Seasonal Mortality Trends in Australia: A Time Series Analysis Using Benchmark and Advanced Models

1st Max Li
Faculty of Science
Thompson Rivers University
Kamloops, Canada
limax231@mytru.ca

2nd Mengyu Liu
Faculty of Science
Thompson Rivers University
Kamloops, Canada
lium24@mytru.ca

3rd Jiahui Yan
Faculty of Science
Thompson Rivers University
Kamloops, Canada
yanj231@mytru.ca

***Abstract*—This report investigates the use of time series models to analyze and forecast Australian mortality rates, emphasizing trends, seasonality, and demographic differences. The dataset includes weekly mortality counts and rates, categorized by age and sex, over an eight-year period. After comprehensive preprocessing and visualization, we applied benchmark models (Mean, Seasonal Naïve, Drift), a regression-based Time Series Linear Model (TSLM), and a Seasonal Auto-Regressive Integrated Moving Average (SARIMA) to identify and capture patterns in the data. Cross-validation was used to evaluate model robustness to guarantee a reliable assessment of predictive performance across all demographic groups. Metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) were used to compare models. The analysis reveals distinct seasonal mortality peaks, particularly mid-year, aligning with Australia's winter months. These insights provide valuable guidance for policymakers and healthcare planners in optimizing resource allocation and improving public health strategies. The report highlights the limitations of statistical models and proposes future directions, including dynamic and machine learning-based approaches, to enhance mortality forecasting.**

***Index Terms*—Mortality Forecasting, SARIMA, TSLM, Time Series Analysis, Cross-Validation**

I. Introduction

Australia's population dynamics are unique compared to other developed nations, with net positive immigration from nearly every country except the United States. Despite having vast land resources, most Australians reside along coastal regions, leading to population pressures in urban centers and challenges in resource allocation. Managing these pressures requires an understanding of key population indicators, especially mortality trends.

Many developed countries, including Australia, face challenges from population aging. The situation creates increasing demand for policies that support elder care, healthcare capacity, and pensions. Forecasting mortality among elders (65+) can precisely meet the needs. Similarly, understanding mortality rates in people between 15 and 64, who are fundamental contributors to the workforce, can help shape immigration policies and address labor shortages for a more sustainable economic development.

Gender-specific mortality forecasts are equally important. A balanced gender ratio supports population growth, stable birth rates, and workforce planning. If gender imbalances emerge, an accurate forecast can help refine immigration strategies and stabilize the population structure.

Mortality forecasts can also guide resource allocation in densely populated cities and economic hubs. Insights into population trends can inform policies on housing, education, and job markets, ensuring sustainable growth and equitable resource distribution.

Another critical aspect of mortality is its seasonal variation. In Australia, mortality rates peak during mid-year, coinciding with winter in the Southern Hemisphere. Understanding this seasonality can help mitigate deaths from seasonal illnesses and support proactive measures, such as vaccination campaigns, emergency planning, and budget allocation for particular periods.

This report analyzes Australian mortality data by gender and age using statistical time series models. We begin with three benchmark models—mean, seasonal naïve, and drift—to provide a baseline for comparison. Then, we utilize more mathematically advanced models, such as time series linear regression (TSLM) and Autoregressive Integrated Moving Average (ARIMA)—to capture more complex trends. Using the auto ARIMA algorithm, we automatically identify seasonality and incorporate it into the model of each demographic group, and therefore, we are more feasible to capture both long-term trends and seasonal variations in mortality rates.

II. Data

A. Dataset Overview

This project utilizes the Australian mortality dataset (*aus_mortality*) from the *fpp3* package [2]. The dataset is collected from the Short-Term Mortality Fluctuations (STMF) data series [3]. It contains Australian weekly death counts and mortality rates from Week 1 of 2015 to Week 12 of 2023. The dataset is structured as a time series tibble, containing $7,740$ observations across five key variables, the details are provided in Table I. Within the listed variables, Mortality represents

the mortality rate per thousand people per week in Australia and is our primary variable of interest. It enables standardized comparisons across different demographic segments of the Australian population.

