Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 10, Number 2, 2026
DOI: 10.30829/zero.v10i2.31283
Page: 743-755

Comparative Analysis of Classical and Time-Varying Lee–Carter Models for Mortality Forecasting in Indonesia and Japan

[1]Syipa Arista Widya
Department of Mathematics, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

[2]Ismail Husein
Department of Mathematics, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

Article Info

Article history:

Accepted : 13 August 2026

Keywords:

Forecasting performance;
Lee-Carter model;
Mortality forecasting;
Mortality modeling;
Time-Varying Lee-Carter.

ABSTRACT

Reliable mortality forecasting is essential for demographic planning and actuarial decision-making. This study compared the forecasting performance of the Classical and Time-Varying Lee-Carter models using World Population Prospects (WPP) 2024 mortality data for Indonesia and Japan (1970–2023). Both models were estimated using 1970–2015 data and evaluated on 2016–2023 through out-of-sample validation based on MAE, RMSE, and MAPE. For Indonesia, the Time-Varying model yielded a slightly lower MAPE (7.262%) than the Classical model (8.086%), whereas the Classical model produced slightly lower forecasting errors for Japan. Overall, the differences between the models were modest. The findings suggest that, in this study, forecasting performance was associated more closely with population-specific mortality characteristics than with model complexity alone.

Corresponding Author:

Syipa Arista Widya,
Department of Mathematics
Universitas Islam Negeri Sumatera Utara
Email: syipa0703222050@uinsu.ac.id

1. INTRODUCTION

Mortality is a fundamental demographic indicator that provides a basis for population projections, actuarial analysis, and social security planning [1], [2]. Over recent decades, global mortality levels have generally declined alongside improvements in healthcare and socioeconomic conditions [3]. However, the COVID-19 pandemic disrupted this long-term pattern in 2020–2021, when mortality differed substantially from previous trends [2], [4]. Such changes highlight the importance of evaluating mortality forecasting models across different mortality conditions.

Accurate mortality forecasts are important for evidence-based decision-making in demography, actuarial science, and public health. Forecasting errors may affect actuarial assessments, pension funding, and healthcare resource allocation [5], [6]. Evaluating model performance is therefore important for understanding the suitability of mortality forecasting approaches for different population contexts.

The Lee-Carter model remains a widely used framework for mortality forecasting. It represents the logarithm of age-specific mortality rates using an age profile (a_x), age sensitivity (b_x), and a mortality index (k_t), conventionally estimated using Singular Value Decomposition (SVD) and projected using a Random Walk with Drift [7]. Previous studies have demonstrated its usefulness for populations with relatively stable mortality trends [8]. Recent applications include forecasting life expectancy in Medan City [9], comparing Random Walk with Drift and Simple Exponential Smoothing for mortality forecasting in Indonesia [10], and constructing life tables in different demographic contexts [11]. However, the assumption that mortality is time-invariant may become less suitable when mortality patterns change substantially over time [8], [12], [13]. This has motivated the development of extensions based on structural changes, time-varying factor loadings, and machine-learning approaches [14], [15], [16], [17].

One such extension models k_t a latent state within a state-space framework estimated using the Kalman Filter and Kalman Smoother [18]. This formulation provides a framework for representing changes in the latent mortality index over time [19], [20]. Similarly, kernel-smoothed Time-Varying Lee–Carter models have been investigated for mortality patterns that vary over time [21]. However, direct empirical comparisons between the Classical and Time-Varying Lee–Carter models under a consistent out-of-sample validation framework remain limited.

Model performance may also vary across demographic contexts. Previous studies have reported that the Classical Lee–Carter model can perform well under relatively stable mortality trends, while structural changes may affect its forecasting performance [22], [23], [24]. Global evidence also indicates differences in mortality trajectories across countries [3], providing a basis for comparing Indonesia and Japan. According to WPP 2024 estimates used in this study, Indonesia shows greater temporal variation in mortality over the study period, whereas Japan exhibits comparatively stable mortality patterns [4], [25], [26].

In Indonesia, previous studies have predominantly applied the Classical Lee–Carter model without an explicit training–testing split [10]. This study therefore compares the Classical and Time-Varying Lee–Carter models under a consistent out-of-sample validation framework. Both models use the same RWD forecasting procedure but differ in how they estimate the mortality index. The Classical model estimates the index through SVD, whereas the Time-Varying model uses a state-space framework with Kalman filtering and smoothing. Keeping the forecasting procedure consistent allows the comparison to focus on differences associated with the estimation framework.

This study uses WPP 2024 mortality estimates for Indonesia and Japan to compare the Classical and Time-Varying Lee–Carter models under a consistent out-of-sample framework. Forecasting performance is evaluated using MAE, RMSE, and MAPE, complemented by prediction interval analysis and annual MAPE during the testing period, including the COVID-19 years (2020–2021). The analysis examines whether the two estimation frameworks perform differently across populations with contrasting mortality dynamics while keeping the forecasting procedure consistent.

2. RESEARCH METHOD

The predictive performance of the Classical and Time-Varying Lee–Carter models was assessed using a quantitative comparative approach. Secondary mortality data were obtained from the United Nations Department of Economic and Social Affairs' World Population Prospects (WPP) 2024, which provides age-specific death counts and corresponding population exposures for Indonesia and Japan from 1970 to 2023. The analysis included age groups from 0 to 100+, with the open-ended age group at 100+ years and older retained as reported in WPP 2024. No additional age truncation or smoothing was applied. Estimates for the oldest age group should therefore be interpreted with caution, as mortality rates at very old ages may be associated with greater uncertainty.

Table 1. Selected age-specific mortality rates ($m_{x,t}$) derived from WPP 2024 death counts and population exposures.

Age	1970	1990	2010	2020	2023
			Indonesia		
0	0.12396	0.06621	0.02834	0.01926	0.01733
20	0.00403	0.00212	0.0017	0.0014	0.00133
40	0.00692	0.00397	0.0034	0.00323	0.00278
60	0.02384	0.01887	0.01766	0.01974	0.0155
80	0.14188	0.1281	0.11301	0.11662	0.09923
100+	0.50549	0.55985	0.54634	0.76274	0.5249
			Japan		
0	0.01377	0.00453	0.00233	0.00178	0.00162
20	0.00098	0.00056	0.00039	0.00031	0.00032
40	0.00242	0.00123	0.00106	0.00078	0.00075
60	0.01334	0.00811	0.00573	0.00447	0.00455
80	0.11013	0.06231	0.03949	0.03189	0.03121
100+	0.57073	0.4925	0.4154	0.39023	0.41895

Note: Only selected ages and calendar years are presented. The complete mortality matrices (ages 0–100+ years, 1970–2023) were used in model estimation and evaluation.

