

Article

Hybrid of the Lee-Carter Model with Maximum Overlap Discrete Wavelet Transform Filters in Forecasting Mortality Rates

Nurul Aityqah Yaacob ^{1,2}, Jamil J. Jaber ³, Dharini Pathmanathan ^{1,*}, Sadam Alwadi ³ and Ibrahim Mohamed ¹

¹ Faculty of Science, Institute of Mathematical Sciences, University Malaya, Kuala Lumpur 50603, Malaysia; nurulaityqah@gmail.com (N.A.Y.); imohamed@um.edu.my (I.M.)

² Faculty of Computer and Mathematical Sciences, Universiti Teknologi Mara Cawangan Negeri Sembilan, Kuala Pilah 72000, Malaysia

³ Department of Finance, Faculty of Business, The University of Jordan/Aqaba branch, Aqaba 77110, Jordan; jameljaber2011@hotmail.com (J.J.J.); sadam_alwadi@yahoo.co.uk (S.A.)

* Correspondence: dharini@um.edu.my

Abstract: This study implements various, maximum overlap, discrete wavelet transform filters to model and forecast the time-dependent mortality index of the Lee-Carter model. The choice of appropriate wavelet filters is essential in effectively capturing the dynamics in a period. This cannot be accomplished by using the ARIMA model alone. In this paper, the ARIMA model is enhanced with the integration of various maximal overlap discrete wavelet transform filters such as the least asymmetric, best-localized, and Coiflet filters. These models are then applied to the mortality data of Australia, England, France, Japan, and USA. The accuracy of the projecting log of death rates of the MODWT-ARIMA model with the aforementioned wavelet filters are assessed using mean absolute error, mean absolute percentage error, and mean absolute scaled error. The MODWT-ARIMA (5,1,0) model with the BL14 filter gives the best fit to the log of death rates data for males, females, and total population, for all five countries studied. Implementing the MODWT leads towards improvement in the performance of the standard framework of the LC model in forecasting mortality rates.

Keywords: MODWT; DWT; BL14; Coiflet; least asymmetric; wavelet



Citation: Yaacob, N.A.; Jaber, J.J.; Pathmanathan, D.; Alwadi, S.; Mohamed, I. Hybrid of the Lee-Carter Model with Maximum Overlap Discrete Wavelet Transform Filters in Forecasting Mortality Rates.

Mathematics **2021**, *9*, 2295. <https://doi.org/10.3390/math9182295>

Academic Editors: Eric Ulm and Budhi Surya

Received: 11 August 2021

Accepted: 14 September 2021

Published: 17 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Mortality studies are essential in understanding the demographic structure and indicating the health status of a population. The analysis of mortality and its historical trends enables a country to comprehend its population dynamics and serves as a foundation for formulating economic and social policies [1]. Actuaries used mortality forecasts to project cash flows and evaluate premiums as well as reserves in life insurance and pension plans. The Lee-Carter (LC) model [2] significantly contributed to the development of various extensions. A popular extension of this model is by Hyndman and Ullah [3], which used a functional data paradigm combined with nonparametric smoothing (penalized regression splines). Currie [4] extended the LC model to a generalized linear model (GLM) framework where the LC model and its extensions were fitted following the GLM framework in the Poisson and binomial settings. Neves et al. [5] considered five probability models (Poisson, binomial, negative binomial, Gaussian, and beta) based on the generalized autoregressive score (GAS) model to estimate the LC parameters and forecast mortality rates.

Apart from the extensions discussed, the original LC model is still widely used in mortality forecasting in many countries due to its simplicity and robustness. The LC model consists of three main parameters. The first and second parameters are age-specific parameters representing every age group while the third parameter, $k(t)$ defines the time-varying effect such that the tendency of all age-specific central death rates has the same pattern of stochastic evolution over time [6]. In the original LC model, the ARIMA

model is used to forecast $k(t)$. However, according to Nigri et al. [7], the standard ARIMA approach has limited ability to recognize unknown and unidentified patterns in future mortality trends over time. Hainaut and Denuit [8] reported that wavelets are great tools for investigating mortality trends across time with numerical illustrations. This motivated the integration of wavelets to enhance the performance of the ARIMA model in forecasting mortality rates.

Wavelets are commonly used in time series analysis, especially in signal processing, engineering, economics, and finance (see Percival and Walden [9] for an excellent review on the use of wavelets in time series). Wavelets decompose the original univariate time series into a group of time series (detail and smooth coefficients) that have an explicit hierarchical structure [10]. Wavelet-based methods have significant advantages in terms of denoising and are robust to outliers [11–13]. To date, research related to the application of wavelets in mortality is limited and patchily available. Morillas et al. [14] pioneered the use of wavelets and piecewise polynomial harmonic interpolation to develop a two-stage method for grading mortality rates and compared it to kernel grading. The wavelet technique can improve smoothness, fit, and oscillations more effectively than the conventional technique. Hainaut and Denuit [8] demonstrated that only a small number of wavelets are required to reconstruct all the mortality curves in the Belgian population from 1965 to 2015. Wavelet coefficients display clear trends in the Belgian population and are therefore straightforward to forecast.

