Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/140402

Seasonal Autoregressive Integrated Moving Average with Exogenous Variables Intervention Analysis: Application to the South African Tourism Industry

Amos M. Mphanya[1], Sarah L. Mahlangu[1], Thabiso E. Masena[1] and Sandile C. Shongwe[1,]*

[1] Department of Mathematical Statistics and Actuarial Science, Faculty of Natural and Agricultural Sciences, University of the Free State, Bloemfontein, 9301, South Africa.

Received: 23 Jan. 2025, Revised: 14 Mar. 2025, Accepted: 5 May. 2025.
Published online: 1 Jul. 2025.

Abstract: Tourism revenue forms an important component of many countries' gross domestic product such that a drop in tourists arrivals causes a significant increase in unemployment in certain sectors of the economy, e.g. hotels, bed and breakfast, etc. While several economic sectors have fully recovered from the recent COVID-19 pandemic, in this paper, we use the famed time series analysis' Box-Jenkins methodology to illustrate that the South African tourism accommodation income has not fully recovered from its negative effect. Additionally, the seasonal autoregressive integrated moving average with exogenous variables (SARIMAX) intervention model with a pulse function covariate vector incorporated through trial-and-error was used to fully model and quantify the negative ramifications the pandemic on the South African tourism accommodation income dataset. Using March 2020 as the intervention point, the South African tourism sector experienced a loss of ZAR 99,009 million in revenue in the 52-months intervention period from March 2020 to June 2024. More importantly, at the end of the study period (June 2024), the tourism accommodation income series had not recovered to its pre-intervention levels. The 1-year out-of-sample forecasts from the best fitting $\text{SARIMAX}(1,1,2)(0,1,1)_{12}$ intervention model estimates that the tourism accommodation income series will not recover to its pre-COVID-19 intervention levels by June 2025 if rescuing efforts are not taken to boost income within the sector. Such efforts include providing resources such as capital financing, skills development and mentorship for aspiring entrepreneurs within the tourism industry.

Keywords: Box-Jenkins Methodology, COVID-19, Forecasting, Interrupted Time Series, Intervention, Maximum Likelihood Estimation, Seasonal Autoregressive Integrated Moving Average (SARIMA), Time Series Analysis.

1 Introduction

Majority of business enterprises (small, medium and large), including national parks, tourism accommodation, national and local tourism attraction sites and all tourism-related activities in R.S.A. (Republic of South Africa) had to adhere to the stringent lockdown regulations because the sector was not categorised as an essential services provider, except hotels housing international guests and essential services workers [1]. Furthermore, [1] argued that the national lockdown restrictions implemented by the R.S.A. government were relatively too strict compared to those implemented by other countries. As a result, [2] identified elements of mistrust between local tourism enterprises in some parts of R.S.A. and the government. Some hotels were rendered to the R.S.A. government to be used as quarantine zones [3]. Majority of hotels were at the brink of collapse, while some were at a high risk of bankruptcy due to limited cashflow, visitor cancellations and economic uncertainty [4]. This is concerning as the R.S.A.'s tourism industry is regarded as one of the largest and best developed in the Southern Africa [5].

Authors like [6] hinted out that the adverse impacts of the COVID-19 intervention were not only economic and socio-political but were catastrophic for tourist demand. Specific to the tourism sector are its social implications due to movement and travel restrictions [7]. In another study, [8] highlighted that the COVID-19 pandemic had far-reaching economic, social, and psychological effects on hospitality businesses in Turkey, as many hotel managers developed anxiety as top international hotel chains shut down all business operations. In a different study, [9] asserted that the pandemic's negative impact on the sustainable transformation of the tourism industry in areas such as tourist perception and government crisis management is irreversible. Scholars in tourism research were concerned about over-tourism prior to 2020. However, in the post-COVID-19 era, the focus shifted to under-tourism due to economic ramifications experienced by business enterprises and significant changes in consumer and market demands in the sector [10]. This solidifies a different perspective by [11], which implies that the pandemic will significantly transform the long-term structural composition of the tourism industry. As a result, tourism demand forecasting has become an integral part of tourism management research in the aftermath of the COVID-19 pandemic [12]. Nyawo et al. [13] reported an estimated loss of 8 million jobs in Africa and in another study, [14] estimated that 3 million employees lost their

*Corresponding author e-mail: shongwesc@ufs.ac.za

jobs from February 2020 to April 2020 in R.S.A. due to the COVID-19 pandemic. According to [15], the majority of individuals who lost jobs were young black women with little or no formal education from low-income households. Additionally, the small and medium enterprises as well as households that depend on income generated from the surrounding tourism attraction sites were the most severely affected by the COVID-19 pandemic [16].

Makoni et al. [17] highlighted that the tourism industry has the potential to lower poverty and stimulate economic growth in Africa. Therefore, to promote the expansion of tourism-related activities and strategize ways to regain losses that the tourism sector experienced during the COVID-19 intervention, additional research with a specific focus on the tourism sector should be conducted [18]. Moreover, statistical models and forecasts assist business enterprises with managing operations, developing effective marketing strategies, improving service delivery, efficiently allocating resources, and implementing proper budgetary planning to boost profitability and growth within the sector. This encourages entrepreneurship and reduces poverty by improving job security and the attractiveness of the tourism sector to local and international investors [18]. Prilistya et al. [19] argued that accurate forecasting techniques play a crucial role in ensuring the sustainability of the tourism industry as they inform policy-building frameworks.

Chipumuro and Chikobvu [18] used the Box-Jenkins seasonal autoregressive integrated moving average (SARIMA) model to assess the impact of the COVID-19 intervention on the number of monthly tourist arrivals from other countries to R.S.A. from January 2009 to February 2020. The study noted a sustained upward trend and seasonality in the number of tourist arrivals to R.S.A. before the emergence of COVID-19, and the counterfactual forecasts from the pre-intervention period exhibited the same stochastic periodicity and estimated a 90% loss in monthly tourist arrivals from March 2020 to March 2021 because of the pandemic. Chipumuro et al. [20] further compared the differences between the counterfactual forecasts from the fitted pre-COVID-19 intervention autoregressive integrated moving average (ARIMA) model and the actual number of tourist arrivals to R.S.A. to emphasise the severe negative impact of the intervention on the R.S.A.'s tourism sector from March 2020. In a similar study, [21] highlighted the negative impact of the COVID-19 pandemic on the growth rate of the tourism and hospitality industry using the number of monthly international tourist arrivals in India using the Box-Jenkins ARIMA modelling. The post-intervention forecasts exhibited a sustained growth in India's total number of tourist arrivals from September 2023 to August 2025. In the Zimbabwean context, [17] and [22] forecasted international tourist arrivals using the Box-Jenkins SARIMA model, where [17] highlighted the potential benefits of the forecasts from the Box-Jenkins model in tourism marketing and planning. For two other studies that applied the basic Box-Jenkins methodology on the wholesale and retail sales data for R.S.A., see [23] and [24], respectively.

Song et al. [25] reviewed 211 studies published from 1968 to 2018 on tourism demand modelling and forecasting. Their study found that no model seemed to perform better than others, as studies were conducted under different circumstances but, the seasonal autoregressive integrated moving average with exogenous components (SARIMAX) model seemed to have gained popularity among researchers in tourism forecasting. Duan [26] showed that the prediction accuracy in tourism demand forecasting can be improved by incorporating search engine data through the SARIMAX model. In their study, the SARIMAX model, where tourists' search interests and intentions were used as exogenous variables, performed better than the traditional SARIMA model. Jula and Jula [27] analysed the net occupancy rate of tourism accommodations using the counterfactual forecasts from the SARIMAX model with a dummy variable representing calendar months (1 to 12 for January to December, respectively) to assess the impact of COVID-19 on tourism industry of 32 European countries. Their study showed a significant drop in the overall occupancy rate in tourism accommodations from March 2020, with April 2020 experiencing a 90% reduction in most countries included in the study. Mendieta-Aragon et al. [28] used the SARIMA and the SARIMAX model with an exogenous variable (tweets) to forecast tourism demand for a pilgrimage destination in northwestern Spain. The findings of their study showed that the forecasts from the SARIMAX model produced the best forecasts. As a result, [28] emphasised that forecasting methods should be frequently refined by using the latest available digital data to improve the reliability of forecasts. Wu et al. [29] asserted that including multi-source data in SARIMAX model building significantly improves the accuracy of tourism arrivals forecasts.

Prilistya et al. [19] evaluated the impact of the COVID-19 pandemic on Indonesian tourist arrivals was assessed using the ARIMA and SARIMA models as well as the autoregressive integrate moving average with exogenous variables (ARIMAX) and the SARIMAX model with Google search query data as exogenous variables. The study [19] found that the SARIMAX model had the lowest prediction errors (root mean squared error, mean absolute percentage error and mean absolute error) compared to the fitted ARIMA, SARIMA and ARIMAX models, highlighting the significance of using search query data as exogenous variables to improve the prediction accuracy of time series models. A similar study by [30] reported that the ARIMA intervention model was able to quantify the size of the COVID-19 effect on Indonesian tourist arrivals and produced more accurate forecasts than the ARIMAX model with Google trends data as an exogenous variable. Forty-three additional studies on the impact of the COVID-19 pandemic on tourism industries from multiple countries are contained in [8].

This study uses the SARIMAX intervention model augmented with a pulse function covariate vector incorporated through trial-and-error to fully model and quantify the negative ramifications of COVID-19 on the R.S.A.'s tourism accommodation income from March 2020 to June 2024. The uniqueness of this study is rooted in the use of a pulse function covariate vector for intervention analysis instead of deriving mathematical equations which may be too complex to compute when applied to real-life data. Moreover, earlier studies in the context of the R.S.A.'s tourism industry used the traditional SARIMA model to assess the disruption in the number of tourist arrivals since the onset of the COVID-19 intervention ([18] and [20]). However, the latter studies did not use any intervention analysis technique to explicitly assess and quantify the total loss in the number of tourist arrivals to R.S.A. as a result of the pandemic. They conducted the pre-intervention analysis only and relied on forecasts from the pre-intervention model to comment on how the pandemic impacted tourism arrivals without using intervention variables during post-intervention period to approximate the relative effect for each of the affected months. Tourism accommodation income depends on tourism arrivals. For instance, as the total number of tourist arrivals into a country increase, the income or revenue generated by business enterprises within the tourism accommodation sector will also increase. Therefore, the SARIMAX intervention model augmented with a pulse function incorporated through trial-and-error presented in this study will lay a foundation on intervention effects quantification using tourist arrivals data. This is a necessary novel contribution because there exist many time series analysis studies based on data from different countries/places but with very little to no new novelty ideas. Stated differently, most authors just apply basic Box-Jenkins methodology on different datasets without adding any new novelty techniques. Therefore, this study will assist empirical tourism researchers or practitioners in approaching the analysis of tourism data using an approach called 'trial-and-error' for post-intervention analysis and quantifying the 'loss in the economy' with an easy to interpret lingering effect plot to check whether the sector has fully recovered from the effect of some intervention.

