

Sheng-Feng Luo, Huei-Wen Teng* and Yu-Hsuan Lee

Forecasting Mortality using Imputed Data: The Case of Taiwan

DOI 10.1515/apjri-2015-0011

Abstract: Mortality forecasting plays an essential role in designing welfare policies and pricing aged-related financial derivatives. However, most prevailing models do not perform well in mortality forecasting particularly for the elder people. Indeed, the problem of missing category for the elderly is a typical feature in developing countries, because people are shorter-lived in earlier times and hence the mortality is recorded up to fewer age categories. For example, in Taiwan, the mortality is recorded up to an age of 95 before 1997, but as the improvement of life expectancy, the mortality is recorded up to an age of 100 afterwards. This paper proposes several approaches for data imputation to alleviate this systematic missing data problem of the mortality data. Motivated by Lee, and Carter. 1992. "Modelling and Forecasting the Time Series of US Mortality." *Journal of the American Statistical Association* 87:659–71 and Renshaw, and Haberman. 2006. "A Cohort-Based Extension to the Lee-Carter Model for Mortality Reduction Factors." *Insurance: Mathematics and Economics* 38:556–70, we employ factor models, in which age, period, and cohort are employed as useful effects. Simulation study and an empirical study using mortality data of Taiwan demonstrate the improvement in forecasting using a suitable data augmentation technique.

Keywords: mortality, missing data, students, longevity, forecasting, data imputation

1 Introduction

Mortality forecasting or demography projection is a challenging issue in actuarial science. One clear reason is that the population trend can be altered by unforeseen events, such as the spread of some epidemic, the outbreak of war, or even the coming of new vaccines. Indeed, improving medical treatment expands

*Corresponding author: Huei-Wen Teng, Graduate Institute of Statistics, National Central University, Taoyuan City, Taiwan, E-mail: venteng@gmail.com

Sheng-Feng Luo, Chung Yuan Christian University, Taoyuan City, Taiwan

Yu-Hsuan Lee, Global Life Insurance, Co., LTD, Taipei City, Taiwan

the age spectrum of human beings, which also results in not only the endogenous structure change in mortality rate but the incomplete data set of death numbers. This constitutes one aspect of difficulty in mortality modelling.

Since the classical models proposed by Gompertz (1825) and Makeham (1860), extensive studies on mortality models have been considered. Among them, a factor model presented in Lee and Carter (1992) attributes the mortality to only two factors: age and period (time). Modifications on this two-factor (AP) model include Lee and Miller (2001), Booth et al. (2002), Hyndman and Ullah (2007), and others. This two-factor consideration has been widely used in practice as well; e.g., Wilmoth (1996) for Japan, and Tuljapurkar et al. (2000) for the G7.

On the other hand, while investigating the data deeper, Kupper et al. (1985) found that, in addition to the previous two factors, there is another significant one in determining the mortality rate. It is the so-called cohort effect, or simply, the year of birth. Renshaw and Haberman (2006) then directly extended the Lee and Carter (1992) (LC) model to incorporate such cohort factor, which becomes the Age-Period-Cohort (APC) model.

Other related studies include, for example, Coale and Gao (1989) and Coale and Kisker (1990) focusing on elderly mortality forecast method, Denuit and Goderniaux (2005) with a polynomial extrapolation technique, and Cairns et al. (2006) employing the stochastic mortality rate model. As for small area mortality forecasting, Li and Lee (2005), Cairns et al. (2009), Cairns et al. (2011), Dowd et al. (2010), Jarner and Kryger (2011), and Zhou et al. (2013) borrowed and combined mortality information with other population used in Brass Logit model, Bayesian mothod, and graduation methods.

In contrast, this paper contributes to the literature on mortality forecasting by proposing useful imputation techniques in order to include the incomplete mortality data for model estimation. The incompleteness of death data accompanied by the increase in the life expectancies of human beings has been rarely addressed in literature. Indeed, the phenomenon of longevity alters the way of data collecting, and the age categories are extended to a longer life span nowadays than before. As a consequence, the missing data problem arises in combining the early and the recent data sets to build up a mortality model. This problem is crucial particularly for developing countries, because the recorded mortality data is relatively few.

For example, according to the statistics from Taiwan Ministry of the Interior, the death data is only available after 1992. More importantly, the number of deaths is only recorded up to the age at 95 and up (95+) before 1998 and 100+ afterwards. Therefore, when trying to establish a mortality model for ages 0–99, one merely has a 15-year complete data set available (1998–2012) at the time we start this research, which is relatively short. This motivates us to explore the

information of the data, by using both the early incomplete 6-year data (1992–1997) and the complete 15-year data,¹ and investigate whether the forecasting performance on the mortality can be improved.

Note that incorporating the earlier 6-year data with the later 15-year data results in 6-year missing values for age groups from 95 to 99, due to nature change in demographics recording conventions. For this reason, this paper aims at improving the forecasting ability for the mortality models, using Taiwan data, via a proper method to impute unavailable death data of the elder people. As remark, although the Human Mortality Database provides mortality data of a longer study period, its data appears to be extrapolated and may not reflect the realistic pattern of human mortality.

Actually, the way of modeling in Renshaw and Haberman (2006) allows for data missing by considering over-dispersion in the Poisson error structure. Yet, there is no formal assessment about their such claim. Both our simulation and empirical studies show that their approach even cannot beat a model calibrated by the late, complete 16-year data, in terms of one prediction error. Thereby, we instead consider three imputation methods to deal with the missing data. Our finding is that the imputation can generally help to improve forecasting, especially by treating missing values as parameters to be estimated.

The rest of this paper is organized as follows. Section 2 reviews the AP and APC models and addresses the estimation approach. Section 3 presents three imputation methods. Section 4 tests these methods via simulations, and Section 5 conducts an empirical analysis using Taiwan's mortality data. The final section concludes.

