Background

The data used for the purposes of completing this analysis is based on a monthly survey of both public and private accommodation businesses involved in short-stay accommodation across South Africa. These businesses include a range of establishments such as hotels, motels, bed-and-breakfasts (BnBs), guest houses, caravan parks, camping sites, and lodges. The dataset spans from January 2007 to June 2024 and covers businesses registered for Value-Added Tax (VAT) and income tax, using data from a stratified sample of over 1,000 enterprises.

Due to travel booms that were experienced during the 2010 FIFA World Cup, which may have influenced earlier data, this analysis focuses on the period starting from January 2013 to ensure greater stability and reliability in the dataset.

Data Analysis

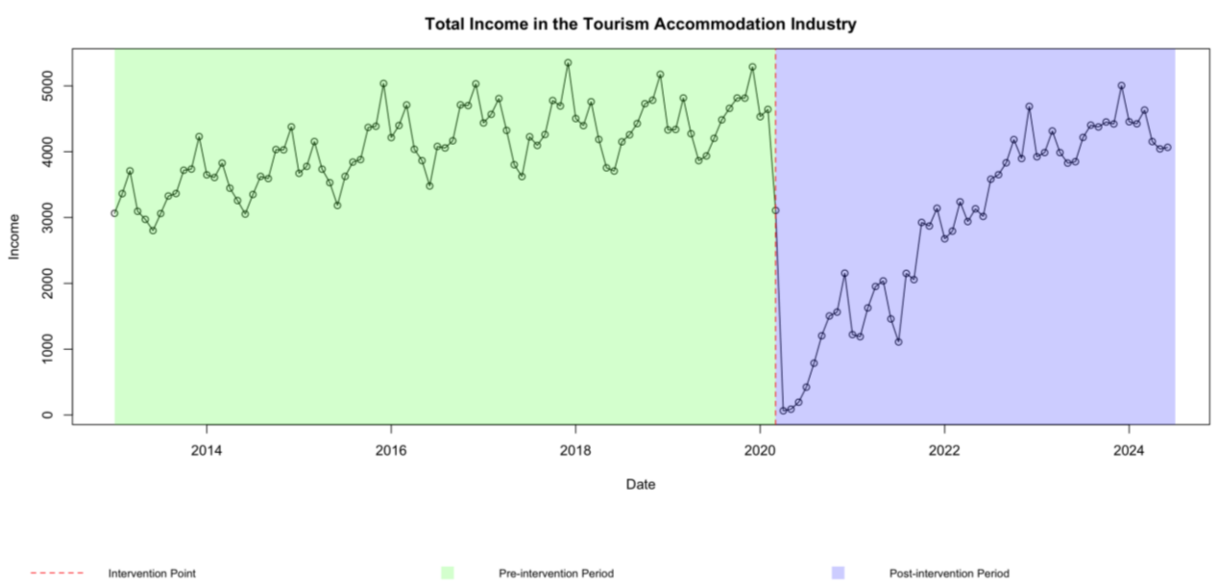


Figure 1: Time Series Plot with pre & post intervention periods noted

Variability

The plot shows moderate variability in income levels throughout the period from January 2013 to June 2024. There are regular fluctuations, suggesting seasonal trends in the tourism accommodation industry, with income levels generally oscillating between R2000 and R5000 million.

Trend

There is a clear upward trend in income from January 2013 to March 2020, indicating steady growth in the tourism accommodation industry during the pre-intervention period. However, the trend dramatically shifts in March 2020, which serves as the intervention starting point, with a sharp decline followed by a slow recovery. After this extreme event, the upward trend resumes from March 2020 to June 2024, albeit at a slower pace.

Extreme Events

The most prominent feature is the extreme drop in tourism income around March 2020, coinciding with the global impact of the COVID-19 pandemic, which caused widespread travel restrictions and a significant decline in tourism. This intervention had an immediate negative impact on income levels in the tourism accommodation industry, as evidenced by the drastic decrease in revenue.

Following this extreme drop in tourism income, there is a notable recovery trend in the post-intervention period (from March 2020 to June 2024), although the variability remains high as income gradually returns to pre-2020 levels. This suggests that while the recovery is underway, the effects of the pandemic continue to influence the industry landscape.

Five Number Summary

```
##      Minimum          1st Quartile       Median      3rd Quartile            Maximum
##      64.7                3238.7          3909.8         4396.5                5352.8
```

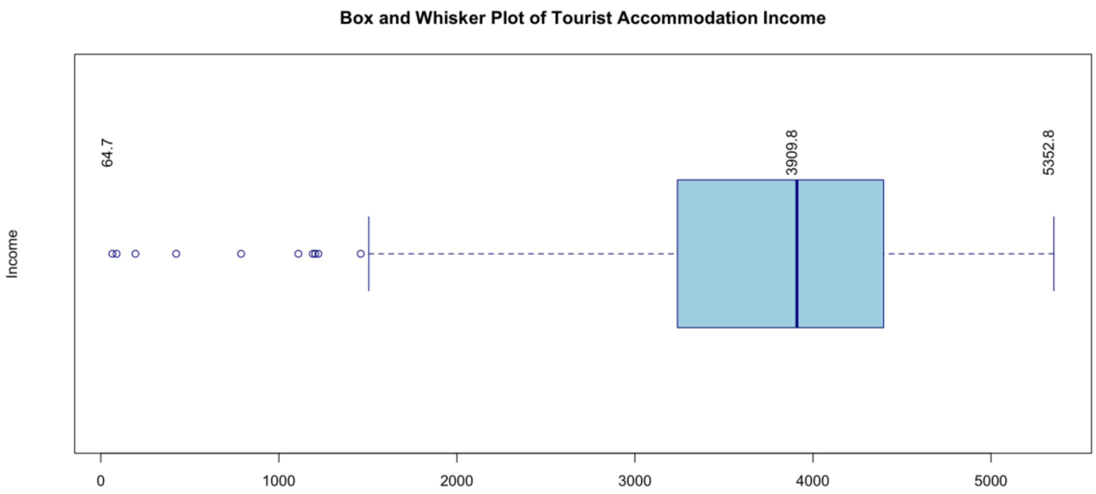


Figure 2: Box and Whisker Plot

The income range spans from R64.7 million to R5352.8 million. The median income of R3909.8 million is significantly higher than the minimum, indicating a skew towards higher income values among tourist accommodations. Additionally, the interquartile range (IQR) of R1157.8 million suggests that the middle 50% of incomes are fairly clustered, indicating stability in income levels among many accommodations.

Pre-intervention Analysis

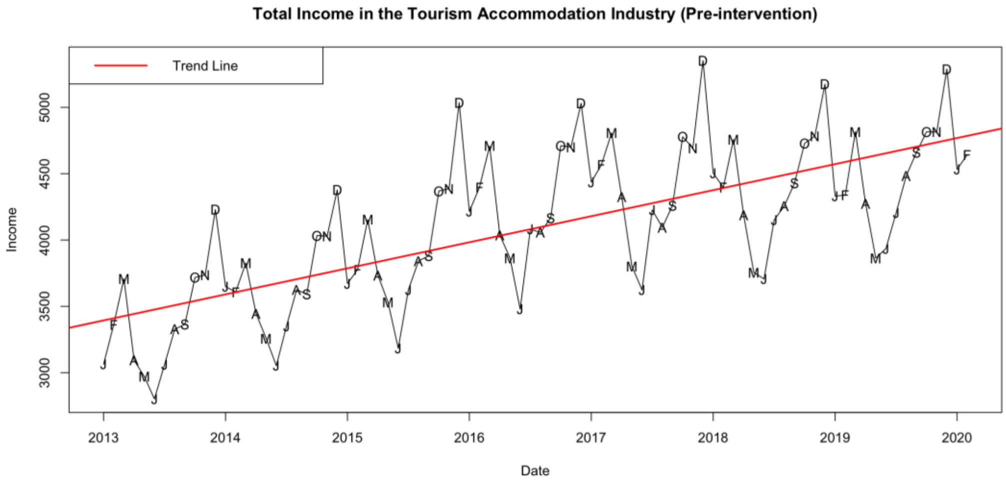


Figure 3: Pre-intervention Time Series Plot

The time series plot of the pre-intervention data reveals several important characteristics of the underlying process. First, the plot clearly indicates an upward trend in the data over time, which indicates stationarity. Additionally, there are signs of seasonality, with earlier months in the year (March - June) typically showing lower income and later months (September - December) exhibiting higher income.