The rest of the paper is structured as follows: In Section 2, the theoretical Box-Jenkins' SARIMAX methodology is discussed, and the corresponding empirical analysis is conducted in Section 3. Mitigating strategies are discussed in Section 4 and finally, concluding remarks of the study are provided in Section 5.

2 Methodology

2.1 SARIMA model

The seasonal autoregressive integrated moving average (SARIMA) model of the form SARIMA$(p,d,q)(P,D,Q)_s$ is expressed as [31]:

$$\phi(B)\Phi(B)(1-B)^d(1-B^s)^D N_t = \theta(B)\Theta(B)\Theta(B)\varepsilon_t \tag{1}$$

where N_t denotes the uninterrupted tourism accommodation income series in the pre-intervention period; $\phi(B)$ and $\theta(B)$ represent the non-seasonal autoregressive (AR) and moving average (MA) operators with parameters ϕ_p and θ_q, respectively. $\Phi(B)$ and $\Theta(B)$ represent the seasonal AR and MA operators with parameters, Φ_P and Θ_Q, respectively. P and p respectively denote the seasonal and non-seasonal AR orders, whilst q and Q respectively denote the seasonal and non-seasonal MA orders. D and d represent the seasonal and non-seasonal orders of differencing, respectively; where s is the number of data points in a single seasonal period and ε_t is the random error component. Equation (1) can be written using characteristics equations as follows,

$$\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$$

$$\Phi(B) = 1 - \Phi_1 B^s - \cdots - \Phi_P B^{Ps}$$

$$\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$$

$$\Theta(B) = 1 - \Theta_1 B^s - \cdots - \Theta_Q B^{Qs}.$$

2.2 Outliers

This study considers two types of outliers, namely (i) additive outliers (AO) and (ii) innovative outliers (IO). AO affects the series at only 1 point while outlier effects of the IO where it is initially detected spill over on succeeding points by weights of θ_q and Θ_Q [32]. The pre-intervention N_t in Equation (1) is amended to incorporate outliers as follows,

$$\phi(B)\Phi(B)(1-B)^d(1-B^s)^D N_t = \theta(B)\Theta(B)\Theta(B)(\varepsilon_t + IO) + AO \tag{2}$$

where $IO = \omega_1 I_1^{T_{IO}} + \cdots + \omega_t I_t^{T_{IO}}$ and $AO = \omega_1 I_1^{T_{AO}} + \cdots + \omega_t I_t^{T_{AO}}$; ω is the estimated size of the outlier; $I_t^{T_{IO}}$ and $I_t^{T_{AO}}$ are binary indicator variables used to indicate the presence or absence of an outlier at a particular point in the series by assigning it a dummy variable $=$ 1 at the data point where an outlier is detected and 0 otherwise [33]. If only one IO is detected at $t = 53$, then IO in Equation (2) becomes $\omega_{53} I_{53}^{T_{IO}}$, where w_{53} is the estimate of the size of the outlier and $I_{53}^{T_{IO}}$ is assigned a value of 1 at $t = 53$ when fitting a the pre-intervention model.

2.3 SARIMAX intervention model

The general form of the SARIMAX intervention model used in this study is given as [34]

$$Y_t = \frac{\omega(B)}{\delta(B)} I_t^{T_k} + N_t \tag{3}$$

where Y_t is the total interrupted tourism accommodation income at t. T_k represents the intervention period (for the tourism dataset considered here, this is 52 months, i.e. from March 2020 to June 2024), $I_t^{T_k}$ is the intervention indicator variable for a pulse function ($P_t^{T_k}$), $\omega(B) = \omega_0 - \omega_1 B - \cdots - \omega_s B^s$ (moving average operator) and $\delta(B) = 1 - \delta_1 B - \cdots - \delta_r B^r$ (autoregressive operator) where s and r represent the intervention duration and decay pattern, respectively. In this study, we express the pulse function as

$$P_t^{T_k} = \begin{cases} 0, where\ t < T_k \\ covariate\ vector\ , where\ t = T_k \end{cases} \tag{4}$$

where a covariate vector is a set of integers selected by trial-and-error using the following steps [35, 36]:

- **Step 1**: Select the intervention point.
- **Step 2**: Select the recovery point, which is where the interrupted series recovers to its counterfactual forecasts. The period from the intervention point to the recovery point is used as the intervention period. However, if the interrupted series does not recover, all data points from the intervention point until the end of the study period is used as the intervention period.
- **Step 3**: Then, the pre-intervention model is extended into the intervention period.
- **Step 4**: A covariate vector with a fixed value of 1 is fitted in the intervention period and the resulting model becomes a SARIMAX intervention model.
- **Step 5**: The SARIMAX intervention model supplemented with a covariate vector in Step 4 is adjusted by trial and error to produce a near to a perfect fit on the interrupted series.
- **Step 6**: We check how the fitted values from the SARIMAX intervention model with a covariate vector in Step 5 compares to the actual values of the interrupted series. The procedure continues until a combination of covariate vector components that produces estimated values from the SARIMAX intervention model with a near to a perfect fit to the interrupted series is obtained.
- **Step 7**: The final SARIMAX intervention model in Step 6 is then used to compute estimated loss of tourism accommodation income for each month in the intervention period and out of sample forecasts.

2.4 Data transformation and stationarity

This study uses the Box-Cox transformation to evaluate the necessary transformation on N_t [32]. The Augmented Dickey Fuller (ADF) test is used to assess the null hypothesis that N_t is non-stationary and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test is used to assess the null hypothesis that N_t is trend-stationary at the 5% significance level [37,38].

2.5 Box-Jenkins Methodology

The SARIMA model in Section 2.1 is adopted from the three-step Box-Jenkins methodology which is comprised of model selection, parameter estimation and model diagnostics [37].

2.5.1 Model Selection and Accuracy Metrics

The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are used to select the orders of multiple candidate SARIMA$(p, d, q)(P, D, Q)_s$ pre-intervention models. The most appropriate model has the least value of the Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC). The AIC and BIC values are computed using the following equations,

$$AIC = -2\log(\mathrm{L}) + 2k, and\ BIC = -2log(\mathrm{L}) + klog(n) \tag{5}$$

where L is the likelihood function of the series, n is the number of observations and $k = p + q$ [37]. The root mean squared error (RMSE) and mean absolute percentage error are used to assess the prediction accuracy of the chosen model [39,40],

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{Y}_t\right)^2} \text{ and } MAPE = \left(\frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t-\hat{Y}_t}{Y_t}\right|\right) \times 100 \quad (6)$$

where Y_t and $\hat{Y}_t$ represent the actual and predicted values, respectively.

2.5.2 Parameter Estimation

The best possible parameter estimates of the selected model are evaluated using maximum likelihood estimation (MLE) as per the following log-likelihood function [41],

$$\hat{\psi}_n = arg \max_{\psi \in \Psi} L_n(Y_t; \psi) = arg \max_{\psi \in \Psi} L_n(\psi) \quad (7)$$

where $\hat{\psi}$, denotes the n^{th} estimated parameter. Parameter estimates are obtained by solving for the derivative of the log-likelihood function.

2.5.3 Model diagnostics

The Ljung-Box and Box-Pierce tests are used to assess the null hypothesis that the standardised residuals from the fitted model are not autocorrelated [37]. Additionally, the Shapiro-Wilk and Jarque-Bera tests are used to assess the null hypothesis that the standardised residuals from the fitted model are normally distributed [42,43].

2.6 R software packages

The analysis in this study was carried out in R statistical software version 4.4.1 using the TSA, tseries, forecast, MASS, tsouliers and lmtest packages [37, 44-49].

3 Results and Discussion

3.1 Background of the data

In this study, the data containing the R.S.A.'s monthly tourism accommodation income recorded in ZAR millions from January 2013 to June 2024 (Stats SA, 2024) is used [50]. The dataset is publicly available on the Statistics South Africa® website (https://www.statssa.gov.za/). The pre-intervention period starts from January 2013 to February 2020 (86 months). March 2020 is the intervention point, and the post-intervention period starts from March 2020 to June 2024 (52 months). In the time series plot in Figure 1, the tourism accommodation income series (Y_t) exhibits a highly seasonal behaviour with an overall increasing trend. A sudden drop in Y_t associated with the implementation of national lockdown is observed in March 2020.

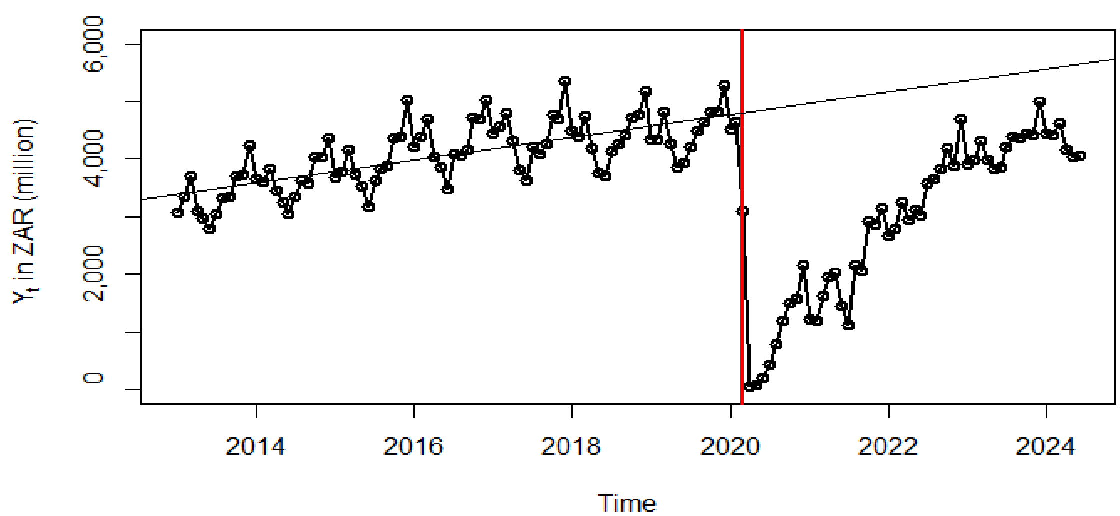


Fig. 1: Time series plot of Y_t

3.2 Pre-Intervention Analysis

The mean, minimum and maximum values of Y_t in the pre-intervention are ZAR 4090 million, ZAR 2802 million and ZAR 5353 million, respectively.

