390	0.00380

From TABLE 2, it is found that all the prediction intervals based on the above model with stochastic parameters have covered the observed future total death rates. Furthermore, all the lengths of prediction intervals are also distinctly shorter than those in the columns "US (Total)", "US (Female-Total)", "Can-US" and "UK-Can-US" (from TABLE 1) for $d = 1,\ldots,10$. This shows that the multivariate time series based on the model with stochastic parameters has significantly improved the predictive ability of the model through shorter prediction intervals lengths while still covering the observed future total death rates.

CONCLUSION

This paper presents a promising application of a multivariate time series model on the United States mortality data where future mortality rates are forecasted based on a model with time-varying stochastic parameters. The results are encouraging as indicated by the prediction intervals' ability in covering the future observed mortality rates with interval lengths that are significantly shorter. As a further work, we may explore applying the above methodology on mortality data that is augmented with data from other countries with similar demographics such as Canada and United Kingdom.

ACKNOWLEDGMENTS

We are grateful to AIP Publishing for the opportunity to publish and share our research work with its research community.

REFERENCES

1. A. Cairns, D. Blake and K. Dowd, J Risk & Insurance **73**, 687-718 (2006).
2. A. Cairns, D. Blake and K. Dowd, Scandinavian Actuarial Journal **2**, 79-113 (2008).

3. A. Cairns, D. Blake, K. Dowd, G. Coughlan and M. Khalaf-Allah, Astin Bulletin **41**, 29-59 (2011).
4. A. Hunt and D. Blake, North American Actuarial Journal **18**, 116-138 (2014).
5. A. Pooi, Applied Mathematical Sciences **6**, 5735-5748 (2012).
6. A. Pooi, Applied Mathematical Sciences **8**, 5039-5051 (2014).
7. A. Renshaw and S. Haberman, Insurance: Mathematics And Economics **38**, 556-570 (2006).
8. C. Tan and A. Pooi, Jurnal Teknologi **78**, 4-4 (2016).
9. C. Tan and A. Pooi, "Prediction of mortality rates in the presence of missing values" in *Proceedings of the 2nd Innovation and Analytics Conference & Exhibition (2015)* (IACE, 2015), Vol. 1691, pp. 050027.
10. I. Yeo, Biometrika **87**, 954-959 (2000).
11. K. Dowd, D. Blake and A. Cairns, Demography **47**, 67-78 (2010).
12. P. Brockett, S. Chuang, Y. Deng and R. MacMinn, Journal Risk And Insurance **80**, 799-826 (2013).
13. R. Lee and L. Carter, Journal Of The American Statistical Association **87**, 659-671 (1992).
14. R. Zhou, Y. Wang, K. Kaufhold, J. Li and K. Tan, North American Actuarial Journal **18**, 150-167 (2014).
15. S. Jarner and C. Dengsøe, The Danish Labour Market Supplementary Pension Fund *(ATP),* (2009).
16. V. D'Amato, S. Haberman, G. Piscopo and M. Russolillo, Insurance: Mathematics And Economics **51**, 694-701 (2012).
17. W. Chan, J. Li and J. Li, North American Actuarial Journal **18**, 38-58 (2014).