This article takes cognizance of the advantages of the maximal overlap discrete wavelet transform (MODWT) and the ARIMA model to overcome the limitations of the ARIMA process in forecasting the future evolution of the $k(t)$ parameter. Wavelet transforms (WT) can be classified as continuous wavelet transforms (CWT) and discrete wavelet transforms (DWT). The MODWT is a modified version of the DWT which avoids the subsampling process, leading to a higher level of information in the resulting wavelet and scaling coefficients. The MODWT was chosen in this study because it can retain down-sampled values at each level of the decomposition and is well defined for all sample sizes [15]. So far, no work has been done to study the performance of MODWT wavelet filters such as the least asymmetric (LA8), best-localized (BL14), and Coiflet (C6) wavelet filters, for modelling and improving the prediction accuracy of mortality trends in the LC model. Hence, it is of interest to undertake this task and investigate the performance of the MODWT-based LC model. In this study, five countries with data on the Human Mortality Database (HMD) for years 1950 to 2016, involving ages 0 to 90+ were considered to verify the effectiveness of the MODWT-based LC model.

This paper is organized as follows: Section 2 presents the methodologies used in constructing the LC model and the MODWT-based LC model, Section 3 illustrates and compares the performances of the LC model and its wavelet counterparts for five countries (Australia, England, France, Japan, and USA), and Section 4 concludes.

2. Methodology

2.1. The Lee-Carter (LC) Model

The Lee-Carter [2] model is as follows:

$$\ln(m(x, t)) = a(x) + b(x)k(t) + \varepsilon(x, t), \quad x = 1, \dots, \omega \quad (1)$$

where $m(x, t)$ is the central death rate of age x at time t , ω is the beginning of the last age interval. Here, $a(x)$ describes the average shape of the age profile, $b(x)$ describes the pattern of deviations from this age profile when $k(t)$ the mortality index at time t varies. $\varepsilon(x, t)$ reflects the age-specific historical influences which are not fully captured by the model which is independent and identically distributed and follows the $N(0, \sigma^2)$ distribution. The parameters in Equation (1) are estimated using a two-stage method by imposing the following restrictions:

$$\sum_t k(t) = 0 \text{ and } \sum_x b(x) = 1, \quad (2)$$

to distinguish a unique solution for the system of equations of the model. The singular value decomposition approach was applied to the matrix of centered age profiles, $\ln(m_{x,t}) - a(x)$, which allows a first estimation of parameters $b(x)$ and $k(t)$. The model in (1) is fitted to the crude mortality rates, $\hat{m}(x, t) = \frac{D(x,t)}{E(x,t)}$ where $D(x, t) > 0$ denotes the number of deaths of age x at time t , and $E(x, t)$ is the matching central exposure of age x at time t . Once $b(x)$ and $k(t)$ are estimated by satisfying (2), a second stage estimate of $k(t)$ is found to ensure that the actual total deaths are identical to the total expected deaths for each t , as a basis for comparing actual and expected deaths. Hence, the parameter estimates satisfy

$$\sum_{x=x_1}^{\omega} D(x, t) = \sum_{x=x_1}^{\omega} E(x, t) \exp(\hat{a}(x) + \hat{b}(x) \hat{k}(t)) \forall t. \tag{3}$$

This adjustment gives more weight to high rates, thus roughly counterbalancing the effect of using a log transformation of the mortality rates. To forecast mortality rates, an appropriate time series was fitted to $k(t)$ using future extrapolation values, i.e., $k(t + n)$. Subsequently, the forecasted mortality rate would be

$$\ln(m(x, t)) = a(x) + b(x)k(t + n). \tag{4}$$

For this method, $a(x)$ and $b(x)$ are fixed. The adjusted $\hat{k}(t)$ is then extrapolated using ARIMA models. Lee and Carter [2] used a random walk with drift model, which can be expressed as

$$k(t) = k(t - 1) + d + e(t), \tag{5}$$

where d is known as the drift parameter and measures the average change according to time t in the series, and $e(t)$ is an uncorrelated error.

2.2. The MODWT-Based LC Model

Wavelets are based on Fourier transform which show any function as the sum of sine and cosine functions. WT is a function of time t that obeys the basic rule known as the admissibility condition [16]:

$$C_{\varphi} = \int_{-\infty}^{\infty} \frac{|\varphi(f)|^2}{f} df < \infty \tag{6}$$

where $\varphi(f)$ is the Fourier transform and a function of the frequency f , of $\varphi(t)$. A father wavelet generates the smooth and low-frequency parts of a signal while a mother wavelet generates the detailed and high-frequency components. The following equations represent the father and mother wavelets, respectively, where $j = 1, 2, 3, \dots, J$ in a J -level wavelet decomposition:

$$\begin{aligned} \phi_{j, k} &= 2^{(\frac{-j}{2})} \phi\left(t - \frac{2^j k}{2^j}\right), \\ \varphi_{j, k} &= 2^{(\frac{-j}{2})} \varphi\left(t - \frac{2^j k}{2^j}\right), \end{aligned} \tag{7}$$

where J denotes the maximum scale sustainable by the number of data points. The father and mother wavelets satisfy:

$$\begin{aligned} \int \phi(t) dt &= 1, \\ \int \varphi(t) dt &= 0. \end{aligned} \tag{8}$$

In any time series data, a function which is an input represented by wavelet transforms can be built as a sequence of projections onto father and mother wavelets indexed by $\{k\} = 2j$ where $k = \{0, 1, 2, \dots\}$, and $\{S\} = 2j$ where $\{j = 1, 2, 3, \dots, J\}$.