TABLE I
VARIABLE DESCRIPTION OF AUSTRALIAN MORTALITY DATASET

Variable	Type	Description
Week	Period	Time indicator in year-week format
Sex	Factor	Sex of the individual (Male, Female, or Both)
Age	Factor	Age group of the individual
Death	Numeric	Weekly death count for the demographic group
Mortality	Numeric	Deaths per thousand people per week

B. Exploratory Data Analysis

The exploratory analysis shows several key temporal and demographic characteristics in the Australian mortality data. As shown in Figure 1, the time series exhibits seasonal fluctuations in mortality rates throughout each year, with roughly yearly variations in older groups and shorter seasonality periods in younger groups. Across the study period from 2015 to 2023, we can observe potential long-term trends in these mortality patterns.

When examining demographic factors, we observed variation across different age groups. older age brackets $(75+)$ show higher and more volatile mortality rates. We also observed that the gender differences in mortality are not uniform across age categories, indicating that there may be interaction effects between age and gender in determining mortality risks.

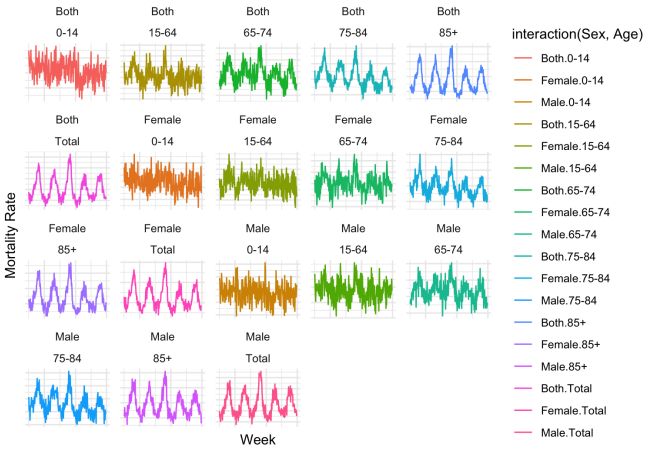

Fig. 1. Weekly Mortality Rates by Demographic Group

III. METHOD

A. Data Preparation and Preprocessing

Before fitting the models, we conducted the following preprocessing steps:

1) Time Index Transformation: The variable *Week* was converted to a time index in the *yearweek* format to follow a regular temporal structure.
2) Train-Test Split: The data was divided into training and testing sets. The training set contains observations from Week 1 of 2015 to Week 29 of 2021, while the testing set includes observations from Week 30 of 2021 onwards.

B. Benchmark Models

1) Mean Mode: The forecast for future values is calculated as the average of all historical observations in the time series. This method assumes the series fluctuates around a constant mean without trend or seasonality. The equation is as follows:

$$\hat{y}_t = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{1}$$

2) Naïve Model: The forecast for future values is based on the most recent observation. This method assumes that the latest value is the best predictor for all future observations. The equation is as follows:

$$\hat{y}_t = y_{t-1} \tag{2}$$

3) Drift Model: The forecast for future values is derived by extending a straight line between the first and last observation in the time series. This method assumes a consistent linear trend over time and continues it forward. The equation is as follows:

$$\hat{y}_t = y_1 + \frac{t-1}{n-1}(y_n - y_1) \tag{3}$$

Among these three baseline models, y is the mortality rate, t is the time, and n is the number of observations.

C. Time Series Linear Model (TSLM)

The Seasonal Time Series Linear Model (TSLM) is created within the linear regression framework [4]. In the context of mortality rate forecasting, the model includes both temporal and seasonal predictors. Specifically, it incorporates a time trend to capture long-term changes in mortality rates and seasonal components to model recurring patterns within annual cycles.

To reflect unique patterns within the Age and Sex group segments, the TSLM is fitted separately for each demographic group. This method allows the model to account for global trends while capturing group-specific variations, enhancing its ability to produce accurate mortality forecasts. The mathematical representation of the TSLM is given by:

$$y_t = \beta_0 + \beta_1 \cdot \text{Trend}_t + \beta_2 \cdot \text{Season}_t + \epsilon_t \tag{4}$$

Where y_t is the mortality rate at time t, β_0, β_1 and β_2 are the intercept, time trend coefficient, and seasonal component coefficient, respectively. The residual error ϵ_t is assumed to be independently and identically distributed.