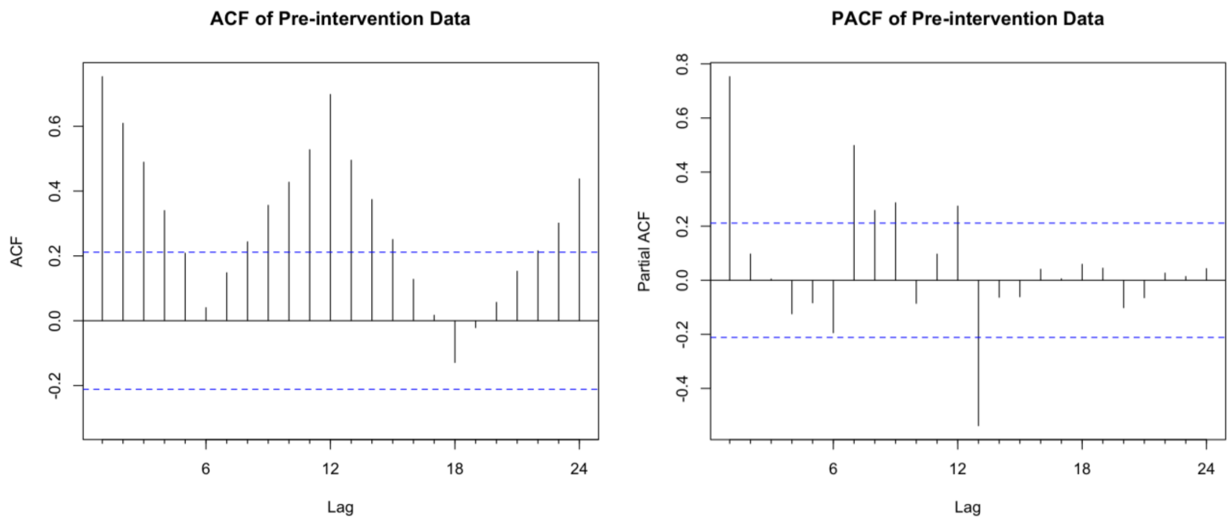


Figure 4: ACF and PACF of Pre-intervention Data

The ACF further supports the observations of made in the pre-intervention time series plot. Significant autocorrelations are evident at several lags, especially at earlier lags and again at lag 12, indicating the presence of seasonality. The slow decay in the ACF suggests that the series is non-stationary and might be an AR(p) process.

The PACF shows a sharp drop-off after lag 1, which suggests that the data may be suitable for an $AR(1)$ model. There's also a spike at lag 12 in the PACF, further solidifying the presence of seasonal patterns.

Transformations

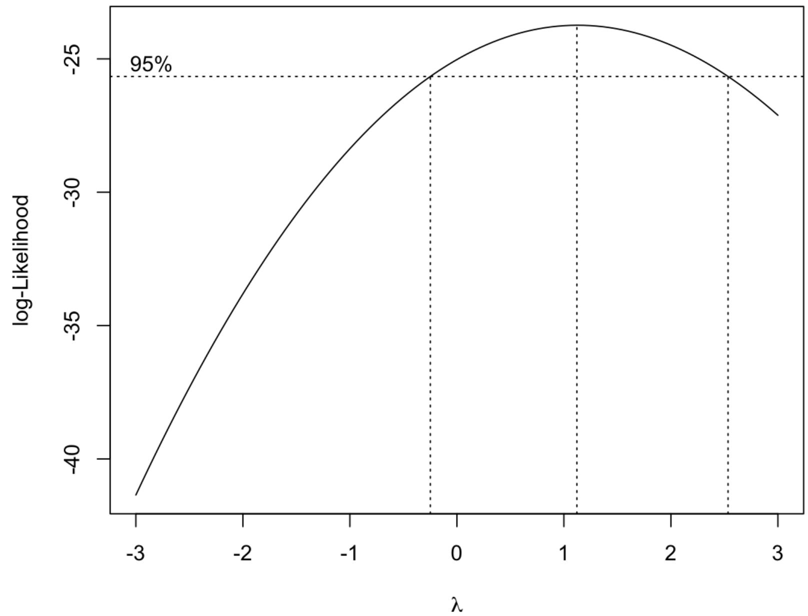


Figure 5: Box-Cox Plot

The lambda value obtained from the Box-Cox transformation was 1, which lies between the 95% confidence interval so this left our data untransformed and thus unchanged. This option was also chosen to aid in a simpler analysis.

Augmented Dickey-Fuller Test	**Phillips-Perron Test**	**KPSS Test**

`Augmented Dickey-Fuller Test`
 	
`data: pre_intervention_ts`
 	
`Dickey-Fuller = -5.407, Lag order = 4, p-value = 0.01`
 	
`alternative hypothesis: stationary`
	`Phillips-Perron Unit Root Test`
	
	`data: pre_intervention_ts`
	
	`Dickey-Fuller Z(alpha) = -40.715, Truncation lag parameter = 3, p-value = 0.01`
	
	`alternative hypothesis: stationary`
		`KPSS Test for Level Stationarity`
		
		`data: pre_intervention_ts`
		
		`KPSS Level = 1.4622, Truncation lag parameter = 3, p-value = 0.01`

Figure 6: Stationarity Tests on Pre-intervention Data

The ADF and PP tests both indicate p-values of 0.01<0.05, which means that we reject the null hypothesis that the data is stationary. The KPSS test gives us a test statistic of 1.4622 and a p-value of 0.01<0.05, which means we reject the null hypothesis that the series is level stationary. There might be hints that some minor non-stationarity might remain. Therefore, a first and seasonal difference of the series was taken.

Augmented Dickey-Fuller Test	Phillips-Perron Test	KPSS Test
Augmented Dickey-Fuller Test data: differenced_pre_intervention_ts Dickey-Fuller = -4.9195, Lag order = 4, p-value = 0.01 alternative hypothesis: stationary	Phillips-Perron Unit Root Test data: differenced_pre_intervention_ts Dickey-Fuller Z(alpha) = -86.781, Truncation lag parameter = 3, p-value = 0.01 alternative hypothesis: stationary	KPSS Test for Level Stationarity data: differenced_pre_intervention_ts KPSS Level = 0.11553, Truncation lag parameter = 3, p-value = 0.1

Figure 7: Stationarity Tests on Differenced Pre-intervention Data

The results from the differenced data suggest that stationarity has been successfully achieved. Both the ADF and PP tests continue to reject the null hypothesis of non-stationarity with p-values=0.01<0.05, while the KPSS test now gives a test statistic of 0.11553 and a p-value=0.1>0.05, which means that we fail to reject the null hypothesis of level stationarity.