The analysis of real discretely sampled data requires creating a lattice for making calculations. Mathematically, it is convenient to use a dyadic expansion as shown in Equation (8). The expansion coefficients are given by the projections:

$$\begin{aligned} S_{j,k} &= \int \phi_{j,k} f(t) dt, \\ d_{j,k} &= \int \varphi_{j,k} f(t) dt. \end{aligned} \tag{9}$$

The wavelet approximation coefficients to $f(t)$, which leads to $k(t)$ in the wavelet framework of (1) is defined by:

$$k(t) = \sum S_{j,k} \phi_{j,k}(t) + \sum d_{j,k} \varphi_{j,k}(t) + \sum d_{j-1,k} \varphi_{j-1,k}(t) + \dots + \sum d_{1,k} \varphi_{1,k}(t), \tag{10}$$

where

$$\begin{aligned} S_j(t) &= \sum S_{j,k} \phi_{j,k}(t), \\ D_j(t) &= \sum d_{j,k} \varphi_{j,k}(t). \end{aligned} \tag{11}$$

WT is used to calculate the wavelet approximation coefficient in (10), for a discrete signal where $S_j(t)$ and $D_j(t)$ introduce the smooth and detailed coefficients, respectively. The smooth coefficients emphasize on the most critical features of the data, and the detailed coefficients detect the main features in the data [9,16].

An orthonormal DWT matrix $k(t)$ can be constructed based on any filter satisfying the properties of a wavelet filter in (8), namely summation to zero and orthonormality [9]. A Daubechies' wavelet filter of even width L has a squared gain function:

$$H^{(S)}(f) = S^{\frac{L}{2}}(f) A_L(f) \tag{12}$$

where $S(f) = 4\sin^2(\pi f)$ defines the squared gain function for the difference filter $\{1, -1\}$ and

$$A_L(f) = \frac{1}{2^{L-1}} \sum_{l=0}^{\frac{L}{2}-1} \binom{\frac{L}{2}-1+l}{l} \cos^{2l}(\pi f) \tag{13}$$

which establishes the squared function of a low-pass filter. The scaling filter $\{g_l\}$ that agrees to the Daubechies wavelet filter has a squared gain function given by

$$G^{(S)}(f) = H^{(S)}\left(\frac{1}{2} - f\right) = 2\cos^L(\pi f) \sum_{l=0}^{\frac{L}{2}-1} \binom{\frac{L}{2}-1+l}{l} \sin^{2l}(\pi f). \tag{14}$$

As L increases, some additional criteria to select a unique wavelet filter or a unique scaling filter may be imposed. For the Daubechies filter, let the scaling filter $\{g_l^{(ep)}\}$ with squared gain $G^S(\cdot)$ such that

$$\sum_{l=0}^m g_l^2 \leq \sum_{l=0}^m [g_l^{(ep)}]^2 \quad \text{for } m = 0, \dots, L-1 \tag{15}$$

where $\{g_l\}$ is any other filter with squared gain $G^{(S)}(\cdot)$ and $g_1^{(ep)}$ denotes the extremal phase (ep) scaling filter.

For a least asymmetric filter, the scaling filter whose transfer function $G(f) = [G^{(S)}(f)]^{1/2} e^{i\theta^{(G)}(f)}$, has a phase function $\theta^{(G)}(\cdot)$. This is as close as possible to that of a linear phase filter. The benefit of the least asymmetric filters is that the \bar{v} of v value minimizes $G(f)$ to match the scaling and wavelet coefficients in such a way that they can be viewed as approximately the output of zero-phase filters [9,16].

The approximate zero phase property is significant as it allows us to link the DWT coefficients meaningfully to different events in the original time series. The least asymmetric wavelet with an excellent general-purpose has a width which is 8 (LA8). The LA8 wavelet

filter strikes a balance between providing smooth approximations with few artifacts and minimal edge effects at data boundaries [9,16].

