Sains Malaysiana 54(3)(2025): 913-926
http://doi.org/10.17576/jsm-2025-5403-22

Forecasting Life Expectancy using Latent and Observable Factors: Effects of Urbanization on Mortality Modelling

(Ramalan Jangka Hayat menggunakan Faktor Terpendam dan Boleh Diperhatikan: Kesan Urbanisasi terhadap Pemodelan Kematian)

NORKHAIRUNNISA REDZWAN[1,2] & ROZITA RAMLI[1,*]

[1]*Department of Mathematical Sciences, Faculty of Science & Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia*
[2]*School of Mathematical Sciences, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia*

Received: 21 May 2024/Accepted: 10 December 2024

ABSTRACT

Sustainable development of the future is influenced by population growth, population ageing, migration, and urbanization. However, urbanization have adverse effects on environmental, social behaviour and health aspects, such as mortality risks due to climate change and infectious diseases. As the urban population is growing rapidly, there is a need to accurately forecasts these risks in anticipation of future ageing population, by developing a model to estimate and forecast mortality. Many existing extrapolative mortality models are described by latent factors, and it is difficult to understand the underlying dynamics of these factors. Therefore, a comprehensive analysis on the relationship between mortality index and urbanization growth were conducted. This study also aims to propose a new model that incorporates urbanization as an observable factor in modelling and forecasting mortality, by extending the Lee-Carter model. Result indicates that there is a possible long-run relationship between mortality and growth of urban population for all countries in the study. The proposed model provides better in-sample fitting for all countries. This study also predicts the life expectancy at birth based on the proposed model and life expectancy is forecast to reach age 85 for selected countries. Results also show that Malaysian adults ages between 20 and 40 years old are more likely to be affected by an increase in urbanization growth. Urbanization and mortality are key factors in planning for sustainable development of the future. Therefore, it is important to develop a mortality forecasting model that can account for the uncertainty surrounding urbanization.

Keywords: Forecasts; Lee-Carter; life expectancy; mortality; urbanization

ABSTRAK

Pembangunan mampan masa depan dipengaruhi oleh pertumbuhan penduduk, penuaan penduduk, penghijrahan dan urbanisasi. Walau bagaimanapun, urbanisasi mempunyai kesan buruk terhadap aspek alam sekitar, tingkah laku sosial dan kesihatan, seperti risiko kematian akibat perubahan iklim dan penyakit berjangkit. Oleh kerana penduduk bandar berkembang pesat, terdapat keperluan untuk meramalkan risiko ini dengan tepat dalam jangkaan penduduk menua masa depan, dengan membangunkan model untuk menganggar dan meramal kematian. Banyak model kematian ekstrapolatif sedia ada digambarkan oleh faktor pendam, yang sukar untuk difahami. Oleh itu, analisis komprehensif mengenai hubungan antara indeks kematian dan pertumbuhan urbanisasi dijalankan. Kajian ini juga bertujuan untuk mencadangkan model baharu yang menggabungkan urbanisasi sebagai faktor boleh cerap dengan faktor pendam dalam pemodelan dan ramalan kematian, sebagai model lanjutan Lee-Carter. Keputusan menunjukkan bahawa terdapat hubungan jangka panjang antara kematian dan pertumbuhan penduduk bandar untuk semua negara dalam kajian ini. Model yang dicadangkan memberi penyuaian sampel yang lebih baik untuk semua negara. Kajian ini turut meramalkan jangka hayat ketika lahir berdasarkan model cadangan dan jangka hayat dijangka mencapai umur 85 tahun bagi negara-negara terpilih. Keputusan kajian juga menunjukkan bahawa orang dewasa Malaysia yang berumur antara 20 dan 40 tahun lebih cenderung terjejas oleh peningkatan pertumbuhan urbanisasi. Urbanisasi dan kematian adalah faktor utama dalam merancang pembangunan mampan masa depan. Oleh itu, adalah penting untuk membangunkan model ramalan kematian yang boleh mengambil kira ketidakpastian dalam urbanisasi.

Kata kunci: Jangka hayat; kematian; Lee-Carter; ramalan; urbanisasi

INTRODUCTION

Ageing population is a global phenomenon that is affecting many countries in the world and it refers to the increase in elderly population as compared to the younger generations. Globally, population of 65 years and older were recorded to be 703 million persons in 2019. It is estimated that 1.5 billion persons in 2050 will consist of elderly population aged 65 years and above (United Nations 2019a). Due to the advancement of technology and health care facilities, many people can have a longer and healthier lives. This scenario plays important role in contributing to population ageing, along with decrease in fertility, increase in life expectancy and migration of people from rural to urban areas. According to the United Nations (2019b), sustainable economic, social, and environmental developments of the future are influenced by population growth, population ageing, international migration, and urbanization.

Urbanization is defined as the action of which substantial amount of population are consistently gathered in relatively small areas to establish cities. As at 2018, 55% of the world population are living in cities and it is estimated to increase to 68% in 2050 (United Nations 2019b). Rising of urban population is caused by migration to urban areas that offers many health and socioeconomic opportunities such as employment, healthcare resources and education (Ahmed, Zafar & Ali 2020; Sajid et al. 2020). It is also due to natural population growth in urban areas and conversion of rural to urban areas (Ahmed, Zafar & Ali 2020). However, urbanization have adverse effects on environmental, social behaviour and health aspects (Bijari, Mahdinia & Mansouri Daneshvar 2021; Garba et al. 2021; Ling et al. 2014). Rapid industrial development in urban areas leads to higher greenhouse gas emissions, air pollution and other pollutants, hence resulting in climate change. Urbanization also caused a reduction in natural resources such as land, water, and food to fulfil the needs of urban population (Ahmed, Zafar & Ali 2020; Antai & Moradi 2010; Bijari, Mahdinia & Mansouri Daneshvar 2021; Liu et al. 2021). For the urban population, more land must be used for waste management, agriculture (including the production of food and cattle), and residential areas. The main causes of social behaviour problems brought on by urbanisation include sedentary lifestyles, high calorie diets, and ongoing stress and pressure. Urban poverty is another social problem brought on by urbanisation since living standards are often better in cities than in rural areas. Financial and socioeconomic challenges endured by urban poor caused them to be less resilient to illness (Antai & Moradi 2010; Burroughs et al. 2015). Another issue related to urbanization is health issues, which include physical and mental health. Due to urban lifestyle, urban populations are more likely to exposed to risks factor of noncommunicable diseases such as diabetes, cardiovascular diseases, and depression (Bijari, Mahdinia & Mansouri Daneshvar 2021; Sajid et al. 2020; Wicke et al. 2022). They are also more exposed to communicable diseases because of overcrowding and high population density of urban areas (Walter & DeWitte 2017). Bijari, Mahdinia and Mansouri Daneshvar (2021) reported that there is a strong correlation between urban population and the confirm cases of the recent COVID-19 pandemic. It is evident that the high population density of urban areas can amplify the risks of infectious diseases.

A major health issues related to urbanization is mortality of urban population. Urban population, especially its most vulnerable groups (children under five years old and female adults) have higher risk of mortality (Antai & Moradi 2010; Walter & DeWitte 2017). Islam et al. (2017) reported that urban female adults have shorter life expectancy than rural female adults from age 20 years old onwards. Male individuals in urban areas between the ages of 65 and 75 are expected to live shorter lives than rural male adults of the same ages. Urban heat island effects are the primary source of heat-related death, which is one example of how climate change contributes to mortality in urban settings. A city is referred to as a 'heat island' when its temperatures are higher than those of nearby rural areas. Continuous exposure to a hot environment can lead to discomfort and bodily ache particularly to high-risk persons such as elderly, children, individuals with health problems and urban poor (Chen et al. 2016; Oleson et al. 2013).

