Journal of Tourism & Adventure (2024), 7(1), 120-143

Journal of
Tourism & Adventure

Trend Analysis of Tourist-arrivals in Nepal Using Auto-Regressive Integrated Moving Average (ARIMA) Model

Bishnu Bahadur Khatri
Associate Professor, Central Department of Rural Development, Tribhuvan University, Kathmandu
bishnu.khatri@cdrd.tu.edu.np

Omkar Poudel
Lecturer, Birendra Multiple Campus, Tribhuvan University
omkar60475@gmail.com

Sarad Chandra Kafle
Lecturer, Birendra Multiple Campus, Tribhuvan University
sarad.kafle49@gmail.com

Pradeep Acharya
Account Specialist at NOWCFO, Kathmandu
pradeep.acharya197@gmail.com

Daya Simkhada
James Madison University, USA
dayasimkhada@gmail.com

Article
Received: August 21, 2024
Revised: August 31, 2024
Accepted: September 1, 2024

Keywords
Tourists-arrivals, trend analysis, auto regressive integrated moving average model, forecasting, Nepal

Corresponding Editor
Ramesh Raj Kunwar
kunwar.dr@gmail.com

Abstract

Nepal, a diverse and geographically varied landlocked country with a rich tapestry of cultures, and natural heritages, boasts iconic attractions such as Mt. Everest as the world's highest peak, Lumbini as the birthplace of Lord Buddha, Pashupatinath Temple as one of the world's most celebrated pilgrimages for the Hindus, Chitwan National Park, home to the one-horned rhino enlisted by United Nations Educational, Scientific and Cultural Organization world's natural treasure, etc. Attracting tourists from all corners of the globe, Nepal's unique blend of landscapes and cultural heritage has made this country a prominent destination. This research delves into a comprehensive trend analysis of tourist arrivals in Nepal from 1964 to 2023, employing the Auto Regressive Integrated Moving Average (ARIMA) model based on the secondary data employed from the

Published by: Janapriya Multiple Campus (JMC), Pokhara, Tribhuvan University, Nepal
ISSN 2645-8683

Ministry of Finance, Government of Nepal. The model provides insights into past trends and forecasts future tourist arrivals, evidenced by the rigorous test using the Root Mean Square Error. The analysis employed several ARIMA models, including ARIMA (2, 1, 3), ARIMA (2, 1, 2), ARIMA (3, 1, 3), and ARIMA (3, 1, 2), with model performance assessed using criteria such as Akaike Information Criterion (AIC) and Mean Absolute Error (MAE). Findings indicate that the ARIMA (2 1, 3) model provides the most accurate forecasts compared to the other models tested. This finding has significant implications for Nepal's tourism industry. The results underscore the importance of adopting flexible and strategic planning approaches to enhance the resilience of Nepal's tourism sector. The study's insights are valuable for policymakers, industry stakeholders, and researchers, contributing to the development of targeted strategies that ensure sustainable growth in the face of ongoing volatility.

Introduction

Nepal has established herself as one of the most preferred destinations for tourists from all around the world. Every year, a huge number of visitors land on this Himalayan country. Due to the vibrant tapestry woven with rich cultural threads, breathtaking landscapes, and unparalleled adventures, Nepal has been able to attract millions of people from across the world in the recent years. Nestled between the steamy jungles of Terai and the icy peaks of the world's highest mountains, this country offers an authentic and mesmerizing experience for passionate travelers. Boasting 8 of the 10 highest mountains globally including the iconic Mount Everest, hosting the pilgrimage of Lumbini as the birth place of Gautam Buddha among the Buddhist communities, having Pashupatinath Temple which is one of the world's most celebrated pilgrimages for the Hindus, Conserving Chitwan National Park, home to the endangered one-horned rhino enlisted by United Nations Educational, Scientific and Cultural Organization (UNESCO) as world's natural treasure, etc. Nepal stands as a hotspot for mountaineers, rock climbers, pilgrimages and those seeking thrilling adventures (*NTB*, 2023). Additionally, the hospitality of the Nepalese people, varieties of the culinary delights deeply rooted in tradition and flavor and the unique festivals observed at times, have also been the reasons behind the choices of the tourists arriving to this nation (*NTB*, 2023).

Crucially, tourism has been serving as a cornerstone for Nepal's economy, contributing significantly to its Gross Domestic Product (GDP) and providing a substantial source of foreign exchange. The World Travel & Tourism Council's research indicates that tourism, on a global scale, generates trillions of dollars annually and employs millions worldwide. In Nepal, the sector accounted for 7.9% of the country's GDP in 2018, supporting over a million jobs and demonstrating its pivotal role in economic growth (Bhattarai, 2023).

However, the tourism sector in Nepal has faced challenges, especially in the wake of the global pandemic of COVID 19 that gripped the world for nearly two years. Despite the setbacks, the industry is gradually recovering, emphasizing the need for strategic thinking and focused efforts from authorities to harness the potential of tourism as a major contributor to foreign revenue and a vital player in reducing currency deficits in Nepal. As the number of tourists and revenue show signs of an upward trajectory in recent years, exploring the current state of tourism in Nepal and its ongoing contributions becomes imperative for a comprehensive understanding of its impact on the country's economic landscape (Bhattarai, 2023).

The number of tourists visiting case looks random in nature and can't be predicted using general linear model like linear regression, but there are many nonlinear models to forecast time series values. Among them, Autoregressive Integrated Moving Average(ARIMA) is suitable method for forecasting the random-natured time series values like number of arrivals, stock prices, variety of production, etc. (Ellis, 2022).

Review of literature

Several studies have shown the success of ARIMA models in predicting tourist arrivals to different countries, including the United States (Edward et al., 2023), Indonesia (Nurhasanah et al., 2022), Surakarta City (Purwanto et al., 2019)and Zimbabwe (Makoni et al., 2023). In these studies, ARIMA models have been found outperforming other forecasting methods, such as Multilayer Perception Models (Jafridin et al., 2021) and Fuzzy time series (Makoni & Chikobvu, 2021). The use of ARIMA models in tourism forecasting is particularly valuable for decision-making and planning purposes, as accurate projections of tourist arrivals can inform resource allocation, marketing strategies, and infrastructure development.

