Scilight

Journal of Modern Applied Statistical Methods

https://jmasm.com

Article

Assessing ARIMA Model Performance in Hierarchical Time Series Forecasting of Tourist Arrivals: The Role of Data Normalization and Bottom-Up Strategy

Madona Yunita Wijaya, Nina Fitriyati * and Najib Ridho Sandika

Study Program of Mathematics, Faculty of Science and Technology, Syarif Hidayatullah Jakarta Islamic State University, Banten 15412, Indonesia

* Correspondence: nina.fitriyati@uinjkt.ac.id

How To Cite: Wijaya, M.Y.; Fitriyati, N.; Sandika, N. R. Assessing ARIMA Model Performance in Hierarchical Time Series Forecasting of Tourist Arrivals: The Role of Data Normalization and Bottom-Up Strategy. *Journal of Modern Applied Statistical Methods* **2025**, *24*(2), 4. https://doi.org/10.53941/jmasm.2025.100004

Abstract: This research evaluates the performance of the ARIMA method in forecasting hierarchical time series data of tourist arrivals in Australia from 1998 to 2016 using a bottom-up strategy. A comparative analysis is conducted between the predicted results and the actual data for both short-term and long-term periods, as well as with various normalization methods for each hierarchical level. The study concludes that ARIMA generally performs better in short-term forecasting at a hierarchical level. However, the evaluation results indicate that SMAPE (Symmetric Mean Absolute Percentage Error) values fluctuate across different forecasting periods, influenced by prediction data generated from various ARIMA models. This study does not determine whether one normalization method is superior to another, as the evaluation results show no significant differences. Nevertheless, this research provides insights into the effectiveness of hierarchical time series forecasting using the ARIMA method and a bottom-up strategy at each hierarchical level for both short-term and long-term periods. It also assesses the performance of various normalization methods used.

Keywords: ARIMA; Australia; bottom-up; forecasting; hierarchical time series; normalization

1. Introduction

Accurate predictions of tourist arrivals are essential for making informed decisions within the industry [1]. Additionally, forecasting tourist arrivals is important due to their impact on the social and cultural aspects of tourism at the destination and within the local community [2]. Tourist arrivals in most countries can fluctuate over time due to various factors such as social and economic conditions, natural disasters, and more. Therefore, accurate predictions are vital to address the uncertainty about tourist arrivals and to assist in the effective management and planning of the tourism sector.

Tourist arrivals can be depicted as hierarchical time series data due to various geographical differences such as regions and states, as well as based on the purpose of the visit. While there has been extensive research on forecasting hierarchical time series data, previous studies often have limitations and primarily focus on evaluating prediction results using the Mean Absolute Percentage Error (MAPE) [3,4]. According to [5], MAPE can face challenges when data values are close to or equal to zero. Moreover, many studies often were concentrated on forecasting and comparing different hierarchical forecasting methods [6–8], but often without addressing the impact of normalization on forecasting accuracy.

Normalization is a crucial aspect in hierarchical time series forecasting, as it helps to address differences in scale and value ranges across various hierarchical levels. By standardizing data, normalization can enhance model

performance and ensure more consistent results across different levels of the hierarchy. Despite its importance, many studies have not thoroughly examined how normalization methods interact with hierarchical forecasting techniques. This study not only evaluates the performance of the Autoregressive Integrated Moving Average (ARIMA) model using a bottom-up strategy but also explores the impact of various normalization methods on forecasting accuracy. The bottom-up approach involves forecasting each disaggregated series at the lowest hierarchical level and aggregating these forecasts to predict higher levels [6].

Moreover, this research seeks to bridge the gap in understanding how different normalization techniques affect the performance of hierarchical time series forecasting models. By comparing several normalization methods, including Z-Score, Min-Max, and Total-Sum normalization, this study aims to provide insights into which techniques offer the best performance for maintaining consistency across hierarchical levels. The findings are expected to contribute to more accurate and reliable forecasting practices, offering valuable guidance for researchers and practitioners in the field of tourism forecasting.