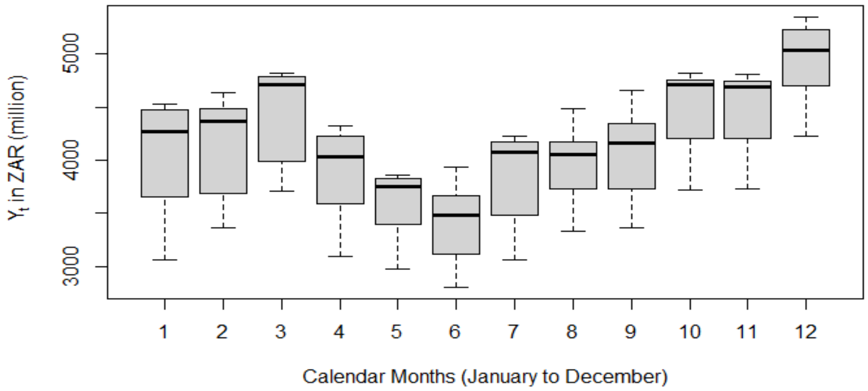


Fig. 2: Seasonal means plot of Y_t in the pre-intervention period

As shown in Figure 2, on average, the R.S.A.'s tourism industry generates more income in December (summer) and lowest income in June (winter).

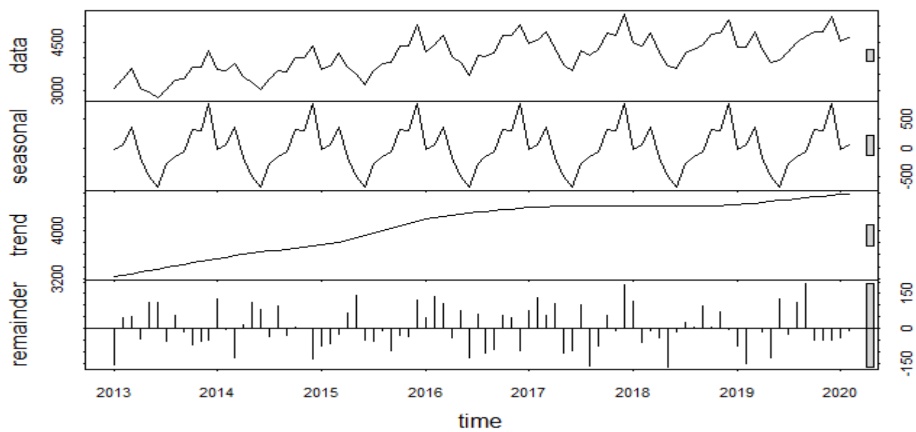


Fig. 3: Decomposition plot of tourism accommodation income in the pre-intervention period

The decomposed time series plot in Figure 3 shows a clear increasing trend and seasonality in the tourism accommodation income. Therefore, first and seasonal differencing of the pre-intervention series are required to detrend and capture the seasonality in the series.

3.3 Data transformation

The lambda value from Box-Cox diagram in Figure 4 is approximately equal to 1, suggesting that the series (Y_t) does not require any transformation.

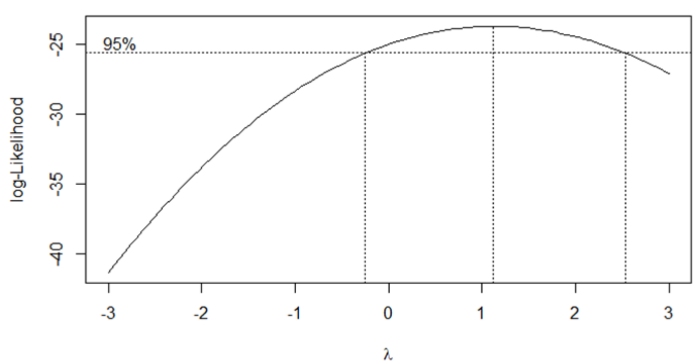


Fig. 4: Box-Cox plot of Y_t in the pre-intervention period

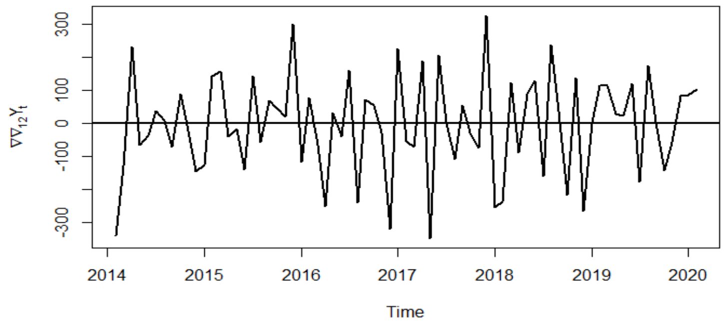
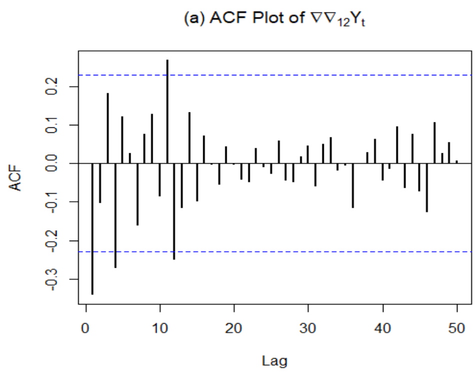


Fig. 5: Plot of $\nabla\nabla_{12}Y_t$

Graphical analysis of the time series plot of the first and seasonally differenced tourism accommodation income ($\nabla\nabla_{12}Y_t$) where $d = D = 1$ and $s = 12$ in Figure 5, suggests that $\nabla\nabla_{12}Y_t$ is stationary since there is no noticeable trend or pattern. In addition, the ADF test on $\nabla\nabla_{12}Y_t$ produced a statistically significant p-value (0.01) at the 5% significance level. As a result, the null hypothesis of non-stationarity is rejected, thus, $\nabla\nabla_{12}Y_t$ is stationary. Additionally, the KPSS test on $\nabla\nabla_{12}Y_t$ produced a statistically insignificant p-value (0.1) at the 5% significance level, suggesting that the null hypothesis of trend stationarity cannot be rejected. Hence, it is concluded that $\nabla\nabla_{12}Y_t$ is trend stationary, see Section 2.4 for more information.

3.4 Model Selection

The ACF plot based on the $\nabla\nabla_{12}Y_t$ presented in Figure 6(a) contains significant lags 1, 4, 11 and 12. The PACF plot of $\nabla\nabla_{12}Y_t$ in Figure 6(b) has significant lags 1, 2, 4 and 11. A candidate model based on the ACF and PACF in Figures 6(a) and (b) is SARIMA$(1,1,3)(0,1,1)_{12}$.



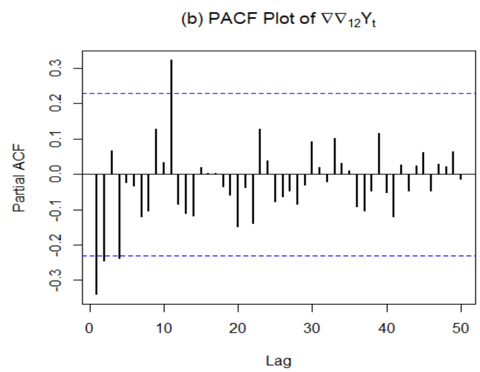


Fig. 6: (a) ACF and (b) PACF plots of $\nabla\nabla_{12}Y_t$

Table 1 summarises the AIC, BIC, RMSE and MAPE values of 11 candidate SARIMA models fitted on the pre-intervention Y_t using the auto.arima function in the R forecast package [46].

Table 1: AIC, BIC, RMSE and MAPE values of SARIMA models fitted on Y_t

Possible models for Y_t	**AIC**	**BIC**	**RMSE**	**MAPE**
SARIMA$(0,1,2)(0,1,1)_{12}$	921.65	930.81	113.6143	1.996%
SARIMA$(0,1,4)(0,1,1)_{12}$	919.26	933	106.7656	1.890%
SARIMA$(1,1,1)(0,1,1)_{12}$	921.6	930.84	113.6164	1.999%
SARIMA$(1,1,1)(0,1,2)_{12}$	921.79	933.24	108.7237	1.893%
SARIMA$(1,1,3)(0,1,1)_{12}$	920.97	934.71	109.2067	1.947%
SARIMA$(1,1,2)(0,1,2)_{12}$	918.92	932.66	103.346	1.806%
SARIMA$(1,1,2)(0,1,1)_{12}$	919.11	930.56	108.0806	1.931%
SARIMA$(2,1,1)(0,1,1)_{12}$	919.36	930.81	108.4047	1.929%
SARIMA$(3,1,1)(0,1,1)_{12}$	920.48	934.22	107.8393	1.927%
SARIMA$(4,1,0)(0,1,1)_{12}$	920.55	934.29	108.4534	1.916%
SARIMA$(2,1,2)(0,1,1)_{12}$	920.34	934.08	108.0735	1.933%

The SARIMA$(1,1,2)(0,1,2)_{12}$ model seems to have an overall best fit on the pre-intervention Y_t, because it produced the lowest AIC, RMSE and MAPE values besides the BIC value in Table 1. Based on Equation (1), the SARIMA$(1,1,2)(0,1,2)_{12}$ is expressed as,

$$(1-\phi_1 B)(1-B)(1-B^{12})Y_t = (1-\theta_1 B-\theta_2 B^2)(1-\Theta_1 B^{12}-\Theta_1 B^{24})\varepsilon_t \quad (8)$$

3.5 Parameter Estimation

The parameters of the SARIMA$(1,1,2)(0,1,2)_{12}$ pre-intervention model provided in Table 2 were estimated using MLE as discussed in subsection 2.5.2.