The age-specific mortality rate is calculated as,

$$m_{x,t} = \frac{D_{x,t}}{P_{x,t}} \tag{1}$$

where $m_{x,t}$ is the mortality rate at age x and year t, $D_{x,t}$ is the number of deaths, and $P_{x,t}$ is the corresponding population exposure [27]. Before model fitting, mortality rates were log-transformed, and the data were divided into a training set (1970–2015) and a testing set (2016–2023) for out-of-sample validation. All analyses were performed in R using the StMoMo, KFAS, and forecast packages.

2.1 Mortality Forecasting Models

The Classical and Time-Varying Lee–Carter models were compared using the same Random Walk with Drift (RWD) forecasting procedure. The Classical Lee–Carter model serves as the baseline, whereas the Time-Varying Lee–Carter model incorporates a state-space representation to allow the latent mortality index to evolve [8], [21]. The formulation of the Classical Lee–Carter model is

$$\ln\left(m_{x,t}\right) = a_x + b_x k_t + \varepsilon_{x,t} \tag{2}$$

where a_x represents the baseline logarithmic mortality level at age x, b_x measures the sensitivity of age-specific mortality to changes in the mortality index, k_t represents the mortality index over time, and $\varepsilon_{x,t}$ captures random deviations from the fitted model [7]. The model parameters were estimated using SVD with the StMoMo package, and the mortality index was forecast using RWD.

$$k_t = k_{t-1} + d + \varepsilon_t \tag{3}$$

where the random disturbance is represented by ε_t and the drift parameter is denoted by d.

To ensure parameter identifiability, the standard Lee–Carter normalization constraints were applied, where the age-specific sensitivity parameters satisfy $\sum_x b_x = 1$ and the mortality index satisfies $\sum_t k_t = 0$. These constraints provide a unique decomposition of the mortality surface and prevent arbitrary scaling between b_x and k_t [7].

The forecast mortality rates were obtained by substituting the projected mortality index derived from Eq. (3) into Eq. (2) [7].

$$\widehat{m}_{x,t} = \exp\left(a_x + b_x \hat{k}_t\right) \tag{4}$$

where $\hat{k}_t$ denotes the forecasted mortality index [7].

For the Time-Varying Lee–Carter model, the mortality index was represented as a latent state within a linear Gaussian state-space framework and estimated using the Kalman Filter and Kalman Smoother [21]. Following the general state-space formulation proposed by Durbin and Koopman, the observation and state equations are expressed as follows [28], [29]:

$$y_t = Z_t \alpha_t + \varepsilon_t \tag{5}$$

and

$$\alpha_t = T_t \alpha_{t-1} + R_t \eta_t \tag{6}$$

where y_t denotes the observation vector, α_t denotes the latent state vector, Z_t is the observation matrix, T_t is the transition matrix, and R_t is the state disturbance matrix. The observation disturbance ε_t and the state disturbance η_t are assumed to be mutually independent and normally distributed as.

$$\varepsilon_t \sim N(0, H),$$

and

$$\eta_t \sim N(0, Q),$$

where H and Q denote the covariance matrices of the observation and state disturbances, respectively. Under these assumptions, the Kalman Filter provides recursive estimates of the latent state, while the Kalman Smoother refines these estimates by incorporating information from the entire observation period [28], [29].

For the Lee–Carter model considered in this study, the observation variable is defined as [7].

$$y_t = \log\left(m_{x,t}\right) - a_x \quad (7)$$

where $m_{x,t}$ denotes the central mortality rate at age x and year t, and a_x represents the average log mortality at age x. The observation matrix was specified as $Z_t = b_x$, where b_x represents the age-specific sensitivity to changes in the mortality index. The latent state corresponds to the mortality index k_t, while the transition and state disturbance matrices were specified as $T_t = 1$ and $R_t = 1$, respectively.

The unknown parameters of the state-space model, including the observation and state disturbance variances, were estimated by maximum likelihood using the KFAS package in R. The latent mortality index (k_t) was first estimated using the Kalman Filter and subsequently refined using the Kalman Smoother to obtain smoothed state estimates [30]. In the Time-Varying Lee–Carter model, a_x and b_x were retained from the Classical Lee–Carter estimation obtained using SVD, while k_t was estimated as a latent state through the state-space framework.

2.2 Prediction Interval Analysis

To quantify the uncertainty associated with mortality forecasts, a 95% prediction interval was constructed for the forecasted mortality index (k_t). A prediction interval provides a range of plausible future values and quantifies the uncertainty associated with point forecasts [31].

Assuming approximately normally distributed forecast errors, the 95% prediction interval for forecast horizon h is given by

$$\hat{k}_{t+h} \pm z_{1-\alpha/2}\sqrt{\mathrm{Var}\left(\hat{k}_{t+h}\right)} \quad (8)$$

where $\hat{k}_{t+h}$ denotes the forecasted mortality index at forecast horizon h, $z_{1-\alpha/2}$ is the critical value of the standard normal distribution corresponding to the selected confidence level (1.96 for a 95% prediction interval), and $\mathrm{Var}\left(\hat{k}_{t+h}\right)$ denotes the variance of the forecast error [32].

Prediction intervals for the Classical Lee–Carter model were obtained using the forecasting procedure implemented in the StMoMo package, while those for the Time-Varying Lee–Carter model were generated using the forecast package. The resulting intervals represent uncertainty in the projected mortality index and do not include additional uncertainty from parameter estimation or Kalman filtering.

2.3 Forecast Accuracy Evaluation

Three commonly used accuracy metrics were used to assess the forecasting accuracy of the Classical and Time-Varying Lee–Carter models: MAE, RMSE, and MAPE [33], [34]. MAE measures the average absolute forecast error, RMSE gives greater weight to larger errors, and MAPE expresses the error relative to the observed mortality rate [35]. Lower values indicate better forecasting performance.