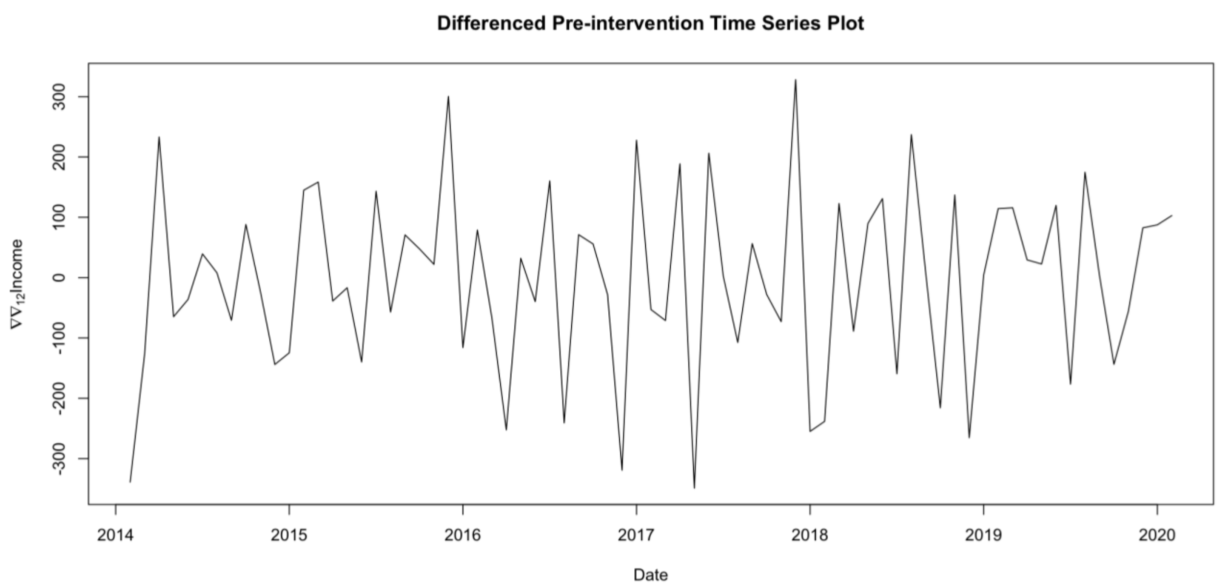


Figure 8: Stationary Pre-intervention Time Series Plot

The resulting time series plot does not seem to exhibit any patterns or trends, indicating that the differencing process has effectively removed any seasonal or long-term trends present in the original series. The data now appears more erratic, with fluctuations around a constant mean.

Model Selection

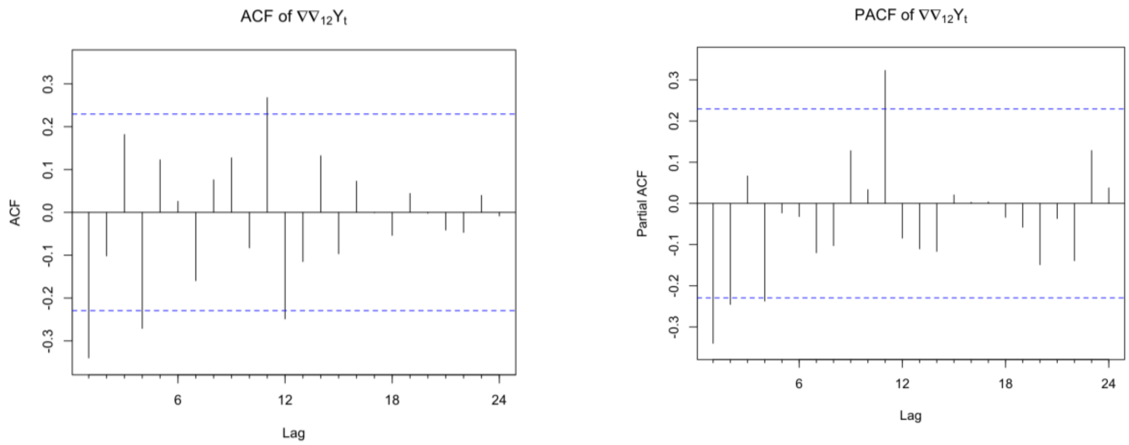


Figure 9: ACF and PACF of Differenced Pre-intervention Data

The ACF cuts off after lag 1, suggesting that an $MA(1)$ model be fit. The ACF also shows significant spikes around lag 12, which indicates the presence of a strong seasonal component in the data with a yearly cycle.

The PACF cuts off after lag 2, suggesting that an $AR(2)$ model be fit.

Given the monthly nature of the data and the patterns in both the ACF and PACF, the model that fits this would likely be an $ARIMA(2,1,1) \times (0,1,1)_{12}$.

Other models to fit that were picked, as suggested by the auto.arima() function are:

- $ARIMA(0,1,1) \times (0,1,1)_{12}$
- $ARIMA(2,1,2) \times (0,1,1)_{12}$

Parameter Estimation

$\boldsymbol{ARIMA(0,1,1) \times (0,1,1)_{12}}$ Model

	Estimate	Std.Error	z value	Pr(>|z|)	
ma1	-0.58006	0.10996	-5.2750	0.0000001328	***
sma1	-0.49215	0.16864	-2.9184	0.003518	**

sigma^2 estimated as 15206: log likelihood = -456.93, aic = 917.85

AIC=915.85 AIC_c=915.91 BIC=918.14

Training set error measures:

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	-15.53756	113.6155	83.9986	-0.424341	2.005787	0.3906666	0.05292667

The $ARIMA(0,1,1) \times (0,1,1)_{12}$ model reveals a significant moving average component and seasonal moving average component with an estimate of -0.58006 and an estimate of -0.49215, respectively.

The estimated variance of the errors is $\sigma^2 = 15206$, which suggests that while the model captures much of the volatility in the data, there remains a level of unexplained variation.

$ARIMA(2,1,1) \times (0,1,1)_{12}$ Model

	Estimate	Std.Error	z value	Pr(>|z|)	
ar1	-1.10266	0.30867	-3.5722	0.0003539	***
ar2	-0.52969	0.10214	-5.1857	0.0000002152	***
ma1	0.67359	0.40163	1.6771	0.0935166	.
sma1	-0.60481	0.18253	-3.3135	0.0009214	***

sigma^2 estimated as 13843: log likelihood = -454.68, aic = 917.36

AIC=911.36 AICc=911.41 BIC=913.65

Training set error measures:

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	-12.8441	108.4047	80.36832	-0.424341	1.929318	0.3737826	-0.001485368

The $ARIMA(0,1,1) \times (0,1,1)_{12}$ model shows a significant autoregressive structure with estimates of -1.10266 for the first component and -0.52969 for the second component. The moving average component, with an estimate of 0.67359, is marginally significant ($p < 0.1$), while the seasonal moving average component, estimated at -0.60481, indicates significant negative effects from seasonal shocks.

The estimated variance of the errors is $\sigma^2 = 13843$, which indicates that while the model captures a substantial portion of the data's volatility, there remains some unexplained variation. Compared to the $ARIMA(0,1,1) \times (0,1,1)_{12}$ model, which had an estimated variance of $\sigma^2 = 15206$, this lower variance suggests that the $ARIMA(2,1,1) \times (0,1,1)$ model may better account for fluctuations in income levels.

In terms of model performance, the AIC for the $ARIMA(2,1,1) \times (0,1,1)_{12}$ model is 911.36, compared to the AIC of 915.85 for the $ARIMA(0,1,1) \times (0,1,1)_{12}$ model. This indicates that the $ARIMA(2,1,1) \times (0,1,1)_{12}$ model provides a slightly better fit for the data, balancing model complexity with goodness-of-fit.