Edward et al. (2023) employed ARIMA hybrids of multilayer perceptron (MLP) models to predict monthly tourist arrivals to the United States and Indonesia, concluding that the ARIMA model outperformed the MLP model in both cases. The study selected the MLP(6,1)-ARIMA(0,1,1)(0,1,1)12 hybrid model as the best model for forecasting monthly tourist arrivals to the United States, and the MLP(12,1)-ARIMA(0,1,1)(0,1,0)12 hybrid model has been recognized as the best model for forecasting monthly tourist arrivals to Indonesia. Similarly, Nurhasanah et al. (2022) utilized the SARIMA model to forecast foreign tourist arrivals in Indonesia, revealing a consistent increase over time, providing valuable insights for the tourism industry stakeholders. Likewise, Makoni et al. (2023) adopted the Box-Jenkins approach and SARIMA model to analyze monthly foreign tourist arrivals in Zimbabwe, offering useful data for marketing strategies and facilities planning. Jafridin et al. (2021) conducted a comprehensive analysis of forecasting methods for tourist arrivals in home stays in Pahang, Malaysia, highlighting the superiority of Fuzzy Time Series over ARIMA in terms of forecast accuracy. Makoni and Chikobvu (2021) employed an ARMA-GARCH process to model the volatility of international tourist arrivals in Zimbabwe, recommending strategies to mitigate the impact of unexpected shocks and found ARMA(1,1) and GARCH(1,1) model as best model to forecast the tourist arrivals. Also, Purwanto et al. (2019) found that the ARIMA (2,1,2) model is the best model for predicting tourist arrivals in Surakarta City. Lastly, Yollanda and Devianto (2020) utilized a SARIMA-ANN hybrid model to forecast tourist arrivals through Minangkabau International Airport, demonstrating good performance with a low MAPE value.

Hence, all these literatures do not give any contribution to the case of tourist arrival trends in Nepal. Therefore, having studied the effectiveness of all the above models, ARIMA seems appropriate in forecasting tourists' arrivals in the years ahead. The objective of this study is to determine the best ARIMA model so as to forecast the tourist arrivals based on annual time series data. This modeling is expected to contribute the academicians and researchers who aspire to apply time series model in studying cases as such. Furthermore, the findings of this research can immensely support the policy makers, industrialists, and other concerned authorizes for future planning and implementation.

Research design and methods

Quantitative research design has been used to complete this research while, descriptive and analytical approaches have been considered in analyzing the data (Poudel, 2023). For data analysis, E-views Statistical Package Version-10 is used.

Nature and sources of data

Secondary data, in time series nature, have been used for the analysis of the number of tourist arrivals in Nepal from 1964 to 2023. The main sources of the data are the different series of "Economic Surveys" published by the Ministry of Finance, Government of Nepal (Ministry of Finance, 2024).The comprehensive historical time series data (1964-2023) strengthens the study's reliability, providing a solid basis for detecting trends.

Model selection

The application of ARIMA modeling, supported by empirical evidence from many research, demonstrates methodological rigor targeted at dependable forecasting. The comparison of several ARIMA configurations guarantees that the best model is chosen based on accuracy and statistical fit.

ARIMA model

The research is based on the Box–Jenkins methodology (Box & Jenkins, 1976) as a suitable technique for time series data forecasting. ARIMA stands for Autoregressive Integrated Moving Average. It is an extension of the ARMA model that includes an additional component called differencing. Differencing is used to make non-stationary time series stationary (Poudel et al., 2024). ARIMA models assume that the time series data is stationary, meaning its statistical properties like mean and variance remain constant over time; this can be checked using plots and statistical tests such as the Augmented Dickey-Fuller test, with transformations like differencing applied if needed. Additionally, the residuals from the model should not exhibit autocorrelation, which can be verified using autocorrelation (ACF) and partial autocorrelation (PACF) plots; if autocorrelation is present, adjusting the model order or using alternatives like SARIMA might be necessary. Lastly, the model assumes a linear relationship between variables, which should be checked with residual plots and statistical tests; if non-linearity is detected, transformations or nonlinear models should be considered.

Box-Jenkins methodology has been employed to study the time series nature of the data on total number tourist arrivals. The ARIMA model, despite its limitations, remains a valuable tool for forecasting in scenarios like Nepal's tourism sector, where data may be influenced by irregular events. ARIMA's strength lies in its ability to model and predict based on historical patterns by differencing the data to remove non-stationarity, which helps in minimizing the impact of cyclical variations. Although it may not fully account for extraordinary events like the Maoist Insurgency, the 2015 earthquake, or the COVID-19 pandemic, the model's capability to handle autocorrelation in time-series data allows it to provide reasonable forecasts under stable conditions. According to Box et al. (2015), ARIMA models are particularly effective when the primary goal is to understand or predict future values in a time series, especially when other explanatory variables are unavailable or unreliable. Thus, even with its limitations, the ARIMA model offers a structured approach to forecasting that can still yield useful insights for Nepal's tourism industry.

Box and Jenkins (1976) developed a four-step method for selecting and estimating ARIMA models:

Identification: This step involves analyzing the time series data to determine whether it is stationary and to identify the orders of the AR, I, and MA components. The first step in the Box and Jenkins method is to identify the orders of the AR, I, and MA components. This can be done by examining the autocorrelation function (ACF) plot and partial autocorrelation function (PACF) plot of the time series data. The ACF plot shows the correlation between the time series and its lagged values. The PACF shows the correlation between the time series and its lagged values after removing the effect of the previous lags. The orders of the AR, I, and MA components can be identified by looking for patterns in the ACF and PACF. If the ACF dies off slowly, then the time series may be non-stationary and require differencing. If the PACF has significant spikes at certain lags, then the AR component may be of a certain order.

Estimation: This step involves estimating the parameters of the ARIMA model. Once the orders of the AR, I, and MA components have been identified, the next step is to estimate the parameters of the ARIMA model. This can be done using a variety of statistical methods.

Diagnostic checking: Once the ARIMA model has been estimated, it is important to check the model to ensure that it is well-fitting and that the residuals are white noise. This can be done by examining the residuals of the model and by performing various statistical tests such as Q Statistic, or Ljung-Box.

Forecasting: Once the ARIMA model has been estimated and checked, it can be used to forecast future values of the time series. This is done by substituting the estimated parameters into the ARIMA model equation and solving for the future values.

An ARIMA model is written as ARIMA (p,d,q) where p is the order of autoregressive terms, d is the order of non-seasonal differences needed for stationary, and q is the number of lagged forecast errors in the prediction equation. The graph, correlograms and statistical test are employed to examine the stationary of the data. The Augmented Dickey–Fuller (ADF), Phillips-Perron (PP) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests is applied as a unit root test, which helps determine if the variables satisfy the condition of stationary.

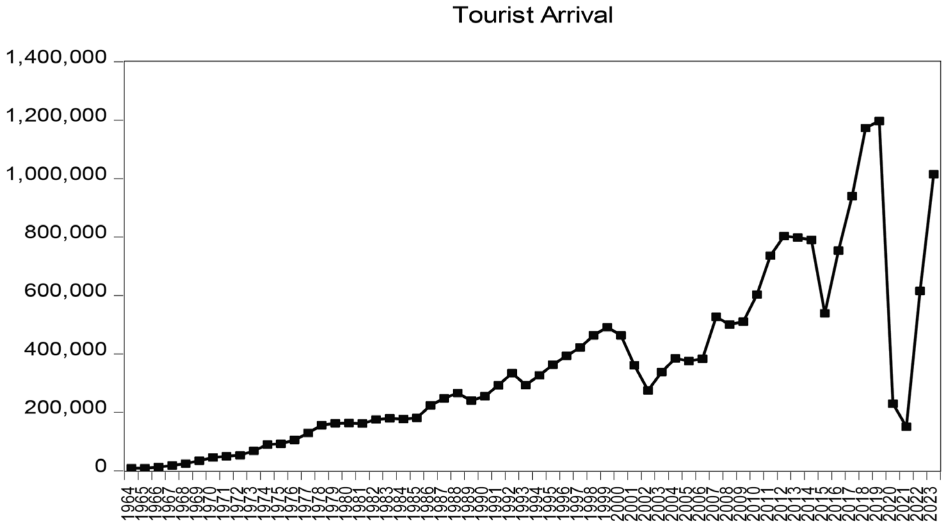
The ARIMA model is then used to forecast future tourist arrivals, with the number of forecasts generated depending on the desired forecast horizon.