The metrics are defined as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |\, m_i - \hat{m}_i| \quad (9)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(m_i - \hat{m}_i)^2} \quad (10)$$

$$\mathrm{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{m_i - \hat{m}_i}{m_i}\right| \quad (11)$$

where m_i and $\hat{m}_i$ denote observed and forecasted mortality rates, respectively, and n is the total number of test observations. Additionally, annual MAPE metrics were calculated for each year across the testing horizon (2016–2023) to assess performance stability over time, particularly during the COVID-19 pandemic (2020–2021).

2.4 Diebold–Mariano Test

The Diebold–Mariano (DM) test was applied to the out-of-sample forecast errors from 2016–2023 to assess whether the forecasting accuracy of the Classical and Time-Varying Lee–Carter models differed significantly. Unlike MAE, RMSE, and MAPE, the DM test assesses the statistical significance of the difference in predictive accuracy between two competing models.

The null hypothesis (H_0) states that both models have equal predictive accuracy, whereas the alternative hypothesis (H_1) states that their predictive accuracies differ. A two-sided test with a significance level of 5% was employed [36].

$$DM = \frac{\bar{d}}{\sqrt{\widehat{Var}(\bar{d})}} \tag{12}$$

Where

$$d_t = L(e_{1,t}) - L(e_{2,t}) \tag{13}$$

where $e_{1,t}$ and $e_{2,t}$ represent the forecast errors of the Classical and Time-Varying Lee–Carter models, respectively. The function $L(\cdot)$ denotes the loss function, $\bar{d}$is the average loss differential between the two competing forecasting models, and $\widehat{\text{Var}}(\bar{d})$ represents the estimated variance of the mean loss differential [36].

In this study, the DM test was performed using all age-specific out-of-sample forecast errors for the testing period (2016–2023). The squared-error loss function ($power = 2$) and a one-step forecast horizon ($h = 1$) were adopted, consistent with the forecast evaluation procedure implemented in the R package *forecast*. A p-value less than 0.05 indicates that the difference in predictive accuracy between the two models is statistically significant [36].

3. RESULT AND ANALYSIS

3.1 Age-Specific Mortality Patterns

The logarithmic age-specific mortality rates for Indonesia and Japan from 1970 to 2023 are shown in Figure 1. The logarithmic scale facilitates comparison of mortality patterns across age groups and calendar years.

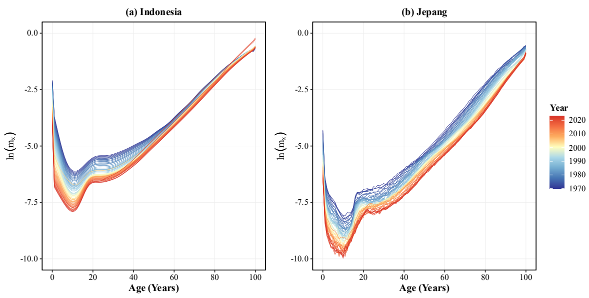

Figure 1. Logarithmic age-specific mortality rates in Indonesia and Japan, 1970–2023.

Both countries exhibited similar age-specific patterns, with mortality highest in infancy, declining through childhood, remaining relatively low during adolescence and early adulthood, and increasing with age thereafter. However, their temporal patterns differed. Indonesia showed greater variation across calendar years, whereas Japan exhibited relatively more stable mortality patterns over the observation period.

3.2 Estimated Lee–Carter Parameters

Building on the descriptive analysis, the Lee–Carter model parameters were estimated to characterize the age-specific mortality profile and its temporal variation. Figures 2–4 present the estimated baseline age effect (a_x), age-specific sensitivity parameter (b_x), and mortality index (k_t) for Indonesia and Japan.

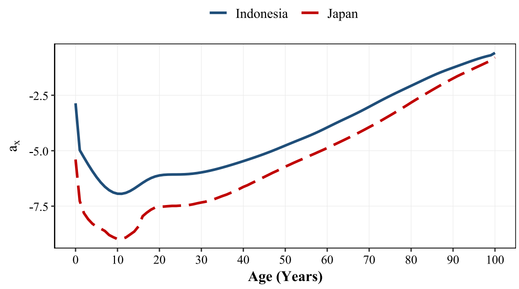

Figure 2. Estimated age-specific average log mortality (a_x) for Indonesia and Japan.

Figure 2 illustrates the estimated baseline age effect (a_x) for Indonesia and Japan. Both countries showed a similar age-dependent pattern, with relatively high estimates during infancy, followed by a sharp decline throughout childhood. The lowest estimates occurred at around 10 years of age, after which a_x increased with age. This pattern indicates lower estimated baseline log mortality during childhood and higher levels at older ages.

Japan generally had lower a_x estimates than Indonesia across most age groups. The difference became smaller at older ages, where the estimated values were more closely aligned. This pattern is consistent with the age-specific mortality profiles shown in Figure 1.

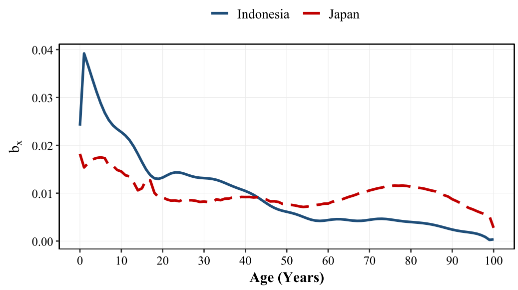

Figure 3. Estimated age-specific sensitivity parameter (b_x) for Indonesia and Japan.

Figure 3 presents the estimated age-specific sensitivity parameter (b_x) for Indonesia and Japan. The b_x profiles varied across age groups in both countries, with differences in their magnitude and age distribution.

In Indonesia, the largest b_x estimates occurred during infancy, followed by a pronounced decline throughout childhood and adolescence. Between approximately 20 and 40 years of age, the parameter remained relatively stable before gradually decreasing toward the oldest ages. This pattern indicates greater sensitivity to changes in the mortality index at younger ages.

Japan showed lower b_x estimates during infancy and childhood, followed by a relatively stable profile during adulthood. A modest increase appeared between approximately 70 and 80 years of age, followed by a slight decline at older ages. Overall, the variation in b_x across age groups was less pronounced in Japan than in Indonesia, indicating a more even age-specific sensitivity to changes in the mortality index.