Let $k(t)$ be associated with the actual time $t_0 + t\Delta t$. Then, the phase properties of the least asymmetric filters dictate associating the wavelet coefficient $W_{j,t}$ with actual time

$$t_0 + \left(2^j(t + 1) - 1 - \left|v_j^{(H)}\right| \bmod N\right)\Delta t, \quad t = 0, \dots, N_j - 1, \tag{16}$$

where

$$\left|v_j^{(H)}\right| = \frac{L_j}{2} + \frac{L}{2} + v - 1, \tag{17}$$

and

$$v = \begin{cases} \frac{-L}{2} + 1, & \text{if } L = 8, 12, 16 \text{ or } 20 \text{ (i.e., } L/2 \text{ is even)} \\ \frac{-L}{2}, & \text{if } L = 10 \text{ or } 18 \\ \frac{-L}{2} + 2, & \text{if } L = 14. \end{cases} \tag{18}$$

For the scaling coefficient $V_{j,t}$, a similar expression is obtained by replacing $\left|v_j^{(H)}\right|$ with

$$\left|v_j^{(G)}\right| = \frac{L_j - 1}{L - 1} |v|, \tag{19}$$

The Coiflet wavelet filters are alternatives to the Daubechies filters that provide better approximations to zero phase filters compared to their least asymmetric counterparts [17]. However, they have a less desirable filter form and less embedded differentiating operations for a specific filter width. Coiflet wavelets and scaling coefficients with actual times use (19) by substituting $v = \frac{-2L}{3} + 1$.

Doroslovacki [18] proposed a best-localized squared gain factorization for the Daubechies scaling filter. This filter refines the least asymmetric idea by using a new linear-phase deviation measure that penalizes departures at low frequencies more severely than those at high frequencies [9]. The best-localized (BL) wavelet filter gives improved results for the scaling function by minimizing the time-localization measure [18]. A proper selection of the wavelet-generating-filter transfer function is crucial to make the scaling functions and wavelets in the binary orthonormal Daubechies form more symmetrical. The Coiflet (C6), least asymmetric (LA8), and best-localized (BL14) wavelet filters will be investigated to find the best approximation for $k(t)$.

3. Results and Discussion

The data for this study is obtained from the Human Mortality Database (HMD). Five developed countries were selected such as Australia, England, Japan, France, and USA by considering sex-specific, as well as total population for analysis. The data for these countries are based on single years of age. The data for older ages (age 90 and above) were grouped to avoid problems associated with erratic rates at these ages. The LC and MODWT-LC models were fitted to log death rates from 1950 to 2016. To fit these models, the data for the five countries were divided into training and test sets. The training set consists of the observed log death rates occurring until 2005, and the test set is chosen from 2006 onwards. Tables 1–3 provide the forecast accuracy based on mean absolute error (MAE), mean absolute percentage error (MAPE), and mean absolute scale error (MASE) for male, female, and total populations for the five countries studied.

The accuracy of two models representing $k(t)$ such as the default ARIMA(0,1,0) and MODWT-ARIMA(5,1,0) by integrating the LA8, BL14, and C6 wavelet filters were evaluated. In selecting the suitable form of ARIMA (p, d, q) for the MODWT-LC model, various values were fitted to p (the number of autoregressive terms), d (the number of non-seasonal differences required for stationarity), and q (the order of moving average). The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test Kwiatkowski et al. [19] was used to test for stationarity and the p -values were significantly lesser than 0.05 for $d = 1$. This indicates that one non-seasonal difference is required to achieve stationarity for the data of

the five countries. Based on the smallest error obtained using MAE, MAPE, and MASE, $p = 5, d = 1$ and $q = 0$ form the best combination in improving $k(t)$ to fit the log of death rates for the five countries investigated. The structure of the data also suggested that no moving average term i.e., q is required.

Table 1. Accuracy measures based on MAE, MAPE, and MASE for total population by country (smallest values are bolded).

Country	MAE				MAPE				MASE			
	LA8	BL14	C6	ARIMA (0,1,0)	LA8	BL14	C6	ARIMA (0,1,0)	LA8	BL14	C6	ARIMA (0,1,0)
Australia	1.9840	1.7816	2.4849	2.3469	8.9306	8.9978	12.4314	19.4595	0.4880	0.4229	0.6129	0.9825
England	1.5697	1.4881	2.2238	1.9533	22.1293	14.8784	15.8047	21.4726	0.4741	0.4251	0.6678	0.9826
France	1.8971	1.8328	2.5823	2.2057	46.7899	13.6233	19.3140	17.7932	0.4749	0.4342	0.6412	0.9826
Japan	2.8555	2.6186	3.9215	3.2803	21.3071	7.2387	25.4780	18.6707	0.4530	0.3940	0.6146	0.9828
USA	1.2028	1.1426	1.4942	1.4685	7.8074	6.0259	11.1851	18.0986	0.4511	0.4114	0.5592	0.9826

Table 2. Accuracy measures based on MAE, MAPE, and MASE for the male population by country (smallest values are bolded).