Results and discussion

Number of tourist arrivals in Nepal from 1964 to 2023

Analyzing the trend of tourist arrival in Nepal from 1964 to 2023, the data reveal the rising trend until 1999.The three years including the millennium recorded a decreasing trend till 2002. Year 2019 welcomed the largest number of tourists in Nepal with 11,97,191 foreign visitors coming from airways and roadways, while the least number of tourist arrival is observed in 1965 with the arrivals of 9,388 international tourists only. Due to the Maoists insurgency, number of tourists arriving Nepal fluctuated between1999 to 2006. It created a highly unstable environment in Nepal, leading to international travel advisories against visiting the country. The tourism sector, which is a crucial part of Nepal's economy, suffered significantly. The reduction in tourist arrivals likely led to substantial revenue losses, affecting hotels, airlines, local businesses, and employment in tourism-related sectors. Tourist arrivals dropped drastically from 790,118 in 2014 to 538,970 in 2015, reflecting the immediate impact of the earthquake that struck Nepal in April 2015. The earthquake caused widespread destruction, including significant damage to heritage sites that are major tourist attractions. This, combined with concerns over safety, led to a sharp decline in tourism. Recovery began in 2016, with tourist arrivals increasing to 753,002 as reconstruction efforts and global awareness campaigns to support Nepal led to a resurgence in tourism. Again, the Nepalese tourism has been immensely affected by the global pandemic of COVID-19. It had a catastrophic impact on global travel and tourism, and Nepal was no exception. Travel bans, lockdowns, and health concerns virtually halted tourism. Since 2021, some improvement can be noticed in the tourists' arrivals in Nepal (See Figure 1).

Figure 1: Trend of tourist arrivals (1964 -2023)



Source: Author's calculations performed using E-Views

Figure 1 clearly illustrates the sensitivity of Nepal's tourism sector to external shocks such as political instability, natural disasters, and global pandemics. Each of these events led to substantial declines in tourist arrivals, highlighting the importance of building a more resilient tourism industry. Economic diversification, improved infrastructure, and better crisis management strategies could help mitigate the impacts of such events in the future. The recovery patterns also suggest that while the tourism sector is vulnerable, it is also capable of rebounding, given the right conditions and interventions.

Correlograms

Correlograms is used to visualize both autocorrelation and cross-correlation. Autocorrelation is the correlation between a variable and itself at different time lags. This can be used to identify patterns in time series data, such as seasonality or trends. Cross-correlation is the correlation between two different variables at different time lags. This can be used to identify relationships between variables, such as lead-lag relationships or causal relationships (Damos, 2016).

Table 1: Correlograms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.785	0.785	38.854	0.000
		2	0.568	-0.125	59.563	0.000
		3	0.553	0.394	79.492	0.000
		4	0.654	0.307	107.90	0.000
		5	0.649	-0.009	136.41	0.000
		6	0.565	0.071	158.39	0.000
		7	0.462	-0.137	173.40	0.000
		8	0.376	-0.207	183.52	0.000
		9	0.351	-0.015	192.52	0.000
		10	0.317	-0.186	200.01	0.000
		11	0.298	0.105	206.75	0.000
		12	0.261	0.026	212.01	0.000
		13	0.205	-0.016	215.33	0.000
		14	0.174	0.168	217.79	0.000
		15	0.158	-0.059	219.84	0.000
		16	0.131	-0.028	221.28	0.000
		17	0.104	0.036	222.22	0.000
		18	0.114	-0.016	223.37	0.000
		19	0.107	0.003	224.42	0.000
		20	0.064	-0.066	224.80	0.000
		21	0.018	-0.073	224.83	0.000
		22	-0.004	-0.042	224.83	0.000
		23	-0.022	-0.135	224.88	0.000
		24	-0.045	-0.028	225.08	0.000
		25	-0.071	-0.024	225.62	0.000
		26	-0.097	-0.011	226.65	0.000
		27	-0.120	0.055	228.28	0.000
		28	-0.151	-0.023	230.94	0.000

Source: Author's calculations performed using E-Views

In Correlogram Table 1 the first column, "AC", shows the autocorrelation coefficients. The second column, "PAC", shows the partial autocorrelation coefficients. The third column, "Q-Stat", shows the Box-Ljung statistic for testing whether the autocorrelation coefficients are significantly different from zero. The fourth column, "Prob", shows the p-value for the Box-Ljung test. The autocorrelation and partial autocorrelation analysis suggests that the time series is non-stationary and that an ARIMA model may be a good fit for the data (See Table 1).

Test of unit root

The unit root test is employed to examine the stationary of the data. The ADF, PP and KPSS tests were utilized as a unit root test, which helps determine if the variables satisfy the condition of stationary (Poudel et al., 2023; Poudel et al., 2024; Upadhyaya et al., 2022).

Table 2: Result of ADF test on tourist arrival series

		ADF H_0: Tourist arrival has a unit root.		PP H_0: Tourist_arrival has a unit root.		KPSS H_0: Tourist arrival is stationary	
		t-Statistic	Prob.	t-Statistic	Prob.	t-Statistic	Prob.
At Level	With constant	-0.8131	0.8073	-1.6224	0.4650	1.0051	***
	With Constant & Trend	-2.6134	0.2763	-4.0281	0.0129**	0.0825	No
	Without Constant & Trend	0.4756	0.8144	-0.0524	0.6613	-	-
At First Difference	With constant	-3.0213	0.0392**	-6.9969	0.0000***	0.2597	No
	With Constant & Trend	-2.8387	0.1904	-7.3508	0.0000***	0.1718	**
	Without Constant & Trend	-2.6244	0.0096***	-5.8546	0.0000***	-	-

Notes: *, ** and *** represent significance levels at 10%, 5% and 1% respectively.

Source: Author's calculations performed using E-Views

The combined results of the ADF, PP, and KPSS tests suggest that the tourist arrival series in Nepal is characterized by a unit root at the level, meaning that shocks to the series (such as the Maoist Insurgency, the 2015 earthquake, or the COVID-19 pandemic) can have long-term effects. However, after first differencing, the series becomes stationary, indicating that changes or growth rates in tourist arrivals tend to revert to a mean or stable trend over time. This combination of tests provides a robust conclusion that the tourist arrival series is stationary only after first differencing (Acharya et al., 2024) (See Table 2). For policymakers and economists, understanding these characteristics is vital. The non-stationarity at the level implies that any interventions (such as marketing campaigns or infrastructure investments) need to be sustained to counteract the long-term effects of negative shocks. Meanwhile, the stationarity of the differenced series suggests that short-term policy measures can be effective in managing year-to-year fluctuations in tourist arrivals.