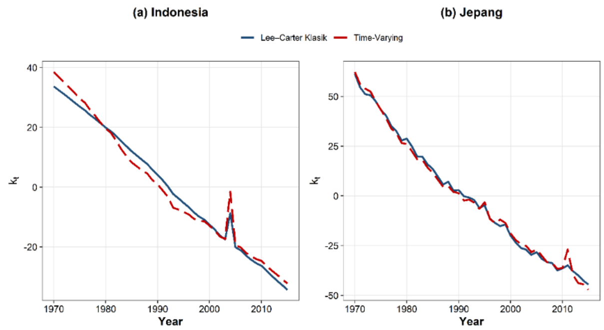

Figure 4. Estimated mortality index (k_t) from the Classical and Time-Varying Lee–Carter models for Indonesia and Japan.

Figure 4 shows the estimated mortality index (k_t) from the Classical and Time-Varying Lee–Carter models. Both countries exhibited declining long-term trajectories, although differences appeared in selected years. In Indonesia, the Time-Varying estimates showed a more noticeable fluctuation around 2004–2005, whereas the Classical estimates followed a smoother path. In Japan, the two estimates remained closely aligned, with a small deviation around 2011.

Table 2. Selected estimated mortality index (k_t) values from the Classical and Time-Varying Lee–Carter models for Indonesia and Japan.

Year	Indonesia (Classical)	Indonesia (TV)	Japan (Classical)	Japan (TV)
1970	33.702	38.505	61.438	62.274
1980	19.928	19.578	28.786	26.095
1990	4.100	0.911	2.910	1.340
2000	-12.559	-12.933	-20.093	-18.880
2010	-26.313	-24.699	-36.418	-36.293
2015	-34.361	-32.259	-44.570	-47.226

Table 2 provides selected estimates to complement the graphical comparison in Figure 4. The values show the same overall decline observed in the plotted trajectories, with differences between the two models at selected years.

3.3 Observed and Fitted Mortality Rates

The fit of the Classical and Time-Varying Lee–Carter models was assessed by comparing the observed mortality rates with their fitted values. Although fitted mortality rates were obtained for all age groups (0–100+ years), only age 50 is presented as an illustrative example to facilitate visual comparison between the two models. Presenting all age groups in a single figure would result in an overly crowded visualization.

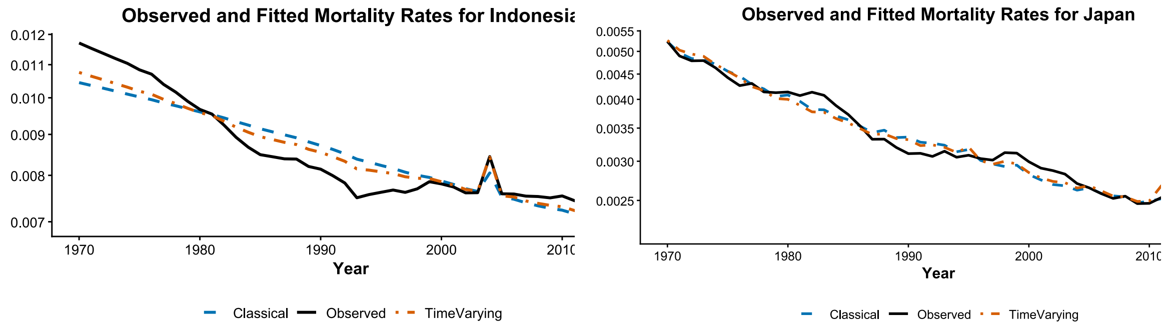

Figure 5. Observed and fitted mortality rates at age 50 for Indonesia and Japan under the Classical and Time-Varying Lee–Carter models.

Figure 5 compares the observed and fitted mortality rates at age 50. In Indonesia, both models captured the overall declining pattern, with deviations during the mid-1980s to mid-1990s and a temporary increase around 2004–2005. In Japan, the fitted rates remained close to the observed values throughout the study period. Overall, the differences between the fitted trajectories were small at the selected age.

3.4 Forecasting Mortality Index

The mortality index was forecast for the out-of-sample period (2016–2023) using the Classical and Time-Varying Lee–Carter models. The resulting forecasts are presented in Figure 6 and Table 3.

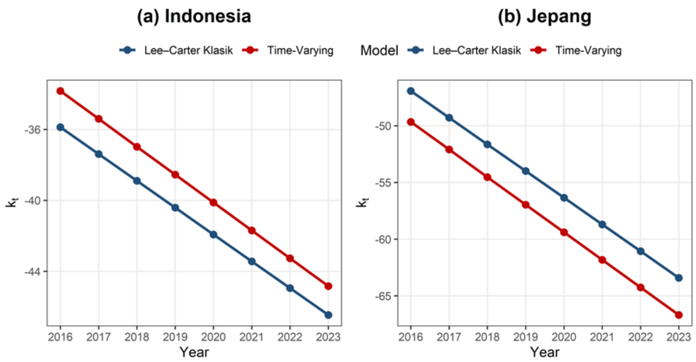

Figure 6. Forecasted mortality index (k_t) for Indonesia and Japan using the Classical and Time-Varying Lee–Carter models, 2016–2023.

Figure 6 and Table 3 show the forecasted mortality index for Indonesia and Japan during the testing period. For both countries, the mortality index generally declined throughout the forecast horizon.

For Indonesia, the forecast trajectories from the two models were closely aligned. The Time-Varying Lee–Carter model produced slightly higher forecasted mortality index values than the Classical model across the forecast period. Despite this difference, the overall trajectories were similar.

A similar pattern was observed for Japan. The mortality index declined throughout the forecast period, with the Time-Varying model producing slightly lower values than the Classical model. The two forecast trajectories remained closely aligned.

Table 3. Forecasted mortality index for Indonesia and Japan using the Classical and Time-Varying Lee–Carter models, 2016–2023.

Year	Indonesia (Classical)	Indonesia (TV)	Japan (Classical)	Japan (TV)
2016	-35.874	-33.831	-46.926	-49.659
2017	-37.386	-35.404	-49.282	-52.092
2018	-38.899	-36.976	-51.638	-54.526
2019	-40.411	-38.549	-53.993	-56.959
2020	-41.924	-40.121	-56.349	-59.392
2021	-43.436	-41.694	-58.705	-61.826
2022	-44.949	-43.266	-61.061	-64.259
2023	-46.461	-44.839	-63.416	-66.692

3.5 Prediction Interval Analysis

To assess the uncertainty associated with the mortality index forecasts, 95% prediction intervals were constructed for both the Classical and Time-Varying Lee-Carter models. Figures 7 and 8 show the forecasted mortality indices and their corresponding prediction intervals for Indonesia and Japan, respectively. The shaded regions represent the 95% prediction intervals, while the dashed lines indicate the forecasted mortality indices during the out-of-sample period (2016–2023).

