

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

Group Project Report

MH6241 - Time Series Analysis, Master of Analytics, SPMS, NTU

Supervisor: Prof. Pan Guangming

Group members (9):

Arya Gupta (G2404206G)

Budonkayala Ishaa (G2404133H)

Ding Juncheng (G2404083G)

Hteik Tin Min Paing (G2406170D)

Hussain Sakarwala (G2404071D)

Mai Xiyang (G2404171F)

Qi Boya (G2404369K)

Shivang Singh (G2404095K)

Ye Shenghua (G2406019J)

Introduction.....	4
Purpose of the Study.....	4
Objective.....	4
Data Description and Exploration.....	5
Source of Data.....	5
Description of the Dataset.....	5
Preliminary Data Exploration.....	5
Plotting the Time Series.....	5
Checking for Seasonality, Trends, and Anomalies.....	6
Initial Model Identification.....	6
Initial analysis:.....	6
Stationarity Checking (using ADF test):.....	7
Data Transformation.....	8
Differencing.....	8
Selection of Initial ARIMA Model Parameters (p, d, q).....	10
Model Estimation and Diagnostics.....	11
Classical Model.....	11
ARIMA(12,1,12).....	11
Residual Plot:.....	11
Ljung-Box Test:.....	11
Model Refinement.....	12
SARIMA(0,1,1).....	12
Residual Plot:.....	12
Ljung-Box Test:.....	12
Ljung-Box Test:.....	13
SARIMA(1,1,1).....	13
Residual Plot:.....	14
Ljung-Box Test:.....	14
Comparison and Summary of Models.....	14
Model Overview and Comparison.....	15
AIC/BIC Comparison.....	15
Residual Analysis and Model Diagnostics.....	15
Ljung-Box Test (Autocorrelation Check).....	15
Jarque-Bera Normality Test.....	16
Best Model Selection.....	16
Forecasting.....	17
Forecasting Visualization and Interpretation.....	17
ARIMA(12,1,12).....	17

Visualization:.....	17
Quantitative Evaluation:.....	18
SARIMA(0,1,1)(0,1,1)[12].....	19
Visualization:.....	19
Quantitative Evaluation:.....	20
SARIMA(1,1,1)(1,1,1)[12].....	20
Visualization:.....	20
Quantitative Evaluation:.....	21
Evaluation of Forecast Accuracy.....	21
Discussion and Conclusion.....	22
Interpretation of Results.....	22
Trends and Patterns Identified.....	22
Forecasting Implications.....	22
Comparison.....	23
Limitations of the Analysis.....	23
Suggestions for Future Work.....	24
Conclusion.....	24
Appendix.....	25
References.....	25

Introduction

Air travel has experienced significant growth over the decades, driven by economic expansion, globalization, and advancements in aviation technology. Understanding historical trends in airline passenger numbers provides valuable insights into seasonal patterns, long-term growth, and demand fluctuations. The dataset under consideration consists of monthly international airline passenger data from January 1949 to December 1960. This time series captures the evolution of air travel during the post-war period, reflecting economic recovery and increasing consumer confidence in aviation. The data is characterized by both seasonal variations and an upward trend, making it an ideal case for forecasting using time series modeling techniques.

Purpose of the Study

The primary goal of this study is to develop an appropriate Autoregressive Integrated Moving Average (ARIMA) model to analyze and forecast future airline passenger numbers. ARIMA is a widely used statistical method for time series forecasting, capable of capturing complex patterns such as trends, seasonality, and autocorrelation. By fitting an ARIMA model to this dataset, we aim to produce accurate predictions that can assist airlines, policymakers, and industry stakeholders in demand forecasting and strategic planning.

Objective

This study aims to demonstrate a comprehensive understanding of time series analysis and ARIMA modeling by:

- Exploring and visualizing the historical airline passenger data to identify trends, seasonality, and stationarity.
- Performing necessary preprocessing, including differencing and transformation, to make the series stationary.
- Determining optimal ARIMA parameters using model selection criteria such as AIC and BIC.
- Evaluating the performance of the fitted ARIMA model through diagnostic checks and residual analysis.