Mortality rates is one of the important demographic indicators that is often used as an indicator of population dynamics, along with fertility and migration rates. Mortality rates and mortality modelling are usually applied to areas of insurance, pensions, and healthcare (Demirel & Basak 2017) by the government, corporations, and individuals. Mortality modelling is important in addressing longevity risk. Longevity risk refers to the risk of unfortunate financial outcome due to people living longer than expected.

As the ageing population is growing rapidly, there is a need to accurately forecasts these risks in anticipation of future ageing population, by developing a model to estimate and forecast mortality. These risks are important in ensuring sufficient pricing and reserving for life insurance and annuity products, and in the risk management process of life insurers and pension plans (Niu & Melenberg 2014). It also assists the government in planning for healthcare and other services for society. Therefore, it is important to develop a mortality model that can account for any uncertainties and have better accuracy (Fung, Peters & Shevchenko 2017).

Mortality forecasting can be classified into expectation, explanatory, and extrapolative methods (Booth & Tickle 2008). Expectation method is based on expert's opinion, explanatory method is based on some certain causes of death with several risk factors and extrapolative method is based on past mortality trends.

The extrapolative approach of mortality forecasting is typically employed by statistical offices, actuaries, and demographers. Compared to other kinds of mortality forecasting models, it is more unbiased and appropriate for long-term forecasting. For example, a mortality forecasting model by Lee and Carter (1992) is widely known for its simplicity and its small number of parameters. It has become one of the prominent mortality models and has since become a base for other mortality forecasting model through modifications and extensions (Boonen & Li 2017; Cairns, Blake & Dowd 2008; Jaber, Yaacob & Alwadi 2023; Li & Lee 2005; Plat 2009). The model introduced by Lee and Carter (1992) describes the pattern of mortality by a latent factor derived from historical data, which is known as the mortality index. However, it is difficult to predict how these patterns work if the underlying dynamics are not fully understood. Lee-Carter model is also known to underestimates probabilities due to its narrow forecast intervals because of the estimated parameters that produced lower variances (Fazle Rabbi & Khan 2022). To overcome this issue, Fung, Peters and Shevchenko (2017) suggested to add other time-varying factors in mortality modelling, while Niu and Melenberg (2014) proposed that mortality model can be described by adding an observable factor. Several researchers have developed mortality models that incorporate both latent and observable factors such as macroeconomic factor (Boonen & Li 2017; Ma & Boonen 2023; Niu & Melenberg 2014), temperature changes (Seklecka, Pantelous & O'Hare 2017), and a combination of both (Dutton, Pantelous & Seklecka 2020).

It is important to consider and incorporate different factors that would impact mortality rates to account for any uncertainties that might occur (Fung, Peters & Shevchenko 2019; Fung, Peters & Shevchenko 2017). While research on the trend of mortality and observable factors is emerging, most studies focused on macroeconomic factors (GDP and CPI) on developed countries. Since urbanization leads to an increase in mortality risks such as unsanitary living conditions and spread of infectious diseases in crowded areas (Cutler, Deaton & Lleras-Muney 2006), we aim to extend the understanding of trend between mortality rates and other demographic factors, such as urban population. We also aim to provide comparison between developed and developing countries. The aim of this study was (1) to describe the trend and long-run relationship of urbanization and mortality index and (2) to propose a new model that incorporates urbanization in modelling and forecasting mortality. The proposed model provides better goodness-of-fit for all countries. This study also predicts the life expectancy at birth based on the proposed model and life expectancy is forecast to reach age 85 for selected countries.

MATERIALS AND METHODS

This study focuses on 6 countries which are Malaysia, United States of America (US), United Kingdom (UK), Finland, Netherlands, and Portugal with high urban population that is above 65%. An urbanized country is defined as country with more than 50% urban population (Sajid et al. 2020). Urban population refers to people living in urban areas as defined by national statistical offices. According to World Bank (2023), in 2020, the percentage of urban population in Portugal and Malaysia were 66% and 77%, respectively. The highest percentage of urban population were in Netherlands at 92%, followed by Finland, UK, and USA at 86%, 84%, and 83%, respectively.

Our period of study was from 1991 to 2020, which covers the early period of the COVID-19 pandemic. The outbreak of COVID-19, also known as Coronavirus Disease 2019, was first recorded in late 2019 which resulted in global public health emergency (Bijari, Mahdinia & Mansouri Daneshvar 2021). Annual mortality rates were obtained from the Department of Statistics Malaysia and Human Mortality Database (Human Mortality Database 2023). The age-specific mortality rates are in 5-year age groups, with a total of 17 age groups from 0-4 years until 80 years and above. Annual urbanization growth rates were obtained from World Bank (2023).

Previous works by Niu and Melenberg (2014) and Seklecka, Pantelous and O'Hare (2017) studied the long-run relationship between Lee-Carter mortality index, k_t, and other observable factors such as gross domestic product and temperature changes. Observable factor refers to data that can be directly measured or observed, and they are represented by the actual data collected in the study. Latent factor, also known as unobserved variable, refers to data that cannot be directly measured or observed, but are inferred from observable variables. This factor is usually abstract and theoretical, and it shows the underlying structure of the unobserved dataset where statistical techniques are required to estimate this factor (Masseran et al. 2024). In this study, we compare mortality models that used latent factor and observable factor. A well-known mortality model that uses latent factor is the Lee-Carter model, however, models with latent factor offers limited economic interpretation (Bai & Ng 2006). Therefore, in this section, we propose mortality model that combines latent and observable factors to provide valuable insights on their effects on projection of mortality.

First, we aim to analyse the trends in mortality index and urbanization growth rate. The latent variable, k_t, and other parameters of the Lee-Carter model (Lee & Carter 1992) for each country were obtained based on the following equation:

$$ln(m_{x,t}) = \alpha_x + \beta_x k_t + \varepsilon_{x,t}^{k}$$
$$\text{for all } x \text{ (age group)} = 1, 2, \ldots, X \text{ and } t \text{ (year)} = 1, 2, \ldots, T. \quad (1)$$

For simplification purposes, the first age group and year are both represented by 1, while X and T refer to the last age group and year, respectively. Mortality rate at age x for year t is denoted by $m_{x,t}$; k_t is the mortality index that describes the changes in the level of mortality over time; β_x is the sensitivity of mortality rates at age x; and $\varepsilon_{x,t}^{k}$ is the error term that is independent and identically distributed normal random variable with mean 0 and variance σ^2. The model is subject to constraints $\sum_{x=1}^{X} \beta_x = 1$ and $\sum_{t=1}^{T} k_t = 0$. By applying the constraint $\sum_{t=1}^{T} k_t = 0$ and minimizing errors, α_x is the average log mortality at age x. Next, by applying singular value decomposition (SVD) to matrix $ln(m_{x,t}) - \alpha_x$, values of β_x and k_t are obtained. Here, k_t is the latent factor as it explains the unobserved underlying structure of mortality rate through statistical technique, which is the SVD method.

