Proceedings of Science and Mathematics
Faculty of Science, Universiti Teknologi Malaysia
Vol. 22, 2024, page 136-144

Utilization of the Fractional Holt-Winters Model for Forecasting Tourist Arrivals from Singapore to Malaysia

Pang Zhen Cong, Ani Shabri*
Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia
*Corresponding author: ani@utm.my

Abstract
The effectiveness of the Holt-Winters model in time-series forecasting is compromised in scenarios involving low data values and pronounced seasonal variations. This issue is particularly significant in the tourism industry, where such data irregularities can lead to substantial inaccuracies in forecasting tourist arrivals, affecting resource planning and strategic decision-making. This study, situated in the field of applied statistics and econometrics within tourism economics, investigates the potential of the Fractional Holt-Winters (FHW) model, a novel adaptation initially proposed for energy forecasting, to address these challenges. The primary objectives are to evaluate the FHW model's effectiveness in forecasting tourist arrivals from Singapore to Malaysia and to compare its accuracy against traditional models, including the conventional Holt-Winters, single exponential smoothing, and double exponential smoothing models. The methodology involves using Visual Studio 2022 to run Python code and Microsoft Excel for forecasting and data visualization. The Low Memory Broyden–Fletcher–Goldfarb–Shanno algorithm is used to optimize the parameters of the FHW model. Results show that simpler models such as Single Exponential Smoothing and Double Exponential Smoothing capture long-term trends but fail to predict short-term seasonal variations accurately, with RMSE values of 114,450.5235 and 89,312.2602, respectively. The conventional Holt-Winters model demonstrates improved accuracy with an RMSE of 81,605.4226, effectively capturing both trend and seasonal patterns. In contrast, the FHW model showed the highest RMSE of 188,760.6593, indicating it is less effective in this context. The findings suggest that while the traditional Holt-Winters model remains the most reliable for forecasting tourist arrivals due to its superior accuracy in capturing both trend and seasonal patterns, the FHW model requires further refinement to be viable in this application. Policymakers and industry stakeholders are advised to rely on the conventional Holt-Winters model for more accurate forecasting in the face of seasonal volatility.
Keywords Fractional Holt-Winters Model; Tourist Arrivals Forecasting; Time-Series Analysis; Exponential Smoothing; Tourism Economics; Forecast Accuracy.

Introduction

The Holt-Winters model is highly effective for time-series forecasting, adeptly capturing trends and seasonality, making it particularly useful for predicting future values in data with similar patterns. However, its handling of multiplicative seasonality is problematic, especially with very low data values, which can significantly impact the accuracy and reliability of forecasts. This limitation is critical in sectors like retail and tourism, where precise demand forecasting is essential for strategic decision-making.

Despite its widespread use, the Holt-Winters model's effectiveness is compromised in scenarios involving low data values. This issue is particularly pronounced in the tourism industry, affecting the accuracy of forecasting tourist arrivals and, consequently, resource planning. Newer models like ARIMA and LSTM are being explored for their potential to handle complex patterns, but they still face challenges in effectively forecasting time series with strong seasonal fluctuations.

This study investigates the potential of the Fractional Holt-Winters (FHW) model, a novel adaptation initially proposed for energy forecasting, to address these shortcomings. By comparing its performance against traditional models like Holt-Winters, single, and double exponential smoothing models, this research aims to determine if the FHW model can provide more accurate forecasts for tourist arrivals from Singapore to Malaysia.

The study focuses on evaluating the effectiveness of the Fractional Holt-Winters model in forecasting tourist arrivals from Singapore to Malaysia and comparing its accuracy against traditional forecasting models. The geographical focus is on tourist arrivals from Singapore to Malaysia, analyzing data specific to this international travel corridor. The time frame considered is from 2012 to 2019, to avoid disruptions caused by the COVID-19 pandemic, ensuring the consistency and reliability of the data set. The methodological approach involves using the FHW model and comparing its performance with conventional forecasting models like the standard Holt-Winters, single exponential smoothing, and double exponential smoothing models. The study utilizes publicly available data on tourist arrivals from the MyTourismData portal, ensuring that the findings are based on reliable and accessible sources. The research primarily focuses on assessing the accuracy of the forecasting models without delving into the economic impacts or policy implications of fluctuating tourist numbers. Limitations of the study are acknowledged, and areas for future research are proposed, including exploring different variations of the model, applying it to different geographical regions, or integrating it with other forecasting tools.