Identification

After finding the data stationary at first difference it is declared that the integration parameter d is one however other parameter p and q required determining. The determination of these parameters p and q is concluded in identification (Poudel et al., 2024).

Table 3: Autocorrelation functions and Partial autocorrelation functions

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.165	0.165	1.6928	0.193
		2	-0.445	-0.485	14.172	0.001
		3	-0.404	-0.289	24.680	0.000
		4	-0.078	-0.253	25.082	0.000
		5	0.251	-0.049	29.275	0.000
		6	0.146	-0.189	30.728	0.000
		7	0.018	0.006	30.750	0.000
		8	-0.111	-0.114	31.612	0.000
		9	-0.082	-0.016	32.098	0.000
		10	-0.054	-0.171	32.310	0.000
		11	0.045	-0.016	32.465	0.001
		12	0.080	-0.108	32.953	0.001
		13	-0.027	-0.124	33.010	0.002
		14	-0.005	-0.052	33.012	0.003
		15	0.031	-0.000	33.093	0.005
		16	-0.019	-0.107	33.124	0.007
		17	-0.062	-0.084	33.454	0.010
		18	0.055	0.058	33.720	0.014
		19	0.105	0.039	34.705	0.015
		20	0.025	0.046	34.761	0.021
		21	-0.056	0.079	35.053	0.028
		22	-0.046	0.139	35.255	0.036
		23	-0.027	0.032	35.326	0.048
		24	-0.003	0.076	35.327	0.064

Source: Author's calculations performed using E-Views

ACF plot suggest that the parameter q of MA could be 2 or 3 as after third lag the spikes are significant. This indicates that tourist arrivals (or the variable in question) at these points are strongly negatively correlated with the values from one and two periods prior. Again PACF plot suggests that the parameter p of AR could be 2 or 3 because after third lag spike inside the significant line. Thus the purposed model is ARIMA (2, 1, 2), ARIMA (2, 1, 3), ARIMA (3, 1, 2) ARIMA (3, 1, 3) (Table 3). The best model will be selected based on Root Mean Square Error Criteria.

Estimation

Once the orders of the AR, I, and MA components have been identified, the next step is to estimate the coefficients of the ARIMA model, the coefficients are estimated and presented in the below tables (Table 4, Table 5, Table 6, and Table 7).

Table 4: Estimates of coefficients for ARIMA (2, 1, 2)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	14429.25	1763.946	8.180094	0.0000
AR(2)	-0.024475	0.177396	-0.137968	0.8908
MA(2)	-1.000000	280.6965	-0.003563	0.9972
SIGMASQ	1.37E+10	1.91E+12	0.007129	0.9943
R-squared	0.494546	Mean dependent var		17039.93
Adjusted R-squared	0.466976	S.D. dependent var		165747.6
S.E. of regression	121009.8	Akaike info criterion		26.42798
Sum squared resid	8.05E+11	Schwarz criterion		26.56883
Log likelihood	-775.6253	Hannan-Quinn criter.		26.48296
F-statistic	17.93769	Durbin-Watson stat		1.908627
Prob(F-statistic)	0.000000			
Inverted AR Roots	-.00+.16i	-.00-.16i		
Inverted MA Roots	1.00	-1.00		

Source: Author's calculations performed using E-Views

Table 5: Estimates of coefficients for ARIMA (2, 1, 3)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	12383.56	15934.46	0.777156	0.4404
AR(2)	-0.514825	0.272465	-1.889507	0.0641
MA(3)	-0.405334	0.224040	-1.809204	0.0759
SIGMASQ	1.70E+10	1.50E+09	11.33731	0.0000
R-squared	0.369919	Mean dependent var		17039.93
Adjusted R-squared	0.335551	S.D. dependent var		165747.6
S.E. of regression	135107.1	Akaike info criterion		26.55123
Sum squared resid	1.00E+12	Schwarz criterion		26.69208
Log likelihood	-779.2612	Hannan-Quinn criter.		26.60621
F-statistic	10.76344	Durbin-Watson stat		2.215849
Prob(F-statistic)	0.000011			
Inverted AR Roots	-.00+.72i	-.00-.72i		
Inverted MA Roots	.74	-.37+.64i	-.37-.64i	

Source: Author's calculations performed using E-Views

Table 6: Estimates of coefficients for ARIMA (3, 1, 2)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	14404.76	1202.228	11.98172	0.0000
AR(3)	-0.390520	0.125816	-3.103902	0.0030
MA(2)	-1.000000	179.0301	-0.005586	0.9956
SIGMASQ	1.18E+10	1.06E+12	0.011147	0.9911
R-squared	0.562491	Mean dependent var		17039.93
Adjusted R-squared	0.538627	S.D. dependent var		165747.6
S.E. of regression	112583.2	Akaike info criterion		26.28538
Sum squared resid	6.97E+11	Schwarz criterion		26.42623
Log likelihood	-771.4186	Hannan-Quinn criter.		26.34036
F-statistic	23.57055	Durbin-Watson stat		2.150190
Prob(F-statistic)	0.000000			
Inverted AR Roots	.37+.63i	.37-.63i	-.73	
Inverted MA Roots	1.00	-1.00		

Source: Author's calculations performed using E-Views

Table 7: Estimates of coefficients for ARIMA (3, 1, 3)

Variable	**Coefficient**	**Std. Error**	**t-Statistic**	**Prob.**
C	12951.36	26905.27	0.481369	0.6322
AR(3)	-0.839055	0.496570	-1.689701	0.0967
MA(3)	0.419053	0.633207	0.661794	0.5109
SIGMASQ	2.05E+10	2.78E+09	7.395438	0.0000
R-squared	0.239641	Mean dependent var		17039.93
Adjusted R-squared	0.198167	S.D. dependent var		165747.6
S.E. of regression	148418.9	Akaike info criterion		26.74652
Sum squared resid	1.21E+12	Schwarz criterion		26.88737
Log likelihood	-785.0222	Hannan-Quinn criter.		26.80150
F-statistic	5.778094	Durbin-Watson stat		2.192306
Prob(F-statistic)	0.001653			
Inverted AR Roots	.47+.82i	.47-.82i	-.94	
Inverted MA Roots	.37-.65i	.37+.65i	-.75	

Source: Author's calculations performed using E-Views

Coefficients of four models: ARIMA (2, 1, 2), ARIMA (2, 1, 3), ARIMA (3, 1, 2) ARIMA (3, 1, 3) for forecasting tourist arrivals in Nepal are estimated and the best model can be identified by the comparison of different criteria.