In this study, first we want to examine the trend behaviour of the Lee-Carter mortality index k_t, and natural logarithm of urbanization growth rate u_t. The stationarity for k_t and u_t were analyzed using Phillips-Perron test (Phillips & Perron 1988) with null hypothesis of nonstationarity because it is a more robust method for unit root testing (Gokmenoglu, Azin & Taspinar 2015). Cointegration analysis is performed to analyse the long-run relationship between k_t and u_t by using Johansen (1988) cointegration test with the null hypothesis of no cointegration ($r = 0$) and null hypothesis of one cointegrating vector ($r \leq 1$).

We then compare the Lee-Carter model (Equation 1) with a model that describes mortality rates with an observable factor as in Equation (2).

$$ln(m_{x,t}) = \theta_{0,x} + \theta_{1,x} u_t + \varepsilon_{x,t}^{u}$$
$$\text{for all } x \text{ (age group)} = 1, 2, \ldots, X \text{ and } t \text{ (year)} = 1, 2, \ldots, T. \quad (2)$$

Here, u_t is the observable factor that was obtained from World Bank database. It is the urbanization growth rate in natural logarithm at year t. $\theta_{0,x}$ is the average log mortality at age x when u_t is zero, $\theta_{1,x}$ is the sensitivity of mortality rates at age x to the changes in u_t. Error term is denoted by $\varepsilon_{x,t}^{k}$ that is independent and identically distributed normal random variable with mean 0 and variance σ^2. The parameters were obtained using ordinary least squares (OLS) method.

Based on analysis in the previous section, we propose a mortality forecasting model that incorporate both latent and observable factors, as follows:

$$ln(m_{x,t}) = a_x + b_x k_t + c_x u_t + \varepsilon_{x,t}$$
$$\text{for all } x \text{ (age group)} = 1, 2, \ldots, X \text{ and } t \text{ (year)} = 1, 2, \ldots, T. \quad (3)$$

This proposed model in Equation (3) follows Niu and Melenberg (2014) model setup and estimation (see Niu and Melenberg 2014 for details). Here $m_{x,t}$ is the mortality rates for age group x at year t; a_x is the average log mortality at age x; k_t is the mortality index which describes the variation in the level of mortality over time; b_x is the sensitivity of mortality rates at age x to the changes in k_t; u_t is the growth of urbanization in natural logarithm; c_x is the sensitivity of mortality rates at age x to urbanization growth; and $\varepsilon_{x,t}$ is the error term.

Equation (3) is fitted to age-specific mortality rates and annual observable factor which is the urbanization growth rate. Parameters were obtained by using the Newton-Raphson method where a process of iteration optimization is involved by minimizing the sum of squares of residuals (See Verbeke and Cools 1995 for details on Newton-Raphson method).

The goodness-of-fit of our model is compared using BIC and mean absolute percentage error (MAPE) in period of between 1991 and 2020. We compare the fitting of our proposed model with the Lee-Carter model. BIC is defined as follows (Gujarati & Porter 2009):

$$BIC = k \ln(n) + n \ln(RSS) - n \ln(n) \quad (4)$$

Here k is the number of independent variables (including intercept); n is the number of observations; and RSS is the residual sum of squares.

MAPE is defined as follows:

$$MAPE = \left[\frac{\sum_{x=1}^{17} \sum_{t=1}^{30} \left| \frac{ln(m_{x,t}) - ln(\hat{m}_{x,t})}{ln(m_{x,t})} \right| \times 100}{n} \right] \quad (5)$$

Here $ln(m_{x,t})$ is the observed logarithm mortality rates; and $ln(\hat{m}_{x,t})$ is the estimated logarithm mortality rates.

In this section, out-sample mortality forecasts were performed for three different fitting periods, which are 24 years, 27 years and 29 years. We classified this as Scenario A (in-sample: 24 years, out-sample: 6 years), Scenario B (in-sample: 27 years, out-sample: 3 years) and Scenario C (in-sample: 29 years, out-sample: 1 year). For each scenario, we re-estimate the parameters based on its respective in-sample periods. The latent and observable factors, k_t and u_t, in both Equation (3) and Lee-Carter model were forecast using ARIMA(0,1,0) with drift model. Forecast results are compared using root mean square error (RMSE) and MAPE. RMSE is defined as follows:

$$RMSE = \sqrt{\frac{\sum_{x=1}^{17}\sum_{t=1}^{6}\left(ln(m_{x,t}) - ln(\hat{m}_{x,t})\right)^2}{n}} \quad (6)$$

Here $ln(m_{x,t})$ is the observed logarithm mortality rates, and $ln(\hat{m}_{x,t})$ is the estimated logarithm mortality rates. As a supplementary analysis, we conduct a hypothesis testing for the difference in residual means for all three different fitting periods with null hypothesis of no difference in residual means.

Mortality rates were forecast for the next 15 years. Forecasts of mortality rates are based on the forecasts of k_t and u_t, which were determined by autoregressive integrated moving average (ARIMA) model with the smallest AIC values. By using the forecast mortality rate, we determine the expected life expectancy at birth for the next 15 years, which is from 2021 to 2035.

RESULTS AND DISCUSSION

Trends in mortality index and urbanization growth rate are shown by the historical plot of both factors as in Figure 1. In general, the mortality index k_t, for each country show a decreasing trend from 1991 to 2020. Portugal experienced a deeper decline of mortality index as compared to other countries. Urbanization growth rate, $u_{t,}$ for Malaysia, USA, Netherlands, and Portugal exhibited a decreasing trend, while u_t for UK and Finland, exhibited a more fluctuating trend. The stationarity for k_t and u_t were analyzed using Phillips-Perron test (Phillips & Perron 1988) with null hypothesis of non-stationarity. Results are shown in Panel A and Panel B of Table 1.

Results of Phillips-Perron test for Malaysia, UK, Netherlands and Portugal for both k_t and u_t shows that the series are not stationary at levels. For k_t, we can only reject the null hypothesis of nonstationarity at 1% significance level for Finland. For u_t, we can only reject the null hypothesis of nonstationarity at 10% significance level for USA. Therefore, the results shows that k_t and u_t are nonstationary. Phillips-Perron test were performed on the first difference of k_t and u_t as shown in Panel C of Table 1. The results indicates that time series for k_t and u_t for each country is integrated of order one I(1).

Cointegration analysis is performed to analyse the long-run relationship between k_t and u_t by using Johansen (1988) cointegration test with the null hypothesis of no cointegration ($r = 0$) and null hypothesis of one cointegrating vector ($r \leq 1$). Results of Johansen cointegration test is shown in Table 2.

Results of the cointegration test are shown in Table 2. Critical values for the null of $r = 0$ are 17.85, 19.96 and 24.6 for $p < 0.10, p < 0.05$ and $p < 0.01$, respectively. Critical values for the null of $r \leq 1$ are 7.52, 9.24, and 12.97 for $p < 0.10, p < 0.05$ and $p < 0.01$, respectively. The null of no cointegration is rejected at 5% significance level for Netherlands, and 1% significance level for all other countries. The null of one cointegrating vector is rejected at 5% significance level for Finland and Netherlands, and at 10% level of significance for Malaysia and Portugal. However, the null of one cointegrating vector cannot be rejected at 10% level of significance. In general, the result indicates that there is a possible long-run relationship between mortality and growth of urban population for all countries. Following this, we further study the implications of k_t and u_t on future life expectancies. We compare the Lee-Carter model (Equation 1) with a model that describes mortality rates with an observable factor as in Equation (2). Table 3 shows the best ARIMA model to forecast k_t and u_t.