By addressing these objectives and scopes, this study aims to offer a more reliable tool for policymakers and industry stakeholders in the face of seasonal volatility.

The Root Mean Square Error (RMSE) is a robust metric largely unaffected by the number of trials, enabling the identification of poorly fitting models, calculation of confidence intervals, and performance of power analyses. Through two simulation studies, it was determined that RMSE effectively discriminates between well-fitting and poorly fitting models, except when the trial numbers are very small. Additionally, the impact of measurement noise, the presence of outliers, and the number of estimated parameters on RMSE values was analyzed [1].

Studies on forecasting tourist arrivals in Kenya utilizing statistical time series modeling techniques such as Double Exponential Smoothing and Auto-Regressive Integrated Moving Average (ARIMA) have shown that forecasting plays a vital role in decision-making processes, such as inventory replenishment and staffing adjustments to meet future service demands. Error measures like MAPE and RMSE were used to ascertain the best model, with Double Exponential Smoothing performing better than ARIMA models [2].

The ARIMA and Seasonal ARIMA (SARIMA) models, typical in statistical econometrics, are prevalent for energy forecasting. Modifications and combinations of these models, such as fractional difference ARIMA and Reg-SARIMA-GARCH, have been adapted for specific data characteristics and have shown promising results in fields like energy consumption where data exhibit seasonal fluctuations [3].

In a study evaluating the forecasting performance of SARIMA, Holt-Winters, and Grey models for foreign tourist arrivals in India from 2003 to 2015, it was found that SARIMA and Modified Holt-Winters models outperformed others according to the MAPE criterion. However, the Grey model, while less effective for seasonal data, showed better performance for de-seasonalized series. This study highlighted the importance of model selection based on data characteristics, as evidenced by turning point analysis and U-statistics [4].

Syafei et al. (2018) state that seasonal and trend data patterns are incorporated into the model. The Exponential Smoothing Holt Winter model integrates these two aspects. The Holt Winter model has been applied to the prediction of electrical power requirements, which has daily data showing seasonal patterns and trends to enhance performance. This model does not consider the stationarity of data but instead uses repetitive steps and past weighting values to obtain new predictive values. Daily data for air pollutant concentrations show seasonal patterns and trends. The similarity in the seasonal patterns and trends of electricity requirements and air pollutant concentrations data indicates that the Holt Winter model can be used to predict the concentrations of air pollutants [5].

The Broyden-Fletcher-Goldfarb-Shanno (BFGS) method is categorized as the Quasi-Newton (QN) method which has the Davidon-Fletcher-Powell (DFP) formula attribute of having a positive definite Hessian matrix. The BFGS method requires large memory in executing the program so another algorithm to decrease memory usage is needed, namely Low Memory BFGS (LBFGS). This is further supported by Zhou et al. (2022) as they have shown that the largest advantage of this algorithm is that it can deal with nonlinear problems in their studies where the superiority of the LBFGS algorithm in a Quasi-Newton method in dealing with nonlinear sequence is fully verified [6].

Kotsialos et al. (2005) state the basic advantage of Holt-Winter's model is that they may capture the unknown nonlinear structure of the process to be modelled. However, their basic disadvantages are the high number of parameters to be calibrated and the black-box approach that renders the plausible interpretation of the modeling structure very difficult and excludes the possibility of ad-hoc parameter choice [7].

Grey system models are advantageous for predictions due to their minimal data requirements and flexibility with data types, which are often constrained by the volume and distribution of data. These models employ a grey operator prior to prediction, laying the foundational framework for the grey prediction model. The grey one-order accumulation generation operator, a core component in grey prediction theory, effectively smooths the original data, allowing for a more effective discovery of underlying data patterns. The focus on refining grey operators to enhance predictive accuracy has seen a steady uptick recently. In the spirit of giving precedence to fresh information, Wu et al. introduced the concept of fractional accumulation operators. These operators were developed to address the contradiction between new data's lower-impact solutions and the emphasis on 'new information priority'. Following this advancement, fractional accumulation operators have been integrated into a variety of models, such as GM(1,1), GM(2,1), and NDGM(1,1), gaining significant attention in the field. Zeng et al. then incorporated fractional accumulation operators into the GM(2,1) model, yielding promising results in both numerical simulations and practical applications [8].