Table 8: Models' comparison-1

	ARIMA(2,1,2)	ARIMA(2,1,3)	ARIMA(3,1,2)	ARIMA(3,1,3)
Significant Coefficient	1	1	2	1
SIGMASQ (Minimum)	1.37E+10	1.70E+10	1.18E+10	2.05E+10
Adjusted R-squared (Maximum)	0.466976	0.335551	0.538627	0.198167
Akaike info criterion (Minimum)	26.42798	26.55123	26.28538	26.74652
Schwarz criterion (Minimum)	26.56883	26.69208	26.69208	26.88737
Hannan-Quinn criterion (Minimum)	26.48296	26.60621	26.34036	26.80150

Source: *Author's calculations performed using E-Views*

Among the four ARIMA models compared, ARIMA(3,1,2) emerges as the best option. It stands out by having two significant coefficients, which is higher than the other models. Additionally, it has the lowest SIGMASQ value (1.18E+10), indicating that it has the smallest error variance. The model also boasts the highest Adjusted R-squared value (0.538627), signifying that it explains the largest proportion of variance in the data.

Furthermore, ARIMA(3,1,2) has the lowest Akaike Information Criterion (AIC) at 26.28538, suggesting it is the most efficient model that best fits the data. Although its Schwarz Criterion (BIC) value is not the absolute minimum, it remains relatively low and comparable to the others. The model also has the lowest Hannan-Quinn Criterion (HQIC) at 26.34036, further supporting its superiority. Overall, ARIMA(3,1,2) strikes the best balance between model complexity and goodness of fit. This model is better suited to capture the dynamics of the data, making it a valuable tool for economic analysis and policy formulation. For instance, policymakers could use the forecasts generated by this model to plan marketing campaigns or infrastructure development aimed at boosting tourism, especially in the aftermath of disruptive events such as earthquakes or pandemics. By accurately capturing the persistence and patterns in tourist arrivals, this model helps in making informed decisions that can stabilize and grow the tourism sector.

Table 9: Models' Comparison-2

	ARIMA(2, 1, 2)	ARIMA(2, 1, 3)	ARIMA(3, 1, 2)	ARIMA(3, 1, 3)
Root Mean Squared Error	177088.6	164978.5	176356.9	165524.9
Mean Absolute Error	125444.4	102626.9	123629.9	104337.9
Mean Absolute Percentage Error	53.37605	38.08983	49.43217	36.85989
Theil Inequality Coefficient	0.182207	0.183064	0.180835	0.179680
Bias Proportion	0.169483	0.004277	0.149283	0.013964
Variance Proportion	0.075476	0.249590	0.086719	0.208622
Covariance Proportion	0.755042	0.746133	0.763999	0.777414
Theil U2 Coefficient	1.523496	1.139553	1.397375	1.064723

The choice between these models ultimately depends on the specific economic application and the relative importance of the metrics. For instance, if minimizing percentage error is most crucial for the application (e.g., budgeting or planning for tourism-related revenue), ARIMA(3,1,2) might be preferred. On the other hand, if absolute accuracy in predictions is more critical (e.g., ensuring adequate accommodation or transportation capacity), ARIMA(2,1,3) could be the better choice.

Diagnostic

The Q statistic in residual analysis of ARIMA models serves as a diagnostic tool to ensure that the model captures and removes all significant autocorrelation in the residuals, helping to validate the adequacy of the chosen model. The null hypothesis of this The Q statistic, or Ljung-Box Q statistic is the residuals are effectively white noise, and the model has successfully accounted for the temporal dependencies in the data (Mahan et al., 2015)assessing brain networks by calculating pairwise correlations of time series generated from different areas of the brain. The assessment of these relations relies, in turn, on the proper calculation of interactions between time series, which is achieved by rendering each individual series stationary and nonautocorrelated (i.e., white noise, or to \" prewhiten \" the series.

Table 10: Q-statistic Probabilities Adjusted for 2 ARMA Terms Based on ARIMA(3, 1, 3)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.091	-0.091	0.5120	
		2	0.068	0.060	0.8039	
		3	-0.018	-0.007	0.8256	0.364
		4	-0.125	-0.133	1.8561	0.395
		5	0.289	0.277	7.4271	0.059
		6	-0.055	0.001	7.6306	0.106
		7	0.038	-0.013	7.7302	0.172
		8	-0.113	-0.118	8.6325	0.195
		9	-0.100	-0.055	9.3528	0.228
		10	-0.079	-0.176	9.8092	0.279
		11	-0.075	-0.079	10.235	0.332
		12	0.043	0.015	10.374	0.408
		13	-0.113	-0.067	11.365	0.413
		14	0.001	-0.005	11.365	0.498
		15	-0.003	0.077	11.365	0.580
		16	0.023	0.066	11.411	0.654
		17	-0.020	-0.082	11.445	0.720
		18	0.061	0.087	11.774	0.759
		19	0.070	0.067	12.214	0.787
		20	0.022	-0.018	12.260	0.834
		21	-0.013	-0.099	12.277	0.873
		22	-0.024	0.008	12.335	0.904
		23	-0.021	-0.081	12.378	0.929
		24	-0.019	-0.082	12.414	0.948

Source: Author's calculations performed using E-Views

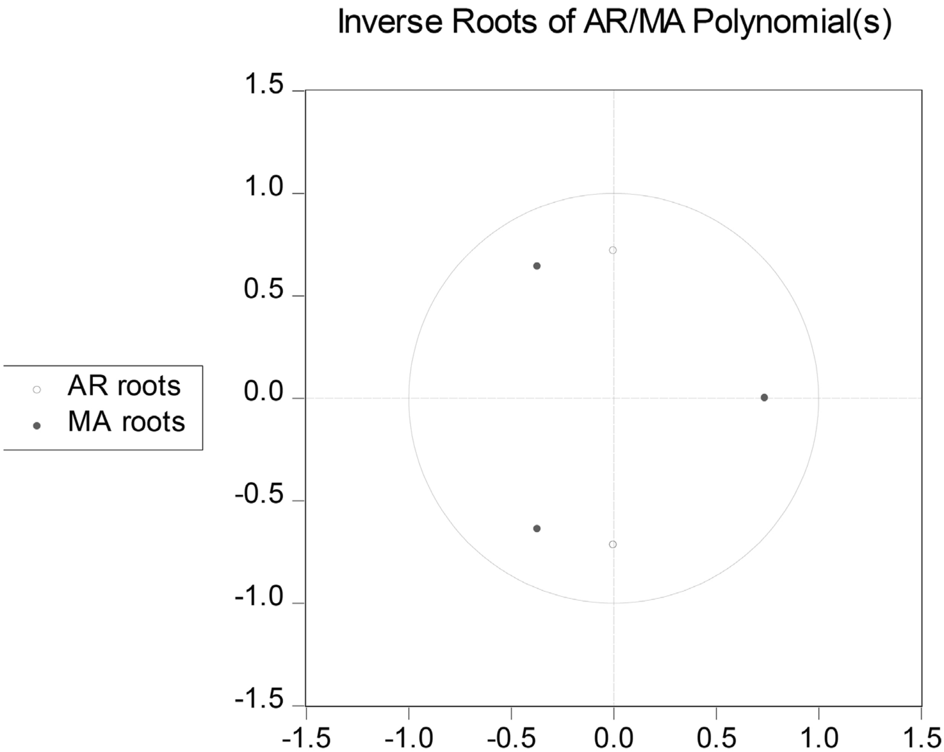
Table 11: Q-statistic Probabilities Adjusted for 2 ARMA Terms Based on ARIMA(2, 1, 3)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.108	-0.108	0.7236	
		2	-0.136	-0.149	1.8865	
		3	-0.028	-0.063	1.9358	0.164
		4	-0.193	-0.234	4.3855	0.112
		5	0.211	0.153	7.3534	0.061
		6	0.012	-0.015	7.3627	0.118
		7	-0.013	0.030	7.3746	0.194
		8	-0.117	-0.156	8.3325	0.215
		9	-0.049	0.000	8.5085	0.290
		10	-0.090	-0.202	9.1061	0.333
		11	-0.057	-0.116	9.3513	0.405
		12	0.059	-0.090	9.6204	0.474
		13	-0.060	-0.078	9.9040	0.539
		14	-0.029	-0.134	9.9716	0.618
		15	0.015	-0.022	9.9909	0.695
		16	0.012	-0.029	10.003	0.762
		17	-0.029	-0.102	10.076	0.815
		18	0.076	0.002	10.589	0.834
		19	0.097	0.077	11.441	0.833
		20	0.026	0.032	11.503	0.872
		21	-0.010	-0.047	11.512	0.905
		22	-0.005	0.031	11.515	0.932
		23	-0.045	-0.071	11.721	0.947
		24	-0.002	-0.073	11.721	0.963