2. Method

This study utilizes data sourced from Kaggle on tourist arrivals in Australia. The data consists of quarterly records from January 1998 to December 2016 and has a hierarchical structure. It includes the number of overnight trips (in thousands) across Australia. The hierarchical levels in the data are represented by state, region, and purpose of travel. The analysis process begins with identifying the levels in the data. At the top of the hierarchy is "Total" denoted as Y_t, which represents level 0 and the fully aggregated series. Level 1 is the first level of disaggregation, with the hierarchy extending down to level K, which contains the most detailed, disaggregated time series. The data in this study shows a hierarchical structure with 4 levels. The top level of the hierarchy in this study represents the total tourist arrivals within the country (level 0). Levels 1 to 3 represent the tourist arrivals at the state, region, and purpose of travel level, respectively.

The bottom-up strategy analyses data from the lowest level and then aggregates each series at each level to obtain forecasts at higher levels. By performing the analysis at the lowest level first, the risk of losing information can be minimized [9]. For example, for the top level, the bottom-up strategy can be represented mathematically as follows:

$$\hat{Y}_t = \sum_{i=1}^{n} \hat{Y}_{i,t} = \hat{Y}_{1,t} + \hat{Y}_{2,t} + \cdots + \hat{Y}_{n,t}.$$

In addition to having a hierarchical structure, the data used in this study also contains time series patterns at each hierarchical level and requires an appropriate method to analyse these patterns. The Autoregressive Integrated Moving Average (ARIMA) is a time series forecasting method that combines the autoregressive and moving average components. The general form of the ARIMA model is expressed as follows:

$$\left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right)(1 - B)^d Y_t = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right) e_t.$$

Estimating the unknown parameters in the ARIMA model can be challenging, especially when forecasting multiple time series simultaneously. However, the auto-ARIMA method addresses this issue by automatically testing various combinations of parameters combinations to select the model with the smallest value according to AIC (Akaike Information Criterion).

To measure the accuracy of the selected model and the forecast results, the Symmetric Mean Absolute Percentage Error (SMAPE) [10,11] is used, which is mathematically represented as follows:

$$SMAPE = \frac{1}{n} \sum_{t=1}^{n} \frac{\left|Y_t - \hat{Y}_t\right|}{\left(\frac{|Y_t| + \left|\hat{Y}_t\right|}{2}\right)} \times 100\%.$$

Finally, data normalization is applied as an initial step. This process is used to eliminate size effects and differences in scale [12]. The normalization techniques employed in this study include z-score normalization [13], min-max normalization [9,14], and total-sum normalization [15].

3. Results & Discussion

The dataset for tourist arrivals in Australia is organized into four hierarchical levels. First, at the national level (level 0), there is one series representing the total number of tourists across the entire country. Next, at the state level (level 1), there are seven series representing each state in Australia. At the regional level (level 2), the tourist data is divided into 77 series that represent different regions within each state, resulting in a total of 77

series at this level. Finally, at the purpose of visiting level (level 3), there are four series representing each purpose of visit within each region, resulting in a total of 308 series at this level. With each series observed over 76 quarters, the total number of observations at the lowest level is 23,408.

Figure 1 displays tourist arrivals by state, highlighting significant differences between states. Figure 2 provides a more detailed view by showing tourist arrivals across regions within each state. Each line in the panels represents a region, and the color variations allow comparison of their trends. The figure clearly reveals considerable differences between regions, with some regions consistently having higher number of tourist arrivals than others within the same state. However, Figure 3 presents a slight variation by showing tourist arrivals by purpose of visit. Within each purpose, the different colored lines represent the trends for each region. While differences between regions are evident, the gaps are less pronounced as compared to the state- and region-level variations shown in Figures 1 and 2. As a result, this study applies several normalization methods to address the scale differences caused by these significant variations in tourist arrivals.