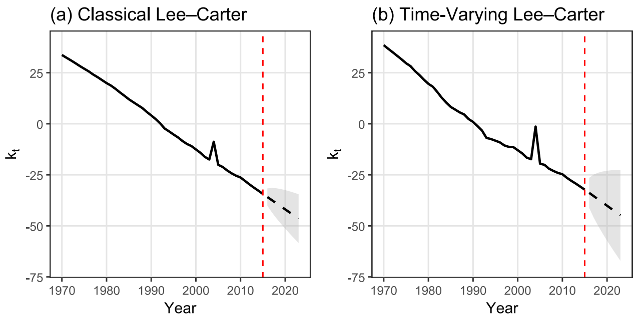

Figure 7. Forecasted mortality index (k_t) with 95% prediction intervals for Indonesia using the (a) Classical Lee–Carter and (b) Time-Varying Lee–Carter models.

For Indonesia, both models showed similar downward forecast trajectories. The prediction intervals widened as the forecast horizon increased. The Time-Varying model produced wider intervals than the Classical model throughout the forecast period.

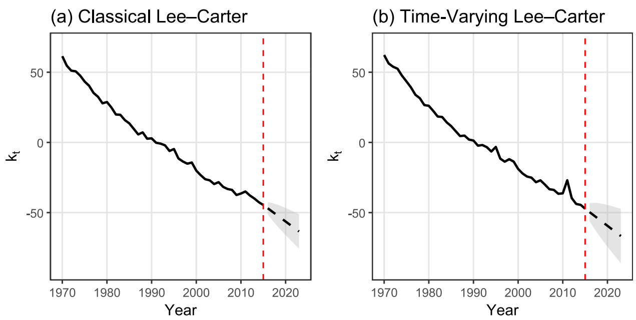

Figure 8. Forecasted mortality index (k_t) with 95% prediction intervals for Japan using the (a) Classical Lee–Carter and (b) Time-Varying Lee–Carter models.

A similar pattern was observed for Japan. The forecast trajectories declined over the testing period, while the prediction intervals widened with the forecast horizon. The Time-Varying model also produced wider intervals than the Classical model.

3.6 Forecasted Mortality Rates

Using the forecasted mortality index (k_t) from the Classical and Time-Varying Lee-Carter models, mortality rates were projected for the testing period (2016–2023). Although forecasts were generated for all age groups (0–100+ years), only age 50 is presented in Figure 9 as an illustrative example to facilitate visual comparison between the two models. Presenting all age groups in a single figure would result in an overly crowded visualization.

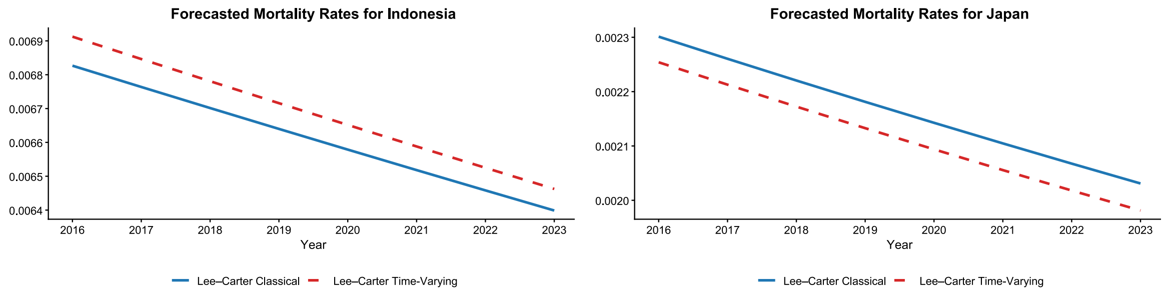

Figure 9. Forecasted mortality rates at age 50 for Indonesia and Japan using the Classical and Time-Varying Lee–Carter models, 2016–2023.

Figure 9 shows declining projected mortality rates for both countries over the forecast horizon, consistent with the forecasted mortality index. For Indonesia, the Time-Varying Lee–Carter model produced slightly higher projected mortality rates than the Classical model, whereas the opposite pattern was observed for Japan. The differences between the two models were small in both countries.

Although Figure 9 illustrates projections at age 50, forecast accuracy was evaluated across all age groups (0–100+ years). Figure 10 complements the age-50 illustration by presenting age-specific MAPE profiles for the full age range.

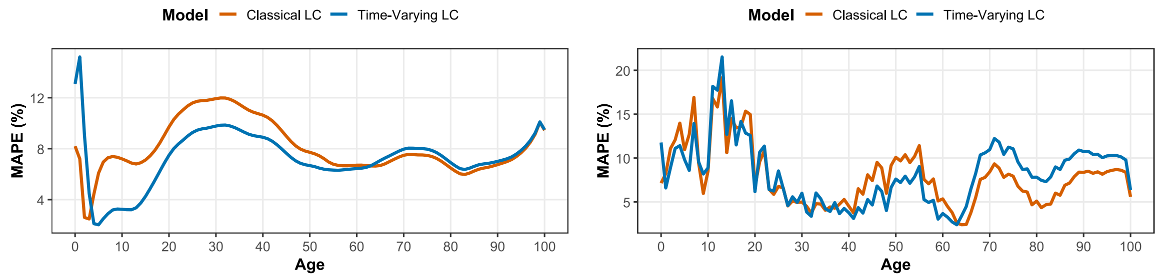

Figure 10. Age-specific MAPE profiles of the Classical and Time-Varying Lee–Carter models during the testing period (2016–2023): (a) Indonesia and (b) Japan.

Figure 10 shows that forecasting errors varied across age groups in both countries. In Indonesia, the Time-Varying model generally produced slightly lower MAPE values across several adult age groups, although the differences remained modest. In Japan, the Classical model tended to yield lower MAPE values at several older ages. Overall, neither model consistently produced lower forecasting errors across all age groups or both countries.

3.7 Forecasting Performance Evaluation

The predictive accuracy of the Classical and Time-Varying Lee–Carter models was examined based on out-of-sample data covering the period from 2016 to 2023. Model performance was quantified using three evaluation metrics: MAE, RMSE, and MAPE. A detailed comparison of the forecasting results for Indonesia and Japan is presented in Table 4.