- Forecasting future passenger numbers and interpreting the results for practical decision-making.

Data Description and Exploration

Source of Data

The dataset is sourced from Box & Jenkins (1976), O.D. Anderson (1976), and O'Donovan (1983). It is widely referenced in time series analysis and is frequently used to study long-term trends, seasonality, and forecasting techniques in airline passenger traffic.

Description of the Dataset

The dataset records monthly international airline passenger numbers over 12 years, providing valuable insights into post-war economic recovery and aviation industry growth. It contains:

- Period: From January 1949 to December 1960
- Frequency: Monthly observations
- Unit of Measurement: Thousands of passengers
- Total Observations: 144

This dataset provides a time series representation of air travel demand, making it an ideal case study for trend analysis, seasonality detection, and forecasting models such as ARIMA and SARIMA.

Preliminary Data Exploration

To gain an initial understanding of the dataset, we performed an exploratory analysis, including time series visualization and statistical checks.

Plotting the Time Series

A time series plot of the passenger numbers was generated to observe patterns and variations over time. The visualization highlights:

- A strong upward trend indicates a steady increase in air travel demand over the years.
- Seasonal fluctuations, where passenger numbers peak around the same months annually, confirm a 12-month seasonality cycle.

- No sudden breaks or disruptions, suggesting a stable data collection process without missing values or irregularities.

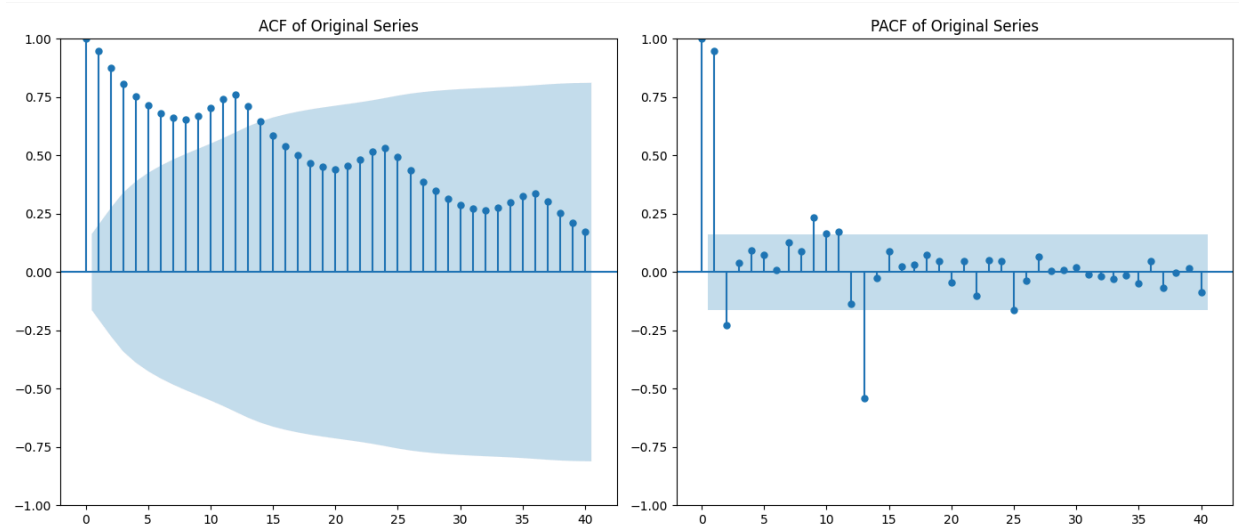
Checking for Seasonality, Trends, and Anomalies

- Trend Analysis:
 - The dataset is non-stationary due to the consistent upward growth in passenger numbers.
 - Increasing variance over time suggests a need for log transformation to stabilize fluctuations.
- Seasonality Detection:
 - A clear annual pattern is observed, with passenger traffic rising significantly in mid-year months, likely due to holidays and summer vacations.
 - Autocorrelation analysis confirms a 12-month periodicity, making seasonal differencing necessary for forecasting models.
- Anomaly Detection:
 - No extreme outliers or sharp declines were observed, indicating a smooth data distribution.
 - Minor variations in seasonal peak magnitudes, but no significant irregularities.