2 Models and Estimation

Let x denote an age, and t denote the time in year. Let $D_{x,t}$ be a random variable to denote the number of deaths in a population at age x and time t . The actual data set takes the form of pair $(d_{x,t}, e_{x,t})$, for integer age $x \in \mathcal{A} = \{0, 1, \dots, k\}$ and calendar year $t \in \mathcal{T} = \{t_1, t_2, \dots, t_n\}$ with $t_{i+1} = t_i + 1$. In our notations, k is a dummy variable indicating the eldest age recorded for mortality, and t_1 indicates the first year of the study period. Here $d_{x,t}$ stands for “the number of deaths during the year t aged x last birthday” and $e_{x,t}$ is “the average population during year t aged x last birthday.”

¹ Nevertheless, we only use 1992–2010 data to build up the forecasting model in our analyses for the last two-year data will be used for out-of-sample testing.

Following the general setting in actuarial modelling, we consider the Poisson regression model with the response variable equal to the death number. Therefore, the model is

$$D_{x,t} = \text{Poisson}(e_{x,t}\mu_{x,t}),$$

where $\mu_{x,t}$ is the force of mortality. In addition, we allow for over-dispersion with scale parameter ϕ , and the first two moments are

$$E(D_{x,t}) = e_{x,t}\mu_{x,t}, \quad V(D_{x,t}) = \phi E(D_{x,t}),$$

where $E(\cdot)$ is the expectation operator and $V(\cdot)$ is the variance operator.

Under the log-link, we begin with the age-period (AP) model, which models the force of mortality $\mu_{x,t}$ by

$$\log(\mu_{x,t}) = \alpha_x + \beta_x \kappa_t, \quad [1]$$

where α_x captures the main age effect while $\beta_x \kappa_t$ incorporates the age specific period trends. This setting is equivalent to the original LC model.

Notice that the parameters in [1] are not identifiable: the structure is invariant with respect to the transformation:

$$\beta_x \leftarrow c\beta_x \quad \text{and} \quad \kappa_t \leftarrow \frac{1}{c}\kappa_t, \quad \forall c \in \mathbb{R}/\{0\}$$

In other words, α_x and κ_t are only determined up to a linear transformation, and β_x is determined up to a multiplicative constant. Thereby, we impose further conditions on these parameters, which are $\sum_x \beta_x = 1$ and $\sum_t \kappa_t = 0$.

To incorporate the cohort effect, the second Age-Period-Cohort (APC) model considers

$$\log(\mu_{x,t}) = \alpha_x + \beta_x^1 \kappa_t + \beta_x^0 t_{t-x}, \quad [2]$$

in which the additional bilinear terms $\beta_x^0 t_{t-x}$ represent the cohort effects. Similarly, to make parameters in [2] identifiable, we impose further conditions on the parameters: $\sum_x \beta_x^0 = \sum_x \beta_x^1 = 1$, $\sum_t \kappa_t = 0$, and $\kappa_{t_1} = t_{t_1 - k} = 0$. It is clear that the AP model in [1] is a special case of the APC model in [2], when $\beta_x^0 = 0$ for all x .

To estimate the above models, the goal is to find the estimators so that the value of the total deviance is the least. Take the AP model for illustration. Let $\hat{d}_{x,t}$ denote the fitted number of deaths given estimators $\hat{\alpha}_x$, $\hat{\beta}_x$, and $\hat{\kappa}_t$:

$$\hat{d}_{x,t} = e_{x,t} \hat{\mu}_{x,t} = e_{x,t} \exp(\hat{\alpha}_x + \hat{\beta}_x \hat{\kappa}_t). \quad [3]$$

The deviance here stands for a proxy of the true log likelihood function (in negative value) under Poisson distribution, which is given by

$$\begin{aligned}
 Deviance(d_{x,t}, \hat{d}_{x,t}) &= \sum_{x,t} dev(x, t) = \sum_{x,t} 2\omega_{x,t} \int_{\hat{d}_{x,t}}^{d_{x,t}} \frac{d_{x,t} - u}{V(u)} du \\
 &= \sum_{x,t} 2\omega_{x,t} \left[d_{x,t} \log \left(\frac{d_{x,t}}{\hat{d}_{x,t}} \right) - \left(d_{x,t} - \hat{d}_{x,t} \right) \right], \tag{4}
 \end{aligned}$$

where $\omega_{x,t}$ is zero for empty cell and one otherwise. The minimization can be implemented by a standard iterative Newton-typed method similar to Renshaw and Haberman (2006).

Note that an R package “ilc” can be used to estimate the above mentioned model (Butt and Haberman, 2009). Also, since the force of mortality $\mu_{x,t}$ is unobservable, we will use the central mortality $m_{x,t}$ as a proxy for model calibration. The central mortality is defined by

$$m_{x,t} = \frac{d_{x,t}}{e_{x,t}}.$$

3 Imputation Methods

This section begins with the description on the data structure used for model estimation, and proposes some possible methods to impute the missing data.

As we mentioned in the beginning, the convention of death-number recording will change as humans live longer. Figure 1 presents mortality data structure of Taiwan. Note that before 1998, the death record of a person who passed away at age over 95 was included in the category “95.” In other words, the data cells at age 95 for years 1992–1997 do not imply the actual mortality rate at exactly age 95. Thus, the cells in that period with age 95–99 can be viewed as missing when we look back in year 2013. Also, we here have only a small sample size of 21-year data, which is quite common for developing countries.

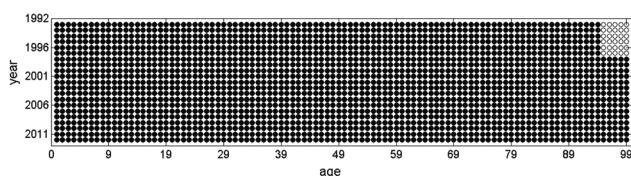


Figure 1: Taiwan mortality data under cross-classification with age (from 0 to 99) and year (from 1992 to 2012). Solid circles denote available data cells, while empty circles denote missing cells.