Table 4. Forecasting performance of the Classical and Time-Varying Lee–Carter models.

Country	Model	MAE	RMSE	MAPE (%)
Indonesia	Classical Lee–Carter	0.0065	0.0250	8.086
	Time-Varying Lee–Carter	0.0066	0.0249	7.262
Japan	Classical Lee–Carter	0.0026	0.0078	7.857
	Time-Varying Lee–Carter	0.0032	0.0090	8.254

For Indonesia, the evaluation metrics showed mixed results. The Time-Varying Lee–Carter model yielded a lower MAPE (7.262%) than the Classical Lee–Carter model (8.086%), whereas the Classical model produced a slightly lower MAE (0.0065 versus 0.0066). The RMSE values were nearly identical for the two models (0.0249 for the Time-Varying model and 0.0250 for the Classical model). Thus, the results did not show a uniform advantage for either model across all three metrics.

For Japan, the Classical Lee–Carter model produced lower MAE (0.0026 versus 0.0032), RMSE (0.0078 versus 0.0090), and MAPE (7.857% versus 8.254%) than the Time-Varying model. The annual MAPE results also showed higher errors for the Time-Varying model in 2022–2023, as discussed in Section 3.9.

These findings are broadly consistent with [21], who reported that the Time-Varying Lee–Carter model is more suitable for evolving mortality patterns, and with [22], who found that the Classical Lee–Carter model remains effective when mortality changes gradually.

Overall, the Time-Varying model yielded a lower MAPE for Indonesia, whereas the Classical model produced lower values for all three evaluation metrics in Japan. These results indicate that the relative performance of the two models varied across countries and evaluation metrics. The statistical significance of the differences was subsequently examined using the Diebold–Mariano test.

3.8 Diebold–Mariano Test Results

The statistical significance of the differences in forecasting accuracy between the Classical and Time-Varying Lee–Carter models was further evaluated using the Diebold–Mariano (DM) test. The results are presented in Table 5.

Table 5. Diebold–Mariano test results comparing the forecasting accuracy of the Classical and Time-Varying Lee–Carter models for Indonesia and Japan.

Country	Compared Models	DM Statistic	p-value	Interpretation
Indonesia	Classical Lee–Carter Time-Varying Lee–Carter	3.995	<0.001	Significant difference
Japan	Classical Lee–Carter Time-Varying Lee–Carter	-8.500	<0.001	Significant difference

For Indonesia, the DM test produced a p-value below 0.001, indicating a statistically significant difference in predictive accuracy between the two models. The same result was obtained for Japan, with a p-value below 0.001. Thus, the null hypothesis of equal predictive accuracy was rejected for both countries.

The statistical significance should be interpreted together with the magnitude of the differences observed in the MAE, RMSE, and MAPE results. Because the DM test uses the pooled age-by-year error series, the analysis includes 808 age-by-year forecast-error observations for each country (101 age groups × 8 testing years), which provides substantially more observations than the eight annual observations alone. This increases the ability of the test to detect relatively small differences between the models. Therefore, statistical significance does not by itself imply that the differences are large in practical forecasting terms.

3.9 Annual Forecast Performance with Emphasis on the COVID-19 Period

To further evaluate forecasting robustness under unusual mortality conditions, annual MAPE values during the out-of-sample period (2016–2023) were examined, with particular attention to the COVID-19 pandemic (2020–2021). This analysis compares forecasting accuracy before, during, and after the pandemic.

Table 6. Annual out-of-sample MAPE (2016–2023), highlighting the COVID-19 years (2020–2021).

Period	Year	Indonesia		Japan	
		Classical	Time-Varying	Classical	Time-Varying
Pre-COVID	2016	4.2915	3.5485	6.5110	5.8592
	2017	4.3668	3.6430	7.2899	6.8534
	2018	5.0848	4.2393	7.2856	7.1477
	2019	4.6434	3.9369	7.0069	7.3544
COVID-19	2020	13.8114	12.6608	7.4374	7.8239
	2021	21.5882	20.5223	8.2406	8.8528
Post-COVID	2022	5.3451	4.6629	10.2345	11.8191
	2023	5.5573	4.8828	8.8501	10.3217

Table 6 shows that forecasting errors for Indonesia increased markedly during the COVID-19 period. The MAPE reached 13.8114% for the Classical model and 12.6608% for the Time-Varying model in 2020, increasing to 21.5882% and 20.5223%, respectively, in 2021. After the pandemic period, the errors decreased substantially in 2022 and remained lower in 2023.

For Japan, the annual MAPE values changed less sharply during the COVID-19 years. The Classical and Time-Varying models produced MAPE values of 7.4374% and 7.8239% in 2020 and 8.2406% and 8.8528% in 2021, respectively. In the post-COVID period, the MAPE increased in 2022 before declining in 2023. The Classical model had lower MAPE values than the Time-Varying model in both post-COVID years.

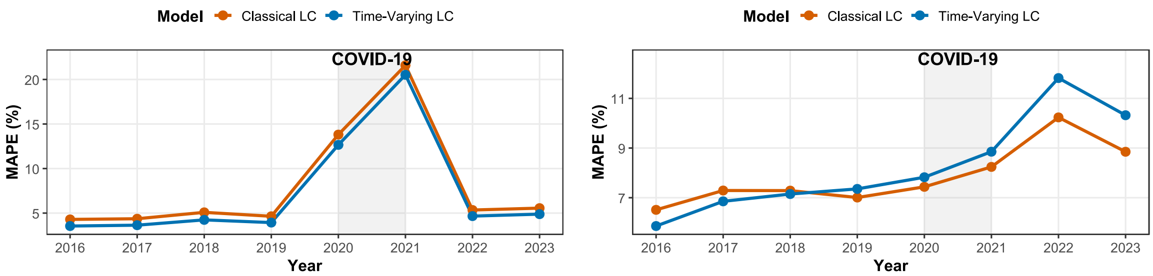

Figure 11. Annual MAPE comparison between the Classical and Time-Varying Lee–Carter models during the testing period (2016–2023) for (a) Indonesia and (b) Japan. The shaded area represents the COVID-19 period (2020-2021).