Figure 4 illustrates the series trend after applying normalization. Each normalization method shows only minor differences compared to non-normalized data, except for the z-score normalization, which presents more uniform data with smaller gaps between states. Nonetheless, a closer look reveals that normalization methods do have a notable impact on the tourist data, though gaps still persist in the min-max and total-sum normalization graphs.

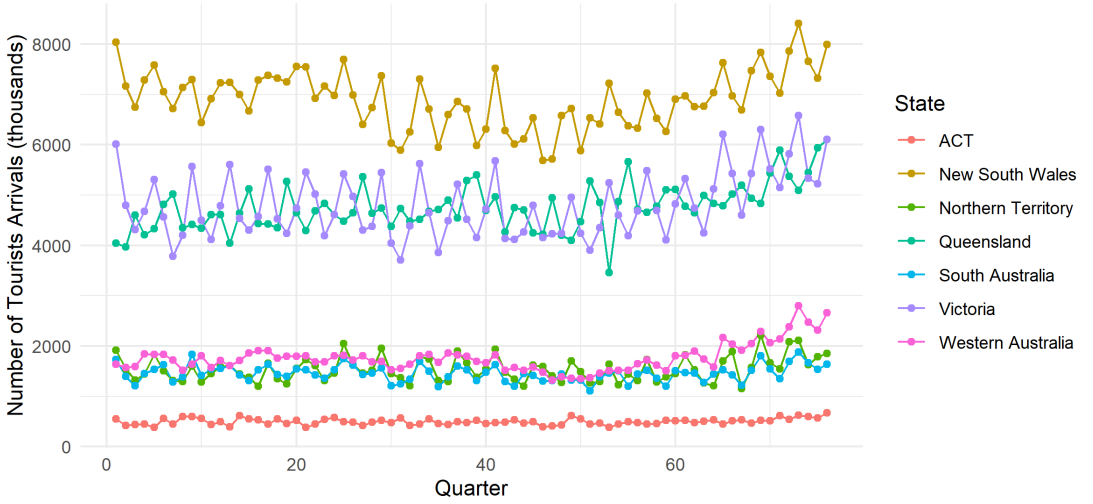

Figure 1. Plot of tourist arrivals in Australia by state.

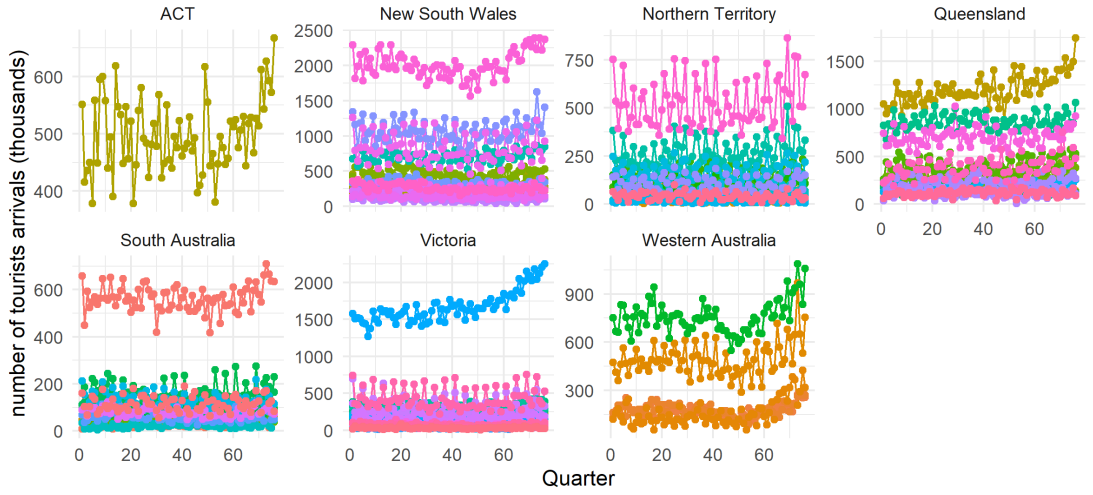

Figure 2. Plot of tourist arrivals in Australia by region and state.