Projection of the life expectancy at birth are as shown in Figure 2. Based on Figure 2, the expected life expectancy at birth for most countries such as UK, Finland, Netherlands, and Portugal exceed 85 years old. USA and Netherlands have similar mean forecasts for both models. For the other countries, the mean forecasts based on Equation (2) are lower than mean forecasts based on Lee-Carter model. Forecasts of Equation (2) have wider intervals as compared to Lee-Carter model. For US, UK, Netherlands and Portugal, Equation (2) provides higher variance in estimated parameter, while forecast intervals by Lee-Carter model have lower variance in the estimated parameters. Lee-Carter model has narrow forecasts because it underestimates the variance in life expectancies (Lee & Carter 1992; Shang 2012). Therefore, we extend our study to incorporate both k_t and u_t in modelling and forecasting mortality rates. We used the same dataset of the six countries as in the previous section. Plot of estimated parameter are shown in Figure 3.

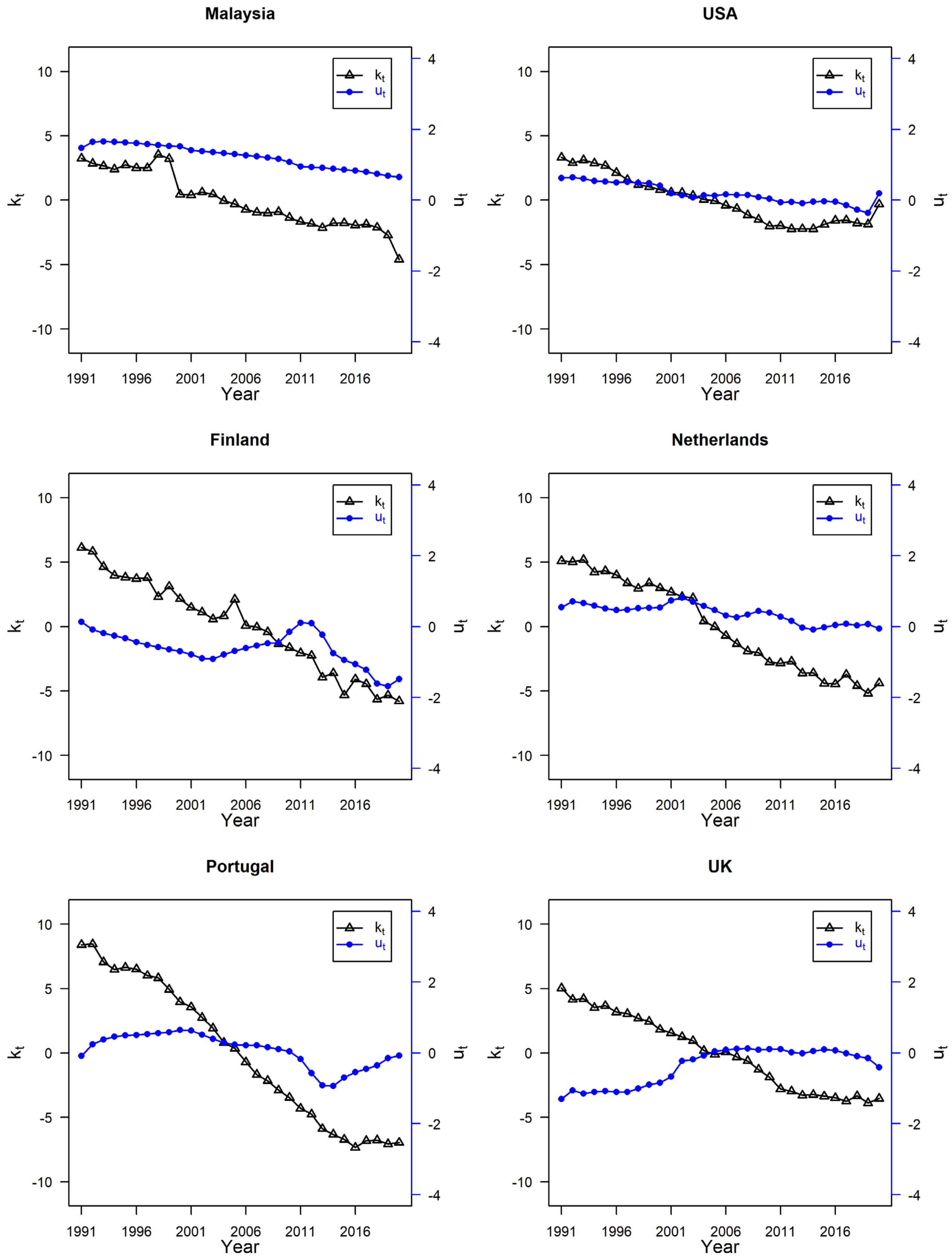

FIGURE 1. Mortality index k_t, (left) and urbanization growth rate u_t, (right) from 1991 to 2020