Figure 11 illustrates the annual MAPE patterns for both countries. Indonesia showed relatively low errors before the pandemic, followed by a marked increase in 2020–2021 and a subsequent decline. Japan exhibited smaller changes during the pandemic, although the errors increased in 2022 before decreasing in 2023. Overall, the annual results show that forecasting errors varied across periods and countries, with the largest increase occurring for Indonesia during 2020–2021.

3.10 Study Limitations and Future Research

This study has several limitations. First, mortality data for both sexes combined were used. Although WPP 2024 provides sex-specific mortality estimates, the availability and quality of the underlying empirical data may vary across countries and periods, which can affect mortality estimation. The combined series was therefore selected to ensure a consistent population-level comparison between Indonesia and Japan. Sex-specific analyses were beyond the scope of this study and may be considered in future research. Second, WPP 2024 mortality rates are model-based estimates rather than directly observed vital-registration data; thus, the out-of-sample results should be interpreted within the WPP estimation framework. Finally, estimates for the open-ended 100+ age group may be subject to greater uncertainty because of the relatively small population exposed to the risk of death. Future research may extend the analysis using sex-specific data, alternative mortality databases, and methods that account for uncertainty at very old ages.

4. CONCLUSION

This study compared the forecasting performance of the Classical Lee–Carter and Time-Varying Lee–Carter models using mortality data for Indonesia and Japan from the World Population Prospects 2024. Overall, the differences in MAE, RMSE, and MAPE were relatively small, although the Diebold–Mariano test indicated statistically significant differences in predictive accuracy for both countries. Thus, statistical significance should not be interpreted as evidence of a substantial practical advantage of one model over the other.

For Indonesia, the Time-Varying Lee–Carter model achieved a lower MAPE, while the Classical Lee–Carter model produced a slightly lower MAE and a nearly identical RMSE. For Japan, the Classical Lee–Carter model produced lower MAE, RMSE, and MAPE. These results indicate that the relative performance of the two models varied across the populations and evaluation metrics considered. The additional flexibility of the state-space formulation did not consistently result in lower forecasting errors.

From a practical perspective, the findings suggest that model selection should consider the characteristics of the mortality data and the objectives of the forecasting analysis rather than model complexity alone. In the populations examined in this study, the Classical Lee–Carter model provided lower errors for Japan, whereas the Time-Varying model showed a modest advantage in MAPE for Indonesia. These findings should not be interpreted as establishing general superiority of either model.

This study is limited by the use of both-sexes-combined mortality data and model-based mortality estimates from WPP 2024. Consequently, the out-of-sample validation should be interpreted within the WPP estimation framework rather than as validation against fully independent observations. Future research may consider sex-specific mortality data, alternative mortality databases, additional countries, longer forecasting horizons, and other state-space forecasting approaches.

5. REFERENCES

[1] R. Duerst, S. Jonas, and C. Bohk-Ewald, "A Validation Workflow for Mortality Forecasting," Rostock, Germany, WP-2023-020, 2023. https://doi.org/10.4054/MPIDR-WP-2023-020 .

[2] A. Nalmpatian, C. Heumann, and S. Pilz, "Forecasting Mortality Trends: Advanced Techniques and the Impact of COVID-19," *Stats*, vol. 7, no. 4, pp. 1172–1188, 2024, https://doi.org/10.3390/stats7040069.

[3] D. Atance, M. Mercè Claramunt, X. Varea, and J. M. Aburto, "Convergence and divergence in mortality: A global study from 1990 to 2030," *PLoS ONE*, vol. 19, no. 1, pp. 1–25, 2024, https://doi.org/10.1371/journal.pone.0295842.

[4] United Nations, "World Population Prospects 2024: Methodology of the United Nations population estimates and projections," New York, 2024.

[5] L. De Mori, S. Haberman, P. Millossovich, and R. Zhu, "Mortality Forecasting via Multi-Task Neural Networks," *ASTIN Bulletin*, vol. 55, no. 2, pp. 313–331, 2025, https://doi.org/10.1017/asb.2025.10.

[6] J. Wakefield and V. Knutson, "Excess Mortality Estimation," *Annual Review of Statistics and Its Application*, vol. 12, no. 1, pp. 45–68, 2025, https://doi.org/10.1146/annurev-statistics-112723-034236.

[7] R. D. Lee and L. R. Carter, "Modeling and Forecasting U.S. Mortality," *Journal of the American Statistical Association*, vol. 87, no. 419, pp. 659–671, 1992, https://doi.org/10.1080/01621459.1992.10475265.

[8] U. Basellini, C. G. Camarda, and H. Booth, "Thirty years on: A review of the Lee–Carter method for forecasting mortality," *International Journal of Forecasting*, vol. 39, no. 3, pp. 1033–1049, 2023, https://doi.org/10.1016/j.ijforecast.2022.11.002.

[9] Fitriani, H. Cipta, and Rina Widyasari, "Penerapan Model Lee-Carter untuk Peramalan Angka Harapan Hidup Berdasarkan Usia di Kota Medan," *Leibniz: Jurnal Matematika*, vol. 5, no. 02, pp. 129–140, 2025, https://doi.org/10.59632/leibniz.v5i02.509.

[10] R. A. L. Kafi and A. A. Sihombing, "The comparison of random walk with drift and SES in forecasting Indonesian mortality rate with Lee-Carter model," *Jurnal Riset dan Aplikasi Matematika (JRAM)*, vol. 06, no. 01, pp. 73–79, 2022, https://doi.org/10.26740/jram.v6n1.p73-79.

[11] Z. P. C. Kertayanda, T. Yulita, D. T. Julianty, A. Ginting, and M. N. A. Simanjuntak, "Life Table Prediction Using the Lee-Carter Model," *Journal of Actuarial, Finance and Risk Management (JAFRM)*, vol. 3, no. 1, 2024, https://doi.org/10.33021/jafrm.v3i1.5429.

[12] A. U. Shelleng, H. G. Dikko, J. Garba, and B. B. Alhaji, "Assessing the suitability of the Lee-Carter model in modelling mortality data with varying bx parameter," *Fudma Journal of Sciences*, vol. 7, no. 3, pp. 293–296, 2023, https://doi.org/10.33003/fjs-2023-0703-1874.

[13] F. Novkaniza, R. Syach, and M. Novita, "Modified Lee–Carter Model Using Extreme Value Theory for Forecasting Mortality Rates Amidst Extreme Events During COVID-19 Era in Indonesia," *Communications in Mathematical Biology and Neuroscience*, pp. 1–15, 2024, doi: 10.28919/cmbn/8656.