In economic forecasting, particularly in sectors like tourism, which might have cyclical or seasonal patterns, it is crucial to have a model that captures these patterns without leaving significant serial correlation unaddressed. A model with significant remaining autocorrelation (as indicated in Table 10) could result in biased or unreliable forecasts, which can lead to poor decision-making in areas like resource allocation, budget planning, and policy implementation. On the other hand, the model from Table 11, which shows no significant serial correlation, is likely to provide more reliable and accurate forecasts, which are essential for strategic planning and risk management in economic applications.

ARIMA(2, 1, 3) represents the model that is more suitable for forecasting, as it shows a well-specified model with minimal remaining serial correlation, leading to more accurate and reliable forecasts.

Figure 2: Unit circle



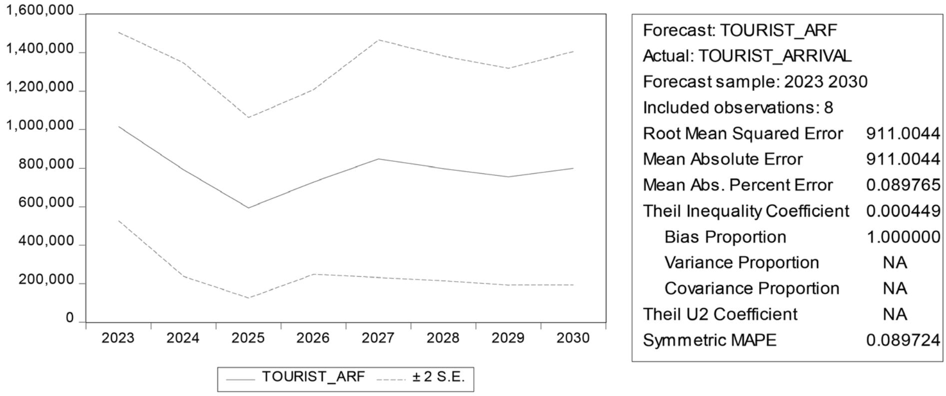
Source: Author's calculations performed using E-Views

Above figure shows that AR roots as well as MA roots are inside the unit circle. This shows that the estimated ARMA process is stationary. Together with test of stationary, the model diagnostic process is completed and the process verified the ARIMA (2, 1, 3) model is appropriate. Now the model can be used for forecasting the annual tourist arrival in Nepal for the future years.

Forecasting

In time series model building with EViews, the Root Mean Squared Error (RMSE) and Theil Inequality Coefficient serve as pivotal metrics for evaluating predictive accuracy. RMSE quantifies the average difference between predicted and observed values using the model fitted, with lower values indicating more accurate predictions. Simultaneously, the Theil Inequality Coefficient provides a comprehensive measure, considering both bias and dispersion of forecast errors, offering insights into the model's ability to capture trends and patterns (Poudel et al., 2024).

Figure 3: Root Mean Squared Error (RMSE) and the Theil Inequality Coefficient

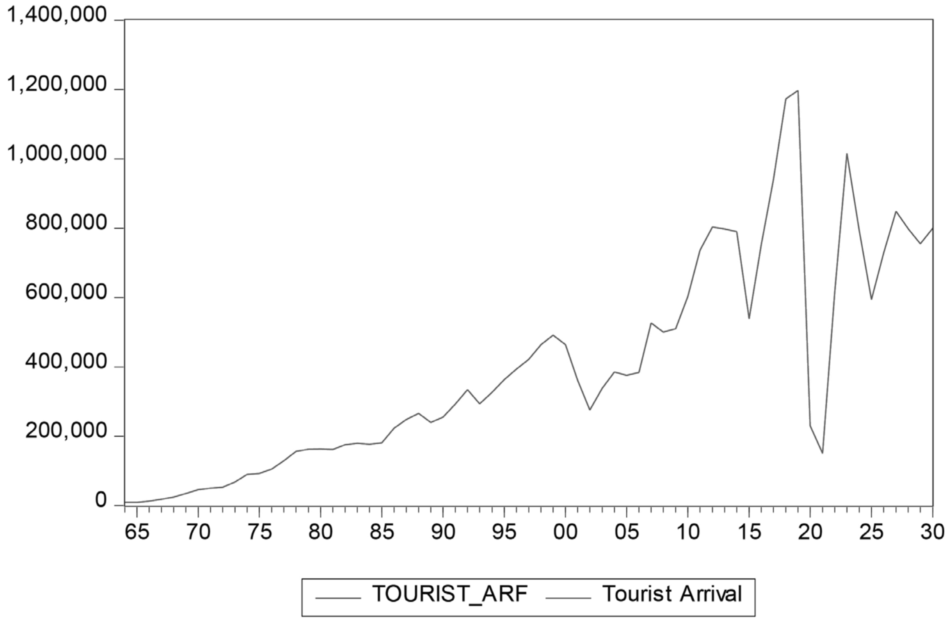


Source: Author's calculations performed using E-Views

Figure 3 illustrates the forecast of tourist arrivals in Nepal from 2023 to 2030, showing the model's predicted values along with the actual historical data. The RMSE is quite low at 911.0044, indicating a high level of precision in the model's predictions. The extremely low Theil Inequality Coefficient (0.000449) further reinforces the model's accuracy, implying that the forecasted values closely match the actual values. In economic terms, such accuracy is crucial for the tourism sector, where precise forecasts can help mitigate the risks associated with demand fluctuations, enabling better preparedness for future tourism seasons. These metrics are not only statistical tools but also have significant economic applications, aiding policymakers, businesses, and stakeholders in making informed decisions that can enhance the resilience and growth of Nepal's tourism sector.