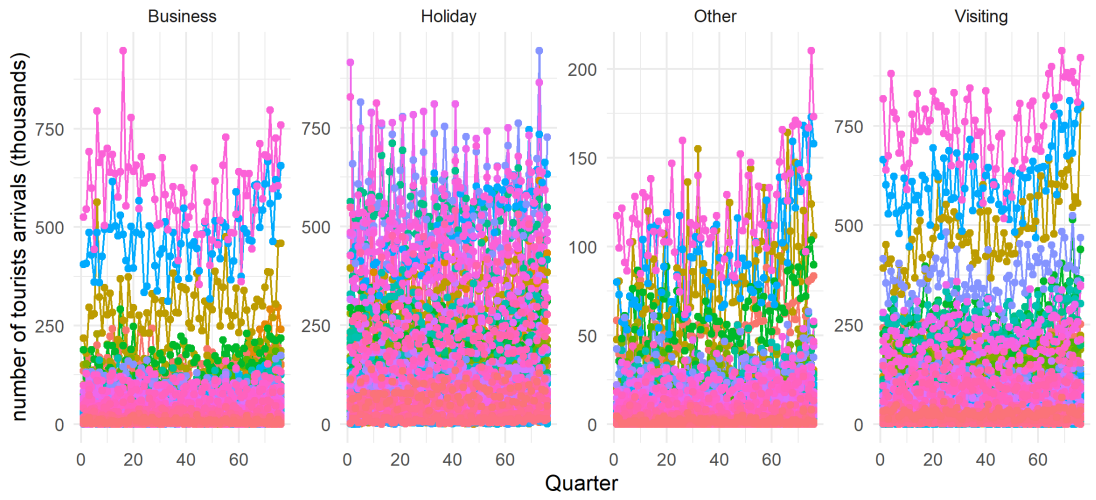

Figure 3. Plot of tourist arrivals in Australia by purpose of visit.

Figure 4. Plot of tourist arrivals in Australia before and after normalization.

Each series of tourist arrivals at the lowest level, both before and after normalization, is divided into two parts: testing and training sets. Each training set consists of the first 76 quarters, while the remaining 12 quarters constitute the test set. The ARIMA model is built on the training set for each series, resulting in 308 different fitted models. The fitted models are then used to forecast each test series at the lowest level. The forecasted series are subsequently aggregated at the next level summing up the forecasted series from the level below. The performances of the predictive model are summarised using SMAPE as tabulated in Table 1. The normalization method appears to yield better forecasting results across all levels, particularly when utilizing the z-score method as compared to other normalization approaches. However, the performance shows only a slight improvement compared to results without normalization. Particularly at the regional level, non-normalization approach performs better for short-term forecasts in quarters 1 and 2. In general, the model provides greater forecasting precision for short-term horizons, but prediction errors increase as the forecasting horizon extends, regardless of the normalization methods used. Additionally, the table shows that as the hierarchical level increases, the SMAPE values tend to decrease, indicating improved performance at higher levels.

Although it appears that a particular normalization method performs better at certain hierarchical levels, the differences in SMAPE values between normalization methods are not notable at each hierarchical level across all forecasting periods. This indicates that it cannot be conclusively determined that one normalization method is superior to others used in this study. Across all hierarchical levels, it is observed that as the forecasting horizon lengthens, SMAPE values generally increase at each level, except at the lowest level. For short-term forecasts, the evaluation results show lower values. This is consistent with the findings of studies [3,16], which report higher

SMAPE values for long-term forecasts. Thus, it can be concluded that, for the case study discussed, the performance of the ARIMA method with a bottom-up strategy diminishes for long-term forecasting.

Table 1. SMAPE (%) for test set forecasting of different normalization approaches to tourist arrivals in Australia.