D. Auto-Regressive Integrated Moving Average (ARIMA)

The Auto-Regressive Integrated Moving Average (ARIMA) model is designed to capture both non-seasonal and seasonal trends, as well as mathematical dependencies in time series data [5]. ARIMA models consist of three primary components: auto-regressive (AR), differencing (I), and moving average (MA). When seasonality is present in the data, the model can be extended to include seasonal auto-regressive (SAR), seasonal differencing, and seasonal moving average (SMA) components, forming the Seasonal ARIMA (SARIMA) model. The general model is represented by:

$$\Phi(B^s)(1-B^s)^D\phi(B)(1-B)^d y_t = \theta(B)\Theta(B^s)\epsilon_t \quad (5)$$

Where:

- y_t is the forecasted mortality rate at time t.
- B represents the backshift operator such as $B*_t = *_{t-1}$.
- $\phi(B)$ is the autoregressive polynomial of degree p to model the dependencies of lagged motality rate observations.
- $\theta(B)$ is the moving average polynomial of degree q to model the dependencies of past residuals.
- $(1-B)^d$ is the differencing operator to remove trends.
- $\Phi(B^s)$ is the seasonal AR polynomial of degree P for seasonal dependencies of mortality rate.
- $\Theta(B^s)$ is the seasonal moving average polynomial of degree Q for past errors.
- $(1-B^s)^D$ is the seasonal differencing operator to account for recurring patterns.
- ϵ_t is the residual error, assumed to follow a white noise process.

The `auto.arima()` function in `R` [6] was used to automatically identify the optimal model parameters (p, d, q) and seasonal components $(P, D, Q)_s$ This function minimizes the Akaike Information Criterion (AIC) to select the best-fitting model. By leveraging `auto.arima()`, we incorporated both seasonal and non-seasonal patterns in the mortality data, capturing long-term trends through differencing and annual seasonality with seasonal components.

E. Evaluation and Validation

1) Cross-Validation: In our project, we used the Rolling-Origin Cross-Validation (ROCV) method to simulate forecasting scenarios in real-life situations. The future values are predicted using past observations only to preserve the time series structure.

We started by splitting the dataset into training and testing sets in a rolling manner. We started with an initial training set containing one years of data and added subsequent observations to expand the training window by two years. Then, we used the next fixed-length one-year window as the test set to validate the fit. The process was repeated until the entire dataset had been covered. We implemented the `stretch_tsibble()` function from the fable package in R for cross-validation [7].

2) Evaluation Metrics: This project used the following error metrics for each forecasting method to compare model performance across different demographic groups:

1) Mean Absolute Error (MAE): MAE measures the average magnitude of absolute error.

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t| \quad (6)$$

2) Root Mean Square Error (RMSE): RMSE is another metric that penalizes larger errors more heavily by squaring.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \quad (7)$$

3) Mean Absolute Percentage Error (MAPE): MAPE expresses forecasting accuracy as a percentage, providing a relative measure of the prediction error.

$$\text{MAPE} = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \quad (8)$$

IV. Results

A. Results of Benchmark Models (MEAN, SNAIVE, DRIFT)

The `gg_tsresidual()}` function was not employed to generate ACF and residual plots due to the presence of 18 groups of time series data in the dataset, which would result in an excessive number of plots. Additionally, we have not included the results of the Portmanteau tests for these benchmark models, as they generally fail to capture sufficient information. This lack of model fit leads to strong autocorrelation in the residuals, making these tests less informative. We applied the models to the training data and used them to forecast the test data. The actual observed values are represented in the plots as the gray area. The forecasts are shown in the following Figure 2 - 4.