For the societies nowadays, the fact of longevity is critical in designing social insurance systems and welfare policies. As a result, developing a suitable statistical inference particularly for the elder population becomes a major concern. As a response to this fashion, we focus on the modelling with the spectrum up to age 99.² In the following, we present and investigate several methods related to the issue of missing data to see which one can perform better in mortality forecasting.

3.1 Data Structure

3.1.1 The Truncated Data

A simple and coarse way to forecast mortality for age more than 95 is to just wipe out the data subsample from year 1992 to 1997, which only covers mortality up to age 95. Therefore, we call the remaining data the truncated data. Although this approach is simple, it results in a fewer sample size and the lost of a respectable, 6-year data information for a majority of age groups from 0 to 94. This will serve as the benchmark of comparison.

3.1.2 The Full Data

The full data can be used for statistical inference with slight modifications by the weighted method adopted in Renshaw and Haberman (2006). Note in the full data, we include mortality up to age groups from 0 to 94 in 1992–1997, and include mortality from 0 to 99 from 1998 to 2012. In our data analysis later, one can see that in the standpoint of forecasting, using the full data does not work well in general, which is even worse than solely using the truncated data (especially true for the elder population).

3.1.3 The augmented data

In statistics, a common technique to improve the quality of statistical inference with missing values is to impute data; i.e, filling the empties by proper proxies. Thereby, for the rest of this work, we will try to seek proper data augmentation approaches that can help to improve forecasting.

² So far, the mortality data of Taiwan is still aggregated over age groups above 100. This is why we choose [0, 99] as the age spectrum here.

3.2 Imputing Data

In the following, we propose three possible and simple ways to impute the original full data. They are, what we call, the simple method, the two-step method, and the Minimum Deviance (MD) method. Note that for ease of reading, we only take the AP model to illustrate the idea.

3.2.1 The simple method

The simple method imputes the missing data merely using the average of the final age category 95+. To be more specific, for age at $x = 95, 96, 97, 98, 99$ in calendar year 1992, 1993, 1994, 1995, 1996, 1997, the augmented number of deaths is given by:

$$\check{d}_{x,t} = \frac{1}{5} d_{95+,t}, \quad [5]$$

where $d_{95+,t}$ denotes the original data recording the total number of deaths at age equaling or above 95. Note that here we implicitly assume $d_{100+,t} = 0$ for $t = 1992, \dots, 1997$. This method is simple to implement, yet it can not capture the variation with the specific age.

3.2.2 The Two-step Method

The two-step method adopts the idea of so-called plug-in principle (Efron and Tibshirani, 1993). The procedure is listed below:

1. Use the estimators obtained from the weighted method to infer the parameters belonging to the empty data cells. In other words, we have to make inferences about

$$\mu_{x,t} = \exp(\alpha_x + \beta_x \kappa_t) \quad \text{for } x = 95, \dots, 99 \text{ and } t = 1992, \dots, 1997.$$

For those ages, we can directly plug in the estimators $\hat{\alpha}_x$ and $\hat{\beta}_x$, while for those periods, we instead need to induct backwards. Specifically, we establish a AR(1) model³ for those estimators $\hat{\kappa}_t$ with $t \geq 1998$ and then extrapolate $\check{\kappa}_t$ for $t = 1992, \dots, 1997$.

³ We found that, for Taiwanas AR(1) fit the period trend κ_t or the cohort effect ι_{t-x} better than ARIMA(1, 1, 0), which is suggested by Renshaw and Haberman (2006)

2. Fill in the empty cells by drawing a sample $\check{d}_{x,t}$ from the matching Poisson distribution with mean parameter

$$\dot{d}_{x,t} = e_{x,t} \exp\left(\hat{\alpha}_x + \hat{\beta}_x \bar{k}_t\right),$$

where the population sizes $e_{x,t}$ are directly available from the data set.

3. Then re-estimate the whole parameters as if there were no missing values; i.e., all the weights $\omega_{x,t}$ are now equal to 1.

Note that we have considered directly using $\dot{d}_{x,t}$ to augment the incomplete data, rather than further simulating, yet similar results were obtained for Taiwan's data. However, the advantage of current practice is that we can easily manipulate $\check{d}_{x,t}$ such that

$$\sum_{x=95}^{99} \check{d}_{x,t} \leq \check{d}_{95+,t} \text{ for all } t = 1992, \dots, 1997. \quad [6]$$

3.2.3 The Minimum Deviance Method

The minimum deviance (MD) method regards the missing number of deaths as parameters to be estimated. As mentioned earlier, the deviance is a substitute proxy for the true likelihood, and hence the MD method here exploits the most information of the data. An iterative Newton method will be applied to find out the optimal estimators $\dot{d}_{x,t}$, with $x = 95, \dots, 99$ and $t = 1992, \dots, 1997$, that help to minimize the deviance.

Note that all the methods involved in this section can be carried out by using the same R package “ilc”, except the MD method here. Thereby, we will specially code the algorithm of the MD method by MATLAB. Besides, thanks to the experience from the 2-step method, we will not further consider the constraint like here, for the purpose of fast convergence.

4 Simulation Studies

In the following, we investigate the model performance in terms of in-sample fit and out-of-sample prediction when using the truncated data, the full data, and the augmented data. For augmented data, the simple method, the two-step method, and the MD method, are conducted.

To assess the performance of in-sample fit and out-of-sample forecasting, we follow Yang et al. (2010) to adopt the mean absolute percentage error (MAPE) as the criteria, which is defined as

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \bar{y}_i}{y_i} \right| \times 100\%$$

Note that in this definition, y_i denotes the logarithm of the observed central mortality, \bar{y}_i can be the *fitted* counterpart \hat{y}_i (referring to in-sample fitting performance) or the *predicted* counterpart \tilde{y}_i (referring to out-of-sample forecasting performance), and n is the number of observations of interest.