Table 2: MLE parameters for the SARIMA$(1,1,2)(0,1,2)_{12}$ model fitted on Y_t

Parameters	Estimate	Standard Error	z value	p-values
ϕ_1	-0.74148	0.16477	-4.5002	6.79×10^{-6} *
θ_1	0.31429	0.19675	1.5974	0.11018
θ_2	-0.60655	0.12018	-5.0472	4.48×10^{-7} *
Θ_1	-0.55092	0.25282	-2.1791	0.02932 *
Θ_2	-0.21748	0.17158	-1.2675	0.20497

* *Statistically significant at the 5% significance level*

All model parameters in Table 2 are statistically significant at a 5% significance level, except θ_1 and Θ_2. Removing these insignificant parameters lead to SARIMA$(1,1,1)(0,1,1)_{12}$ which has relatively higher AIC, BIC, RMSE and MAPE values than the chosen SARIMA$(1,1,2)(0,1,2)_{12}$ model.

Outlier detection on the SARIMA$(1,1,2)(0,1,2)_{12}$ model was conducted using the tsouliers R package [48]. Two outliers were detected, (i) an additive outlier at $t = 48$ corresponding to December 2016 with a statistically significant p-value $= 0.001$ at the 5% significance level and an estimated negative impact of ZAR 300 million and, (ii) an innovative outlier at $t = 53$ corresponding to May 2017 with a statistically insignificant p-value $= 0.135$ and an estimated negative impact of ZAR 122 million. These outliers are incorporated into the original SARIMA$(1,1,2)(0,1,2)_{12}$. Based on Equation (2), the SARIMA$(1,1,2)(0,1,2)_{12}$ with AO_{48} and IO_{53} is expressed as,

$$(1-\phi_1 B)(1-B)(1-B^{12})Y_t = (1-\theta_1 B-\theta_2 B^2)(1-\Theta_1 B^{12}-\Theta_1 B^{24})(\varepsilon_t + IO_{53}) + AO_{48} \quad (9)$$

The SARIMA$(1,1,2)(0,1,2)_{12}$ with AO_{48} and IO_{53} has AIC = 916.86, BIC = 935.18, RMSE = 98.86468 and MAPE = 1.761112. However, only the AO_{48} is statistically significant at the 5% significance level. This is not an ideal model for forecasting.

Table 3: MLE Parameter estimates of SARIMA$(1,1,2)(0,1,2)_{12}$ with AO_{86} and IO_{53} outlier model

Parameters	Estimate	Standard Error	z value	p-values
ϕ_1	0.52415	0.49399	1.0610	0.288
θ_1	-0.99126	0.51109	-1.9395	0.052
θ_2	0.18852	0.30412	0.6199	0.535
Θ_1	-0.41776	0.36535	-1.1435	0.252
Θ_2	-0.40548	0.25657	-1.5804	0.114
AO_{48}	-299.65672	86.13096	-3.4791	0.001*
IO_{53}	-122.47252	81.95913	-1.4943	0.135

* *Statistically significant at the 5% significance level*

The second-best model in Table 1 is the SARIMA$(1, 1, 2)(0, 1, 1)_{12}$ with an AIC = 919.11, BIC = 930.56, RMSE = 108.08 and MAPE = 1.93%. Based on Equation (1), the SARIMA$(1,1,2)(0,1,1)_{12}$ is expressed as,

$$(1-\phi_1 B)(1-B)(1-B^{12})Y_t = (1-\theta_1 B-\theta_2 B^2)(1-\Theta_1 B^{12})\varepsilon_t \quad (10)$$

Table 4: MLE Parameter estimates of SARIMA$(1,1,2)(0,1,1)_{12}$

Parameters	Estimate	Standard Error	z value	*p*-values
ϕ_1	-0.790034	0.086314	-9.1530	2.2×10^{-16} *
θ_1	0.371548	0.176557	2.1044	0.035*
θ_2	-0.628450	0.137592	-4.5675	4.936×10^{-6} *
Θ_1	-0.456162	0.175500	-2.5992	0.009*

* *Statistically significant at the 5% significance level*

The fitted $\text{SARIMA}(1,1,2)(0,1,1)_{12}$ has no outliers detected and all its parameter estimates are statistically significant at the 5% significance level. It has the least necessary number of parameters (parsimonious) with slight differences in the AIC, BIC, RMSE and MAPE values compared to the $\text{SARIMA}(1,1,2)(0,1,2)_{12}$. Therefore, the analysis going forward will use only the $\text{SARIMA}(1,1,2)(0,1,1)_{12}$ model.

3.6 Model Residual Analysis

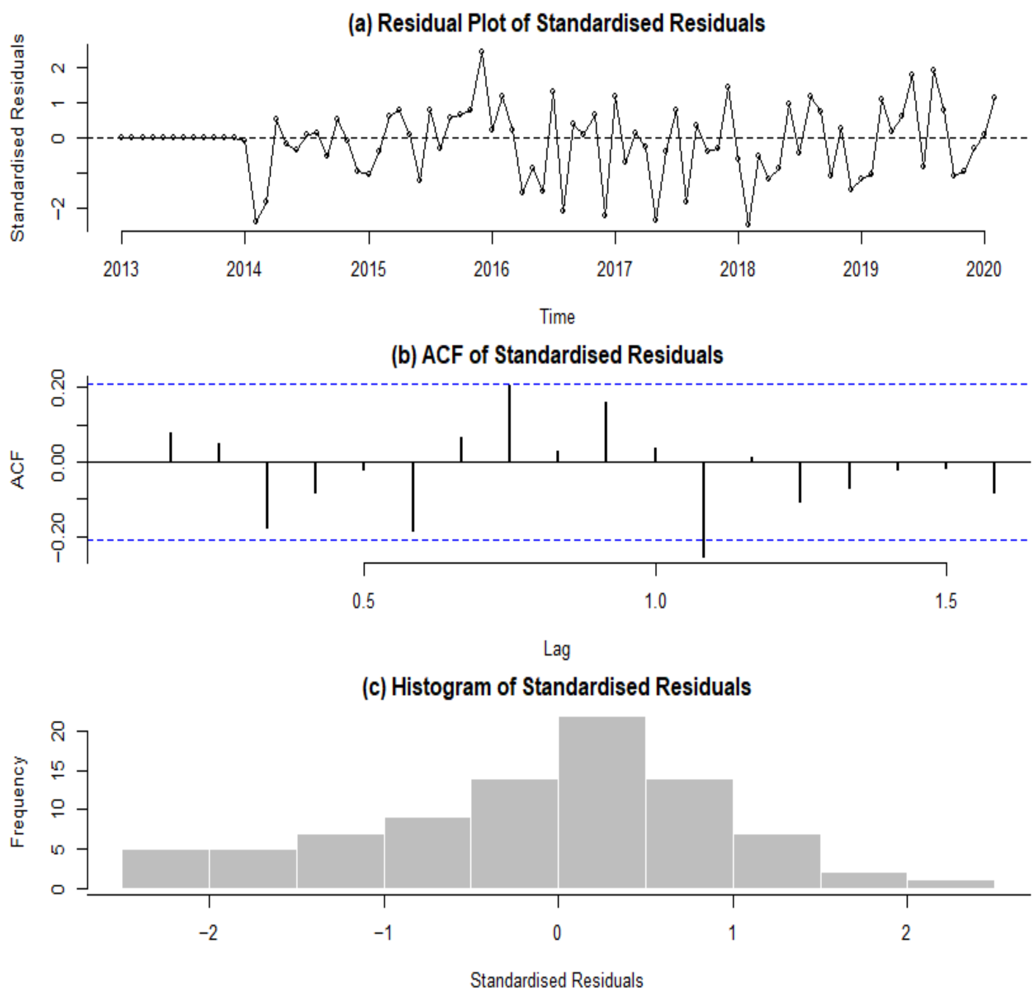


Fig. 7: (a) Residual plot, (b) ACF and (c) the histogram of the standardised residuals of $\text{SARIMA}(1,1,2)(0,1,1)_{12}$ model

The standardised residuals of the in Figure 7(a) have no apparent trend and the ACF in Figure 7(b) suggests that there is no autocorrelation on the standardised residuals of the $\text{SARIMA}(1,1,2)(0,1,1)_{12}$ model, with only one slightly significant lag. The *p*-values from The Ljung-Box (0.1007) and Box-Pierce (0.1677) tests respectively have statistically insignificant *p*-values of 0.1007 and 0.1677 at the 5% significance level, solidifying the null hypothesis that the standardised residuals from the selected $\text{SARIMA}(1,1,2)(0,1,1)_{12}$ model are uncorrelated. This shows that the slightly significant lag in the ACF plot in Figure 7(b) resulted from random sampling error. The histogram in Figure 7(c) slightly mimics that of normal distribution. The *p*-values from The Shapiro-Wilk (0.2281) and Jarque-Bera (0.6495) tests respectively produced statistically insignificant *p*-values of 0.2281 and 0.6495 at the 5% significance level. Thus, the standardised residuals of the $\text{SARIMA}(1,1,2)(0,1,2)_{12}$ model do not violate the model assumption of normality. Hence it is concluded that the standardised residuals from the fitted $\text{SARIMA}(1,1,2)(0,1,1)_{12}$ are white noise. The Ljung-Box, Box-Pierce, Shapiro-Wilk and Jarque-Bera tests are discussed in subsection 2.5.3.

3.7 Data vs Fitted

The comparative analysis of Y_t versus the fitted values in Figure 8 suggests that the selected $\text{SARIMA}(1,1,2)(0,1,1)_{12}$ model has a good fit on the pre-intervention Y_t data. Therefore, it is ideal for forecasting future tourism accommodation income.