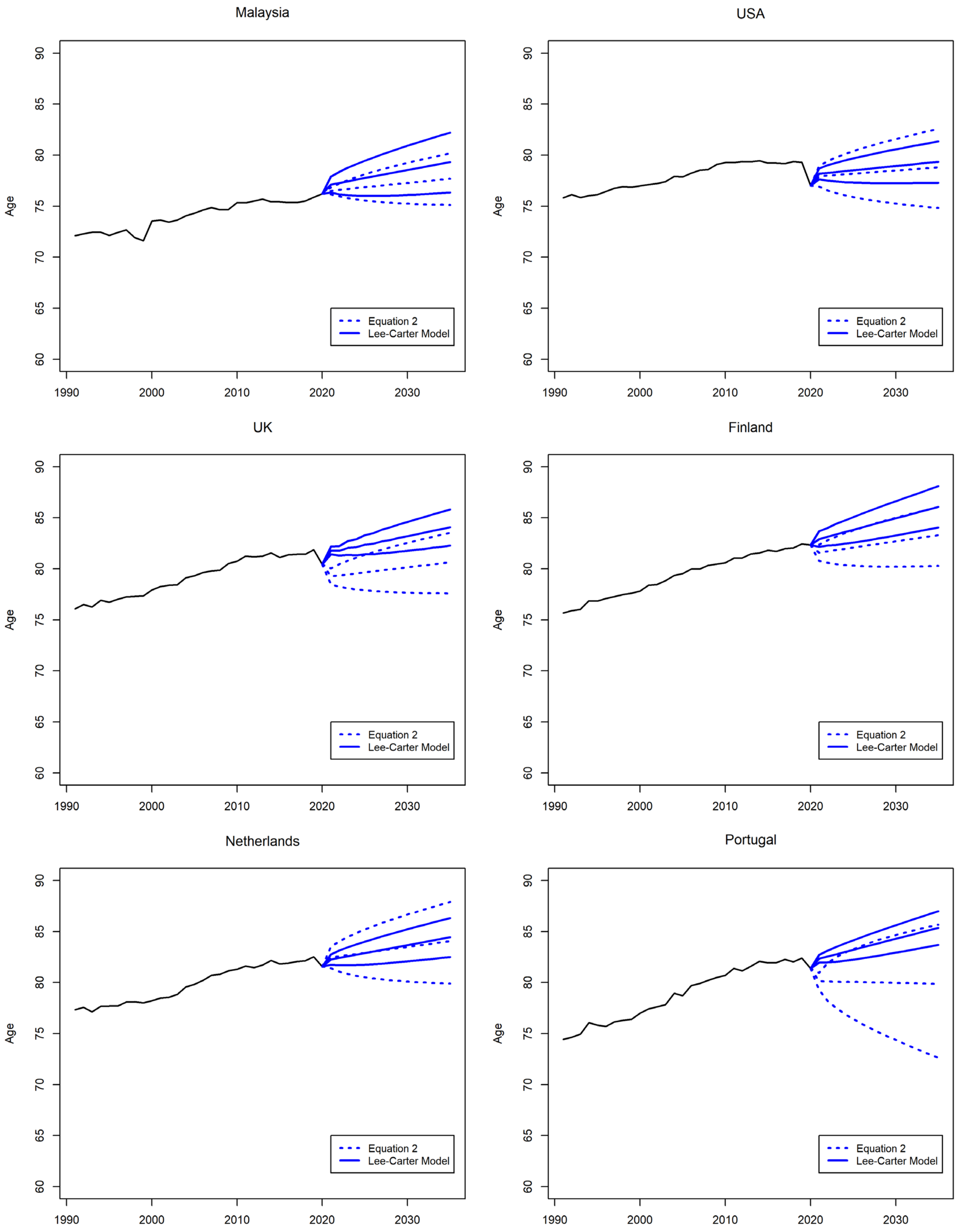

FIGURE 2. Historical and forecast life expectancies at birth with 95% prediction intervals for Equation 2 (dashed lines) and Lee-Carter model (solid lines)

TABLE 1. Results of unit root tests: Test statistics

	Malaysia	USA	UK	Finland	Netherlands	Portugal
Panel A: Levels of the time series (Phillips-Perron test)						
k_t	-13.9	5.94	-5.37	-26.1***	-8.07	-2.19
u_t	-15.0	-17.5*	1.03	-5.32	-10.6	-3.08
Panel B: First difference of time series (Phillips-Perron test)						
k_t	-22.1**	-21.5**	-41.4***	-39.2***	-32.8***	-27.4***
u_t	-23.2***	-23.3***	-24.1***	-12.6*	-19.1**	-11.8*

Note: *p-values < 0.10; **p-values < 0.05; ***p-values < 0.01

TABLE 2. Johansen cointegration test statistic

	Malaysia	USA	UK	Finland	Netherlands	Portugal
Stage 2: $r \leq 1$	8.02*	6.09	6.17	11.12**	9.70**	7.83*
Stage 1: $r = 0$	27.27***	27.56***	27.66***	35.66***	22.32**	39.80***

Note: *p-values < 0.10; **p-values < 0.05; ***p-values < 0.01

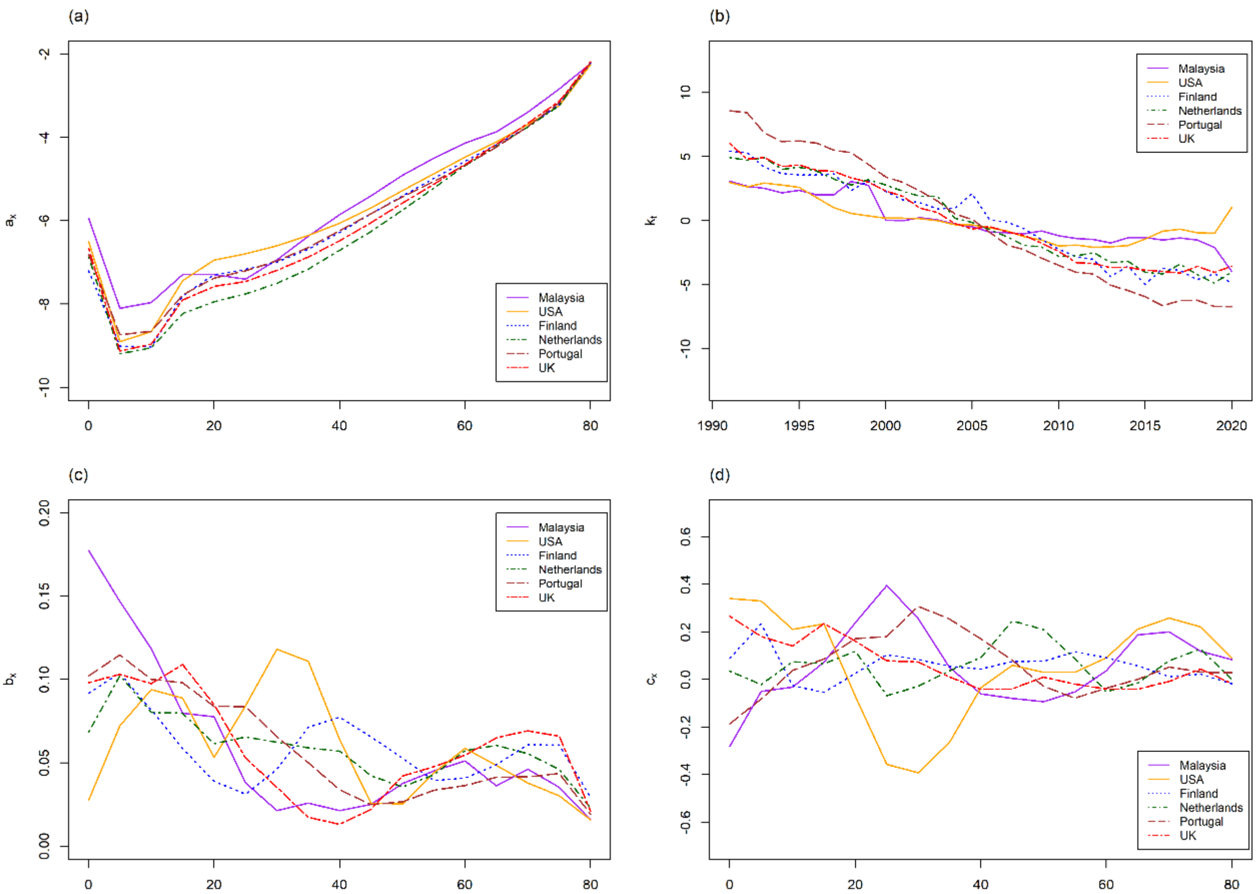

FIGURE 3. Parameters of proposed model

Based on Figure 3, the patterns of a_x, which measures the average mortality, are similar for all countries. The values of b_x, varies across countries with the highest b_x b_x at younger and older ages are in Malaysia and UK, respectively. This means that in Malaysia, mortality rates at younger ages are more sensitive to changes in the mortality index, while in UK, mortality rates at older ages are more sensitive to changes in the mortality index. Lowest b_x at both younger and older ages are in the USA. The latent factor k_t, in the proposed model is smaller as compared to k_t in Lee-Carter model in most cases. The values of $c_x c_x$, that measures the sensitivity of mortality rates toward urbanization growth are almost similar in most countries, except USA. Generally, c_x have a positive relationship with mortality rate. In the US, relationship between u_t and mortality rate is negative for ages between 20 and 40 years old. In Malaysia and UK, there is a positive relationship between u_t and mortality rate for ages between 20 and 40 years old. This means that an increase in u_t will lead to an increase in mortality rates in these countries. This finding supports Islam et al. (2017) that concluded urban adults are exposed to increased risk of mortality, for example, due to climate change and infectious diseases (Bijari, Mahdinia & Mansouri Daneshvar 2021; Chen et al. 2016). Furthermore, comparison of model fitting between LC model and the proposed model were done using Bayesian Information Criterion (BIC) and Mean Absolute Percentage Error (MAPE). Results are shown in Table 4.