The article by Hu (2021) provides valuable insights into the application of fractional order models within grey system theory, which is relevant to understanding the theoretical underpinnings of the FHW model. The fractional order grey model (FGM) discussed in the article extends the traditional grey model by incorporating fractional calculus, enhancing its flexibility and accuracy in handling time series data with irregular trends and seasonal variations. This model has demonstrated superior performance in various forecasting scenarios, such as energy consumption and financial markets, suggesting its potential applicability in tourism forecasting. The article highlights that FGM outperforms traditional models like Single Exponential Smoothing (SES), Double Exponential Smoothing (DES), and standard grey models in terms of accuracy and robustness. The use of root mean square error (RMSE) as a performance metric in the article aligns with your proposed method for evaluating the accuracy of the FHW model in predicting tourist arrivals [9].

The article by Hoang-Sa Dang, Thuy-Mai-Trinh Nguyen, Chia-Nan Wang, Jen-Der Day, and Thi Minh Han Dang (2020) provides valuable insights into the application of hybrid forecasting models, combining traditional and modern techniques to improve forecasting accuracy. The study emphasizes the importance of integrating different forecasting methods to handle the complexities of real-world data. Specifically, the authors discuss the hybridization of grey models with other time series forecasting techniques, highlighting how these models can effectively address the limitations of single-method approaches. The article explores the use of grey system theory in conjunction with exponential smoothing methods to enhance forecasting accuracy. This approach is particularly relevant for datasets with incomplete or uncertain information, where traditional models may struggle. The authors demonstrate that hybrid models, which incorporate elements of grey system theory, can outperform standard forecasting methods like SES, DES, and HWM in terms of accuracy and robustness. This finding underscores the potential of the Fractional Holt-Winter's Model (FHW), which integrates grey system theory, to provide more reliable forecasts for complex time series data, such as tourist arrivals [10].

The article by Ma, Liu, and Wang (2018) introduces a novel nonlinear multivariate grey Bernoulli model (NMGBM) designed to predict tourist income in China. This innovative model combines grey system theory with Bernoulli processes to enhance forecasting accuracy, especially in scenarios characterized by limited and uncertain data. The NMGBM's development involves integrating the grey Bernoulli model with nonlinear multivariate analysis, addressing the limitations of traditional grey models that typically handle linear relationships and single-variable datasets. By incorporating nonlinear dynamics, the NMGBM can capture more complex patterns and interactions within the data. In their study, Ma, Liu, and Wang applied the NMGBM to predict China's tourist income, utilizing historical data on tourist arrivals, income levels, and other relevant economic indicators. The case study with real-world data demonstrates the practical application of the NMGBM in a tourism context. Performance evaluation using mean absolute percentage error (MAPE) as the primary metric shows that the NMGBM significantly outperforms traditional grey models and other forecasting techniques. This superior performance is attributed to the model's ability to handle nonlinear relationships and multivariate interactions more effectively. The NMGBM offers several advantages, particularly in situations where data is scarce, incomplete, or highly uncertain—a common scenario in tourism forecasting. Its capacity to incorporate multiple variables and nonlinear dynamics makes it highly adaptable to different forecasting contexts, including economic and tourism data. The successful application of the NMGBM to predict tourist income in China underscores its potential utility in other areas of tourism forecasting, such as predicting tourist arrivals or spending patterns. The model's adaptability and accuracy suggest it could be a valuable tool for tourism planners and policymakers, providing more reliable forecasts to support strategic decision-making [11].

Fractional Holt-Winter's Model

Step 1: Suppose that $X^{(0)} = \left(x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)\right)$ is a nonnegative primitive sequence with seasonality and nonlinearity, and the period is L (L=4 *or* L=12). Then, the r-order accumulation sequence can be obtained by fractional periodic accumulation operator, which is recorded as $X^{(r)} = \left(x^{(r)}(1), x^{(r)}(2), \dots, x^{(r)}(n)\right)$. The fractional periodic accumulation operator is expressed as:

$$x^{(r)}(k) = \sum_{i=\left(\left\lceil\frac{k}{L}\right\rceil-1\right)L+1}^{k} \frac{\Gamma(r+k-i)}{\Gamma(k-i+1)\Gamma(r)} x^{(0)}(i), k = 1,2,3,\dots,n$$

where $\left\lceil\frac{k}{L}\right\rceil$ represents an integer not less than k/L and r represents fractional order.