$ARIMA(2,1,2) \times (0,1,1)_{12}$ Model

	Estimate	Std.Error	z value	Pr(>|z|)	
ar1	-0.99260	0.25678	-3.8656	0.0001108	***
ar2	-0.33633	0.27912	-1.2050	0.2282191	
ma1	0.52419	0.28238	1.8563	0.0634067	.
ma2	- 0.27371	0.31333	-0.8736	0.3823584	
sma1	-0.56722	0.19366	-2.9290	0.0034009	**

sigma^2 estimated as 13758: log likelihood = -454.17, aic = 918.34

```
AIC=910.34     AICc=910.39    BIC=913.63

Training set error measures:

               ME          RMSE       MAE        MPE          MAPE      MASE       ACF1
Training set   -13.85782   108.0735   80.79539   -0.3723718   1.933472  0.3757689  0.0373446
```

The $ARIMA(0,1,1) \times (0,1,1)_{12}$ model shows a significant first autoregressive component (ar1 = -0.99260, $p < 0.001$) and a seasonal moving average component (sma1 = -0.56722, $p < 0.01$), indicating substantial dynamics within the data. However, the second autoregressive component (ar2) and the second moving average component (ma2) are not statistically significant.

In terms of model performance, the AIC for the $ARIMA(2,1,2) \times (0,1,1)_{12}$ model is 910.34, compared to the AIC of 911.36 for the $ARIMA(2,1,1) \times (0,1,1)_{12}$ model. This indicates that the $ARIMA(2,1,1) \times (0,1,1)_{12}$ model provides a slightly better fit for the data, balancing model complexity with goodness-of-fit.

Upon conducting our DetectAO and DetectIO tests to investigate any outliers using the TSA package, we found that our data didn't contain any additive nor innovative outliers. Further, we tried using the tsoutliers function and our output came back null.

In conclusion, we chose the $ARIMA(2,1,1) \times (0,1,1)_{12}$ model, because it has the least necessary parameters (principle of parsimony). It also has the lowest estimated variance of the errors and the lowest AIC and BIC out of the other models, which suggest a lower variability and a better fit overall. The RMSE and MAPE is also lower for this chosen model, which shows that it is the more suitable model.

Model Diagnosis

$\boldsymbol{ARIMA(0,1,1) \times (0,1,1)_{12}}$ Model

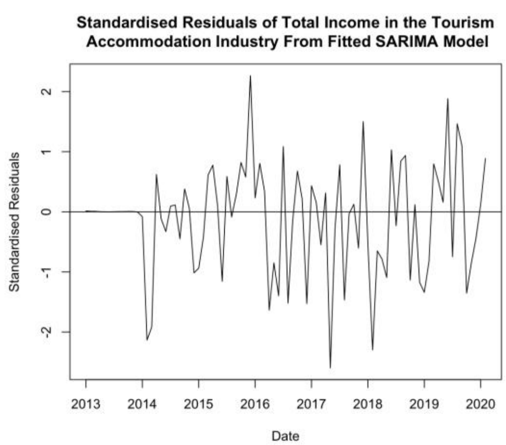
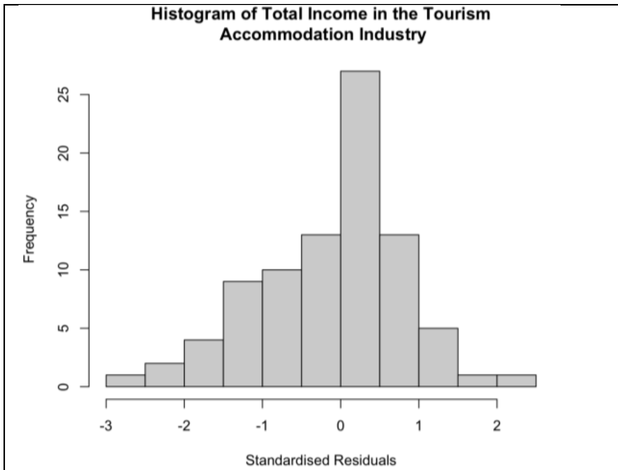
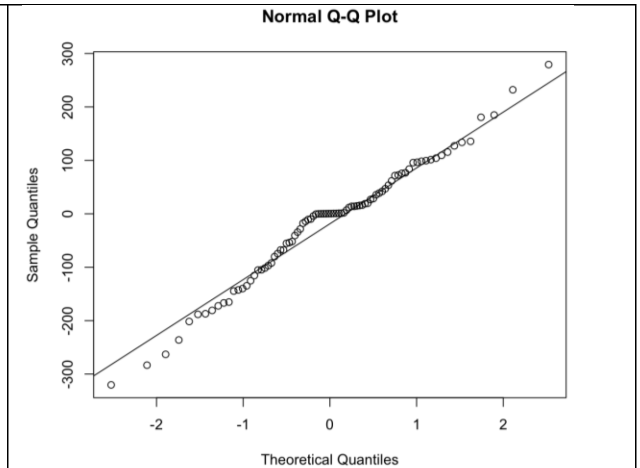


Figure 10: Standardized Residuals of $ARIMA(0,1,1) \times (0,1,1)_{12}$

The standardized residuals from the $ARIMA(0,1,1) \times (0,1,1)_{12}$ model appear to be randomly distributed, with no clear patterns or systematic deviations from the mean.





Shapiro-Wilk normality test	Jarque Bera Test
data: residuals(underfitting_model)	data: residuals(underfitting_model)
W = 0.98052, p-value = 0.2208	X-squared = 1.3108, df = 2, p-value = 0.5192

Figure 11: Normality Tests For Underfitted Model

The histogram of the residuals is centered around zero, but the distribution appears to be more peaked than a normal distribution, with some skewness on the left.

Many of the points of the QQ plot don't lie on the line, which means that we cannot be sure that the residuals have a normal distribution.

The p-value for the Shapiro-Wilk Normality Test is 0.2208 which is greater than 0.05, meaning we do not reject the null hypothesis that the residuals are normally distributed. Thus, the test suggests that the residuals are normally distributed.

The p-value for the Jarque-Bera Test is 0.5192 which is greater than 0.05, meaning we do not reject the null hypothesis that the residuals are normally distributed. Thus, the test suggests that the residuals are normally distributed.

Box-Ljung test	Box-Pierce test
data: residuals from underfitting_model	data: residuals from underfitting_model
X-squared = 19.124, df = 10, p-value = 0.03872	X-squared = 16.789, df = 10, p-value = 0.07916

Figure 12: Autocorrelation Tests For Underfitted Model

The Ljung-Box test suggests that there is significant autocorrelation in the residuals of the $ARIMA(0,1,1) \times (0,1,1)_{12}$ model. This violates the assumption of no autocorrelation in the residuals. The Box-Pierce test suggests otherwise, where it says that there is insignificant autocorrelation in the residuals of the model.