Initial Model Identification

Initial analysis:

The slow decay in ACF confirms the presence of a trend, indicating non-stationarity in the data. Noticeable peaks at lags 12, 24, and 36 suggest a seasonal pattern with a 12-month period. High correlations at multiple lags indicate the presence of strong seasonal autocorrelations. With PACF, we can find that there is a significant spike at lag 1, suggesting an AR(1) process, meaning the data follows an autoregressive model with one lag.



Stationarity Checking (using ADF test):

ADF Statistic: 0.8153688792060498

p-value: 0.991880243437641

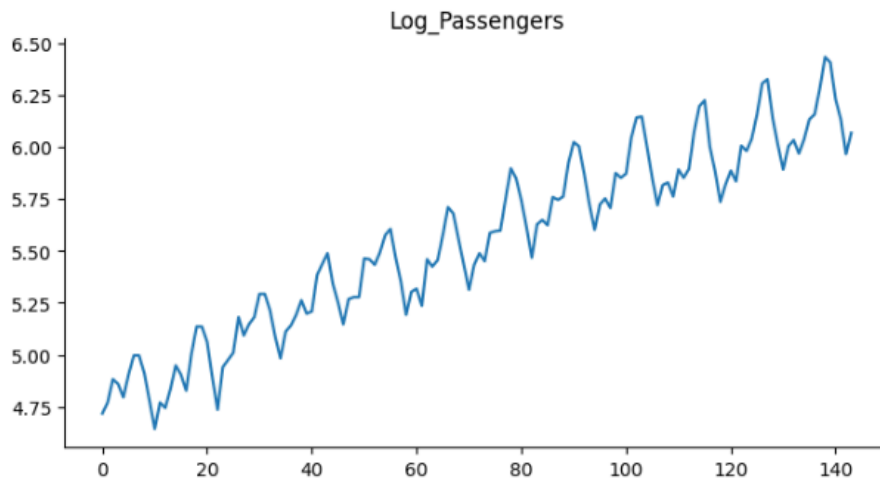
Critical Values: 1% -3.4816817173418295

Critical Values: 5% -2.8840418343195267

Critical Values: 10% -2.578770059171598

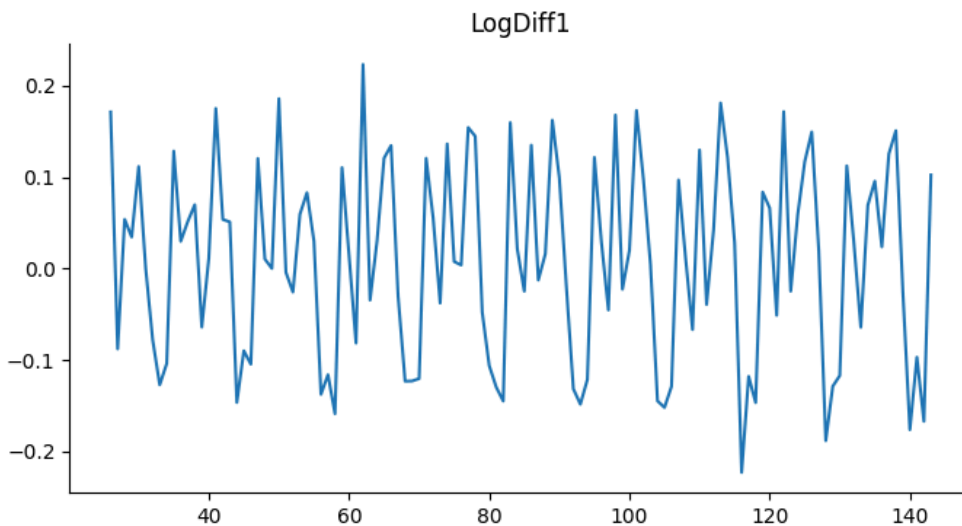
So usually, the p-value needs to be lower than 0.05 to say that it is stationary. Now it is obviously not stationary. The P-value should be below the significance level (0.05) for us to reject. Now it is above. That's why we need to do differencing. But before that let's transform the data to stabilize the variance.