The corresponding reduction formula can be expressed as:

$$x^{(0)}(k)\begin{cases} \left(x^{(r)}\right)^{(-r)}(k) = \sum\limits_{i=0}^{(k\bmod L)-1} (-1)^i \frac{\Gamma(r+1)}{\Gamma(i+1)\Gamma(r-i+1)} x^{(r)}(k-i) & k\bmod L \neq 0,1 \\ \left(x^{(r)}\right)^{(-r)}(k) = \sum\limits_{i=0}^{(k\bmod L)+3} (-1)^i \frac{\Gamma(r+1)}{\Gamma(i+1)\Gamma(r-i+1)} x^{(r)}(k-i) & k\bmod L = 0,1 \\ x^{(r)}(k) & k\bmod L = 1 \end{cases}$$

Step 2: Establish the model

$$S_k = \alpha \frac{x^{(r)}(k)}{C_{k-L}} + (1-\alpha)(S_{k-1} + b_{k-1}), 0 < \alpha < 1$$
$$b_k = \beta(S_k - S_{k-1}) + (1-\beta)b_{k-1}, 0 < \beta < 1, \qquad k = L+1, L+2, \dots, n$$
$$C_k = \gamma \frac{x^{(r)}(k)}{S_k} + (1-\gamma)C_{k-L}, 0 < \gamma < 1$$

The initial value of the model is

$S_L = x^{(r)}(L);$

$$C_i = \frac{x^{(r)}(i)}{\frac{\sum_{j=1}^{L} x^{(r)}}{L}}, i = 1, 2, 3, \dots, L;$$

$$b_L = \frac{1}{L}\left(\frac{x^{(r)}(L+1) - x^{(r)}(1) + x^{(r)}(L+2) - x^{(r)}(2) + \dots + x^{(r)}(L+L) - x^{(r)}(L)}{L}\right)$$

where α is data smoothing factor, β is trend smoothing factor, γ is the seasonal change smoothing factor, *S* denotes level, *b* denotes trend, and *C* denotes seasonal.