$ARIMA(\mathbf{2},\mathbf{1},\mathbf{1}) \times (\mathbf{0},\mathbf{1},\mathbf{1})_{\mathbf{12}}$ Model

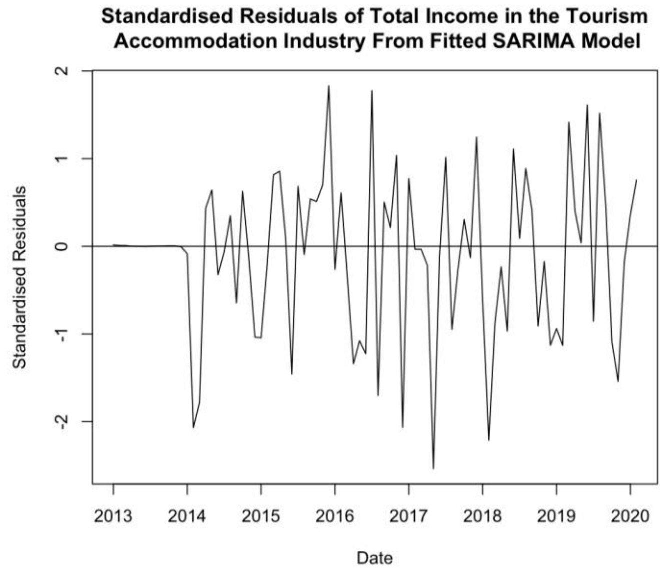
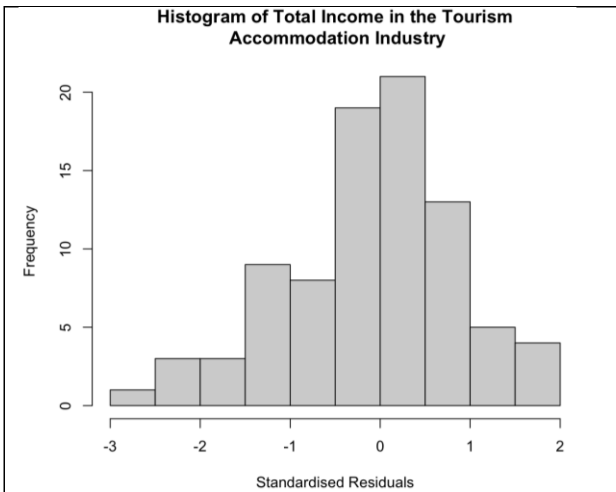
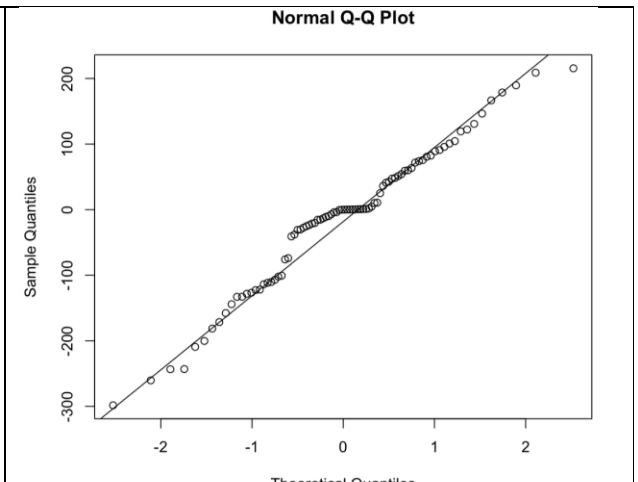


Figure 13: Standardized Residuals of $ARIMA(2,1,1) \times (0,1,1)_{12}$

The standardized residuals from the $ARIMA(2,1,1) \times (0,1,1)_{12}$ model appear to be randomly distributed, with no clear patterns or systematic deviations from the mean.





Shapiro-Wilk normality test	Jarque Bera Test
`data:  residuals(proposed_model_1)`	`data:  residuals(proposed_model_1)`
`W = 0.97479, p-value = 0.09058`	`X-squared = 1.7294, df = 2, p-value = 0.4212`

Figure 14: Normality Tests For Proposed Model

The histogram of the residuals appears to be approximately centered around 0, but there is some skewness to the left and possibly some kurtosis (heavier tails than expected in a normal distribution).

Some of the points of the QQ plot don't lie on the line, which means that we cannot be sure that the residuals have a normal distribution, thus we run additional tests.

The p-value for the Shapiro-Wilk Normality Test is 0.09058 which is greater than 0.05, meaning we do not reject the null hypothesis that the residuals are normally distributed. Thus, the test suggests that the residuals are normally distributed.

The p-value for the Jarque-Bera Test is 0.4212 which is greater than 0.05, meaning we do not reject the null hypothesis that the residuals are normally distributed. Thus, the test suggests that the residuals are normally distributed.

Box-Ljung test	**Box-Pierce test**
data: residuals from proposed_model_1	data: residuals from proposed_model_1
X-squared = 15.104, df = 12, p-value = 0.2358	X-squared = 13.041, df = 8, p-value = 0.1104

Figure 15: Autocorrelation Tests For Proposed Model

The Ljung-Box test suggests that there is no autocorrelation in the residuals of the $ARIMA(2,1,1) \times (0,1,1)_{12}$ model. The Box-Pierce test agrees with the Ljung-Box test that there is no autocorrelation. This suggests that the model has adequately captured the structure in the data and that the residuals behave as expected.

$ARIMA(2,1,2) \times (0,1,1)_{12}$ Model

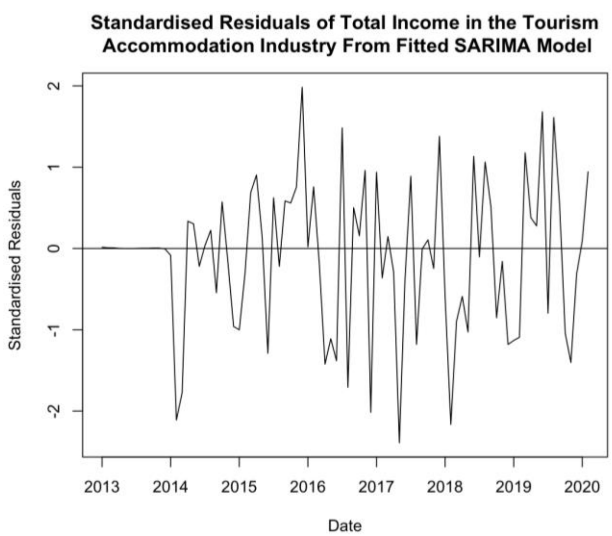
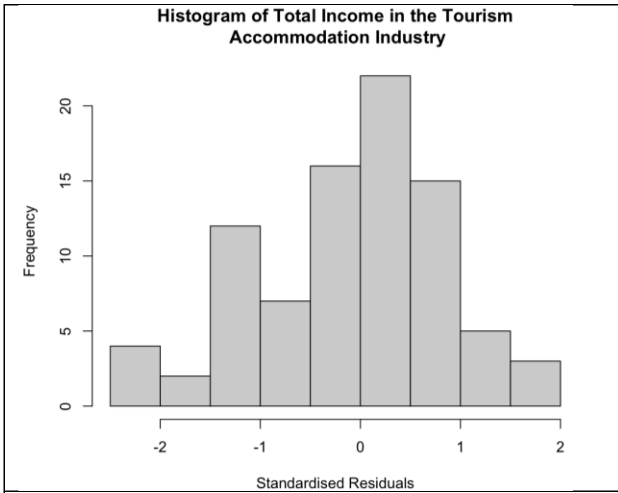
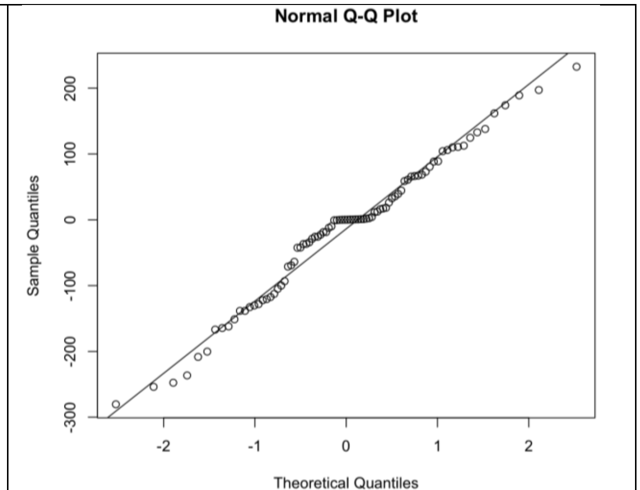


Figure 16: Standardized Residuals of $ARIMA(2,1,2) \times (0,1,1)_{12}$

The standardized residuals from the $ARIMA(2,1,2) \times (0,1,1)_{12}$ model appear to be randomly distributed, with no clear patterns or systematic deviations from the mean.