Country	MAE				MAPE				MASE			
	LA8	BL14	C6	ARIMA (0,1,0)	LA8	BL14	C6	ARIMA (0,1,0)	LA8	BL14	C6	ARIMA (0,1,0)
Australia	1.9442	1.7009	2.5810	2.3013	11.2313	8.7570	21.5385	29.1528	0.4996	0.4180	0.6659	0.9825
England	1.4885	1.3998	2.1093	1.8621	8.8632	8.0906	16.1702	23.6105	0.4698	0.4195	0.6629	0.9826
France	1.7029	1.6557	2.3716	2.0757	8.3331	11.6279	13.2880	20.6962	0.4690	0.4304	0.6456	0.9826
Japan	2.5935	2.3471	3.5118	2.9440	17.2278	7.5573	17.9236	18.1469	0.4626	0.3960	0.6186	0.9828
USA	1.2260	1.1177	1.4552	1.4022	9.8111	6.7211	11.9714	22.2012	0.4686	0.4099	0.5564	0.9826

Table 3. Accuracy measures based on MAE, MAPE, and MASE for female population by country (smallest values are bolded).

Country	MAE				MAPE				MASE			
	LA8	BL14	C6	ARIMA (0,1,0)	LA8	BL14	C6	ARIMA (0,1,0)	LA8	BL14	C6	ARIMA (0,1,0)
Australia	2.1886	1.9695	2.6622	2.6321	16.3362	9.4682	17.7354	19.9199	0.4992	0.4393	0.6092	0.9826
England	1.7878	1.6691	2.4163	2.2765	11.6679	8.6624	13.7441	20.9444	0.4881	0.4283	0.6537	0.9826
France	2.3191	2.1789	2.8169	2.5817	9.6123	8.0206	12.4539	25.6141	0.4932	0.4409	0.5942	0.9826
Japan	3.2771	3.0345	4.6137	3.7966	22.1074	7.3877	21.8618	27.0103	0.4473	0.3912	0.6227	0.9828
USA	1.3180	1.2149	1.8348	1.6363	9.3555	6.7910	22.2544	20.9945	0.4623	0.4068	0.6390	0.9826

Note: for Tables 1–3, LA8, BL14, and C6 represent the filters used in the MODWT ARIMA (5,1,0) model representing $k(t)$.

Tables 1–3 show that the MODWT-ARIMA (5,1,0) model gives the smallest MAE, MAPE, and MASE values for the total, male and female populations of the countries studied. It can be seen from Tables 1–3 that the BL14 filter consistently outperforms its counterparts. In the LC model, $k(t)$ plays a significant role in capturing the mortality trend over time. The application of the BL14 MODWT-ARIMA) to model $k(t)$ for the log death rates data decomposes them into various resolution levels that reveal their essential structure and generates detailed coefficients at every level. The MODWT with filters were used to capture the pattern of the $k(t)$ series over time. The MODWT-based decomposition (Alenezy et al. [20] and Cornish et al. [15]) is an effective approach for revealing variations, magnitudes, and phases of the data. In capturing the pattern of the $k(t)$ series, the BL14 filter has markedly better phase properties than the LA8 and C6 filters. The best localized filter penalizes deviations at low frequencies more heavily than those at high frequencies [18].

For the total population of Australia (Table 1) and the male population of France (Table 2), the LA8 filter appeared to have a slight edge when using the MAPE measure. Simulation is not required to prove the effectiveness of the BL14 filter because the proper

choice of ARIMA model equipped with the MOWDT-BL14 filter is versatile in capturing extreme values.

Figures 1–3 show values of $k(t)$ from 1950 to 2005 and forecasted values of $k(t)$ from 2006 to 2016 for the Japanese total, male, and female populations. The forecasted values using BL14 MODWT-ARIMA (5,1,0) is closer to the observed $k(t)$ compared to its ARIMA(0,1,0) counterpart. On the other hand, it is apparent from Figures 4–6 that the BL14 MODWT-ARIMA(5,1,0) model gives a good fit for Japanese log of death rates in 2016 compared to ARIMA (0,1,0) and MODWT-ARIMA (0,1,0).

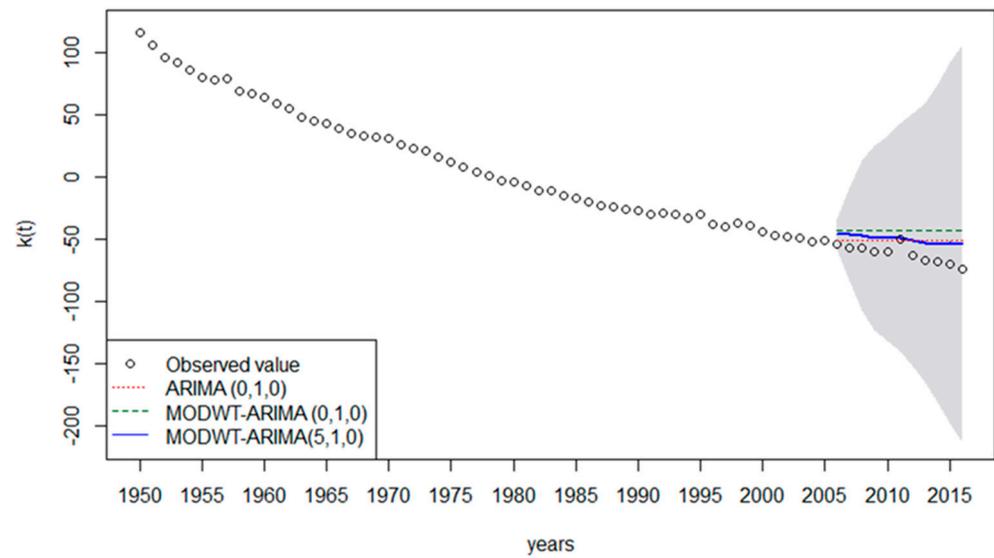


Figure 1. Actual and forecasted values of $k(t)$ for Japanese (total population).