Results show that the proposed model provide better quality of fitting for all countries. The most improved quality of fitting is for Portugal and Malaysia, which recorded the highest difference in BIC values. Results for out-sample forecast accuracy using the proposed model is as shown in Table 5.

TABLE 3. Best ARIMA model forecast k_t and u_t

	k_t	u_t
Malaysia	ARIMA(0,1,0) with drift	ARIMA(1,1,0) with drift
USA	ARIMA(0,1,0) with drift	ARIMA(0,1,0) with drift
UK	ARIMA(2,1,0) with drift	ARIMA(0,1,0) with drift
Finland	ARIMA(1,1,0) with drift	ARIMA(0,1,0) with drift
Netherlands	ARIMA(0,1,0) with drift	ARIMA(0,1,0) with drift
Portugal	ARIMA(1,1,0) with drift	ARIMA(0,1,1) with drift

TABLE 4. Results for BIC and MAPE

	BIC		MAPE (%)	
	LC	Proposed model	LC	Proposed model
Malaysia	-2861.002	-3017.18	0.89	0.78
USA	-2879.33	-3038.77	0.71	0.65
UK	-3223.06	-3344.23	0.56	0.49
Finland	-2467.82	-2498.57	1.07	1.002
Netherlands	-3056.33	-3081.88	0.59	0.57
Portugal	-2636.07	-2772.05	0.89	0.76

TABLE 5. Out-sample forecast accuracy

	RMSE		MAPE (%)	
	Proposed model	LC	Proposed model	LC
Scenario A: 6 years forecast				
Malaysia	0.1309	0.1035	1.87	1.45
US	0.2148	0.1820	2.54	2.19
Finland	0.2147	0.1444	3.29	1.90
Netherlands	0.1689	0.0932	2.41	1.23
Portugal	0.2497	0.1908	3.84	2.38
UK	0.2919	0.0996	4.69	1.53
Scenario B: 3 years forecast				
Malaysia	0.1241	0.1264	1.57	1.58
US	0.1487	0.1526	2.04	2.07
Finland	0.1843	0.1553	2.11	1.56
Netherlands	0.1013	0.0875	1.36	1.06
Portugal	0.2113	0.1337	2.82	1.72
UK	0.1632	0.0950	2.10	1.43
Scenario C: 1 year forecast				
Malaysia	0.1567	0.1596	1.87	1.64
US	0.1685	0.2196	3.16	3.72
Finland	0.1254	0.1299	1.37	1.31
Netherlands	0.1202	0.1051	1.93	1.69
Portugal	0.1263	0.0973	2.12	1.58
UK	0.1432	0.1185	2.47	2.11

TABLE 6. Difference in residual means: Test statistics

	Scenario A	Scenario B	Scenario C
Malaysia	-1.43	-0.80	-0.17
US	-0.06	0.42	0.59
Finland	-3.30*	-1.53	-0.29
Netherlands	-6.24*	-2.92*	-0.61
Portugal	5.59*	-2.39	-0.71
UK	0.94	0.98	-0.42

*reject null hypothesis at 1% level of significance

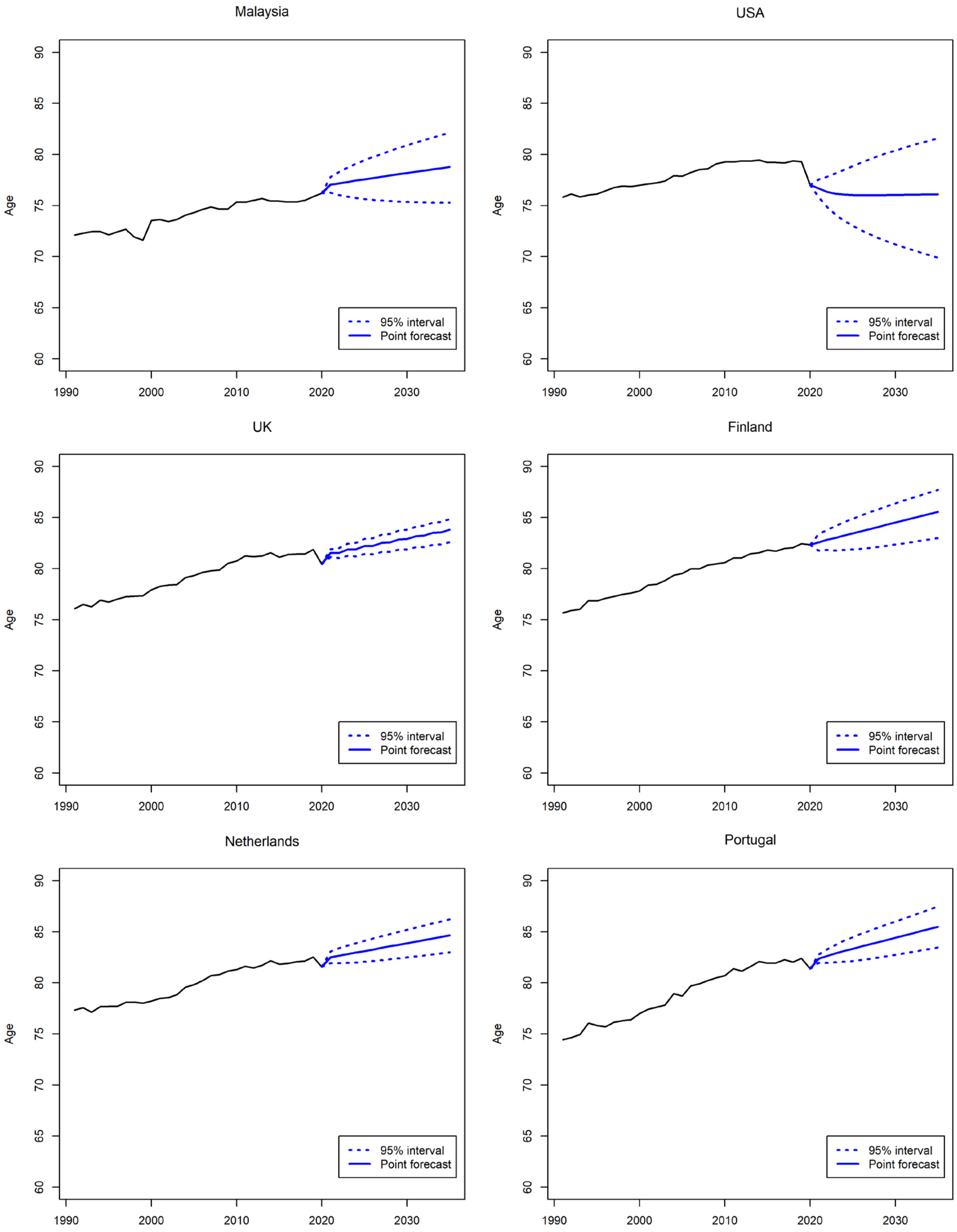

FIGURE 4. Historical and forecast life expectancies at birth with 95% prediction intervals based on the proposed model

Based on Table 5, the proposed model outperforms LC model in Malaysia and US for 3 years out-sample forecast, while it outperforms Malaysia, US and Finland for 1 year out-sample forecast. The proposed model incorporates an additional factor to model mortality, which is urbanization growth. Hence, the proposed model provides additional information as compared to the LC. It also provides more interpretable scenarios about mortality, for example it explains how urbanization helps to model future trends of mortality. Although in other cases it seems that LC performs better than the proposed model, LC only consists of latent factor which challenges the understanding and interpretation of future trends of mortality particularly to individuals with limited statistical knowledge. It is interesting to note that the difference in both RMSE and MAPE of both models are rather small. As a supplementary analysis, we conducted hypothesis testing on the difference between residual means of LC model and the proposed model. Results are shown in Table 6.