Data Transformation

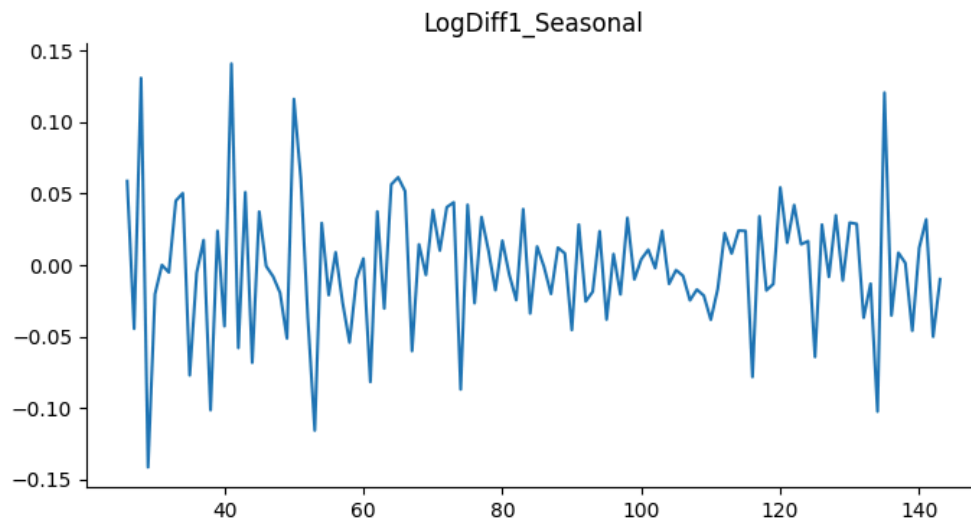


Before the transformation, the original series exhibited increasing variance the log transformation reduces this heteroskedasticity, making the variance more uniform over time. The upward trend is still present, indicating that further transformationsThe seasonal pattern remains, suggesting that seasonal differencing may be necessary for stationarity.

Differencing



We can still observe that after the differencing there is seasonality. So we did seasonal differencing. After the seasonal differencing, further eliminated seasonal patterns.



ADF Statistic: -4.4433249418311425

p-value: 0.00024859123113838495

Critical Values: 1% -3.4870216863700767

Critical Values: 5% -2.8863625166643136

Critical Values: 10% -2.580009026141913

Now as we can see the p-value is way less than 0.05 and the ADF stat is also lower than critical values. It means it is stationary.

Selection of Initial ARIMA Model Parameters (p, d, q)