Figure 4: Actual and forecast trend line of tourist arrivals in Nepal

Year	2023	2024	2025	2026	2027	2028	2029	2030
Tourist Arrival Forecasted	1015793	791910	593884	727903	848611	798373	754989	799611



Source: Author's calculations performed using E-Views

The figure shows a clear depiction of the historical trend of tourist arrivals, characterized by periods of growth and decline. Notable dips in tourist numbers correspond with significant events such as political instability, natural disasters (e.g., the 2015 earthquake), and the COVID-19 pandemic. These disruptions highlight the sector's vulnerability to external shocks.

The forecast starts with a high value in 2023 (1,015,793) but shows a significant decline over the next two years, with the lowest forecasted value in 2025 (593,884). This decline might suggest that the model anticipates a short-term reduction in tourist arrivals, possibly due to economic, political, or global factors that could negatively impact tourism during these years.

After the drop in 2025, the forecast indicates a recovery in 2026, with tourist arrivals increasing to 727,903 and continuing to grow to 848,611 in 2027. This could imply a recovery phase where the tourism sector rebounds from the earlier decline, potentially due to improved conditions or successful promotional and recovery strategies.

The forecasted values for 2028 to 2030 show some fluctuation. After reaching 848,611 in 2027, the numbers drop again to 798,373 in 2028 and continue to slightly decrease to 754,989 in 2029, before slightly rising again to 799,611 in 2030. This pattern suggests ongoing instability or uncertainty in the factors influencing tourist arrivals, leading to minor ups and downs rather than a steady increase.

The forecasted values do not show a straightforward increasing trend but rather a pattern of fluctuation with periods of decline, recovery, and minor variations. This suggests that while the tourism sector in Nepal may experience growth, it could also face challenges that lead to temporary downturns. Effective planning, strategic investments, and risk management will be crucial in navigating these fluctuations to achieve sustainable growth in the tourism industry.

Discussion

Nepal's tourism sector, a vital pillar of its economy, has exhibited considerable dynamism and resilience over the years. Notably, the sector's contribution to Nepal's GDP was substantial, accounting for 7.9% in 2018 and supporting over a million jobs (Bhattarai, 2023). Despite the robust growth, the tourism industry has faced significant disruptions, most notably from the COVID-19 pandemic, which resulted in a dramatic downturn in tourist arrivals. The time series analysis of tourist arrivals from 1964 to 2023 reveals both historical growth trends and recent fluctuations. The data indicates a significant rise in tourist numbers up until 1999, followed by a downturn through the early 2000s, attributed to political instability. The pandemic further exacerbated these fluctuations, with 2021 showing signs of recovery. This variability underscores the complexity of forecasting in the tourism sector, which can be influenced by numerous unpredictable factors.

The application of the ARIMA model to forecast tourist arrivals proves particularly relevant. The ARIMA (Auto Regressive Integrated Moving Average) methodology, established by Box and Jenkins (1976), is well-suited for handling non-stationary time series data by incorporating differencing to achieve stationarity. The study employed various ARIMA models: ARIMA (2, 1, 2), ARIMA (2, 1, 3), ARIMA (3, 1, 2), and ARIMA (3,1,3) to identify the most effective forecasting model for Nepalese tourism data.

Among the models evaluated, ARIMA (2, 1, 3) emerged as the superior model. It demonstrated the lowest Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE), indicating higher accuracy in predictions compared to the other models. The model's performance is also supported by its higher Adjusted R-squared value, signifying that it explains a greater proportion of the variance in tourist arrivals.

However, the ARIMA(3, 1, 2) model achieved the lowest Akaike Information Criterion (AIC) and Hannan-Quinn Criterion (HQIC) values, suggesting that it provided the best balance between model complexity and goodness of fit. Despite this, diagnostic checks using Q-statistics and residual analysis confirmed the adequacy of the ARIMA (2, 1, 3) model. The residuals of the ARIMA (2, 1, 3) model exhibited characteristics of white noise, indicating that it effectively captured the underlying patterns in the data without significant autocorrelation. Therefore, the ARIMA(2, 1, 3) model was determined to be more suitable than the ARIMA(3, 1, 2) model for forecasting, as it demonstrated a well-specified structure with minimal remaining serial correlation, leading to more accurate and reliable forecasts. Consequently, this research adopted the ARIMA (2, 1, 3) model for forecasting. From this discussion, the ARIMA (2, 1, 3) model offers a robust tool for forecasting tourist arrivals in Nepal.

The forecasted trends of tourist arrivals in Nepal, based on the ARIMA model, show a pattern of ups and downs rather than steady growth. This finding is similar to what other studies have found, highlighting how tourism in areas like Nepal can be unpredictable and easily affected by external factors. For example, research by Edward et al. (2023) and Nurhasanah et al. (2022) showed that ARIMA models are good at capturing these fluctuations in tourist numbers in countries like the United States and Indonesia. Similarly, the ARIMA model in this study shows that Nepal's tourism sector is sensitive to disruptions, meaning that while recovery is possible, it can be interrupted by unexpected events. This suggests that policymakers need to create strategies to make the tourism sector more resilient to these fluctuations, ensuring that recovery is steady and long-lasting.

Compared to previous studies, the ups and downs in the forecasted values for Nepal are different from the more stable trends seen in other regions. For example, Makoni et al. (2023) found a steady increase in tourist arrivals in Zimbabwe, thanks to consistent marketing and better infrastructure. The difference between Nepal and Zimbabwe could be due to the fact that Nepal's tourism is more vulnerable to political instability, natural disasters, and global pandemics. This difference highlights the need for strategies that are specifically designed to address Nepal's unique challenges. For instance, improving crisis management and investing in strong infrastructure could help lessen the impact of future disruptions on tourist arrivals, leading to more consistent growth.

Economically, the fluctuating forecasted trend has important implications for how resources are allocated and how strategic planning is done in Nepal's tourism sector. Unlike the steady growth seen in other studies, the predicted declines in certain years (2024-2025) and the recovery afterward (2026-2027) suggest that stakeholders need to be more flexible in their planning. For example, investments in tourism infrastructure and marketing should be timed to match periods of expected growth, while cost-saving measures might be needed during predicted downturns. Additionally, the government and private sector should work together to diversify tourism offerings and reach new markets, reducing dependence on a narrow visitor base and lowering the risks associated with these fluctuations. This proactive approach, based on a careful analysis of forecasted trends, can help ensure that Nepal's tourism sector continues to be a strong contributor to the national economy, despite its inherent volatility.

Strengths and limitations of the study

The focus of the study on forecasting tourist arrivals in Nepal is highly relevant, given the country's reliance on tourism for economic stability and growth. ARIMA models assume that the time series is linear and can be represented by past values and past errors. However, tourist arrivals may be influenced by non-linear factors such as sudden events (e.g., natural disasters, political instability) or seasonal effects that are not linear, which ARIMA may not capture accurately. ARIMA models are actually effective for capturing trends based on past data, they do not incorporate external factors that might influence tourist arrivals, such as economic conditions, geopolitical events, or promotional campaigns.