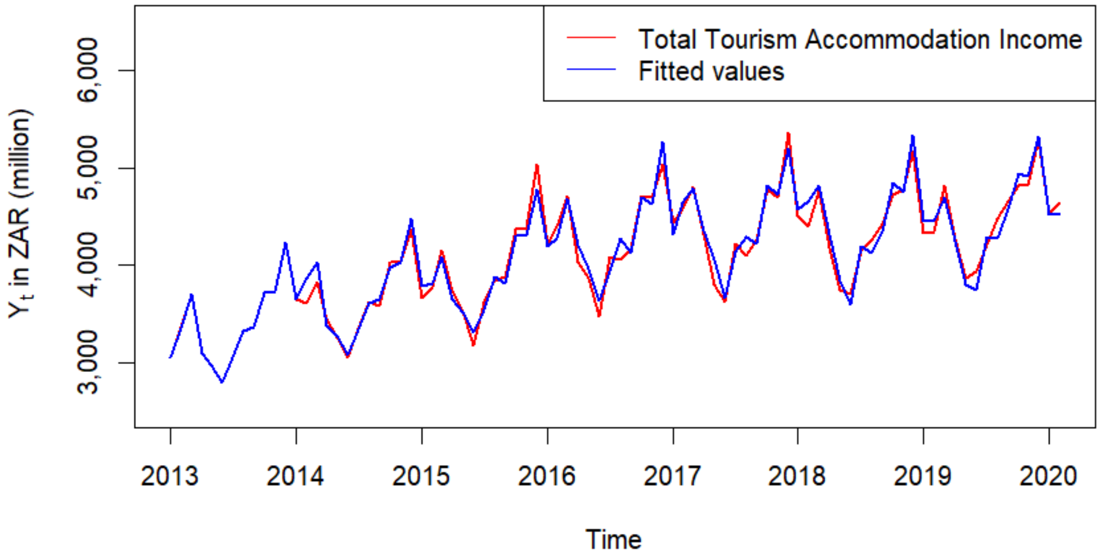


Fig. 8: Actual Y_t versus fitted $\hat{Y}_t$ from (a) SARIMA$(1,1,2)(0,1,1)_{12}$ model

3.8 Forecasting

In Figure 9, the actual values of Y_t during the pre-intervention period are represented by the black series. The green vertical line represents the intervention point in March 2020. The red series shows the interrupted actual values of Y_t from March 2020 to June 2024. The blue line depicts the in-sample forecasts generated from the SARIMA$(1,1,2)(0,1,1)_{12}$ model from March 2020 to June 2024, as well as 1-year out-of-sample forecasts from July 2024 to June 2025. It appears that the interrupted Y_t (red series) in Figure 9 failed to recover to its pre-intervention level, revealing that the pandemic continued to have a negative lingering impact on the income generated from tourism accommodation. The 80% and 95% confidence limits are shown using the dark grey and light grey shading, respectively. Prediction limits in Figure 9 are thin at the start of the in-sample forecasts but get thicker over time due to increased uncertainty. The out-of-sample forecasts imitate the stochastic periodicity depicted in the series from the pre-intervention period. Given that the interrupted Y_t did not recover to its pre-intervention levels, intervention analysis in the next section will be conducted based on the 52-month intervention period. Therefore, the period from March 2020 to June 2024 will be referred to as the "intervention period" in sections to follow.

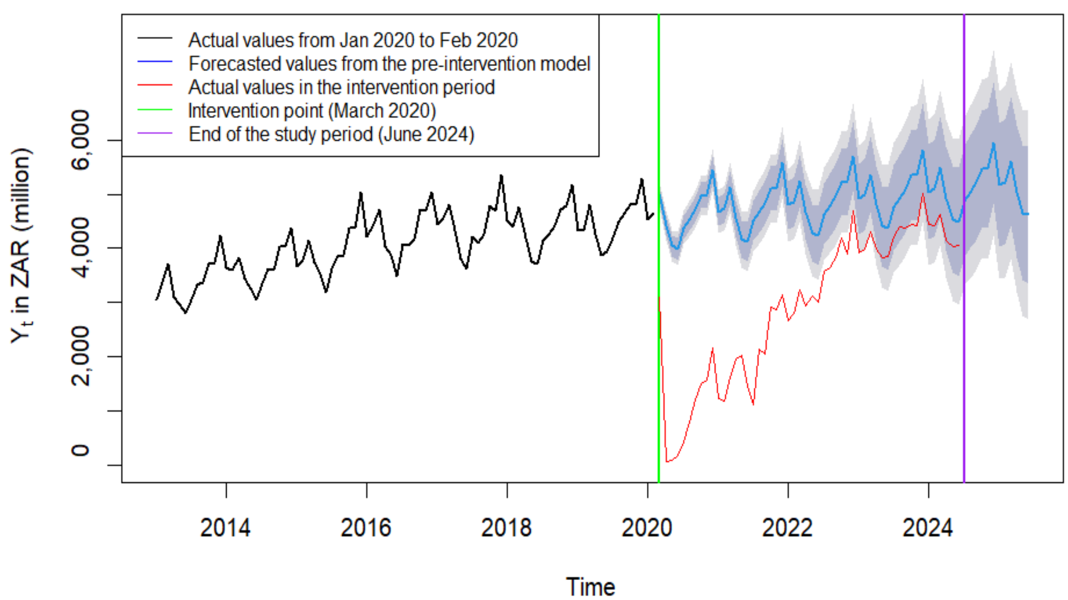


Fig. 9: Forecasted $\hat{Y}_t$ from the SARIMA$(1,1,2)(0,1,1)_{12}$ model

3.9 Intervention Analysis

In this section, the SARIMA$(1,1,2)(0,1,1)_{12}$ model based on the pre-intervention Y_t is expanded into the 52-month intervention period and augmented with a pulse function covariate regression vector as an exogenous component through trial-and-error using the steps outlined in Section 2.3. The resulting model becomes a SARIMAX$(1,1,2)(0,1,1)_{12}$ intervention model. The fitted values from the SARIMAX intervention model in Figure 10 produced a near to a perfect fit to the actual Y_t in the intervention period. Then, the fitted SARIMAX$(1,1,2)(0,1,1)_{12}$ intervention model is used to quantify intervention effects in the next Section 3.10.

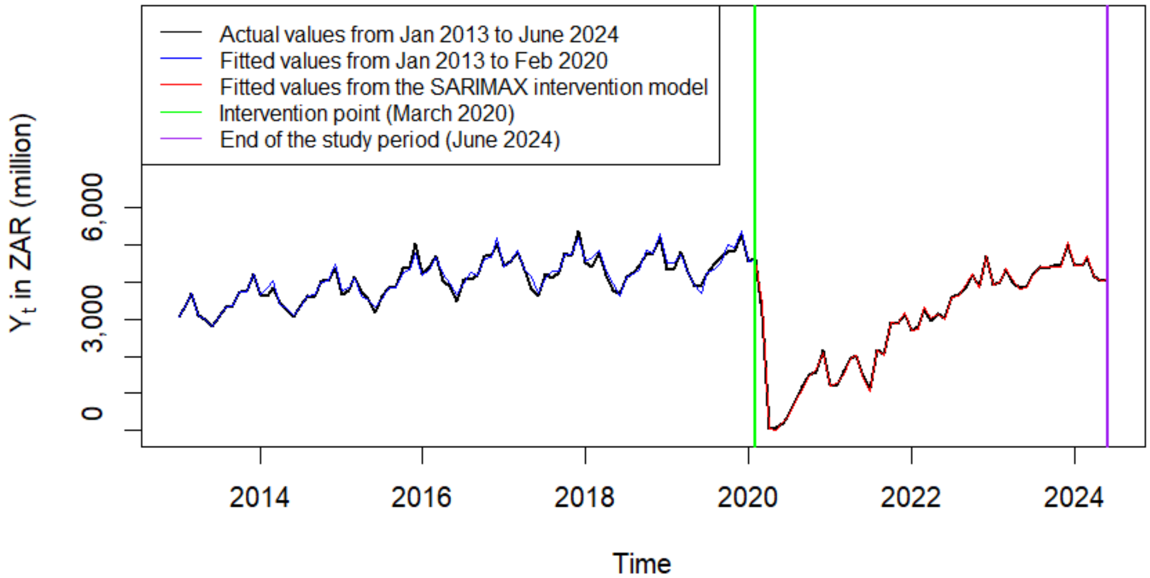


Fig. 10: SARIMAX$(1,1,2)(0,1,1)_{12}$ intervention model augmented with a pulse function covariate vector by trial-and-error

3.10 Quantifying Intervention Effects

Table 5 reports the estimated COVID-19 intervention effects on the tourism accommodation income (Y_t) in the intervention period. The first month into the intervention period (March 2020) recorded total income of ZAR 3,108 million which is ZAR 1,897 million or 37,9% less than the ZAR5,005 million which would have been generated in the absence of COVID-19. The tourism accommodation income was severely affected in April 2020 as the approximations in Table 5 indicate that the tourism industry generated merely ZAR 65 million in April 2020, which is a significantly lower than all other months in the intervention period. This corresponds to 98.5% reduction in expected income when compared to ZAR 5,005 million as predicted by the pre-intervention SARIMA$(1,1,2)(0,1,2)_{12}$ model.

Table 5: Estimated intervention effects of COVID-19 on Y_t in the intervention period

Time	Observed sales vs Predictions from the pre-intervention model						Fitted SARIMAX intervention model values vs Predictions from the pre-intervention model	
	Percentage Change (%)	Predicted values ($\hat{Y}_t$)	Actual Sales (Y_t)	Estimated losses ($Y_t - \hat{Y}_t$)	Covariate vector components	Fitted (Y_t)	Percentage change (%)	Estimated losses (Fitted (Y_t) − $\hat{Y}_t$))
Mar-2020	-37.9%	ZAR 5,005	ZAR 3,108	-ZAR 1,897	1	ZAR 3,380	-32.5%	-ZAR 1,625
Apr-2020	-98.5%	ZAR 4,419	ZAR 65	-ZAR 4,354	2.7	ZAR 30	-99.3%	-ZAR 4,389
May-2020	-97.8%	ZAR 4,045	ZAR 89	-ZAR 3,956	2.5	ZAR 10	-99.8%	-ZAR 4,035
Jun-2020	-95.1%	ZAR 3,995	ZAR 194	-ZAR 3,800	2.3	ZAR 239	-94.0%	-ZAR 3,755
Jul-2020	-90.3%	ZAR 4,386	ZAR 423	-ZAR 3,962	2.45	ZAR 396	-91.0%	-ZAR 3,990
Aug-2020	-82.7%	ZAR 4,543	ZAR 788	-ZAR 3,756	2.3	ZAR 784	-82.8%	-ZAR 3,760
Sep-2020	-74.5%	ZAR 4,718	ZAR 1,204	-ZAR 3,514	2.2	ZAR 1,073	-77.3%	-ZAR 3,646
Oct-2020	-69.7%	ZAR 4,975	ZAR 1,505	-ZAR 3,470	2.2	ZAR 1,477	-70.3%	-ZAR 3,498
Nov-2020	-68.6%	ZAR 4,985	ZAR 1,564	-ZAR 3,422	2.1	ZAR 1,647	-67.0%	-ZAR 3,338
Dec-2020	-60.5%	ZAR 5,450	ZAR 2,155	-ZAR 3,296	2.1	ZAR 2,092	-61.6%	-ZAR 3,358
Jan-2021	-73.9%	ZAR 4,674	ZAR 1,222	-ZAR 3,453	2.2	ZAR 1,188	-74.6%	-ZAR 3,486