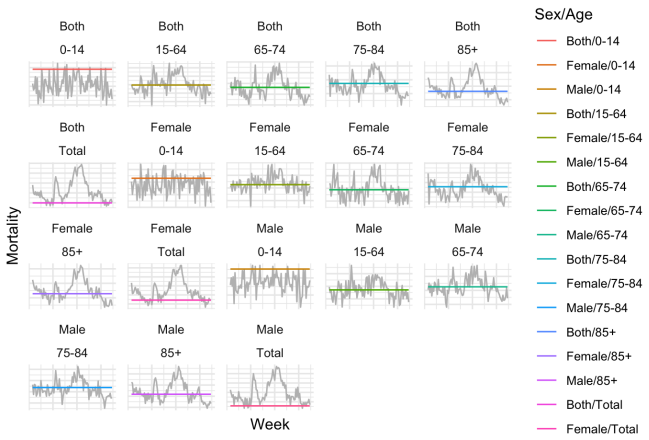

Fig. 2. Mean Method Forecasts by Groups

These benchmark models are unable to adequately capture the underlying structure of the data. Consequently, they are not

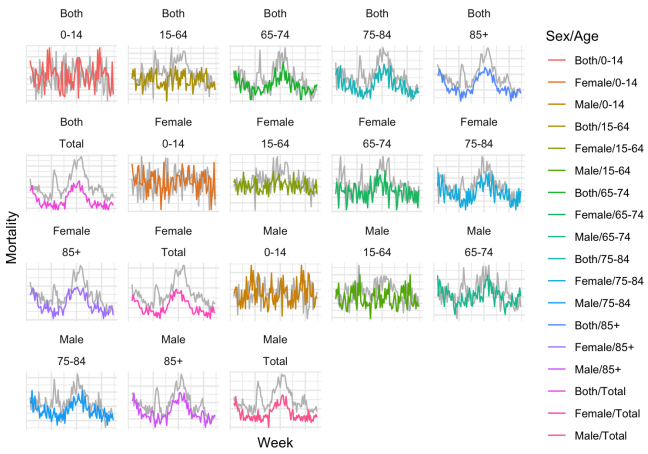

Fig. 3. SNAIVE Method Forecasts by Groups

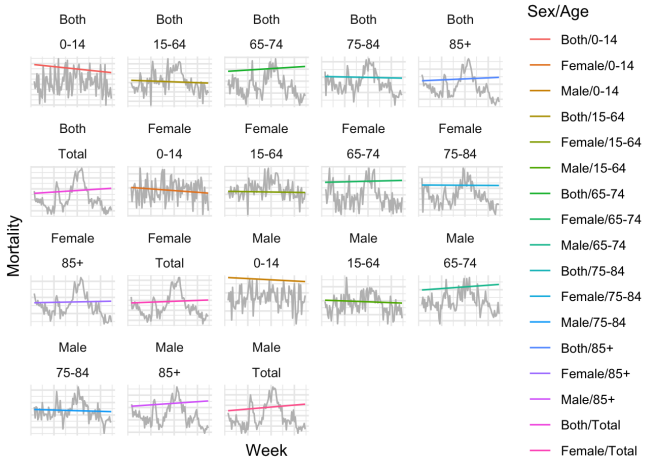

Fig. 4. Drift Method Forecasts by Groups

TABLE II
SELECTED ARIMA MODELS FOR FORECASTING

Sex	Age	ARIMA Model
Both	0-14	$\text{ARIMA}(1,1,1)$
Both	15-64	$\text{ARIMA}(1,1,1)(0,0,1)_{52}$
Both	65-74	$\text{ARIMA}(0,1,1)(1,0,0)_{52}$
Both	75-84	$\text{ARIMA}(1,0,1)(1,1,0)_{52}$ w/ drift
Both	85	$\text{ARIMA}(1,0,1)(2,1,0)_{52}$
Both	Total	$\text{ARIMA}(1,0,1)(2,1,0)_{52}$ w/ drift
Female	0-14	$\text{ARIMA}(0,1,1)(1,0,0)_{52}$
Female	15-64	$\text{ARIMA}(0,1,1)(1,0,0)_{52}$
Female	65-74	$\text{ARIMA}(0,1,1)(1,0,0)_{52}$
Female	75-84	$\text{ARIMA}(2,0,2)(1,1,0)_{52}$ w/ drift
Female	85	$\text{ARIMA}(2,0,2)(1,1,0)_{52}$
Female	Total	$\text{ARIMA}(1,0,1)(1,1,0)_{52}$
Male	0-14	$\text{ARIMA}(0,1,2)$ w/ drift
Male	15-64	$\text{ARIMA}(1,1,2)(0,0,2)_{52}$
Male	65-74	$\text{ARIMA}(1,1,2)_{52}$
Male	75-84	$\text{ARIMA}(1,0,1)(2,1,0)_{52}$
Male	85	$\text{ARIMA}(1,0,1)(1,1,0)_{52}$ w/ drift
Male	Total	$\text{ARIMA}(1,0,1)(1,1,0)_{52}$