	Forecast Horizon (Quarter)												
	1	2	3	4	5	6	7	8	9	10	11	12	Average
Level 0: Australia													
Non-normalization	10.56	9.83	6.92	7.68	9.71	10.21	10.33	11.44	12.88	13.12	13.27	14.16	10.84
Z-Score	**9.66**	**8.87**	**6.18**	**6.90**	**8.87**	**9.34**	**9.52**	**10.62**	**12.03**	**12.25**	**12.44**	**13.33**	**10.00**
Min-Max	10.56	9.83	6.92	7.68	9.71	10.21	10.33	11.44	12.88	13.12	13.27	14.16	10.84
Total-Sum	10.78	9.94	6.91	7.64	9.68	10.18	10.28	11.39	12.83	13.06	13.21	14.10	10.83
Level 1: State													
Non-normalization	10.72	11.24	10.55	10.31	12.12	12.17	12.30	13.49	15.02	15.27	15.38	16.31	12.91
Z-Score	**9.97**	**9.78**	**9.71**	**9.72**	**11.50**	**11.39**	**11.65**	**12.72**	**14.25**	**14.43**	**14.59**	**15.46**	**12.10**
Min-Max	10.72	11.24	10.55	10.31	12.12	12.17	12.30	13.49	15.02	15.27	15.38	16.31	12.91
Total-Sum	10.82	11.29	10.62	10.35	12.15	12.19	12.31	13.49	15.02	15.27	15.37	16.30	12.93
Level 2: Regional													
Non-normalization	**21.36**	**19.75**	19.99	**19.73**	20.77	20.59	21.14	21.50	22.31	22.44	22.85	23.26	21.31
Z-Score	22.14	19.81	**19.97**	19.77	**20.72**	**20.45**	**21.00**	**21.27**	**22.01**	**22.14**	**22.53**	**22.91**	**21.23**
Min-Max	21.36	19.75	19.99	**19.73**	20.77	20.59	21.14	21.50	22.31	22.44	22.85	23.26	21.31
Total-Sum	21.39	19.76	19.99	**19.73**	20.77	20.59	21.14	21.49	22.30	22.43	22.84	23.25	21.31
Level 3: Purpose of travel													
Non-normalization	54.41	50.64	50.66	49.50	48.91	48.59	48.77	48.95	48.94	48.59	48.60	48.64	49.60
Z-Score	**53.76**	**49.85**	**50.05**	**48.97**	**48.45**	**48.07**	**48.18**	**48.31**	**48.27**	**47.91**	**47.95**	**47.95**	**48.98**
Min-Max	54.41	50.64	50.66	49.50	48.91	48.59	48.77	48.95	48.94	48.59	48.60	48.64	49.60
Total-Sum	54.38	50.62	50.65	49.48	48.90	48.58	48.76	48.94	48.93	48.58	48.59	48.62	49.59

Values in bold represent the lowest SMAPE observed at each level.

4. Conclusions

This research focuses on evaluating the performance of ARIMA model using bottom-up strategy and normalization approaches for hierarchical time series of tourist arrivals in Australia from 1998 to 2016. The model-building process and forecasting of future values are carried out at the lowest level, with the forecasted values then aggregated at higher level. The evaluation, using SMAPE, indicates that the ARIMA model performs better for short-term forecasting horizon. Normalization is an effective method to minimize the difference in range or unit when dealing with multiple time series simultaneously that may have varying scales. However, our study reveals that z-score normalization only slightly improves forecasting results. For future research, investigating the performance of other forecasting models, such as machine learning techniques or advanced time series methods, could be valuable to determine if they offer superior results compared to ARIMA, especially for long-term forecasts.