Based on Table 6, the critical values for null hypothesis of no difference in means are 1.645, 1.96, and 2.576 for level of significance 10%, 5%, and 1%, respectively. In Scenario A, the null hypothesis is rejected at 1% level of significance for Finland, Netherlands, and Portugal. From Figure 1, trend of $\boldsymbol{u_t}$ for Finland, Netherlands, and Portugal are centered at 0, while trend of $\boldsymbol{k_t}$ for these countries are sharply decreasing. Hence, there is a difference in the mean of residuals of Lee-Carter and the proposed model. In Scenario B, the null hypothesis is rejected 1% level of significance for Netherlands. Other than that, the result indicates that we fail to reject null hypothesis of no difference in the means of residuals between the proposed model and Lee-Carter models. In Scenario C, we fail to reject the null hypothesis of no difference in the means of residuals for all countries. Ideally, the proposed model needs at least 29 years of fitting periods so that there will be no difference in the means of residuals between the proposed model and Lee-Carter model. Results for forecast life expectancy at birth are shown in Figure 4.

We can see from Figure 4 that all countries show increase in forecast life expectancy for year 2021 to 2035, except for US. After 2011, Djeundje et al. (2022) reported that many developed countries experienced lower mortality improvement rate, including US. It is also interesting to note that the sharp decline in life expectancies in year 2020 due to COVID-19 pandemic. Finland and Portugal exhibited the highest forecast life expectancy at approximately 85 years old in the next 15 years. Wider forecast intervals are illustrated by USA and Malaysia, while UK, Finland, Netherlands, and Portugal show a relatively narrow forecast interval. According to Niu and Melenberg (2014), narrow forecast interval means least volatile, while wider forecast interval means more volatile. The volatility in k_t and u_t affects its uncertainty and is then reflected in the forecast interval. UK have lower variance of the estimated parameters, while estimated parameters in the US have higher variance. This means that UK's mortality and urbanization experience are less volatile as shown by its narrow forecast interval, while USA's mortality and urbanization experience are more volatile which corresponds to its wider forecast interval. Mortality, life expectancy and urbanization are part of the sustainable development goals (SDG) developed by United Nations (Fonseca, Domingues & Dima 2020). The SDG, particularly SDG3 on good health and well-being and SDG11 on sustainable cities and communities. Therefore, the findings in this study are aligned with the SDG mentioned earlier, and it is important to develop empirical connection between urbanization, mortality, and life expectancy.

CONCLUSION

An estimated 1.5 billion people in 2050 will be 65 years of age or older, contributing to the global phenomenon of population ageing that is impacting many nations. The primary causes of the population's ageing are rising life expectancy, declining fertility, and rural-to-urban migration. The latter is a contributing factor to the increasing urban population. Urban population mortality, including mortality risks from infectious diseases and climate change, is a significant health concern associated with urbanisation. This paper examines the relationship between mortality and urbanization based on data from 1991 to 2020 for countries with high urban population, which are Malaysia, USA, UK, Finland, Netherlands, and Portugal. Mortality is represented by mortality index of the Lee-Carter model k_t, and urbanization is represented by urbanization growth u_t. Findings indicate a long-term relationship between urbanisation and mortality based on the cointegration analysis. Therefore, we extend the existing Lee-Carter model by including observable factor that is represented by urbanization growth to provide better mortality estimation and forecasts. In terms of BIC and MAPE error measures, the proposed model outperforms the Lee-Carter model in terms of goodness-of-fit for all nations. This study also indicates that there is no difference in means of residuals for both Lee-Carter and the proposed models for a fitting period of 29 years. Additionally, we provide forecast of life expectancy at birth using the proposed model by forecasting both the latent and observable factors.

In the context of Malaysia population, the proposed model show that young ages in Malaysia are more sensitive to changes in the mortality index. Malaysian adults ages between 20 and 40 years old are more likely to be affected by an increase in u_t, because there is a positive relationship between urbanization growth and mortality rate at these ages. As urbanization rate increases, risk of mortality increases due to the deterioration of health condition and financial capabilities of the urban population. Limited financial resources of the urban population restrict their spending on healthcare which caused them to be less resilient to illness (Antai & Moradi 2010; Burroughs et al. 2015). Urbanization also affects health condition of the urban population, for example, urban lifestyle that increases the exposure to diabetes and cardiovascular diseases (Bijari, Mahdinia & Mansouri Daneshvar 2021; Sajid et al. 2020). Issues of overcrowding and high population density in urban areas increases the transmission of infectious disease such as COVID-19 (Bijari, Mahdinia & Mansouri Daneshvar). The proposed model outperforms Lee-Carter model in forecasting mortality rate in fitting periods of 27 and 29 years. Overall, there is no difference in mean of residuals for both Lee-Carter and the proposed models for all three different fitting scenarios.

Urbanization and mortality are key factors in planning for sustainable development of the future. Therefore, it is important to develop a mortality forecasting model that can account for the uncertainty surrounding urbanization. Findings of this study can provide valuable insights on effects of urbanization on mortality, which is useful information to help policymakers to understand the drivers of mortality and how changes of urbanization affect mortality. Although this study focused on the case of high urbanization country which may limit the generalization of the study, future research can extend and replicate the models used and compare them with low urbanization countries. This research can also be extended to include comparison with other extrapolative mortality forecasting models and to include longer time series data.

ACKNOWLEDGEMENTS

Norkhairunnisa Redzwan would like to acknowledge the Ministry of Higher Education (MOHE), Malaysia and Universiti Teknologi MARA (UiTM) for their support in her PhD studies at Universiti Kebangsaan Malaysia (UKM), under Skim Latihan Akademik Bumiputera (SLAB).

REFERENCES

Ahmed, Z., Zafar, M.W. & Ali, S. 2020. Linking urbanization, human capital, and the ecological footprint in G7 countries: An empirical analysis. *Sustainable Cities and Society* 55: 102064.

Antai, D. & Moradi, T. 2010. Urban area disadvantage and under-5 mortality in Nigeria: The effect of rapid urbanization. *Environmental Health Perspectives* 118(6): 877-883.

Bai, J. & Ng, S. 2006. Evaluating latent and observed factors in macroeconomics and finance. *Journal of Econometrics* 131: 507-537.

Bijari, N.B., Mahdinia, M.H. & Mansouri Daneshvar, M.R. 2021. Investigation of the urbanization contribution to the COVID-19 outbreak in Iran and the MECA countries. *Environment, Development and Sustainability* 23: 17964-17985.

Boonen, T.J. & Li, H. 2017. Modeling and forecasting mortality with economic growth: A multipopulation approach. *Demography* 54(5): 1921-1946.