Feb-2021	-74.8%	ZAR 4,733	ZAR 1,190	-ZAR 3,542	2.2	ZAR 1,267	-73.2%	-ZAR 3,466
Mar-2021	-68.2%	ZAR 5,120	ZAR 1,627	-ZAR 3,492	2.2	ZAR 1,512	-70.5%	-ZAR 3,607
Apr-2021	-57.1%	ZAR 4,549	ZAR 1,952	-ZAR 2,597	1.65	ZAR 1,953	-57.1%	-ZAR 2,596
May-2021	-51.1%	ZAR 4,163	ZAR 2,037	-ZAR 2,126	1.4	ZAR 2,018	-51.5%	-ZAR 2,145
Jun-2021	-64.6%	ZAR 4,122	ZAR 1,460	-ZAR 2,663	1.7	ZAR 1,402	-66.0%	-ZAR 2,721
Jul-2021	-75.4%	ZAR 4,506	ZAR 1,110	-ZAR 3,396	2.2	ZAR 1,057	-76.5%	-ZAR 3,449
Aug-2021	-53.9%	ZAR 4,669	ZAR 2,151	-ZAR 2,519	1.6	ZAR 2,168	-53.6%	-ZAR 2,501
Sep-2021	-57.5%	ZAR 4,840	ZAR 2,057	-ZAR 2,782	1.8	ZAR 2,008	-58.5%	-ZAR 2,832
Oct-2021	-42.6%	ZAR 5,100	ZAR 2,926	-ZAR 2,174	1.45	ZAR 2,922	-42.7%	-ZAR 2,178
Nov-2021	-43.8%	ZAR 5,107	ZAR 2,871	-ZAR 2,236	1.45	ZAR 2,895	-43.3%	-ZAR 2,213
Dec-2021	-43.7%	ZAR 5,575	ZAR 3,141	-ZAR 2,434	1.6	ZAR 3,156	-43.4%	-ZAR 2,418
Jan-2022	-44.2%	ZAR 4,797	ZAR 2,677	-ZAR 2,120	1.4	ZAR 2,699	-43.7%	-ZAR 2,098
Feb-2022	-42.5%	ZAR 4,857	ZAR 2,794	-ZAR 2,063	1.4	ZAR 2,728	-43.8%	-ZAR 2,129
Mar-2022	-38.2%	ZAR 5,242	ZAR 3,239	-ZAR 2,004	1.25	ZAR 3,313	-36.8%	-ZAR 1,930
Apr-2022	-37.1%	ZAR 4,673	ZAR 2,939	-ZAR 1,734	1.1	ZAR 3,023	-35.3%	-ZAR 1,650
May-2022	-26.9%	ZAR 4,286	ZAR 3,132	-ZAR 1,153	0.8	ZAR 3,139	-26.8%	-ZAR 1,147
Jun-2022	-28.9%	ZAR 4,246	ZAR 3,017	-ZAR 1,229	0.8	ZAR 3,002	-29.3%	-ZAR 1,244
Jul-2022	-22.6%	ZAR 4,629	ZAR 3,582	-ZAR 1,046	0.7	ZAR 3,601	-22.2%	-ZAR 1,028
Aug-2022	-23.8%	ZAR 4,793	ZAR 3,651	-ZAR 1,142	0.75	ZAR 3,635	-24.2%	-ZAR 1,158
Sep-2022	-22.8%	ZAR 4,963	ZAR 3,832	-ZAR 1,131	0.7	ZAR 3,902	-21.4%	-ZAR 1,061
Oct-2022	-19.9%	ZAR 5,223	ZAR 4,184	-ZAR 1,039	0.7	ZAR 4,205	-19.5%	-ZAR 1,019
Nov-2022	-25.5%	ZAR 5,231	ZAR 3,896	-ZAR 1,334	0.9	ZAR 3,841	-26.6%	-ZAR 1,390
Dec-2022	-17.7%	ZAR 5,698	ZAR 4,688	-ZAR 1,010	0.7	ZAR 4,690	-17.7%	-ZAR 1,008
Jan-2023	-20.3%	ZAR 4,920	ZAR 3,923	-ZAR 997	0.7	ZAR 3,917	-20.4%	-ZAR 1,004
Feb-2023	-19.9%	ZAR 4,980	ZAR 3,987	-ZAR 993	0.7	ZAR 3,979	-20.1%	-ZAR 1,001
Mar-2023	-19.6%	ZAR 5,366	ZAR 4,315	-ZAR 1,051	0.65	ZAR 4,350	-18.9%	-ZAR 1,015
Apr-2023	-16.9%	ZAR 4,797	ZAR 3,987	-ZAR 810	0.5	ZAR 4,065	-15.3%	-ZAR 732
May-2023	-13.2%	ZAR 4,409	ZAR 3,828	-ZAR 581	0.45	ZAR 3,797	-13.9%	-ZAR 612
Jun-2023	-11.9%	ZAR 4,369	ZAR 3,849	-ZAR 521	0.35	ZAR 3,840	-12.1%	-ZAR 530
Jul-2023	-11.3%	ZAR 4,752	ZAR 4,217	-ZAR 535	0.4	ZAR 4,189	-11.8%	-ZAR 563
Aug-2023	-10.4%	ZAR 4,916	ZAR 4,404	-ZAR 513	0.35	ZAR 4,409	-10.3%	-ZAR 508
Sep-2023	-14%	ZAR 5,086	ZAR 4,376	-ZAR 710	0.45	ZAR 4,407	-13.4%	-ZAR 680
Oct-2023	-16.7%	ZAR 5,347	ZAR 4,451	-ZAR 895	0.65	ZAR 4,412	-17.5%	-ZAR 934
Nov-2023	-17.4%	ZAR 5,354	ZAR 4,422	-ZAR 932	0.65	ZAR 4,410	-17.6%	-ZAR 944
Dec-2023	-14%	ZAR 5,821	ZAR 5,005	-ZAR 816	0.55	ZAR 5,073	-12.9%	-ZAR 748
Jan-2024	-11.7%	ZAR 5,044	ZAR 4,455	-ZAR 588	0.45	ZAR 4,435	-12.1%	-ZAR 608
Feb-2024	-13.3%	ZAR 5,104	ZAR 4,425	-ZAR 679	0.5	ZAR 4,424	-13.3%	-ZAR 680
Mar-2024	-15.6%	ZAR 5,489	ZAR 4,632	-ZAR 857	0.5	ZAR 4,701	-14.4%	-ZAR 788
Apr-2024	-15.6%	ZAR 4,920	ZAR 4,154	-ZAR 766	0.5	ZAR 4,148	-15.7%	-ZAR 772
May-2024	-10.8%	ZAR 4,533	ZAR 4,043	-ZAR 490	0.4	ZAR 4,015	-11.4%	-ZAR 518
Jun-2024	-9.5%	ZAR 4,493	ZAR 4,067	-ZAR 426	0.3	ZAR 4,054	-9.8%	-ZAR 439
	Total			-ZAR 99,006			Total	-ZAR 98,944

This means that April 2020 generated approximately -98,5% (ZAR 4,354 million) lower income than would have been generated in the absence of the pandemic. The total estimated loss of income for businesses in the tourism accommodation in the intervention period based on the sum of differences between the actual values (Y_t) and counterfactual values ($\hat{Y}_t$) from the pre-intervention $\text{SARIMA}(1,1,2)(0,1,1)_{12}$ model amount to ZAR 99,006 million. Similarly, the total estimated loss based on the sum of differences between the fitted values from the $\text{SARIMAX}(1,1,2)(0,1,1)_{12}$ intervention model and counterfactual values ($\hat{Y}_t$) from the pre-intervention $\text{SARIMA}(1,1,2)(0,1,1)_{12}$ model amount to ZAR 98,944 million (difference of 0.06% or ZAR 62 million). This serves as evidence that the chosen $\text{SARIMAX}(1,1,2)(0,1,1)_{12}$ intervention model adequately captured the impact of the pandemic on Y_t in the intervention period and is ideal to forecast future tourism accommodation income.

All relative percentage changes in Figure 11 are negative and display a decaying pattern with a sustained lingering pattern towards June 2024, indicating that the series did not recover to its pre-intervention level. This highlights the severity of the destructive nature of the COVID-19 pandemic on the total income from the R.S.A.'s tourism accommodation businesses. More importantly, the relative percentage changes towards June 2024 are significantly low when compared to those in the early stages of the pandemic.