Author Contributions

M.Y.W.: Conceptualization, Methodology, Formal Analysis, Writing—Original draft preparation, Reviewing and editing; N.F.: Conceptualization, Data curation, Validation, Supervision, Writing—Reviewing and editing; N.R.S.: Software Implementation, Visualization, Writing—Original draft preparation. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The data supporting this study were obtained from a publicly available dataset hosted on Kaggle (Quarterly Tourism In Australia, available at: https://www.kaggle.com/datasets/luisblanche/quarterly-tourism-in-australia). The dataset is open access and can be freely downloaded for research purposes.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Asesh, A. Normalization and Bias in Time Series Data. In *Digital Interaction and Machine Intelligence*, *Proceedings of MIDI'2021–9th Machine Intelligence and Digital Interaction Conference, Warsaw, Poland, 9–10 December 2021*; Biele, C., Kacprzyk, J., Kopeć, W.; et al., Eds.; Lecture Notes in Networks and Systems; Springer International Publishing: Cham, Switzerland, 2021; Volume 440.
2. Athanasopoulos, G.; Ahmed, R.A.; Hyndman, R.J. Hierarchical Forecasts for Australian Domestic Tourism. Available online: http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/ (accessed on 10 August 2025).
3. De Gooijer, J.G.; Hyndman, R.J. 25 years of IIF time series forecasting: A selective review. *Int. J. Forecast.* **2006**, *22*, 443–473. https://doi.org/10.1016/j.ijforecast.2006.01.001.
4. Filzmoser, P.; Walczak, B. What can go wrong at the data normalization step for identification of biomarkers? *J. Chromatogr. A* **2014**, *1362*, 194–205. https://doi.org/10.1016/j.chroma.2014.08.050.
5. Hyndman, R.J.; Ahmed, R.A.; Athanasopoulos, G.; et al. Optimal combination forecasts for hierarchical time series. *Comput. Stat. Data Anal.* **2011**, *55*, 2579–2589.
6. Kartikasari, P.; Suhartono, S. Prediksi penjualan di perusahaan ritel dengan metode peramalan hirarki berdasarkan model variasi kalender. *J. Sains Dan Seni Pomits* **2013**, *2*, 2337–3520.
7. Ma, E.; Liu, Y.; Li, J.; et al. Anticipating Chinese tourists arrivals in Australia: A time series analysis. *Tour. Manag. Perspect.* **2016**, *17*, 50–58.
8. Mahkya, D.A.; Ulama, B.S. Hierarchical time series bottom-up approach for forecast the export value in Central Java. *J. Phys. Conf. Ser.* **2017**, *893*, 012033.
9. Majer, K.K.; Hryniewicz, O. Data-mining approach to finding weights in the model averaging for forecasting of short time series. In *Advances in Fuzzy Logic and Technology*; Springer: Cham, Switzerland, 2017.
10. Mancuso, P.; Piccialli, V.; Sudoso, A.M. A machine learning approach for forecasting hierarchical time series. *Expert Syst. Appl.* **2021**, *182*, 115102.
11. Ali, P.J.M.; Faraj, R.H.; Koya, E.; et al. Data Normalization and Standardization: A Technical Report. *Mach. Learn. Tech. Rep.* **2014**, *1*, 1–6.
12. Parvey, Y.; Shaadid, S. Univariate time series prediction of wind speed with a case study of Yanbu, Saudi Arabia. *Int. J. Adv. Trends Comput. Sci. Eng.* **2021**, *10*, 257–264.
13. Patro, S.G.; Sahu, K.K. Normalization: A Preprocessing Stage. *arXiv* **2015**, arXiv:1503.06462.
14. Soto-Ferrari, M.; Chams-Anturi, O.; Escorcia-Caballero, J.; et al. Evaluation of Bottom-Up and Top-Down Strategies for Aggregated Forecasts: State Space Models and ARIMA Applications. In *Computational Logistics*; Paternina-Arboleda, C., Voß, S., Eds.; Springer: Cham, Switzerland, 2019; Volume 11756.
15. Susilaningrum, D.; Susanti, R. Prediksi penjualan sepeda motor merek 'X' di kabupaten dan kotamadya Malang dengan metode peramalan hierarki. *J. Sains Dan Seni Pomits* **2014**, *3*, 2337–3520.
16. Walach, J.; Filzmoser, P.; Hron, K. Data Normalization and Scaling: Consequences for the Analysis in Omics Sciences. In *Comprehensive Analytical Chemistry*; Elsevier: Amsterdam, The Netherlands, 2018; pp. 165–196.