Shapiro-Wilk normality test	Jarque Bera Test
data: residuals(overfitting_model)	data: residuals(overfitting_model)
W = 0.98042, p-value = 0.2173	X-squared = 1.0654, df = 2, p-value = 0.587

Figure 17: Normality Tests For Overfitted Model

The histogram of the residuals is centered around 0, but there appears to be slight skewness to the left, and the distribution is not perfectly symmetric as expected in a normal distribution.

Many of the points of the QQ plot lie on the line, while a few don't, which means that the residuals might have a normal distribution, so we run additional tests to see.

The p-value *for the* Shapiro-Wilk Normality Test *is* 0.2173 *which is* greater than 0.05, meaning we do not reject the null hypothesis *that* the residuals are normally distributed. Thus, the test suggests that the residuals are normally distributed.

The p-value for the Jarque-Bera Test is 0.587 which is greater than 0.05, meaning we do not reject the null hypothesis that the residuals are normally distributed. Thus, the test suggests that the residuals are normally distributed.

Box-Ljung test	Box-Pierce test
data: residuals from overfitting_model	data: residuals from overfitting_model
X-squared = 16.78, df = 7, p-value = 0.01887	X-squared = 14.562, df = 7, p-value = 0.04205

Figure 18: Autocorrelation Tests For Overfitted Model

The Ljung-Box and Box-Pierce test suggests that there is significant autocorrelation in the residuals of the $ARIMA(2,1,2) \times (0,1,1)_{12}$ model. The Box-Pierce test also agrees that there is significant autocorrelation of the residuals in our model. This violates the assumption of no autocorrelation in our residuals.

Model Selection Outcome

Table	AIC	BIC	RMSE	MAPE	Shapiro-Wilk	Jarque Bera	Ljung-Box	Box-Pierce
$\boldsymbol{ARIMA(0,1,1) \times (0,1,1)_{12}}$	915.85	918.14	113.6155	2.005787	0.03872	0.07916	0.03872	0.07916
$\boldsymbol{ARIMA(2,1,1) \times (0,1,1)_{12}}$	911.36	913.65	108.4047	1.929318	0.09058	0.4212	0.2358	0.1104
$\boldsymbol{ARIMA(2,1,2) \times (0,1,1)_{12}}$	910.34	912.63	108.0735	1.933472	0.2173	0.587	0.01887	0.04205

We summarized the result above and looked at the values side by side. This is where we saw that the $ARIMA(2,1,1) \times (0,1,1)_{12}$ model was the preferred model for our specific data and that it would be a better fit overall. Our selected model performed optimally in the AIC, BIC, RMSE and MAPE categories, being the second preferred model. The model then performed better in the Normality of the residuals test, where we failed to reject the Null hypothesis that the residuals were Normally distributed. This complied with the assumption of Normal residuals. The model outperformed all the others in the no autocorrelation tests, where it was the only model that complied with the assumption of no autocorrelation. The $ARIMA(2,1,1) \times (0,1,1)_{12}$ model is the best fit for our data.

Data vs Fitted SARIMA

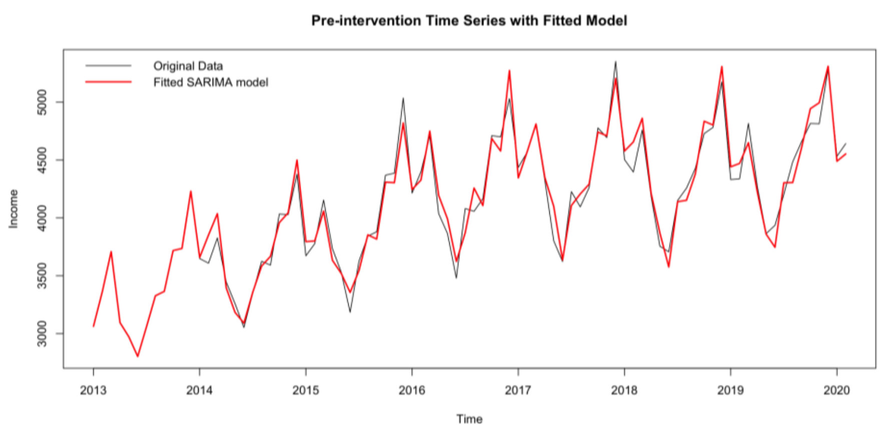


Figure 19: Comparison Between Fitted Model & Pre-intervention Data

The plot above compares the original pre-intervention income data (black) with the fitted SARIMA model (red) from 2013 to early 2020. The SARIMA model captures both the trend and seasonality in the data well, with minimal deviations between the actual and fitted values, indicating a good fit. This suggests the model is effective in approximating the patterns of the pre-intervention period.

Forecast

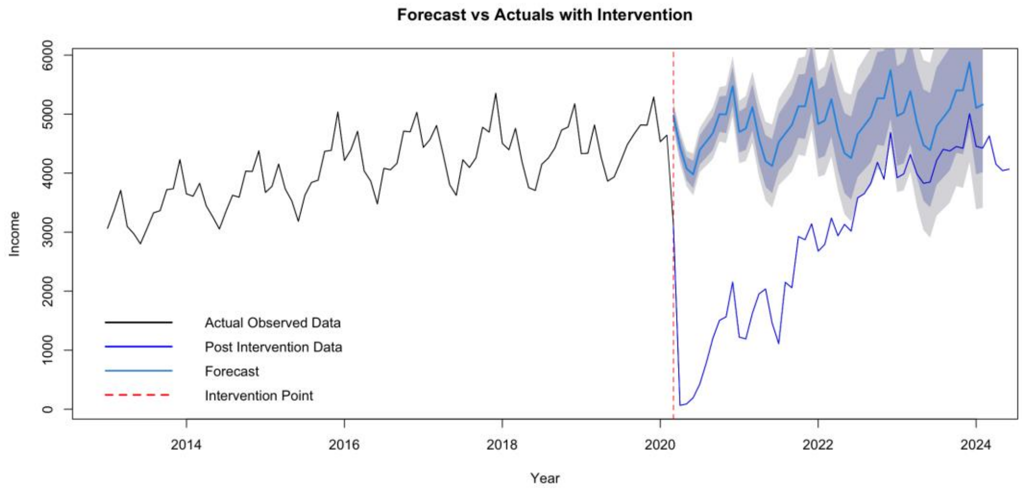


Figure 20: $ARIMA(2,1,1) \times (0,1,1)_{12}$ forecasts

Pre-intervention Period (January 2013 to February 2020):

The time series during the pre-intervention period, modeled by N_t, is displayed in black in the figure. These values represent the actual income data (in millions of Rands) for tourism accommodation businesses in South Africa.

Intervention Point (March 2020):

The intervention point, represented by the orange dashed line, corresponds to March 2020. This marks the beginning of the post-intervention period, during which a significant disruption (likely due to the pandemic) impacted the income from tourism. This is shown by the start of the red line in the figure.

Post-intervention Period (March 2020 to June 2024):

The red line, m_t, represents the observed values of income during the post-intervention period. There is a sharp decline immediately after March 2020, indicating the impact of the intervention. The income starts to recover, but it remains below the pre-intervention levels for a significant period.