Recall the MSE is defined as

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_i)^2 \times 100\%.$$

One major reason to use the MAPE but not MSE in this paper is because the mortality rate varies drastically at different age. Figure 2 shows that the central mortality rate ranges from 10^{-1} to 10^{-8} . For this reason, the MSE will be influenced solely by the larger mortality rate. Similarly, it is rather complicated to interpret a likelihood-based measure, because the mortality rate or the death number varies drastically across age groups. In contrast, the MAPE use the percentage error and can better reflect the assessment of in-sample fitting and out-of-sample forecasting.

As for the criteria AIC and BIC, they are suitable measures for model selection, but are not used for assessing prediction performance. Indeed, our concern is to assess whether using a data set of longer study period by adding a proper imputed data set can help prediction performance, no matter if the model we use is optimal.

Other than the “overall” MAPE, which is aggregated over the whole age groups, we also report four other MAPES aggregated over different age subsets. They are “Child”, “Adult”, “Elder”, and “Over 90”, which are defined as follows:

1. Child: People at age under 15.
2. Adult: People at age equal to and larger than 15 but under 65.
3. Elder: People at age equal to and larger than 65.
4. Over 90: People at age equal to and larger than 90.

Notice that the first three are consistent with what the Department of Statistics, Ministry of the Interior of Taiwan defines. But here we additionally compute the over-90 MAPE from the elder population to see if the imputation for people aged

from 95 to 99 can really improve the model performance for the extreme elderly. In general, the smaller the computed MAPE is, the better the model performs.

We first test our conjecture if data imputation improves forecasting via a simulation study. The data set is simulated from the models [1] and [2], respectively, which mimics the Taiwan's mortality data structure as specified in Figure 1. More specifically, we first create a data set at age ranging from 0 to 99 in year from 1992 to 2012, and then wipe out the data cells labelled with $x = 95, \dots, 99$ and $t = 1992, \dots, 1997$.

To provide reasonable simulation studies, instead of using arbitrary parameters, the parameters here are estimates of a randomly selected data downloaded from the Human Mortality Database, namely, the U.K.'s female mortality. Moreover, we divide the whole data set into two parts according to the time-line. Specifically, we use the 1992~2010 data to estimate the model, and use the remaining 2011~2012 data to test the forecasting performance.

Table 1 reports relevant results, where the MAPE calculation for in-sample fitting is only aggregated over years from 1998 to 2010 to make the comparison among different data sets more reliable and fair.

First of all, augmented data with three methods perform comparably to the truncated data in terms of in-sample fitting. However, augmented data with a suitable imputation method outperforms the truncated data in terms of out-of-sample forecasting. In fact, forecast using augmented data with the simple method produces the smallest overall prediction MAPEs under the AP model, and forecast using augmented data with the MD method produces the smallest prediction MAPEs under the APC models. All in all, using augmented data outperforms using the truncated data and full data from the standpoint of out-of-sample predication.

It is worthy noting that under both the AP and APC models, using the full data generally dose not perform well, no matter in terms of in-sample-fitting or out-of-sample forecasting. Using full data even overall did worse than using simply the truncated data.

5 Empirical Analysis

We now turn to see whether the success of imputation in the simulation study can be reproduced in the real case. Before that, we will first conduct a exploratory analysis of Taiwan's mortality data, which can help us to judge the empirical results more precisely. In addition, the female and male data will be examined separately.

Table 1: Simulation results.

MAPEs	In-sample fitting (1998–2010)				Out-of-sample forecasting (2011–2012)				
	Child	Adult	Elder	Overall	Child	Adult	Elder	Over 90	Overall
The age-period model									
Truncated data	0.980	0.567	1.551	3.986	0.974	1.455	0.735	2.309	5.395
Full data	1.004	0.578	4.033	12.616	1.851	1.356	0.667	3.915	11.825
Augmented data									1.907
The simple method	1.002	0.579	2.236	6.336	1.222	1.363	0.673	1.479	3.196
The two-step method	1.001	0.580	2.093	5.843	1.173	1.367	0.676	2.417	6.420
The MD method	1.001	0.580	1.561	3.979	0.986	1.366	0.675	2.080	5.259
The age-period-cohort model									
Truncated data	1.682	0.496	1.709	4.469	1.098	5.269	0.885	2.525	5.481
Full data	2.529	0.523	1.712	4.453	1.240	11.540	0.821	2.572	6.196
Augmented data									3.042
The simple method	2.529	0.523	1.712	4.453	1.240	11.465	0.821	2.572	6.196
The two-step method	2.529	0.523	1.712	4.453	0.807	11.543	0.821	2.572	6.196
The MD method	0.992	0.521	1.860	4.992	1.060	1.701	0.793	2.503	5.453
									1.528