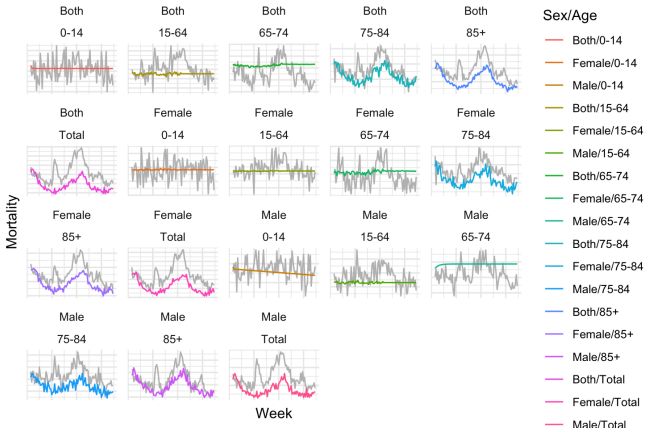

Fig. 5. ARIMA Model Forecasts by Group

suitable for use in final forecasting. However, these models serve as baseline benchmarks, providing a simple reference point and a rough understanding of the dataset's behavior.

B. Results of ARIMA

We used the `auto.arima()` function to automatically determine the optimal order of the ARIMA model. Table II below is an overview of the selected model for the dataset.

The plot of ARIMA forecasts against the test data is shown in Figure 5. Gray lines represent observed mortality rates, while colored lines indicate model forecasts.

The Ljung-Box test results in Table III indicate that the ARIMA model performs relatively well for this dataset. The model effectively captures the underlying patterns for most time series groups, as evidenced by non-significant p-values in most cases. It suggests that the residuals are largely uncorrelated and follow a white noise process. The model adequately explains the variability in the data in general. However, for certain groups such as with low p-values (e.g., 65-74 age group), the model may not fully capture the underlying structure, indicating room for further refinement in those cases.

C. Results of TSLM

The TSLM model forecasts are illustrated in Figure 6. Gray lines represent observed mortality rates, while colored lines indicate model forecasts. From the graphs, the model captures part of the seasonal patterns and trends, particularly in older age groups. In younger age groups (0-14), the TSLM captures more variation than other models. However, the consistently small p-values from the Portmanteau test in Table IV indicate that residuals exhibit significant autocorrelation. The TSLM model is still limited in fully explaining the variability in mortality rates.

D. Time Series Cross Validation

For each model, we calculated the average RMSE, MAPE, and MASE based on the cross-validation results across the 18 groups of data. Table V presents the metrics. The TSLM model achieves the best performance during cross-validation, with the lowest RMSE (0.0023), MAPE (6.4981), and MASE (0.8677). The results demonstrate that TSLM is well-suited for capturing patterns during the training phase.

TABLE III
LJUNG-BOX TEST RESULTS FOR ARIMA MODELS

Sex	Age	lb test statistics	p-value
Both	0-14	104.1322	0.4779
Both	15-64	117.0452	0.1801
Both	65-74	159.0270	0.0004
Both	75-84	140.6085	0.0098
Both	85	69.6291	0.9961
Both	Total	114.2392	0.2315
Female	0-14	125.6272	0.0732
Female	15-64	106.1095	0.4241
Female	65-74	145.4227	0.0046
Female	75-84	134.5923	0.0234
Female	85	111.3708	0.2927
Female	Total	141.7398	0.0082
Male	0-14	86.6159	0.8912
Male	15-64	98.6537	0.6296
Male	65-74	146.0277	0.0042
Male	75-84	107.8767	0.3777
Male	85	124.6903	0.0816
Male	Total	113.6787	0.2428