Post-intervention vs. Forecast Discrepancy:

It's important to note that the post-intervention income levels (red line) do not align with the forecasted values (blue line) from the ARIMA model. The forecast, based on pre-intervention data, predicted a return to pre-intervention income levels over time. However, the actual post-intervention income remains significantly lower, suggesting a sustained decrease in revenue in the tourism accommodation industry following the intervention (likely due to the COVID-19 pandemic). This persistent drop could indicate a structural shift in the industry, with a potential permanent reduction in income post-intervention.

Forecasts (July 2024 to 2026):

The blue line represents the forecasts produced by the $ARIMA(2,1,1) \times (0,1,1)_{12}$ model, extending beyond the observed data from July 2024 to 2026. The model predicts that income would gradually return to pre-intervention levels, though the post-intervention data suggests this is unlikely without further adjustments to the model to account for the observed structural shift.

Forecast Confidence Intervals:

The light grey and dark grey shaded areas represent the 95% and 80% confidence intervals, respectively. These intervals capture the uncertainty of the forecasts. The prediction limits are relatively narrow in the short-term but widen considerably as the forecast moves further into the future, reflecting increased uncertainty.

Model Recovery:

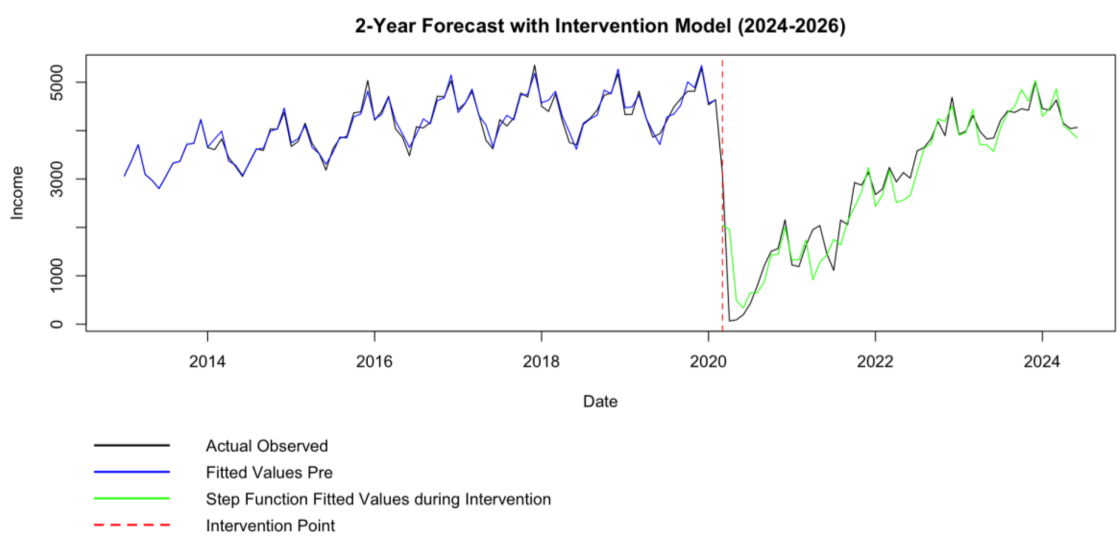
Although the ARIMA model forecasts a recovery, the red line does not fully converge with the blue line, signaling that the tourism accommodation income has not returned to pre-intervention levels, reinforcing the likelihood of a permanent shift.

Forecast Accuracy:

The model successfully captures the seasonal patterns in the data and produces forecasts that follow the expected periodicity of the income. However, the widening gap between the actual post-intervention data and the forecast suggests that the model's assumptions may need refinement to better represent the lasting effects of the intervention.

Intervention Analysis

Covariates Approach



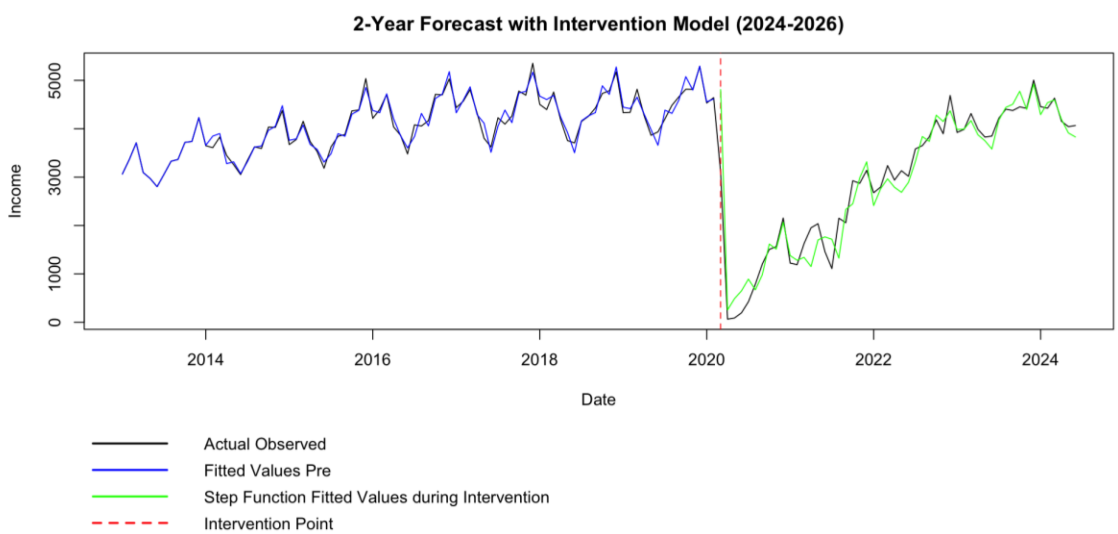
	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Test set	66.84374	424.5792	270.232	-63.15525	76.64324	-0.0107442	2.529732

The model with a step function that, where the covariates approach was applied, suffers from significant overestimation bias and exhibits high error rates across both RMSE and MAE. The particularly high MAPE indicates that its predictions are highly inaccurate in terms of percentage error.

While the residuals do not show strong autocorrelation (ACF1 close to zero), the model's overall predictive performance is weak, as reflected by the high Theil's U, which suggests that a simple naïve forecast would outperform it.

Due to its large errors and overestimation tendency, Model 1 is not suitable for accurate forecasting, and significant improvements are needed, either by adjusting the model parameters or using alternative models.

Ratio Approach

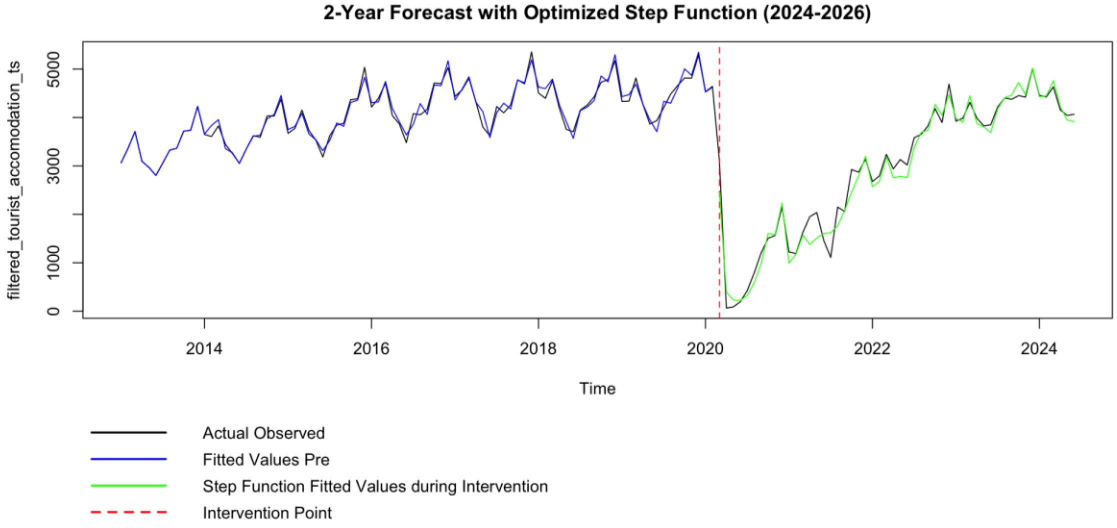


	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Test set	3.556598	366.5996	245.5215	-19.96702	30.02569	0.07521105	3.21024

The model with a step function that where the ratio approach was applied performs reasonably well, especially in terms of low bias (ME) and moderate error rates (MAPE, RMSE), but it is not the most accurate of the models compared. Its major drawback is its high Theil's U, indicating it is outperformed by a simple naïve forecast, which suggests that further refinement is needed.