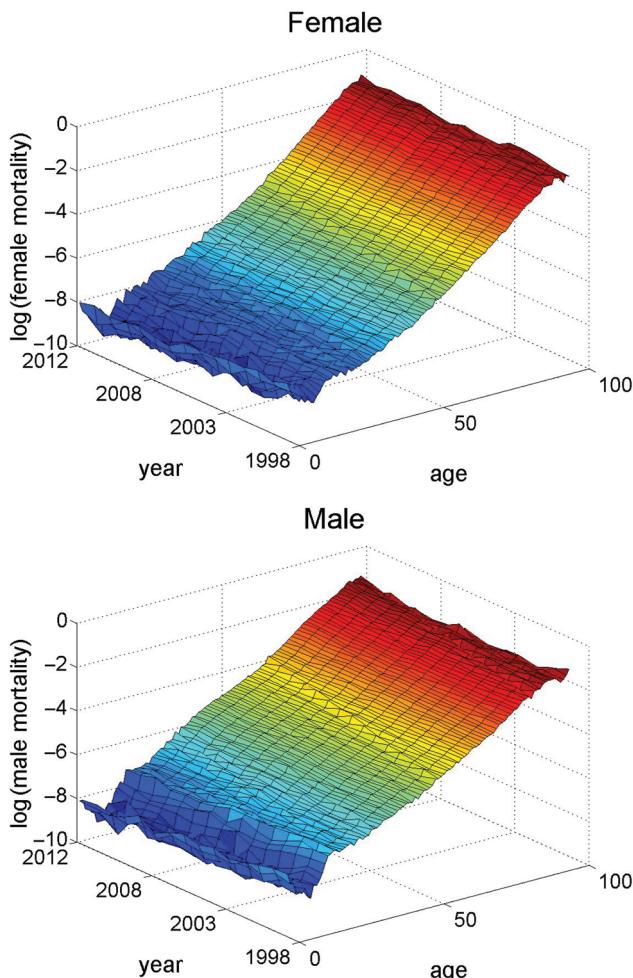


Figure 2: Logarithm of female and male central mortality rate of Taiwan from year 1998 to 2012 at age 0–99.

5.1 Data Exploration

The data of pairs $(e_{x,t}, d_{x,t})$ can be directly downloaded from the website of Taiwan Ministry of the Interior. Figure 2 shows the surface of the logarithm of $m_{x,t}$, for integer age $x \in \mathcal{A} = \{0, \dots, 99\}$ and calendar year $t \in \mathcal{T} = \{1992, \dots, 2012\}$, for both female and male data. One can clearly see that the two surfaces are very

smoothing, which suggests there are no serious, irregular exogenous factors (like natural catastrophe) that drive the mortality to change dramatically during this sample period. It also showed the unstable mortality patterns for the ultra-young and old group. Generally, especially for the middle ages, the mortality is monotonically increasing in age while decreasing in time. In other words, an adult is inclined to die as he/she gets older and such tendency decreases with calendar year because human medical technology improves along with time.

The phenomenon of improved mortality as time passes can also be confirmed by checking the next two ratios. One is the annual improvement rate defined by

$$\frac{m_{x,t} - m_{x,t-1}}{m_{x,t-1}},$$

and the other is the reduction factor defined by

$$\frac{m_{x,t}}{m_{x,t_1}},$$

where t_1 indicates the first year of the study period. In this empirical study, $t_1 = 1992$. The former compares the change of mortality rate for the two consecutive years within the same age group, while the latter demonstrates the whole time series of mortality rate, leveled by the first period, at each age group. The results are depicted in Figures 3 and 4, respectively, where both the area with blue colors denote the improvement in mortality.

Notice that there are significant oblique lines (from top left to bottom right) clearly shown in Figure 3, particularly for the middle age and the elderly. This observation can be seen as an echo of the cohort effect, which also justifies the consideration of APC model. On the other hand, according to Figure 4, although mortality decreases as medical standards improves, it is mostly reflected in the young population. After the age 97, the contour even becomes irregular which give us another hint that the prevailing model consideration is hardly amenable to the mortality data of the oldest-old (i.e., with age 95 +). Yang et al. (2010) utilized the observations from the reduction factor to construct a new extended model of [1], other than the bilinear cohort effect as in [2].

We finally directly examine the population evolution of each cohort. Figures 5 and 6 demonstrate the exposures of different cohorts over our sample period (i.e., 1992–2012) for both female and male, respectively. There are two observations. First, due to immigration and emigration, the population size is not definitely decreasing, particularly for the young period of each cohort. These facts partially increase the degree of difficulty in studying mortality data. Second, the population pattern of each cohort becomes stable and even straight

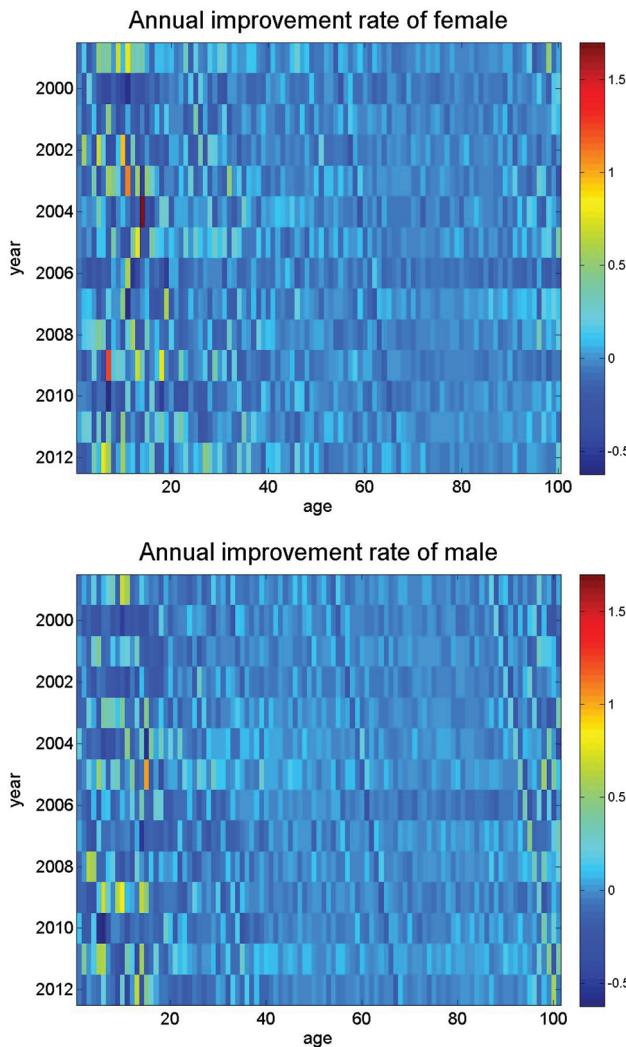


Figure 3: Improvement rates in mortality for Taiwan.

when that cohort gets aged. This somewhat validates our adoption of AR model to predict κ_t and ι_{t-x} .

All in all, from the evidence so far, the raw mortality data of Taiwan can really be comprehended from the three aspects: age, period, and cohort. However, there still seems room to total understanding especially for the children and the oldest-old.