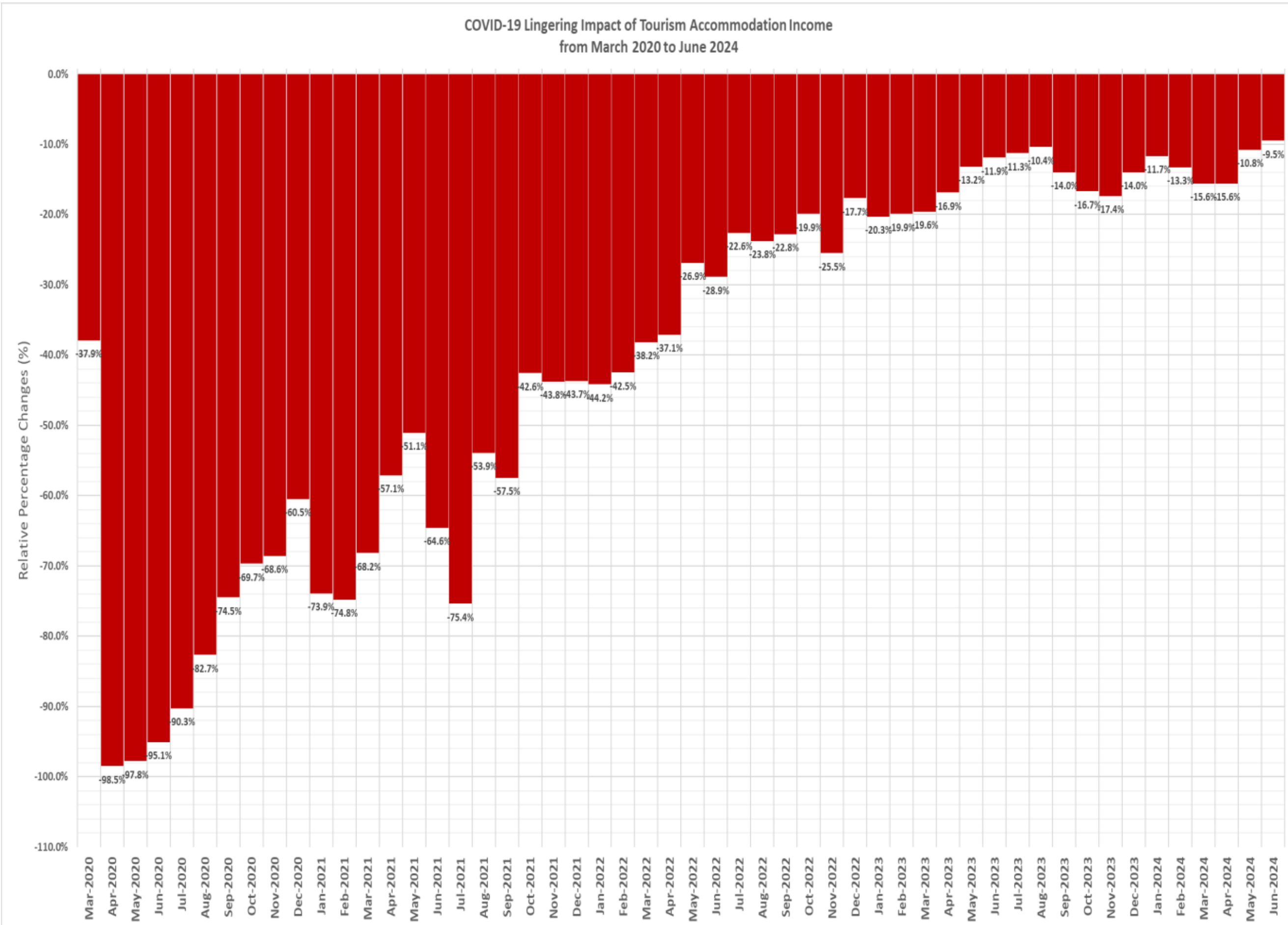


Fig. 11: Estimated Y_t reduction due to the lingering impact of COVID-19

The 1-year out-of-sample forecasts from the SARIMAX$(1,1,2)(0,1,1)_{12}$ intervention model in Figure 12 suggests that all things constant, the income generated by businesses in tourism accommodation will not return to its pre-intervention baseline by June 2025. Therefore, these forecasts serve as a reminder to encourage all associated stakeholders to tighten their rescue efforts and provide more resources to fasten the recovery of business enterprises within the R.S.A. tourism sector.

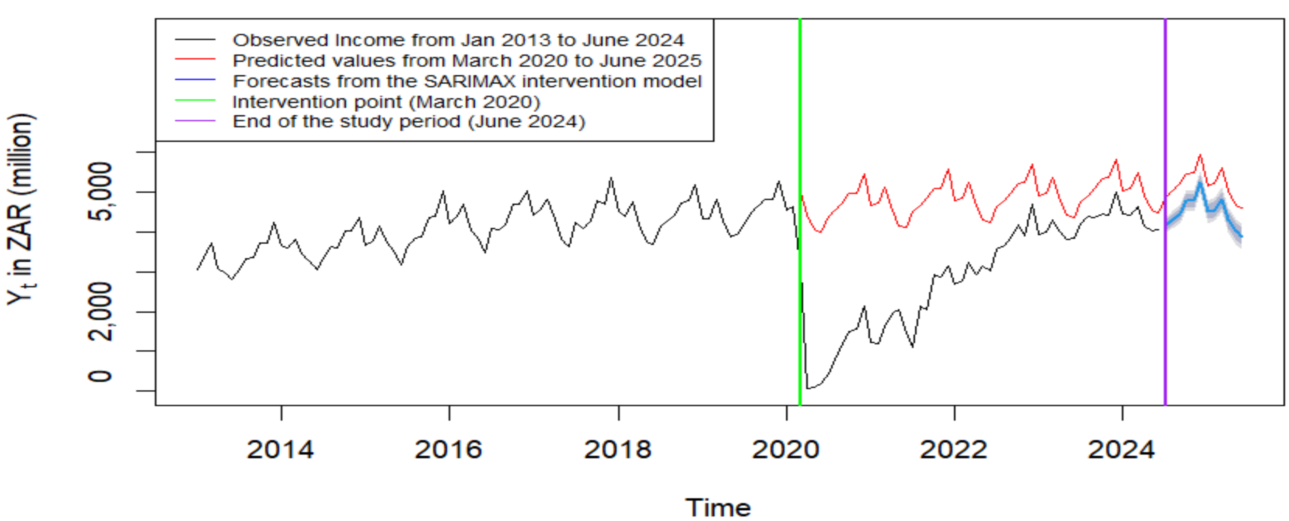


Fig. 12: Forecasted $\hat{Y}_t$ from SARIMAX$(1,1,2)(0,1,1)_{12}$ intervention model

4 Mitigation Strategies and Policy Recommendations

The national lockdown restrictions implemented by the R.S.A. government were relatively too strict compared to those implemented by other countries [1]. As a result, [2] identified elements of mistrust between local tourism enterprises in some parts of R.S.A. and the government. Therefore, fostering a good working relationship and information assimilation between business enterprises and the government may be a good start to resolving conflict in the unfortunate re-emergence of interventions like COVID-19. In an effort to speed up the recovery of the industry, the R.S.A. Minister of Tourism recently announced a "Gimme Summer – Sho't Left" campaign in the Free State province following a similar campaign in 2023, which generated a revenue of R38 billion [51]. The aim was to boost the revenue generated within the tourism sector by encouraging R.S.A. citizens to explore local attractions and support regional tourism economies. The campaign offered up to 50% discounts on travel experiences, making it easier for South Africans to afford travel amid rising living costs. The initiative also aimed to promote lesser-known regions and provide affordable, memorable travel experiences across all nine provinces. Aligning with the National Development Plan, which targets 15 million international arrivals by 2030, the campaign added to the R.S.A.'s commitment to welcoming global tourists [51]. However, the responsibility to ensure a fast recovery of the R.S.A. tourism industry should not sorely rest on the government. As part of their social responsibility, large thriving businesses can meet the government half-way by minimising economic barriers to entry and provide resources such as capital financing, skills development and mentorship for aspiring entrepreneurs within the industry. Good safety and hygiene standards and practices should be normalised to boost customer confidence and enhance the attractiveness of products and services offered across all tourism facilities.

5 Conclusion

Our study adds to the growing field of time series analysis for tourist-related studies by providing a framework for intervention effects analysis and quantification. This study emphasises the importance of explicit quantification of intervention effects for ease of interpretability for all associated stakeholders. The results of this study showed that the selected SARIMAX$(1,1,2)(0,1,1)_{12}$ intervention model successfully captured the negative lingering effect of the pandemic on the R.S.A.'s tourism accommodation income in the 52-month intervention period using the pulse function covariate vector with components estimated by trial-and-error. The estimated ZAR 99,006 million total loss in tourism income from March 2020 to June 2024 is a clear indication of the severity of the pandemic in the tourism industry as jobs and livelihoods of ordinary R.S.A.'s citizens and the profitability of tourism businesses are at stake. The negative ramifications of the pandemic on the income generated by the tourism sector have a direct negative impact on the overall R.S.A. economy. Although the tourism accommodation income had not returned/recovered to their pre-COVID-19 levels at the end of the study period, the relative percentage change of -9.5% in June 2024 shows that increased financial support through government subsidies and investments can boost revenue and speed up recovery of the tourism sector.

Forecasts on tourism data are important because they help managers of tourist sites to have enough time to plan accordingly and ensure that there are enough staff members on duty to cater for their guests. Additionally, clear and detailed estimated losses in this study may assist policymakers in making policy decisions that do not hamper the long-term sustainability of the tourism industry. The dataset in this study is analysed collectively for all nine provinces in R.S.A. as well as different types of accommodation (hotels, bed and breakfast, etc.), there is a possibility that certain provinces or certain types of tourist accommodations may have fully recovered, and others are worse off. Thus, as a future research idea, it would be interesting to conduct a similar analysis; however, focusing on each province or different types of accommodation. Also, machine learning techniques can be used to devise a more efficient algorithm to finding the optimal trial-and-error pulse function covariate vector components for use in longer time series. Further studies can use the multivariate SARIMAX model used in [19] to incorporate both tourist arrivals and tourism accommodation data in a single model and compare the estimates with those obtained in the current study.

Data Availability Statement:

All the data are publicly available on the following links: https://www.statssa.gov.za/

Conflicts of Interest Statement

The authors certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

References

[1] Rogan, M. and Skinner, C. (2020). The COVID-19 Crisis and the South African Informal Economy: 'Locked-out' of Livelihoods and Employment, Cape Town: University of Cape Town, National Income Dynamics Study (NIDS) – Coronavirus Rapid Mobile Survey, Report 10. https://www.datafirst.uct.ac.za/dataportal/index.php/citations/6835.

[2] Rogerson, J.M. (2021). Tourism business responses to South Africa's COVID-19 pandemic emergency. *GeoJournal of Tourism and Geosites*, 35(2), 338–347. http://dx.doi.org/10.30892/gtg.35211-657.

[3] Rogerson, C.M. and Rogerson, J.M. (2020). COVID-19 tourism impacts in South Africa: Government and industry responses. *GeoJournal of Tourism and Geosites*, 31(3), 1083-1091. https://doi.org/10.30892/gtg.31321-544.

[4] Sucheran, R. (2021). Preliminary Economic Impacts of the COVID-19 Pandemic on the Hotel Sector in South Africa. *African Journal of Hospitality, Tourism and Leisure*, 10(1):115-130. https://doi.org/10.46222/ajhtl.19770720-84.

[5] Dube, K. (2021). Implications of COVID-19 Induced Lockdown on the South African Tourism Industry and Prospects for Recovery. *African Journal of Hospitality, Tourism and Leisure*, 10(1), 270-287. https://doi.org/10.46222/ajhtl.19770720-99.

[6] Zenker, S. and Kock, F. (2020). The coronavirus pandemic – A critical discussion of a tourism research agenda. *Tourism Management*, 81, 104164. https://doi.org/10.1016/j.tourman.2020.104164.

[7] Butcher, J. (2020). Let the good times roll (as soon as possible): Why we need a post COVID convivial revolution. *ATLAS Tourism and Leisure Review*, 2020(2), 26-29. https://vbn.aau.dk/ws/portalfiles/portal/335951637/ATLAS_Review_2020_2_Corona_edition.pdf

[8] Demir, M., Demir, Ş. Ş., Dalgıç, A., Ergen, F. D. (2021). Impact of COVID-19 pandemic on the tourism industry: An evaluation from the hotel managers' perspective. *Journal of Tourism Theory and Research*, 7(1), 44-57, 857610. https://doi.org/10.24288/jttr.857610.