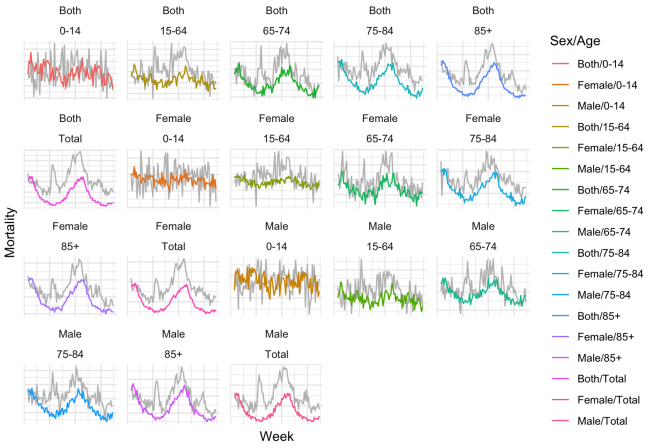

Fig. 6. TSLM Model Forecasts by Groups

E. Testing Accuracy

The average testing accuracy results for all models are summarized in Table VI. The metrics include MAE, RMSE, and MAPE, calculated based on the testing dataset across all 18 demographic groups.

The Mean model demonstrates the lowest MAE (0.0026) and RMSE (0.0034). It shows the ability to provide a stable average forecast. The lowest MAPE (8.5302) indicates that the mean method is also reliable with less relative error.

The Seasonal Naive model performs slightly worse than the Mean model in terms of MAE (0.0028) and RMSE (0.0035). The relatively low MAPE (8.8540) shows that it can approximate seasonality well.

The Drift model exhibits the highest error metrics of MAE (0.0036), RMSE (0.0042), and MAPE: (10.5871). It is not suitable to handle non-linear and seasonal variations in the mortality data.

The ARIMA model performs moderately well, with MAE (0.0031) and RMSE (0.0038) slightly higher than those of the Mean and Seasonal Naive models. Its relatively low MAPE (8.7314) shows its ability to handle seasonality, though it falls short compared to simpler models.

TABLE IV
LJUNG-BOX TEST RESULTS FOR TSLM MODELS

Sex	Age	lb test statistics	p-value
Both	0-14	154.9014	8.99E-04
Both	15-64	559.6124	0.00E+00
Both	65-74	3093.1615	0.00E+00
Both	75-84	7582.5049	0.00E+00
Both	85	9778.961	0.00E+00
Both	Total	9472.084	0.00E+00
Female	0-14	131.682	3.46E-02
Female	15-64	239.4263	1.10E-12
Female	65-74	964.7829	0.00E+00
Female	75-84	4774.2549	0.00E+00
Female	85	8602.571	0.00E+00
Female	Total	8635.7552	0.00E+00
Male	0-14	102.1487	5.33E-01
Male	15-64	268.3943	2.22E-16
Male	65-74	2101.6998	0.00E+00
Male	75-84	5828.6599	0.00E+00
Male	85	8210.2473	0.00E+00
Male	Total	8155.5932	0.00E+00

TABLE V
CROSS VALIDATION RESULT ON TRAINING DATA

Model	Average RMSE	Average MAPE	Average MASE
MEAN	0.0035	8.5592	1.1698
SNAIVE	0.0027	7.7638	1.0215
DRIFT	0.0040	9.2042	1.2749
ARIMA	0.0030	6.8808	0.9504
TSLM	**0.0023**	**6.4981**	**0.8677**

The TSLM model shows the highest MAE (0.0034) and RMSE (0.0040) among all models. Its MAPE (9.3202) also shows that the adjustments for trend and seasonality may not be generalized for the testing data with higher prediction errors.

A visual comparison is provided in Figure 7.

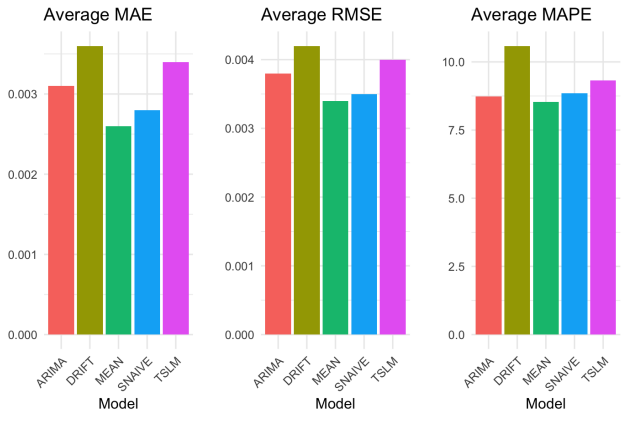

Fig. 7. Accuracy Comparison by Metrics

TABLE VI
ACCURACY RESULT ON TESTING DATA

Model	Average MAE	Average RMSE	Average MAPE
MEAN	**0.0026**	**0.0034**	**8.5302**
SNAIVE	0.0028	0.0035	8.8540
DRIFT	0.0036	0.0042	10.5871
ARIMA	0.0031	0.0038	8.7314
TSLM	0.0034	0.0040	9.3202