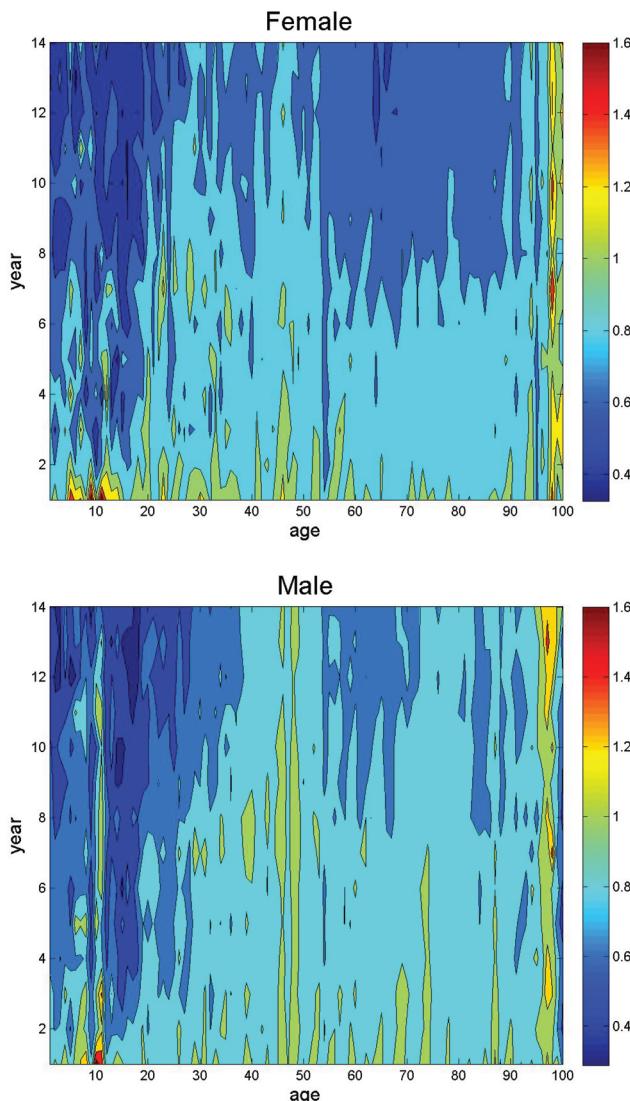


Figure 4: Reduction factors of the mortality for Taiwan.

5.2 Results Comparison

We apply the imputed methods to data in Taiwan. Similar to the simulation studies in Section 4, we divide the whole sample into two parts. The first 19-year

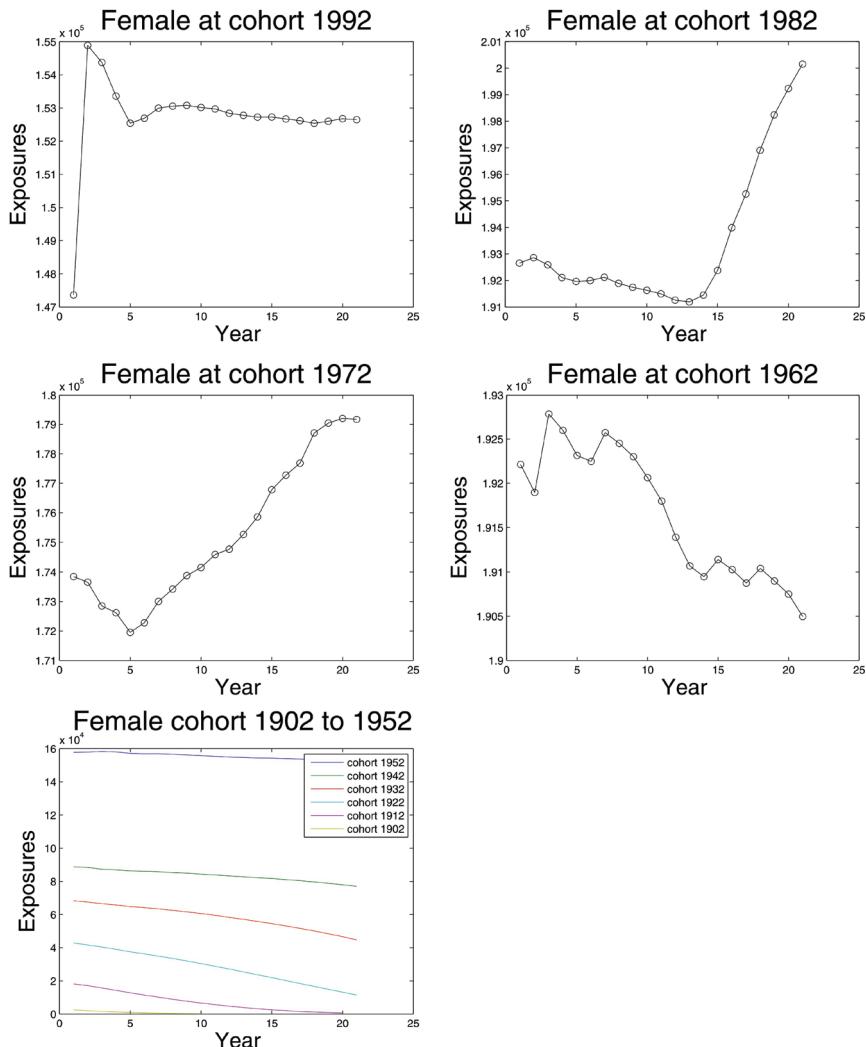


Figure 5: The exposures at various cohorts for Taiwan female data.

data is the training set used to calibrate the model, and the last 2-year data is the testing set used to assess the model. The results are summarized in Table 2.