SARIMAX Results						
Dep. Variable:	Log_Passengers	No. Observations:	118			
Model:	ARIMA(12, 1, 12)	Log Likelihood	214.963			
Date:	Sat, 15 Feb 2025	AIC	-379.926			
Time:	12:33:18	BIC	-310.871			
Sample:	0	HQIC	-351.890			
	- 118					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.0527	0.127	0.414	0.679	-0.197	0.302
ar.L2	-0.0885	0.147	-0.602	0.547	-0.377	0.200
ar.L3	0.0156	0.134	0.116	0.908	-0.248	0.279
ar.L4	-0.0475	0.142	-0.333	0.739	-0.327	0.232
ar.L5	-0.0001	0.130	-0.001	0.999	-0.254	0.254
ar.L6	-0.0543	0.135	-0.403	0.687	-0.318	0.210
ar.L7	0.0016	0.129	0.012	0.990	-0.252	0.255
ar.L8	-0.0990	0.130	-0.761	0.446	-0.354	0.156
ar.L9	-0.0062	0.128	-0.049	0.961	-0.256	0.244
ar.L10	-0.0234	0.135	-0.173	0.862	-0.288	0.241
ar.L11	-0.0141	0.120	-0.118	0.906	-0.250	0.221
ar.L12	0.8873	0.123	7.193	0.000	0.646	1.129
ma.L1	-0.4132	0.164	-2.517	0.012	-0.735	-0.092
ma.L2	0.2448	0.226	1.082	0.279	-0.198	0.688
ma.L3	-0.3523	0.200	-1.762	0.078	-0.744	0.040
ma.L4	0.0747	0.229	0.326	0.744	-0.374	0.524
ma.L5	0.1420	0.267	0.532	0.595	-0.381	0.665
ma.L6	0.0850	0.233	0.366	0.715	-0.371	0.541
ma.L7	-0.1202	0.229	-0.525	0.599	-0.569	0.328
ma.L8	0.1503	0.263	0.571	0.568	-0.366	0.666
ma.L9	0.0122	0.218	0.056	0.956	-0.415	0.439
ma.L10	-0.0323	0.235	-0.137	0.891	-0.492	0.428
ma.L11	0.0954	0.211	0.452	0.651	-0.318	0.509
ma.L12	-0.2403	0.177	-1.356	0.175	-0.588	0.107
sigma2	0.0012	0.000	5.383	0.000	0.001	0.002
Ljung-Box (L1) (Q):	0.08	Jarque-Bera (JB):	7.74			
Prob(Q):	0.78	Prob(JB):	0.02			
Heteroskedasticity (H):	0.53	Skew:	0.58			
Prob(H) (two-sided):	0.05	Kurtosis:	3.50			

ARIMA(0,1,1) AIC: -379.925562758008

ARIMA(0,1,1) BIC: -310.8712143880641

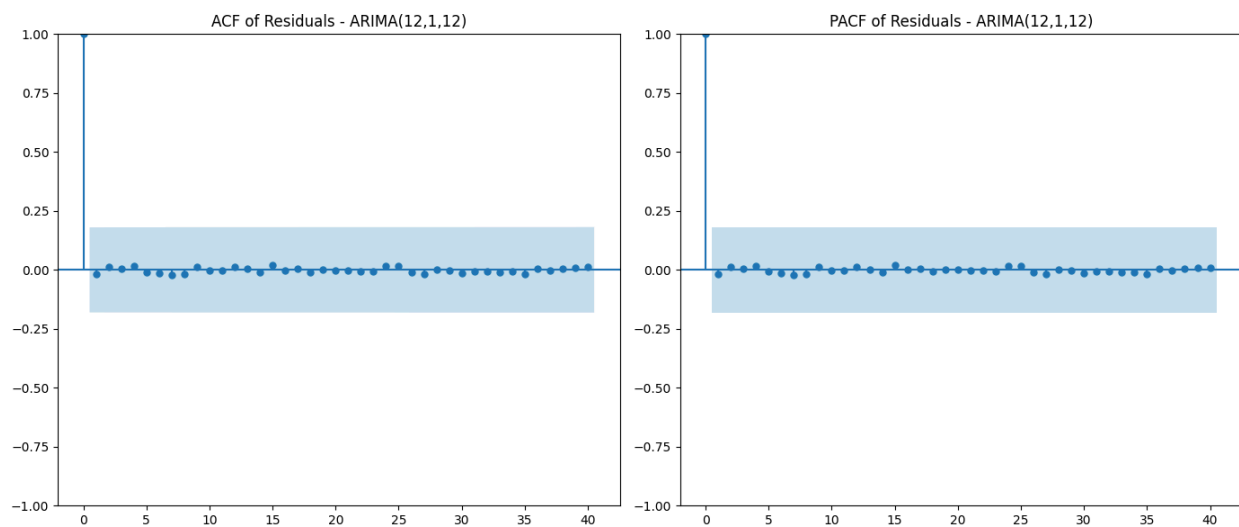
We can find that AIC and BIC are best when $p=12$, $d=1$, and $q=12$, so we choose (12,1,12) for the ARIMA model.