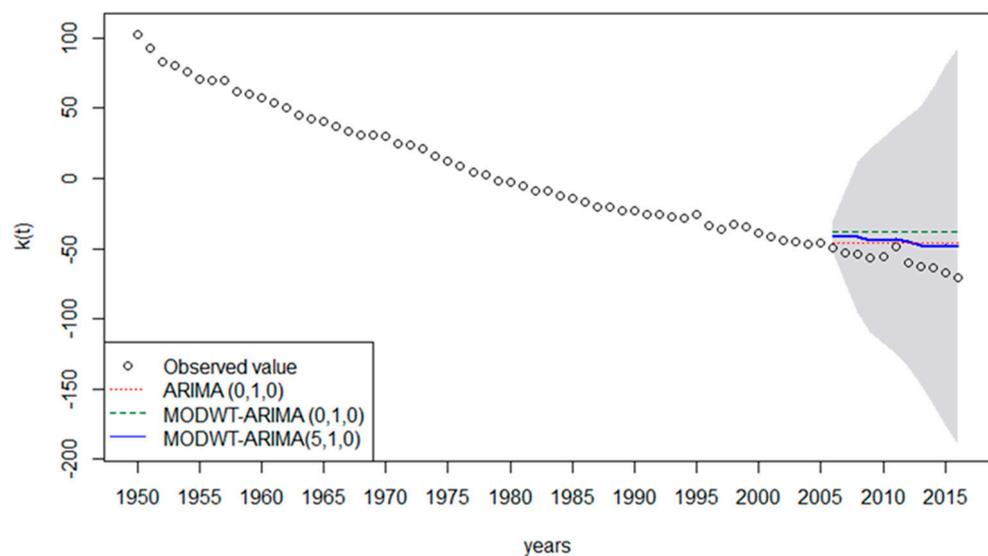


Figure 2. Actual and forecasted values of $k(t)$ for Japanese (male population).

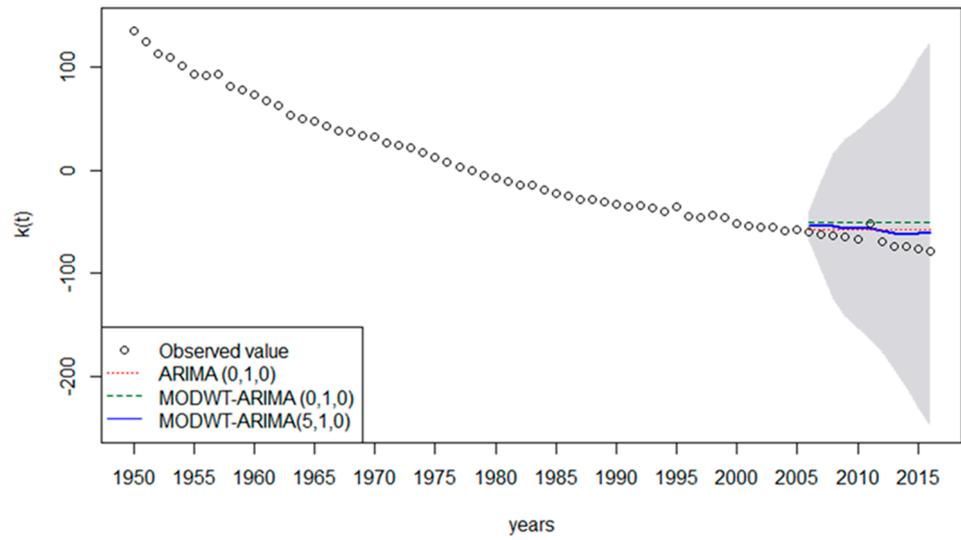


Figure 3. Actual and forecasted values of $k(t)$ for Japanese (female population).