V. DISCUSSION

The analysis of mortality trends using statistical time series models provides valuable insights and reveals limitations in model adequacy. Specifically, the ARIMA and SARIMA models demonstrate room for improvement. The automatic parameter selection by the `auto.arima()` function effectively identifies optimal configurations for each demographic group. However, the residual analysis from the Ljung-Box test shows that some residuals exhibit significant autocorrelation. The result suggests that the models may fail to fully capture the complexities and dynamics of mortality data, particularly for specific age groups or during seasonal variations. For example, demographic groups such as the 65–74 age category displayed stronger evidence of model inadequacy.

The TSLM model was designed to incorporate trend and seasonality through a linear regression framework. It captures seasonal patterns effectively for older age groups but struggles with generalization on testing data. The Portmanteau test results for TSLM show significant autocorrelation in the residuals across nearly all demographic groups. The results reinforce the limitations of linear regression-based methods in fully capturing the non-linear dynamics or volatility present in mortality trends, particularly in younger groups with shorter seasonality periods.

When comparing across models, the benchmark methods (Mean, Seasonal Naive, and Drift) provide simple baselines but lack the sophistication needed to accurately capture the temporal dependencies and seasonal characteristics of mortality data. Interestingly, the Mean model shows the best performance on testing data in terms of RMSE, MAPE, and MAE. This outcome suggests that simple models can sometimes outperform more complex ones when the underlying patterns are stable and the data is not overly volatile.

While the current models provide a strong foundation for understanding Australian mortality trends, their limitations highlight the need for more sophisticated, dynamic approaches. In future studies, we have to explore more dynamic modeling approaches to address the limitations. One promising direction is the incorporation of hybrid models that combine statistical methods with machine learning techniques. For example, neural networks, particularly recurrent architectures like Long Short-Term Memory (LSTM) or Gated Recurrent Units (GRU), have shown the ability to capture non-linear dependencies and long-term temporal dynamics in time series data. The volatility and interactions between demographic factors can be properly handled thereby.

CODE AVAILABILITY

The full project, including scripts, processed data, and supplementary analysis, is publicly available on GitHub: https://github.com/maxjinli/DASC6510-Time-Series/.

REFERENCES

[1] G. D. Greenwade, "The Comprehensive Tex Archive Network (CTAN)," TUGBoat, vol. 14, no. 3, pp. 342–351, 1993.

[2] R. Hyndman, G. Athanasopoulos, M. O'Hara-Wild, N. Palihawadana, S. Wickramasuriya, and R. Hyndman, "Package 'fpp3'," 2024. [Online]. Available: https://pkg.robjhyndman.com/fpp3/ [Accessed: Nov. 20, 2024]

[3] Human Mortality Database, "Short-Term Mortality Fluctuations (STMF) Data Series," 2024. [Online]. Available: https://mortality.org/Data/STMF [Accessed: May 29, 2024]

[4] R. Christensen, Linear Models for Multivariate, Time Series, and Spatial Data, vol. 1. Springer, 1991.

[5] S. L. Ho and M. Xie, "The use of ARIMA models for reliability forecasting and analysis," Computers & Industrial Engineering, vol. 35, no. 1-2, pp. 213–216, 1998.

[6] R. J. Hyndman and Y. Khandakar, "Automatic time series forecasting: the forecast package for R," Journal of Statistical Software, vol. 27, pp. 1–22, 2008.

[7] A. Mosnier, L. Penescu, K. Pérez Guzmán, J. Steinhauser, M. Thomson, C. Douzal, and J. Poncet, "FABLE Calculator 2020 update," International Institute for Applied Systems Analysis (IIASA), 2020.