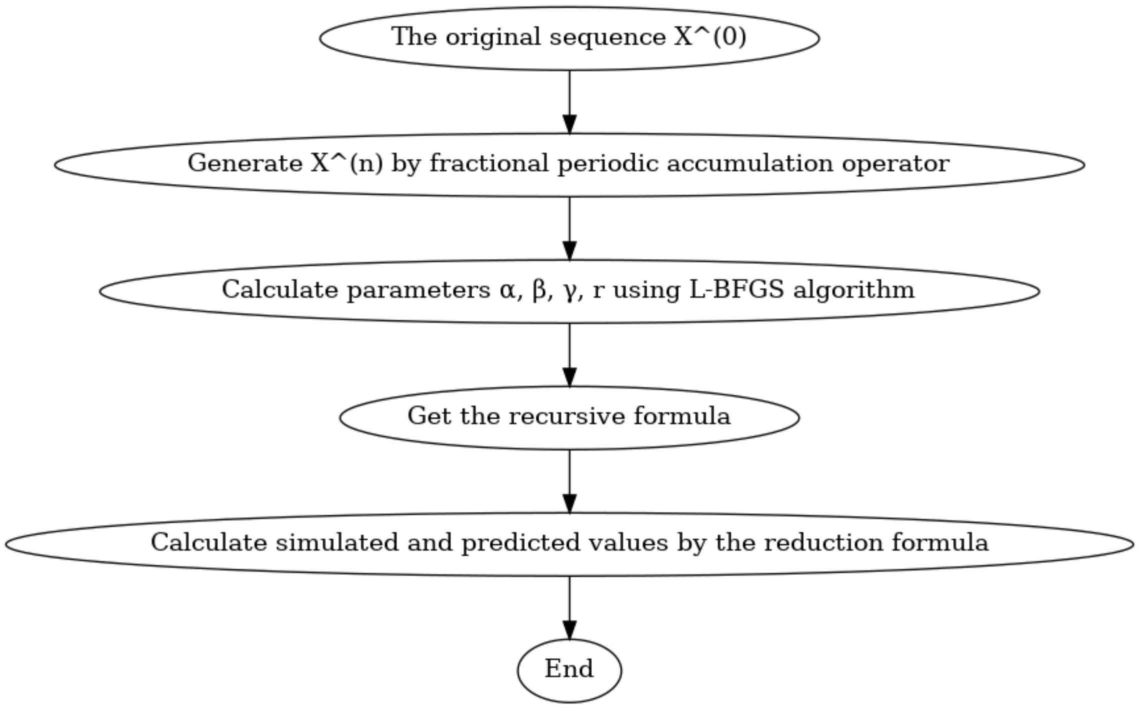

Figure 1 Flow chart of Fractional Holt-Winter's Model

Results and Discussions

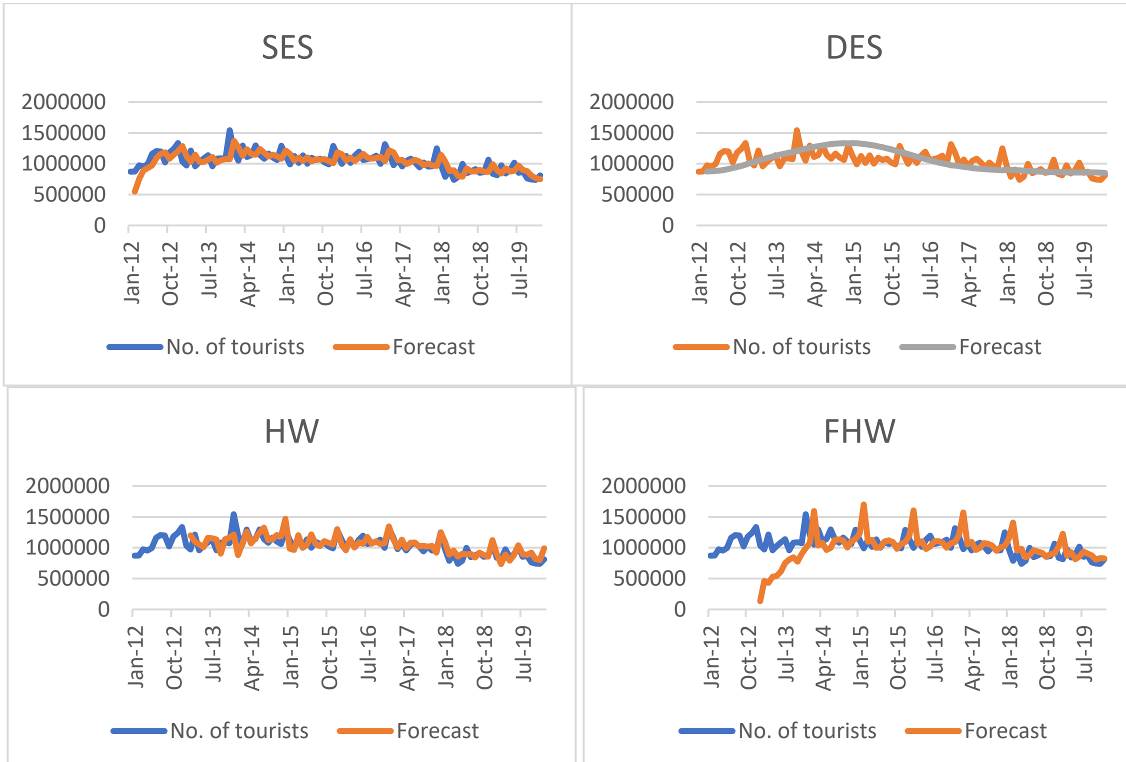

Figure 2 Forecast graphs of four models for monthly tourists arrival from Singapore to Malaysia

RMSE values of different forecasting models are compared to determine their accuracy ranking; the lower the RMSE, the more accurate the model. The table below ranks the models based on their RMSE values, indicating their accuracy.