Indeed, there are two important messages one can tell from Table 2. First, no matter for female or male, the APC model substantially has a better in-sample fitting performance, but a worse forecasting ability than the AP model. This suggests that the APC model may over-fit the Taiwan's mortality data. Because

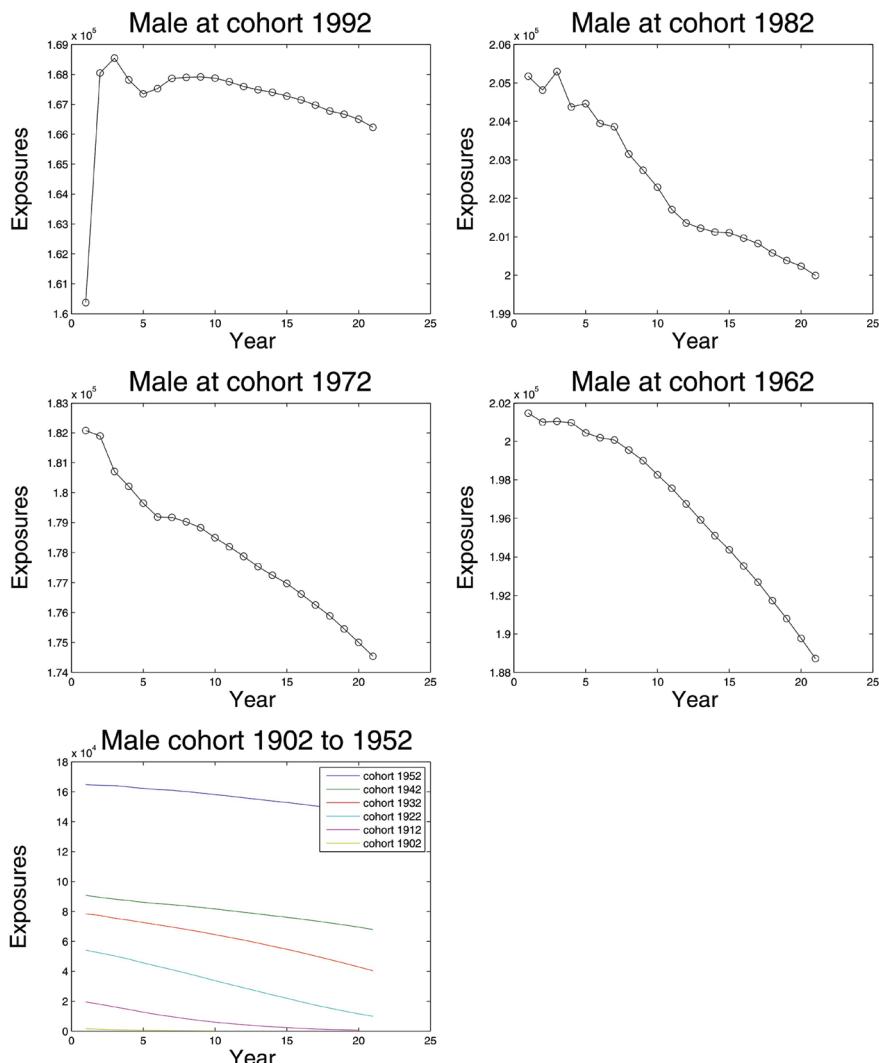


Figure 6: The exposures at various cohorts for Taiwan male data.

the APC model contains a lot more parameters than the AP model, forecast using the APC model is less stable in the case of using full data and augmented data. For these reasons, we simply focus on the AP model for the out-of-sample analysis.

First note that, compared with the truncated data in out-of-sample forecast, the full data performs worse for the female, but better for the male. It appears

Table 2: Empirical results.

MAPEs	In-sample fitting (1998–2010)				Out-of-sample forecasting (2011–2012)				
	Child	Adult	Elder	Over 90	Overall	Child	Adult	Elder	
The age-period model									
Female									
Truncated data	1.845	0.939	1.750	4.125	1.359	2.584	1.307	2.253	4.229
Full data	2.008	1.019	5.572	17.132	2.761	2.210	1.268	6.859	20.788
Augmented data									3.366
The simple method	1.965	0.998	2.512	6.543	1.673	2.293	1.247	3.417	8.167
The two-step method	1.967	1.000	2.643	6.999	1.720	2.279	1.245	3.043	6.933
The MD method	1.958	0.996	1.830	4.155	1.432	2.330	1.250	2.345	4.282
Male									
Truncated data	1.683	0.891	2.245	5.072	1.484	2.213	0.998	4.518	9.938
Full data	1.817	0.934	3.650	9.764	2.017	2.412	1.029	4.003	7.970
Augmented data									2.277
The simple method	1.815	0.931	2.771	6.706	1.707	2.425	1.023	6.125	15.325
The two-step method	1.816	0.929	2.755	6.642	1.701	2.437	1.020	4.029	7.901
The MD method	1.816	0.929	2.284	5.000	1.536	2.440	1.019	4.785	10.335
The age-period-cohort model									
Female									
Truncated data	2.786	0.801	1.626	4.103	1.388	8.958	1.304	1.926	3.935
Full data	3.316	0.860	1.725	4.324	1.531	10.679	1.191	2.089	4.206
Augmented data									2.670
The simple method	3.289	0.861	1.697	4.255	1.518	10.244	1.180	2.126	4.540
The two-step method	3.295	0.859	1.735	4.344	1.531	10.317	1.195	3.309	8.217
The MD method	1.753	0.795	1.598	3.816	1.220	2.656	2.775	2.248	4.425
Male									
Truncated data	2.489	0.717	1.462	3.506	1.243	9.255	0.916	3.570	9.577
Full data	1.774	0.768	2.004	5.277	1.351	432.691	0.825	3.823	9.540
Augmented data									66.654
The simple method	1.748	0.771	1.666	4.085	1.231	570.897	0.813	4.454	12.055
The two-step method	1.776	0.767	1.983	5.206	1.344	NAN	0.820	4.289	11.185
The MD method	1.716	0.766	1.740	4.280	1.250	2.200	0.839	3.929	10.184

that using full data does not consistently produce better forecast than simply using truncated data. We then concentrate on the comparison between truncated and augmented datums.