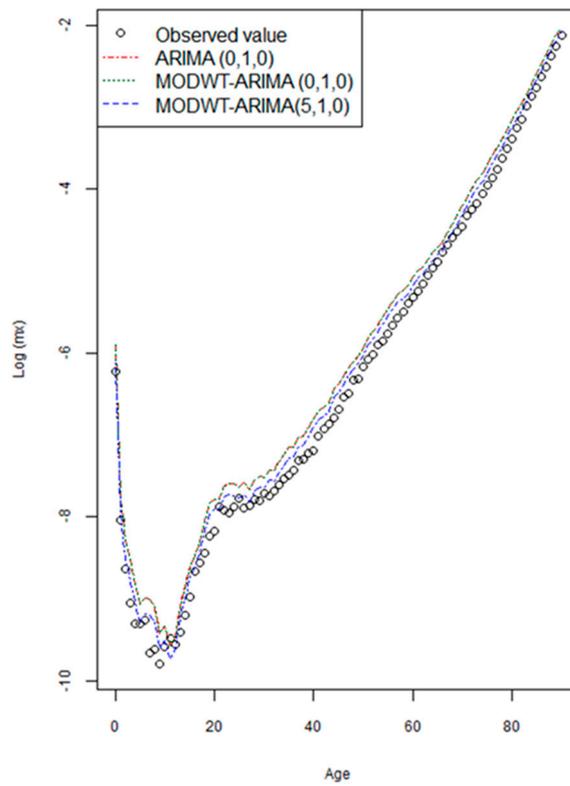


Figure 4. Observed and predicted death rates in Japan on log scale (year 2016) for total population using ARIMA(0,1,0), MODWT-ARIMA(0,1,0) and MODWT-ARIMA(5,1,0) with BL14 filter.

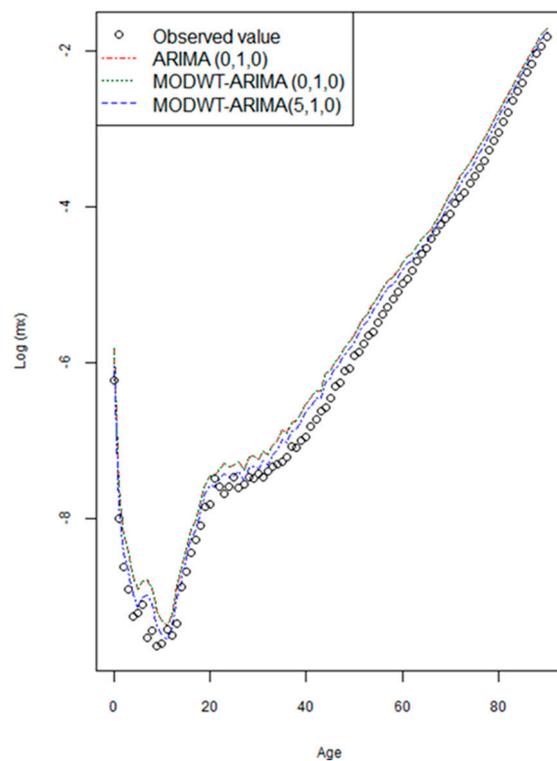


Figure 5. Observed and predicted death rates in Japan on log scale (year 2016) for: male population using ARIMA(0,1,0), MODWT-ARIMA(0,1,0), and MODWT-ARIMA(5,1,0) with BL14 filter.

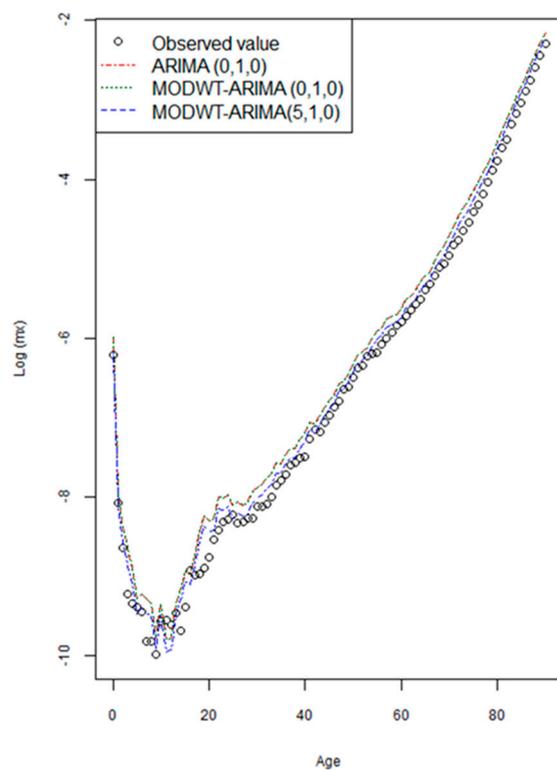


Figure 6. Observed and predicted death rates in Japan on log scale (year 2016) for female population using ARIMA(0,1,0), MODWT-ARIMA(0,1,0), and MODWT-ARIMA(5,1,0) with BL14 filter.

4. Conclusions

This study considered the hybrid of the MODWT with the Lee-Carter model to improve the forecast accuracy of the time-dependent mortality index, $k(t)$. The MODWT is more advantageous than the DWT for mortality modelling. The MODWT can handle any sample size. In each level, there are detail (wavelet) and smooth (scale) coefficients which are associated with zero phase filters. The MODWT also produces a more asymptotically efficient wavelet variance estimator than the DWT. This clearly shows that the MODWT has an edge over its counterparts. Overall, the results of this study show that the hybrid of the MODWT using BL14 filters with the Lee-Carter [2] model entails significant improvement in forecasting accuracy. Our findings show that the minimization of a time-localization measure offered by BL14 wavelet filters can improve results in forecasting $k(t)$. The MODWT-ARIMA-(5,1,0) with the BL14 filter generally shows excellent proximity to the actual log-mortality rates for the five countries studied. For future studies, the extension of wavelet-based neural network models with the LC model is of interest.