Rank	Model	RMSE
1	Holt-Winter's	81605.4226
2	Double Exponential Smoothing	89312.2602
3	Single Exponential Smoothing	114450.5235
4	Fractional Holt-Winter's	188760.6593

Table 1 Accuracy rankings for the four models implemented

The single exponential smoothing model initially starts with a lower forecast but quickly adjusts to align with the actual tourist numbers by mid-2012. From 2013 onwards, it closely follows the general trend and seasonality of the actual data, although it smooths out some pronounced fluctuations. The model captures the main trends and shows peaks and troughs, albeit with some lag during sharp changes, such as the peaks around 2014 and the declines in later years. From 2015 onwards, the model's forecasts remain fairly accurate, effectively capturing the overall pattern without overfitting to short-term noise. However, it occasionally underestimates or overestimates sudden changes, particularly during 2018 and 2019. The single exponential smoothing model provides a reasonable and stable forecast, effectively capturing long-term trends while smoothing out minor short-term variations. This makes it suitable for scenarios where simplicity and general trend capture are prioritized over detailed seasonal adjustments. However, its RMSE of 114,450.5235 indicates it is less accurate than the other models due to its simplistic approach.

Initially, from 2012 to mid-2014, the double exponential smoothing model effectively captures the upward trend and seasonal variations, although it slightly overestimates the number of tourists during peak periods around 2013 and 2014. From mid-2014 onward, the model reflects a more stabilized trend in tourist numbers. It smoothens out many of the fluctuations seen in the actual data, indicating the model's focus on trend capturing rather than detailed seasonal variations. This results in underestimating some of the more pronounced rises and falls, particularly around 2015 and 2016. As the data progresses from 2016 to 2019, the model follows the overall declining trend but smooths out the data significantly, resulting in a less volatile forecast line compared to the actual data. This smoothing effect means the model does not fully capture the smaller, more frequent fluctuations. While the double exponential smoothing model provides a good indication of the overarching trend and mitigates some noise in the data, it sacrifices the ability to accurately predict short-term seasonal variations. This characteristic makes it useful for long-term trend analysis but less effective for detailed short-term forecasting. Its RMSE of 89,312.2602 reflects this trade-off, as it is more accurate than single exponential smoothing but less so than Holt-Winter’s.

From early 2012 until mid-2014, the Holt-Winter’s model captures the general upward trend and seasonal fluctuations in tourist numbers, including several pronounced peaks around the end of each year. However, there are discrepancies between the actual and forecasted values, with significant deviations observed around 2014, where the model overestimates the peak in tourist numbers. Post-2014, the model continues to reflect the stabilized trends in tourist numbers and captures the periodic dips and rises that occur approximately on an annual basis. From 2015 onwards, the forecast aligns more closely with the actual data, suggesting improved accuracy over time. However, there are still instances where the forecast does not fully capture the extent of variability in tourist numbers, particularly during mid-2016 and throughout 2017. In the later years, particularly from 2018 to 2019, the model maintains its ability to follow the general trend but sometimes misses the exact amplitude of changes in tourist numbers. Overall, the Holt-Winter’s model provides a generally accurate forecast, capturing the primary trend and seasonal patterns in the number of tourists. However, discrepancies between actual and forecasted values highlight the challenge of modeling complex real-world data, suggesting potential areas for further refinement. Its RMSE of 81,605.4226, the lowest among the methods, confirms it as the most accurate model due to its comprehensive approach in modeling seasonality, trend, and level.

The Fractional Holt-Winter’s model shows a clear increase in the number of tourists from January 2012 to mid-2014, with regular seasonal patterns and identifiable peaks and troughs. After mid-2014, the actual number of tourists becomes more volatile, showing significant fluctuations but an overall decreasing trend. The forecast sometimes closely matches the actual data, but notable periods of overestimation and underestimation arise. The model captures the general seasonal pattern and aligns reasonably well with the overall trend over longer periods. However, it struggles with high volatility in actual tourist numbers, often smoothing out sharp peaks and troughs, suggesting the model might be too conservative or not responsive enough to sudden changes. Significant deviations between the forecast and actual values indicate weaknesses in predicting short-term variations accurately. Overall, the Fractional Holt-Winter’s model provides a reasonable forecast of seasonal patterns and long-term trends but requires improvements to enhance responsiveness to sudden changes and better capture the declining trend observed in the later years. Its RMSE of 188,760.6593, the highest among the methods, indicates it is the least accurate, suggesting potential misalignment with the data characteristics.

In summary, the Holt-Winter’s method emerges as the most accurate for forecasting tourist numbers, likely due to its comprehensive approach in modeling seasonality, trend, and level. Double Exponential Smoothing, while useful, is less effective without accounting for seasonality. Single Exponential Smoothing is even less accurate due to its simplistic approach, and Fractional Holt-Winter’s method is the least accurate, suggesting potential misalignment with the data characteristics. Each method has its strengths and weaknesses, and the choice of method should be aligned with the specific forecasting requirements and data characteristics.