[14] S. Tang, J. Li, and L. Tickle, "A New Fourier Approach under the Lee-Carter Model for Incorporating Time-Varying Age Patterns of Structural Changes," *Risks*, vol. 10, no. 8, 2022, https://doi.org/10.3390/risks10080147.

[15] L. He, F. Huang, J. Shi, and Y. Yang, "Mortality Forecasting using Factor Models: Time-varying or Time-invariant Factor Loadings?," *Insurance: Mathematics and Economics*, vol. 98, pp. 1–40, 2021, https://doi.org/10.1016/j.insmatheco.2021.01.006.

[16] S. A. Gyamerah, A. A. Mensah, C. Asare, and N. Dzupire, "Improving mortality forecasting using a hybrid of Lee–Carter and stacking ensemble model," *Bulletin of the National Research Centre*, vol. 47, no. 1, 2023, https://doi.org/10.1186/s42269-023-01138-2.

[17] J. Garrido, Y. Shang, and R. Xu, "LSTM-Based Coherent Mortality Forecasting for Developing Countries," *Risks*, vol. 12, no. 2, pp. 1–24, 2024, https://doi.org/10.3390/risks12020027.

[18] M. C. Fung, G. W. Peters, and P. V. Shevchenko, "A State-Space Estimation of the Lee-Carter Mortality Model and Implications for Annuity Pricing," in *Proceedings of the 21st International Congress on Modelling and Simulation (MODSIM2015)*, Gold Coast, Australia: Modelling and Simulation Society of Australia and New Zealand, 2015, pp. 952–958. https://doi.org/10.2139/ssrn.2699624.

[19] M. Matějka and I. Malá, "The Usage of State Space Models in Mortality Modeling and Predictions," *Statistika*, vol. 103, no. 1, pp. 121–131, 2023, https://doi.org/10.54694/STAT.2022.33.

[20] R. Li, R. Zhou, and D. Pitt, "Dynamic Mortality Forecasting via Mixed-Frequency State-Space Models," 2026.

[21] Z. Chen, Y. Shi, and A. Shu, "Managing Mortality and Aging Risks with a Time-Varying Lee–Carter Model," *Healthcare (Switzerland)*, vol. 11, no. 5, pp. 1–20, 2023, https://doi.org/10.3390/healthcare11050743.

[22] B. Koichubekov, B. Omarkulov, M. Mukhanova, and R. Zakirova, "Age-Specific Mortality Forecasting in Kazakhstan: Alternative Approaches to the Lee – Carter Model," *International Journal of Environmental Research and Public Health*, vol. 22, no. 3, p. 346, 2025, https://doi.org/10.3390/ijerph22030346.

[23] N. Redzwan, P. Sivasundaram, and R. Ramli, "Mortality Index Simulation for Forecasting Malaysian Mortality Rates," *ASM Science Journal*, vol. 18, pp. 1–11, 2023, https://doi.org/10.32802/asmscj.2023.1465.

[24] S. Sahin and S. Ozen, "A Quantitative Comparison of Mortality Models with Jumps: Pre- and Post-COVID Insights on Insurance Pricing," *Risks*, vol. 12, no. 3, p. 53, 2024, https://doi.org/10.3390/risks12030053.

[25] GBD 2021 Japan Collaborators, "Three decades of population health changes in Japan, 1990–2021: a subnational analysis for the Global Burden of Disease Study 2021," *The Lancet Public Health*, vol. 10, no. 4, pp. e321–e332, 2025, https://doi.org/10.1016/S2468-2667(25)00044-1.

[26] H. Surendra, D. Paramita, N. N. Arista, A. I. Putri, and A. A. Siregar, "Geographical variations and district-level factors associated with COVID-19 mortality in Indonesia: a nationwide ecological study," *BMC Public Health*, pp. 1–12, 2023, https://doi.org/10.1186/s12889-023-15015-0.

[27] C. T. Pallari *et al.*, "Magnitude and determinants of excess total, age-specific and sex-specific all-cause mortality in 24 countries worldwide during 2020 and 2021: results on the impact of the COVID-19 pandemic from the C-MOR project," *BMJ Global Health*, vol. 9, no. 4, 2024, https://doi.org/10.1136/bmjgh-2023-013018.

[28] J. Durbin and S. J. Koopman, *Time Series Analysis by State Space Methods*, 2nd ed. Oxford: Oxford University Press, 2012. https://doi.org/10.1093/acprof:oso/9780199641178.001.0001.

[29] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960, https://doi.org/10.1115/1.3662552.

[30] J. Helske, "KFAS: Exponential Family State Space Models in R," *Journal of Statistical Software*, vol. 78, no. 10, pp. 1–39, 2017, https://doi.org/10.18637/jss.v078.i10.

[31] A. Léger and S. Rizzi, "Month-to-month all-cause mortality forecasting: a method allowing for changes in seasonal patterns," *American Journal of Epidemiology*, vol. 193, no. 6, pp. 898–907, 2024, https://doi.org/10.1093/aje/kwae004.

[32] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, no. 4, pp. 679–688, 2006, https://doi.org/10.1016/j.ijforecast.2006.03.001.

[33] E. Strøm and O. E. Gundersen, "Performance metrics for multi-step forecasting measuring win-loss, seasonal variance and forecast stability: an empirical study," *Applied Intelligence*, vol. 54, no. 13, pp. 10490–10515, 2024, https://doi.org/10.1007/s10489-024-05715-4.

[34] P. Manandhar, H. Rafiq, E. Rodriguez-Ubinas, and T. Palpanas, "New Forecasting Metrics Evaluated in Prophet, Random Forest, and Long Short-Term Memory Models for Load Forecasting," *Energies*, vol. 17, no. 14, pp. 1–30, 2024, https://doi.org/10.3390/en17143472.

[35] Z. Pala and F. Şevgin, "Statistical Modeling for Long-Term Meteorological Forecasting: A Case Study in Van Lake Basin," *Natural Hazards*, vol. 120, no. 12, pp. 14101–14116, 2024, https://doi.org/10.1007/s11069-024-06747-2.

[36] F. X. Diebold and R. S. Mariano, "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, vol. 13, no. 3, pp. 253–263, 1995, https://doi.org/10.1080/07350015.1995.10524599.