Given its relatively balanced performance, Model 2 could be used in cases where low bias is prioritized, but caution should be exercised in relying on its predictions due to its relatively higher errors and poor comparison to a naïve benchmark.

Trial and Error Approach



	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Test set	62.67941	223.7336	160.3654	-9.805275	20.19676	0.02698585	0.9329116

The model with a step function that where the trial and error approach was applied performs demonstrates strong performance across most key metrics, especially in terms of error minimization and percentage accuracy. The slight overestimation bias (ME) is a minor issue, but it is offset by the model's accuracy and ability to forecast better than a naïve model. Model 3 is the preferred model due to its ability to deliver reliable and precise forecasts with minimal errors, making it suitable for forecasting applications.

Approach Selection Outcome

Model 3 consistently outperforms the other models across multiple key metrics, including RMSE, MAE, MAPE, and Theil's U, making it the best overall model for forecasting in this context.

Model 2 also performs reasonably well, particularly in terms of ME and MPE, indicating that its forecasts are less biased.

Model 1 exhibits the poorest performance, with high errors and a significant overestimation bias, as shown by its ME and MAPE.

As a result, the model with a step function that where the trial and error approach was applied is recommended for future forecasts due to its superior accuracy and lower forecast errors across various metrics.

Intervention Effects

After using the model with a step function that where the trial and error approach was applied the following intervention effects were observed.

Date	Covariate.vector	percentage.change	estimated.loss	original.data	predictions	difference
3/1/2020	1	-51	-2534.08836	3108	4974.577389	1866.577389
4/1/2020	1	-91	-4046.683603	64.7	4441.998201	4377.298201
5/1/2020	1	-94	-3845.898279	88.6	4076.776756	3988.176756
6/1/2020	1	-95	-3759.929485	194.3	3977.857568	3783.557568
7/1/2020	1	-93	-4068.787561	423.4	4393.780719	3970.380719
8/1/2020	1	-87	-3950.027165	787.8	4534.333441	3746.533441
9/1/2020	1	-79	-3696.333983	1204	4679.711186	3475.711186
10/1/2020	1	-68	-3395.455862	1505	4999.066761	3494.066761
11/1/2020	1	-68	-3417.437432	1563.8	4995.411331	3431.611331
12/1/2020	1	-59	-3241.700852	2154.6	5473.443219	3318.843219
1/1/2021	1	-79	-3705.995438	1221.9	4697.595055	3475.695055
2/1/2021	1	-75	-3580.336926	1190.5	4755.875505	3565.375505
3/1/2021	1	-69	-3543.70726	1627.4	5118.549305	3491.149305
4/1/2021	1	-70	-3201.271683	1952.5	4579.522272	2627.022272
5/1/2021	1	-64	-2695.67997	2037.3	4205.203921	2167.903921
6/1/2021	1	-61	-2511.802455	1459.8	4119.730861	2659.930861
7/1/2021	1	-64	-2905.356027	1109.5	4525.646112	3416.146112
8/1/2021	1	-62	-2906.810703	2150.9	4670.111804	2519.211804
9/1/2021	1	-57	-2734.016032	2057.3	4816.475989	2759.175989
10/1/2021	1	-52	-2676.043959	2925.5	5132.671193	2207.171193
11/1/2021	1	-46	-2345.247398	2871.4	5131.978056	2260.578056
12/1/2021	1	-43	-2412.299527	3140.9	5608.417575	2467.517575
1/1/2022	1	-47	-2263.134695	2677.2	4832.756144	2155.556144
2/1/2022	1	-45	-2214.954411	2793.6	4891.674156	2098.074156
3/1/2022	1	-39	-2069.657135	3238.7	5253.546034	2014.846034
4/1/2022	1	-42	-1957.591215	2939	4715.065534	1776.065534
5/1/2022	1	-36	-1556.917231	3132.5	4340.569317	1208.069317
6/1/2022	1	-35	-1493.97713	3017.2	4255.002888	1237.802888
7/1/2022	1	-28	-1285.937256	3582.5	4661.115307	1078.615307
8/1/2022	1	-23	-1115.727588	3650.8	4805.413047	1154.613047
9/1/2022	1	-24	-1209.787752	3831.6	4951.857988	1120.257988
10/1/2022	1	-19	-997.2351125	4184	5268.053109	1084.053109
11/1/2022	1	-23	-1220.259644	3896.3	5267.317288	1371.017288
12/1/2022	1	-22	-1270.495354	4688	5743.803916	1055.803916
1/1/2023	1	-20	-985.3105884	3923.3	4968.113149	1044.813149
2/1/2023	1	-22	-1120.682794	3987.1	5027.038556	1039.938556
3/1/2023	1	-17	-941.2740002	4314.7	5388.917819	1074.217819
4/1/2023	1	-20	-978.4489148	3986.6	4850.425259	863.8252587
5/1/2023	1	-15	-668.6442196	3827.7	4475.938428	648.2384275
6/1/2023	1	-16	-702.1147891	3848.7	4390.368037	541.6680373
7/1/2023	1	-13	-619.2071561	4217.1	4796.479853	579.3798526
8/1/2023	1	-11	-531.3659102	4403.6	4940.780357	537.1803565

9/1/2023	1	-12	-625.1143019	4376	5087.222569	711.2225693
10/1/2023	1	-13	-681.7133128	4451.4	5403.419235	952.0192347
11/1/2023	1	-17	-926.6908511	4422	5402.683156	980.6831559
12/1/2023	1	-15	-859.9817907	5005.2	5879.169251	873.9692506
1/1/2024	1	-13	-676.1719611	4455.1	5103.479208	648.3792085
2/1/2024	1	-14	-716.2321277	4424.6	5162.404098	737.8040982
3/1/2024	1	-14	-764.2883566	4632.4	5524.283547	891.8835471
4/1/2024	1	-15	-765.883566	4154.1	4985.791056	831.691056
5/1/2024	1	-14	-663.9618091	4042.9	4611.30405	568.4040504
6/1/2024	1	-13	-610.4187758	4066.7	4525.733816	459.033816

The highest percentage decrease is 95%, which signifies the biggest drop in our model and this occurred in June 2020. The lowest percentage decrease is 11% and this occurred in August 2023. The percentages are decreasing, which suggests that it could potentially recover even though it seems a bit stable at this point. Currently, as it stands, it has not recovered.

The covid effect gives an estimate of the impact that COVID-19 (the intervention) had on the tourist accommodation income. All the values are negative depicting that COVID-19 had a detrimental effect, reducing tourist accommodation income below the expected forecast. The total effect is -R103668.1 (in millions) indicating an overall loss in income due to the pandemic.