Conclusion

Out of the four forecasting models for predicting tourist arrivals from Singapore to Malaysia evaluated: Single Exponential Smoothing, Double Exponential Smoothing, Holt-Winter's Model, and the Fractional Holt-Winter's Model. The Single Exponential Smoothing model, though effective for long-term trends, had a higher RMSE of 114,450.5235 due to its limitations in reflecting sudden changes. The Double Exponential Smoothing model, while reducing noise and better capturing trends, had an RMSE of 89,312.2602, indicating its struggle with short-term seasonal variations. The Holt-Winter's Model excelled in capturing both trends and seasonal patterns, achieving the lowest RMSE of 81,605.4226, thus proving its robustness for complex real-world data. These results indicate that while simpler models can provide reasonable forecasts, sophisticated methods like the Holt-Winter's Model are more effective for accurately predicting tourist arrival patterns.

Acknowledgement

I would like to express my sincere gratitude to my advisor, Associate Professor Dr. Ani Shabri, for his invaluable guidance and support throughout this research. I also extend my thanks to the Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, for providing the necessary resources and environment to conduct this study. Special thanks to my coursemates and family for their encouragement and motivation.

References

[1] A. L. Schubert, D. Hagemann, A. Voss, and K. Bergmann, "Evaluating the model fit of diffusion models with the root mean square error of approximation," *J Math Psychol*, vol. 77, 2017, doi: 10.1016/j.jmp.2016.08.004.

[2] S. Jere, A. Banda, B. Kasense, I. Siluyele, and E. Moyo, "Forecasting Annual International Tourist Arrivals in Zambia Using Holt-Winters Exponential Smoothing," *Open J Stat*, vol. 09, no. 02, 2019, doi: 10.4236/ojs.2019.92019.

[3] W. Zhou, H. Tao, and H. Jiang, "Application of a Novel Optimized Fractional Grey Holt-Winters Model in Energy Forecasting," *Sustainability (Switzerland)*, vol. 14, no. 5, Mar. 2022, doi: 10.3390/su14053118.

[4] K. Kriti, "Forecasting Foreign Tourist Arrivals in India Using Different Time Series Models," *International Journal of Business and Management*, vol. 15, pp. 38–43, 2016.

[5] A. D. Syafei, N. Ramadhan, J. Hermana, A. Slamet, R. Boedisantoso, and A. F. Assomadi, "Application of exponential smoothing holt winter and ARIMA models for predicting air pollutant concentrations," *EnvironmentAsia*, vol. 11, no. 3, 2018, doi: 10.14456/ea.2018.52.

[6] D. R. S. Saputro and P. Widyaningsih, "Limited memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) method for the parameter estimation on geographically weighted ordinal logistic regression model (GWOLR)," in *AIP Conference Proceedings*, American Institute of Physics Inc., Aug. 2017. doi: 10.1063/1.4995124.

[7] A. Kotsialos, M. Papageorgiou, and A. Poulimenos, "HOLT-WINTERS AND NEURAL-NETWORK METHODS FOR MEDIUM-TERM SALES FORECASTING," 2005.

[8] L. Wu, S. Liu, L. Yao, R. Xu, and X. Lei, "Using fractional order accumulation to reduce errors from inverse accumulated generating operator of grey model," *Soft comput*, vol. 19, no. 2, pp. 483–488, Feb. 2015, doi: 10.1007/s00500-014-1268-y.

[9] Y. C. Hu, "Forecasting tourism demand using fractional grey prediction models with Fourier series," *Ann Oper Res*, vol. 300, no. 2, 2021, doi: 10.1007/s10479-020-03670-0.

[10] H. S. Dang, T. M. T. Nguyen, C. N. Wang, J. Der Day, and T. M. H. Dang, "Grey system theory in the study of medical tourism industry and its economic impact," *Int J Environ Res Public Health*, vol. 17, no. 3, 2020, doi: 10.3390/ijerph17030961.

[11] X. Ma, Z. Liu, and Y. Wang, “Application of a novel nonlinear multivariate grey Bernoulli model to predict the tourist income of China,” *J Comput Appl Math*, vol. 347, 2019, doi: 10.1016/j.cam.2018.07.044.