Author Contributions: Conceptualization, D.P.; data curation, N.A.Y.; formal analysis, N.A.Y., J.J.J.; methodology, N.A.Y., D.P. and J.J.J.; project administration, D.P.; software, N.A.Y. and J.J.J.; supervision, D.P. and I.M.; validation, D.P. and S.A.; writing—original draft, N.A.Y., D.P.; writing—review and editing, D.P., S.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the University of Malaya, Faculty Research Grant [GPF028B-2018].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Human Mortality Database, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany), 2020; dataset available at www.mortality.org, accessed on 7 April 2020.

Acknowledgments: The authors thank the referees for many valuable and helpful suggestions to improve the presentation of this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Díaz-Rojo, G.; Debón, A.; Mosquera, J. Multivariate Control Chart and Lee-Carter Models to Study Mortality Changes. *Mathematics* **2020**, *8*, 2093. [[CrossRef](#)]
- Lee, R.D.; Carter, L.R. Modeling and forecasting United-States mortality. *J. Am. Stat. Assoc.* **1992**, *87*, 659–671.
- Hyndman, R.J.; Ullah, M.S. Robust forecasting of mortality and fertility rates: A functional data approach. *Comput. Stat. Data Anal.* **2007**, *51*, 4942–4956. [[CrossRef](#)]
- Currie, I.D. On fitting generalized linear and non-linear models of mortality. *Scand. Actuar. J.* **2016**, *4*, 356–383. [[CrossRef](#)]
- Neves, C.; Fernandes, C.; Hoeltgebaum, H. Five different distributions for the Lee-Carter model of mortality forecasting: A comparison using GAS models. *Insur. Math. Econ.* **2017**, *75*, 48–57. [[CrossRef](#)]
- Hong, W.H.; Yap, J.H.; Selvachandran, G.; Thong, P.H.; Son, L.H. Forecasting mortality rates using hybrid Lee-Carter model, artificial neural network and random forest. *Complex Intell. Syst.* **2021**, *7*, 163–189. [[CrossRef](#)]
- Nigri, A.; Levantesi, S.; Marino, M.; Scognamiglio, S.; Perla, F. A deep learning integrated Lee-Carter model. *Risks* **2019**, *7*, 33. [[CrossRef](#)]
- Hainaut, D.; Denuit, M. Wavelet-based feature extraction for mortality projection. *ASTIN Bull. J. IAA* **2020**, *50*, 675–707. [[CrossRef](#)]
- Percival, D.B.; Walden, A.T. *Wavelet Methods for Time Series Analysis*; Cambridge University Press: Cambridge, UK, 2000; Volume 4.
- Zhang, K.; Gençay, R.; Yazgan, M.E. Application of wavelet decomposition in time-series forecasting. *Econ. Lett.* **2017**, *158*, 41–46. [[CrossRef](#)]
- Fadili, M.J.; Bullmore, E.T. A comparative evaluation of wavelet-based methods for hypothesis testing of brain activation maps. *NeuroImage* **2004**, *23*, 1112–1128. [[CrossRef](#)] [[PubMed](#)]
- Achard, S.; Salvador, R.; Whitcher, B.; Suckling, J.; Bullmore, E.D. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *J. Neurosci.* **2006**, *26*, 63–72. [[CrossRef](#)] [[PubMed](#)]
- Zhang, Z.; Telesford, Q.K.; Giusti, C.; Lim, K.O.; Bassett, D.S. Choosing wavelet methods, filters, and lengths for functional brain network construction. *PLoS ONE* **2016**, *11*, e0157243. [[CrossRef](#)] [[PubMed](#)]

14. Morillas, F.; Baeza, I.; Pavia, J.M. Risk of death: A two-step method using wavelets and piecewise harmonic interpolation. *Estat. Esp.* **2016**, *58*, 245–264.
15. Cornish, C.R.; Bretherton, C.S.; Percival, D.B. Maximal overlap wavelet statistical analysis with application to atmospheric turbulence. *Bound.-Layer Meteorol.* **2006**, *119*, 339–374. [[CrossRef](#)]
16. Mehra, M. *Wavelets Theory and Its Applications*; Springer: Singapore, 2018.
17. Daubechies, I. *Ten Lectures on Wavelets*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 1992.
18. Doroslovacki, M.L. On the least asymmetric wavelets. *IEEE Trans. Signal Process.* **1998**, *46*, 1125–1130. [[CrossRef](#)]
19. Kwiatkowski, D.; Phillips, P.C.; Schmidt, P.; Shin, Y. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *J. Econom.* **1992**, *54*, 159–178. [[CrossRef](#)]
20. Alenezzy, A.H.; Ismail, M.T.; Wadi, S.A.; Tahir, M.; Hamadneh, N.N.; Jaber, J.J.; Khan, W.A. Forecasting Stock Market Volatility Using Hybrid of Adaptive Network of Fuzzy Inference System and Wavelet Functions. *J. Math.* **2021**, *2021*, 9954341. [[CrossRef](#)]