[9] Persson-Fischer, U., Liu, S. (2021). The Impact of a Global Crisis on Areas and Topics of Tourism Research, *Sustainability*, 13(2), 906. https://doi.org/10.3390/su13020906.

[10] Corbisiero, F., and La Rocca, R.A. (2020). Tourism on demand: A new form of urban and social demand of use after the pandemic event. *TeMA-Journal of Land Use, Mobility and the Environment, (Special Issue)*, 91-104. https://doi.org/10.6092/1970-9870/6916.

[11] Sigala, M. (2020). Tourism and COVID-19: Impacts and implications for advancing and resetting industry and research. *Journal of Business Research*, 117, 312-321. https://doi.org/10.1016/j.jbusres.2020.06.015.

[12] Qiu, R.T.R., Park, J., Li, S. and Song, H. (2020). Social cost of tourism during the COVID-19 pandemic. *Annals of Tourism Research*, 84, 102994. https://doi.org/10.1016/j.annals.2020.102994.

[13] Nyawo, J.C. (2020). Evaluation of Government Responses and Measures on COVID-19 in the Tourism Sector: A Case of Tour Guides in South Africa. *African Journal of Hospitality, Tourism and Leisure*, 9(5), 1144-1160. https://doi.org/10.46222/ajhtl.19770720-74.

[14] Bassier, I., Budlender, J., Zizzamia, R. and Jain, R. (2023). The labour market and poverty impacts of COVID-19 in South Africa. *South African Journal of Economics*, 91(4), 419-445. https://doi.org/10.1111/saje.12356.

[15] Casale, D. and Posel, D. (2020). Gender and the early effects of the COVID-19 crisis in the paid and unpaid economies in South Africa. *National Income Dynamics (NIDS)-Coronavirus Rapid Mobile Survey (CRAM) Wave*, *1*, 1-25.

[16] Rogerson, C.M & Rogerson J.M. (2020b). COVID-19 and Tourism Spaces of Vulnerability in South Africa. *African Journal of Hospitality, Tourism and Leisure*, 9(4), 382-401. http://dx.doi.org/10.46222/ajhtl.19770720-26.

[17] Makoni, T., Mazuruse, G. and 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.

[18] Chipumuro, M. and Chikobvu, D. (2022). Modelling Tourist Arrivals in South Africa To Assess the Impact of the COVID-19 Pandemic on the Tourism Sector. *African Journal of Hospitality, Tourism and Leisure*, 11(4), 1381-1394. https://doi.org/10.46222/ajhtl.19770720.297.

[19] Prilistya, S.K., Permanasari, A.E. and Fauziati, S. (2021, August). The effect of the COVID-19 pandemic and Google trends on the forecasting of international tourist arrivals in Indonesia. In *2021 IEEE Region 10 Symposium*

(TENSYMP) (pp. 1-8). IEEE. https://doi.org/10.1109/TENSYMP52854.2021.9550838.

[20] Chipumuro, M., Chikobvu, D. and Makoni, T. (2024). Statistical Analysis of Overseas Tourist Arrivals to South Africa in Assessing the Impact of COVID-19 on Sustainable Development. *Sustainability*, 16, 5756. https://doi.org/10.3390/su16135756.

[21] Devi, R., Agrawal A. and Dhar, J. and Misra, A.K. (2024). Forecasting of Indian tourism industry using modeling approach, *MethodsX*, 12, 102723. https://doi.org/10.1016/j.mex.2024.102723.

[22] Nyoni, T., 2021. Modeling and forecasting international tourism in Zimbabwe: A bright future for Zimbabwe's tourism industry. *Journal of Applied Economic Sciences*, 16(4), pp.421-434. https://doi.org/10.57017/jaes.v16.4(74).05.

[23] Masena T.E. & Shongwe S.C. (2024). Investigating the lingering effects of the pandemic on wholesale industry sales in South Africa. *ITM Web of Conferences*, 67, 01021. https://doi.org/10.1051/itmconf/20246701021.

[24] Masena T.E. & Shongwe S.C. (2024). The impact of the pandemic on the retail industry sales in South Africa: A Box-Jenkins approach. *ITM Web of Conferences*, 67, 01025. https://doi.org/10.1051/itmconf/20246701025.

[25] Song, H., Qiu, R.T. and Park, J. (2019). A review of research on tourism demand forecasting: Launching the Annals of Tourism Research Curated Collection on tourism demand forecasting. *Annals of Tourism Research*, 75, 338-362. https://doi.org/10.1016/j.annals.2018.12.001.

[26] Duan, Q. (2024). Predicting Tourism Demand by Combining Search Engine Data. *Tourism Management and Technology Economy.* 7, 1-8, 70301. http://dx.doi.org/10.23977/tmte.2024.070301.

[27] Jula, D. and Jula, N.M. (2023). The impact of COVID-19 pandemic on tourism activities. A counterfactual analysis. *HERiPRENEURSHIP*. Proceedings of the Tulcea International Conference, 229–248. https://biblioteca-digitala.ro/reviste/carte/icem/bipp/11_BIPP_Heripreneurship_2023_28.pdf

[28] Mendieta-Aragón A., Navío-Marco J. and Garín-Muñoz T. (2025). Twitter's capacity to forecast tourism demand: the case of way of Saint James. *European Journal of Management and Business Economics*. https://doi.org/10.1108/EJMBE-09-2023-0295

[29] Wu J., Li M., Zhao E., Sun S. and Wang S. (2023). Can multi-source heterogeneous data improve the forecasting performance of tourist arrivals amid COVID-19? Mixed-data sampling approach. *Tourism Management*, 98, 104759. https://doi.org/10.1016/j.tourman.2023.104759.

[30] Rianda F. and Usman H. (2023). Forecasting Tourism Demand During the COVID-19 Pandemic: ARIMAX and Intervention Modelling Approaches. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*., 17(1), 285-294. https://doi.org/10.30598/barekengvol17iss1pp0285-0294

[31] Min J.C., Lim C. and Kung H.H. (2011). Intervention analysis of SARS on Japanese tourism demand for Taiwan. *Quality & Quantity*, 45, 91-102. https://doi.org/10.1007/s11135-010-9338-4.

[32] Ahmar A.S., Guritno S., Rahman A., Minggi I., Tiro M.A., Aidid M.K., Annas S., Sutiksno D.U., Ahmar D.S., Ahmar K.H. and Ahmar, A.A. (2018). Modeling data containing outliers using ARIMA additive outlier (ARIMA-AO). *Journal of Physics: Conference Series* (Vol. 954, 012010). IOP Publishing. https://doi.org/10.1088/17426596/954/1/012010.

[33] Akpan E.A. and Moffat I.U. (2019). Modeling the effects of outliers on the estimation of linear stochastic time series model. *International Journal of Analysis and Applications*, 17(4), 530-547. https://doi.org/10.28924/2291-8639.

[34] Aliffia, N., Sediono, S., Suliyanto, S., Mardianto, M.F.F. and Amelia, D. (2024). Modelling of world crude oil price based on pulse function intervention analysis approach. *Media Statistika*, 16(2), 136–47. https://doi.org/10.14710/medstat.16.2.136-147.

[35] Masena T.E., Shongwe S.C. and Yeganeh A. (2024). Quantifying Loss to the Economy Using Interrupted Time Series Models: An Application to the Wholesale and Retail Sales Industries in South Africa. *Economies*, 12(9), 249. https://doi.org/10.3390/economies12090249.

[36] Masena T.E., Mahlangu S.L. and Shongwe S.C. (2024). Time Series Perspective on the Sustainability of the South African Food and Beverage Sector. *Sustainability*, 16(22), 9746. https://doi.org/10.3390/su16229746.

[37] Cryer J.D. and Chan K. (2008). Time series analysis with applications in R. New York: Springer. https://doi.org/10.1007/978-0-387-75959-3.

[38] Su J.J., Amsler C. and Schmidt P. (2012). A note on the size of the KPSS unit root test. *Economics Letters*, 117(3), 697-

699. https://doi.org/10.1016/j.econlet.2012.08.019.

[39] Hodson T.O. (2022). Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development Discussions*, 15(14), 5481-5487. https://doi.org/10.5194/gmd-15-5481-2022.

[40] Shah, J., Vaidya, D. and Shah, M. (2022). A comprehensive review on multiple hybrid deep learning approaches for stock prediction. *Intelligent Systems with Applications*, 16, 200111. https://doi.org/10.1016/j.iswa.2022.200111.

[41] Brockwell, P.J. and Davis, R.A. (1996). *Introduction to Time Series and Forecasting*. Springer Nature: Dordrecht, The Netherlands. https://doi.org/10.1007/978-3-319-29854-2.

[42] Box G.E.P.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.) John Wiley & Sons: Hoboken, NJ, USA.

[43] Gujarati D.N. and Porter D.C. (2008). Basic Econometrics (5th ed.). McGraw-Hill Irwin: New York, NY, USA.

[44] R Core Team. (2024). R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Austria, Available online: https://www.rproject.org/ (accessed on 26 November 2024).

[45] Trapletti A. and Hornik K. (2018). tseries: Time Series Analysis and Computational Finance. R Package Version 0.10-43. Available online: https://cran.rproject.org/package=tseries (accessed on 26 November 2023).

[46] Hyndman R.J. and Khandakar Y. (2008). Automatic time series forecasting: the forecast package for R. Journal of Statistical Software, 27, 1-22. https://doi.org/10.18637/jss.v027.i03.

[47] Venables, W. and Ripley, B. (2002). *Modern Applied Statistics with S (4th ed.)*. Springer: New York, NY, USA. https://doi.org/10.1007/978-0-387-21706-2.

[48] López-de-Lacalle, J. (2019). tsoutliers: Detection of Outliers in Time Series. R Package Version 0.6-10. Available online: https://cran.r-project.org/web/packages/tsoutliers/index.html (accessed on 17 December 2024).

[49] Zeileis, A., and Hothorn, T. (2002). Diagnostic checking in regression relationships. *R News*, 2(3), 7–10. Available online: https://cran.r-project.org/doc/Rnews/ (accessed 17 December 2024).

[50] Stats SA. (2024). Statistical Release – Tourism accommodation (Preliminary). Available online: https://www.statssa.gov.za/?page_id=1854&PPN=P6410&SCH=73712, (accessed on 17 December 2024).

[51] Khumalo, J. (2024). SA's Summer Campaign Launch Sparks Local and International Tourism Boost. Central News. Available online: https://centralnews.co.za/sas-summer-campaign-launch-sparks-local-and-international-tourism-boost/ (accessed on 20 January 2025).