Focusing on the out-of-sample forecast, the overall prediction MAPEs for the augmented data using the MD method and the two-step method are smaller than the truncated one for the female and the male, respectively. The improvement largely comes from the population of age under 65 for the female and from the elder for the male. We therefore conclude that the augmented data with a suitable method helps to improve out-of-sample prediction in general, in comparison with the truncated data. Also, the imputation method does make up the deficiency of the weighted method, in addressing the missing-data problem.

6 Conclusion

The missing data problem of Taiwan arises from the change in data recording convention due to increasing life expectations. We provide three data augmentation methods to impute the missing mortality data of Taiwan based on the AP model and the APC model. Our empirical analysis shows mortality forecasting can be improved when a proper data augmentation method is used. In particular, regarding missing values as parameters provides a great complementary alternative to the weighted method advocated by Renshaw and Haberman (2006).

Although the APC model can explain more variations than the AP model in terms of in-sample fitting, we found it does not perform well in prediction. One reason is that the APC model demands a great deal of parameters to incorporate the cohort effect and hence may cause the over-fitting problem to the data, compared with the AP model. This observation hints the need of other ways for cohort modelling, which is one focus of our future studies.

Funding: Ministry of Science and Technology, Taiwan (Grant / Award Number: “102-2410-H-033-053”, “103-2633-M-008-001”.

References

- Booth, H., J. Maindonald, and L. Smith. 2002. “Applying Lee-Carter under Conditions of Variable Mortality Decline.” *Population Studies* 56 (3):325–36.
Butt, Z., and S. Haberman. 2009. “ilc: A collection of R functions for fitting a class of Lee-Carter mortality models using iterative fitting algorithms,” *Actuarial Research Paper No.190*.

- Cairns, A., D. Blake, and K. Dowd. 2006. "A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration." *Journal of Risk and Insurance* 73:687–718.
- Cairns, A., D. Blake, K. Dowd, G. Coughlan, D. Epstein, A. Ong, and I. Balevich. 2009. "A Quantitative Comparison of Stochastic Mortality Models Using Data From England & Wales and the United States." *North American Actuarial Journal* 13:1–35.
- Cairns, A., D. Blake, K. Dowd, G. Coughlan, and M. Khalaf-Allah. 2011. "Bayesian Stochastic Mortality Modelling for Two Populations." *Astin Bulletin* 41:29–59.
- Coale, A., and G. Gao. 1989. "Revised Regional Model Life Tables at Very Low Levels of Mortality." *Population Index* 55:613–43.
- Coale, A., and E. Kisker. 1990. "Defects in Data on Old-Age Mortality in the United States: New Procedures for Calculating Schedules and Life Tables at the Highest Ages." *Asian and Pacific Population Forum* 4:1–31.
- Denuit, M., and A. -C. Goderniaux. 2005. "Closing and Projecting Lifetables Using Log-Linear Models." *Bulletin de l'Association Suisse des Actuaires*, 1:29–49.
- Dowd, K., D. Blake, A. Cairns, G. Coughlan, D. Epstein, and M. Khalaf-Allah. 2010. "Backtesting Stochastic Mortality Models: An Ex-Post Evaluation of Multi-Period-Ahead Density Forecasts." *North American Actuarial Journal* 14:281–98.
- Efron, B., and R. J. Tibshirani. 1993. *An Introduction to the Bootstrap*. New York: Chapman & Hall/CRC.
- Gompertz, B. 1825. "On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies." *Philosophical Transactions of the Royal Society of London* 115:513–83.
- Hyndman, R. J., and M. S. Ullah. 2007. "Robust Forecasting of Mortality and Fertility Rate: A Functional Data Approach." *Computational Statistics and Data Analysis* 51:4942–56.
- Jarner, S., and E. Kryger. 2011. "Modelling Adult Mortality in Small Populations: The Saint Model." *Astin Bulletin* 41:377–418.
- Kupper, L. L., J. M. Janis, A. Karmous, and B. G. Greenberg. 1985. "Statistical Age-Period-Cohort Analysis: A Review and Critique." *Journal of Chronic Diseases* 38:811–30.
- Lee, R., and L. R. Carter. 1992. "Modelling and Forecasting the Time Series of US Mortality." *Journal of the American Statistical Association* 87:659–71.
- Lee, R., and T. Miller. 2001. "Evaluating the Performance of the Lee-Carter Method for Forecasting Mortality." *Demography* 38:537–49.
- Li, N., and R. Lee. 2005. "Coherent Mortality Forecasts for a Group of Populations: An Extension of the Lee-Carter Method." *Demography* 42:575–94.
- Makeham, W. 1860. "On the Law of Mortality." *Journal of the Institute of Actuaries* 13:325–58.
- Renshaw, A. E., and S. Haberman. 2006. "A Cohort-Based Extension to the Lee-Carter Model for Mortality Reduction Factors." *Insurance: Mathematics and Economics* 38:556–70.
- Tuljapurkar, S., N. Li, and C. Boe. 2000. "A Universal Pattern of Mortality Decline in G7 Countries." *Nature* 405:789–92.
- Wilmoth, J. R. 1996. "Mortality Projections for Japan: A Comparison of Four Methods." In *Health and Mortality among Elderly Populations*, edited by G. Caselli and A. D. Lopez, eds., International Studies in Demography, 266–87. New York: Oxford University Press, chapter 13.
- Yang, S. S., J. C. Yue, and H. C. Huang. 2010. "Modeling Longevity Risks Using a Principal Component Approach: A Comparison With Existing Stochastic Mortality Models." *Insurance: Mathematics and Economics* 46:254–70.
- Zhou, R., J. S. -H. Li, and K. S. Tan. 2013. "Pricing Standardized Mortality Securitizations: A Two-Population Model with Transitory Jump Effects." *Journal of Risk and Insurance* 80:733–74.