Booth, H. & Tickle, L. 2008. Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science* 3(1-2): 3-43.

Burroughs Pena, M.S., Bernabé-Ortiz, A., Carrillo-Larco, R.M., Sánchez, J.F., Quispe, R., Pillay, T.D., Málaga, G., Gilman, R.H., Smeeth, L. & Miranda, J.J. 2015. Migration, urbanization and mortality: 5-year longitudinal analysis of the PERU MIGRANT study. *Journal of Epidemiology and Community Health* 69(7): 715-718.

Cairns, A.J., Blake, D. & Dowd, K. 2008. Modelling and management of mortality risk: A review. *Scandinavian Actuarial Journal* 2008(2-3): 79-113.

Chen, K., Zhou, L., Chen, X., Ma, Z., Liu, Y., Huang, L., Bi, J. & Kinney, P.L. 2016. Urbanization level and vulnerability to heat-related mortality in Jiangsu Province, China. *Environmental Health Perspectives* 124(12): 1863-1869.

Cutler, D., Deaton, A. & Lleras-Muney, A. 2006. The determinants of mortality. *Journal of Economic Perspectives* 20(3): 97-120.

Demirel, D.F. & Basak, M. 2017. Human mortality modeling with a fuzzy approach based on singular value decomposition technique. In *Computational Intelligence: International Joint Conference*, IJCCI 2015 Lisbon, Portugal, November 12-14, 2015, Revised Selected Papers 2017. Springer International Publishing. pp. 197-215.

Djeundje, V.B., Haberman, S., Bajekal, M. & Lu, J. 2022. The slowdown in mortality improvement rates 2011-2017: A multi-country analysis. *European Actuarial Journal* 12(2): 839-878.

Dutton, L., Pantelous, A.A. & Seklecka, M. 2020. The impact of economic growth in mortality modelling for selected OECD countries. *Journal of Forecasting* 39(3): 533-550.

Fazle Rabbi, A.M. & Khan, H.T. 2022. Stochastic mortality forecasts for Bangladesh. *PLoS ONE* 17(11): e0276966.

Fonseca, L.M., Domingues, J.P. & Dima, A.M. 2020. Mapping the sustainable development goals relationships. *Sustainability* 12(8): 3359.

Fung, M.C., Peters, G.W. & Shevchenko, P.V. 2019. Cohort effects in mortality modelling: A Bayesian state-space approach. *Annals of Actuarial Science* 13(1): 109-144.

Fung, M.C., Peters, G.W. & Shevchenko, P.V. 2017. A unified approach to mortality modelling using state-space framework: Characterisation, identification, estimation and forecasting. *Annals of Actuarial Science* 11(2): 343-389.

Garba, T.E., Richard, R.L., Thani, N.E.A., Majid, M.A.A., Lawal, M. & Yelwa, N.A. 2021. Geological effects on water quality: A review of issues and challenges in Malaysia. *Sains Malaysiana* 50(7): 1857-1870.

Gokmenoglu, K., Azin, V. & Taspinar, N. 2015. The relationship between industrial production, GDP, inflation and oil price: the case of Turkey. *Procedia Economics and Finance* 25: 497-503.

Gujarati, D.N. & Porter, D.C. 2009. *Basic Econometrics*. 5th ed. New York: Prentice Hall.

Human Mortality Database. 2023. Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France). www.mortality.org (Accessed on 15th September 2023).

Islam, M.S., Tareque, M.I., Mondal, M.N., Fazle Rabbi, A.M., Khan, H.T. & Begum, S. 2017. Urban-rural differences in disability-free life expectancy in Bangladesh using the 2010 HIES data. *PLoS ONE* 12(7): e0179987.

Jaber, J.J., Yaacob, N.A. & Alwadi, S. 2023. Hybrid Lee-Carter model with adaptive network of fuzzy inference system and wavelet functions. *Sains Malaysiana* 52(3): 1011-1021.

Johansen, S. 1988. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12(2-3): 231-254.

Lee, R.D. & Carter, L.R. 1992. Modeling and forecasting US mortality. *Journal of the American Statistical Association* 87(419): 659-671.

Ling, O., Leh, H., Musthafa, S.N.A.M. & Rasam, A.R.A. 2014. Urban environmental health: Respiratory infection and urban factors in urban growth corridor of Petaling Jaya, Shah Alam and Klang, Malaysia. *Sains Malaysiana* 43(9): 1405-1414.

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

Liu, X., Xu, Y., Engel, B.A., Sun, S., Zhao, X., Wu, P. & Wang, Y. 2021. The impact of urbanization and aging on food security in developing countries: The view from Northwest China. *Journal of Cleaner Production* 292: 126067.

Ma, Q. & Boonen, T.J. 2023. Longevity risk modeling with the Consumer Price Index. *North American Actuarial Journal* 28(3): 593-610.

Masseran, N., Farouk, S.N.A., Nur-Firyal, R. & Mansor, M.M. 2024. Latent hidden factor model for Malaysian consumer price index. *Sains Malaysiana* 53(4): 969-982.

Niu, G. & Melenberg, B. 2014. Trends in mortality decrease and economic growth. *Demography* 51(5): 1755-1773.

Oleson, K.W., Monaghan, A., Wilhelmi, O., Barlage, M., Brunsell, N., Feddema, M., Hu, L. & Steinhoff, D.F. 2013. Interactions between urbanization, heat stress, and climate change. *Climatic Change* 129: 525-541.

Phillips, P.C. & Perron, P. 1988. Testing for a unit root in time series regression. *Biometrika* 75(2): 335-346.

Plat, R. 2009. On stochastic mortality modeling. *Insurance: Mathematics and Economics* 45(3): 393-404.

Sajid, M.R., Muhammad, N., Zakaria, R. & Bukhari, S.A. 2020. Modifiable risk factors and overall cardiovascular mortality: Moderation of urbanization. *Journal of Public Health Research* 9(1893): 410-416.

Seklecka, M., Pantelous, A.A. & O'Hare, C. 2017. Mortality effects of temperature changes in the United Kingdom. *Journal of Forecasting* 36(7): 824-841.

Shang, H.L. 2012. Point and interval forecasts of age-specific life expectancies: A model averaging approach. *Demographic Research* 27: 593-644.

United Nations, Department of Economic and Social Affairs, Population Division. 2019a. *World Population Ageing 2019*. New York: United Nations.

United Nations, Department of Economic and Social Affairs, Population Division. 2019b. *World Urbanization Prospects 2018: Highlights*. New York: United Nations.

Verbeke, J. & Cools, R. 1995. The Newton-Raphson method. *International Journal of Mathematical Education in Science and Technology* 26(2): 177-193.

Walter, B.S. & DeWitte, S.N. 2017. Urban and rural mortality and survival in Medieval England. *Annals of Human Biology* 44(4): 338-348.

Wicke, F.S., Ernst, M., Otten, D., Werner, A., Dreier, M., Brähler, E., Tibubos, A.N., Reiner, I., Michal, M., Wiltink, J., Münzel, T., Lackner, K.J., Pfeiffer, N., König, J., Wild, P.S. & Beutel, M.E. 2022. The association of depression and all-cause mortality: Explanatory factors and the influence of gender. *Journal of Affective Disorders* 303: 315-322.

World Bank. 2023. Urban population (% of total population). (n.d.). https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS (Accessed on 19th September 2023).

*Corresponding author; email: itaramli@ukm.edu.my