Conclusion and implications

The ARIMA model's forecast of tourist arrivals in Nepal underscores the inherent volatility of the country's tourism sector, reflecting its sensitivity to external shocks such as political instability, natural disasters, and global pandemics. Unlike the stable growth patterns observed in other regions, Nepal's forecasted trends exhibit significant fluctuations, necessitating a strategic and flexible approach to tourism planning and resource allocation. Based on the results and discussion, it is recommended that Nepalese authorities enhance marketing efforts to restore confidence in tourism, invest in infrastructure improvements to attract more visitors, and offer targeted incentives for tourism-related businesses. Additionally, developing contingency plans for future disruptions can help stabilize the sector. The policymakers and industry stakeholders must prioritize the development of resilient infrastructure, effective crisis management, and diversified tourism offerings to mitigate the impact of these fluctuations. By adopting such measures, Nepal can enhance the stability and sustainability of its tourism sector, ensuring its continued contribution to the national economy. This study highlights the critical need for tailored strategies that address the unique challenges faced by Nepal's tourism industry, providing valuable insights for future research and policy development.

Future research should focus on these areas to provide a more comprehensive and robust analysis, ensuring that the forecasts and recommendations are based on well-validated and consistent data. By addressing these gaps, Nepal's tourism sector can better prepare for future challenges and maximize its contribution to the national economy. Future research should consider exogenous factors influencing tourist influx, such as economic circumstances, geopolitical incidents, and advertising strategies. The integration of these variables into a multivariate analytical framework would result in a more comprehensive understanding of their impact on tourist dynamics. Such a methodology could improve the accuracy of forecasting models and lead the development of more effective tourism initiatives.

Institutional review board statement: Not applicable.

Funding: None

Informed consent: Not applicable.

Data availability statement: Not applicable.

Conflicts of interest: The authors declare no conflicts of interest.

References

Acharya, B., Poudel, O., Upadhyaya, Y. M., & Acharya, P. (2024). A comprehensive analysis of remittances and basic level government school enrollment in Nepal. *International Research Journal of MMC*, *5*(3), 86-103. DOI: https://doi.org/10.3126/irjmmc.v5i3.68503

Acharya, B., Kharel, K. R., Poudel, O., Upadhyaya, Y. M., & Sigdel, S. (2024). Unlocking opportunity: Examining the relationship between remittances and private school enrollment in Nepal. *Oikos Nomos: Jurnal Kajian Ekonomi dan Bisnis*, *17*(1), 51-65. https://doi.org/10.37479/jkeb.v17i1.26221

Bhattarai, M. (2023). *Tourism and its contribution in economy of Nepal*. https://economicsansar.com/tourism-and-its-contribution-in-economy-of-nepal/

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control*. Holden-Day.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.

Damos, P. (2016). Using multivariate cross correlations, Granger causality and graphical models to quantify spatiotemporal synchronization and causality between pest populations. *BMC Ecology*, *16*, 1–17. https://doi.org/10.1186/s12898-016-0087-7

Edward, E., Misengo, Dedy, D., Prastyo, & Heri, K. (2023). Modeling and forecasting monthly tourist arrivals to the United States and Indonesia using ARIMA hybrids of multilayer perceptron models. *Nucleation and Atmospheric Aerosols*. https://doi.org/10.1063/5.0105680

Ellis, C. (2022). *Crunching the Data*. https://crunchingthedata.com/when-to-use-arima/

Jafridin, M. A., Fauzi, N. F., Alias, R., Halim, H. Z. A., Bakhtiar, N. S. A., Khairudin, N. I., & Shafii, N. H. (2021). A comparison of fuzzy time series and ARIMA to forecast tourist arrivals to homestay in Pahang. *Journal of Computing Research and Innovation*, *6*(4), Article 4. https://doi.org/10.24191/jcrinn.v6i4.235

Mahan, M., Chorn, C., & Georgopoulos, A. (2015). *White noise test: Detecting autocorrelation and nonstationarities in long time series after ARIMA modeling*. https://doi.org/10.25080/Majora-7b98e3ed-00f

Makoni, T., & Chikobvu, D. (2021). *Modelling international tourist arrivals volatility in Zimbabwe using a GARCH process*. http://hdl.handle.net/11660/11237

Makoni, T., Mazuruse, G., & Nyagadza, B. (2023). International tourist arrivals modelling and forecasting: A case of Zimbabwe. *Sustainable Technology and Entrepreneurship*, *2*(1), 100027. https://doi.org/10.1016/j.stae.2022.100027

Ministry of Finance. (2023). https://mof.gov.np/public/site/publication-category/21

NTB. (2023). *Everest: World's highest peak.* https://ntb.gov.np/everest

Nurhasanah, D., Aurielle, M. S., & Mujiati, D. K. (2022). Forecasting international tourist arrivals in Indonesia Using SARIMA model. *Enthusiastic*, 19–25. https://doi.org/10.20885/enthusiastic.vol2.iss1.art3

Poudel, O. (2023). Relationship between defense expenditure and economic growth in Nepal. *Unity Journal, 4*(1), 208–226. https://doi.org/10.3126/unityj.v4i01.52242

Poudel, O., Acharya, P., & Simkhada, D. (2023). Examining the drivers of economic growth in Nepal: A macroscopic econometric analysis. *Shaheed Smriti Journal,12*(9),37-53. https://ssmcchitwan.edu.np/wp-content/uploads/2023/12/2023-journal.pdf#page=31

Poudel, O., Kharel, K. R., Acharya, P., Simkhada, D., & Kafle, S. C. (2024). ARIMA modeling and forecasting of national consumer price index in Nepal. *Interdisciplinary Journal of Management and Social Sciences, 5*(1), 105-118. https://doi.org/10.3126/ijmss.v5i1.62666

Purwanto, Sunardi, Julfia, F. T., & Paramananda, A. (2019). Hybrid model of ARIMA-linear trend model for tourist arrivals prediction model in Surakarta City, Indonesia. *AIP Conference Proceedings*, *2114*(1), 060010. https://doi.org/10.1063/1.5112481

UNWTO. (2008). International recommendations for tourism statistics. https://www.unwto.org/tourism-statistics/on-basic-tourism-statistics-irts-2008

Vanhove, N. (2022). *The economics of tourism destinations: Theory and practice*. Routledge. https://scholar.google.com/scholar_lookup?&title=The%20Economics%20of%20Tourism%20Destinations&publication_year=2011&author=Vanhove%2CN

United Nations World Tourism Organization. (2008). *International recommendations for tourism statistics 2008*. www.statistics.unwto.org

Upadhyay, H. P., Pradhan, B. L., Poudel, O., Sapkota, K. R., & Simkhada, D. R. (2022). Impact of remittance, money supply and inflation on economic growth of Nepal: Using autoregressive distributed lag model (ARDL). *BMC Journal of Scientific Research, 5*(1), 52-59. https://doi.org/10.3126/bmcjsr.v5i1.50673

Yollanda, M., & Devianto, D. (2020, January). Hybrid model of seasonal ARIMA-ANN to forecast tourist arrivals through Minangkabau International Airport. In *Proceedings of the 1st International Conference on Statistics and Analytics, ICSA 2019, 2-3 August 2019, Bogor, Indonesia*. https://doi.org/10.4108/eai.2-8-2019.2290473