Model Estimation and Diagnostics

Classical Model

ARIMA(12,1,12)

Residual Plot:



We can observe that the residuals closely resemble white noise which indicates our model is a good fit.

Ljung-Box Test:

Lag12:

test statistics: 0.256963 p-value: 1.0

Lag24:

test statistics: 0.401256 p-value: 1.0

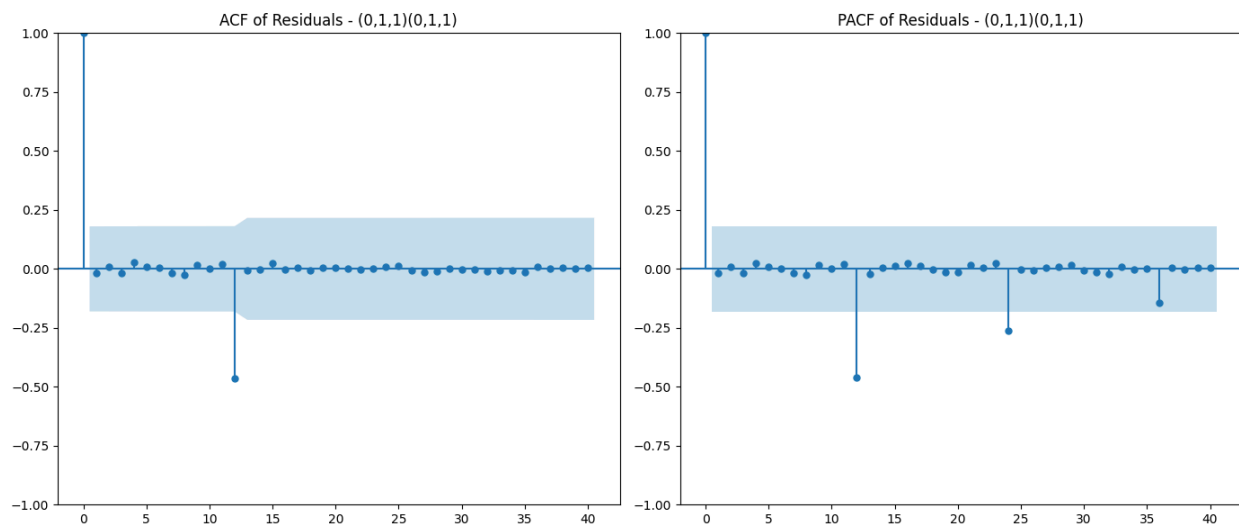
The Ljung-Box test results for the ARIMA(12,1,12) model show test statistics of 0.256963 and 0.401256 with p-values of 1.0 at lags 12 and 24, respectively. These high p-values indicate no significant autocorrelation in the residuals, suggesting that the residuals resemble white noise. This implies that the model fits the data well, and no further adjustments are needed based on this diagnostic test.

Model Refinement

SARIMA(0,1,1)

Next, we are fitting the model SARIMA(0,1,1).

Residual Plot:



Ljung-Box Test:

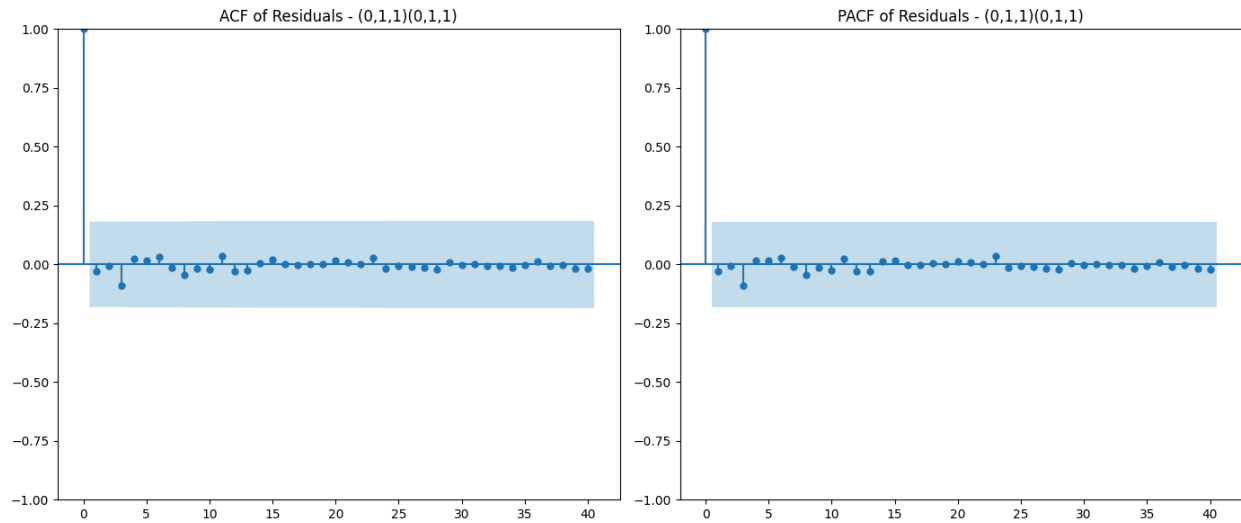
Lag12:

test statistics: 28.992960 p-value: 0.003949

Lag24:

test statistics: 29.106246 p-value: 0.216158

According to the plot, we can identify an outlier at x=24. This outlier contributes to the anomaly where the first point also has a high residual, leading to significant autocorrelation. Therefore, we will drop the first point and recalculate the ACF and PACF.



Ljung-Box Test:

Lag12:

test statistics: 1.946723 p-value: 0.999483

Lag24:

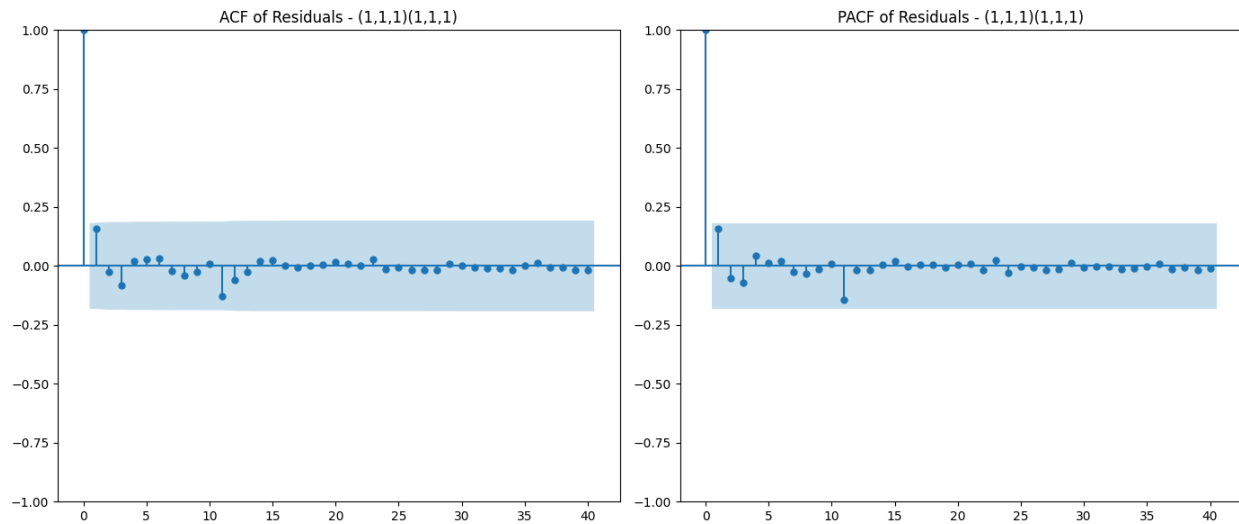
test statistics: 2.346358 p-value: 1.000000

The Ljung-Box test results for the ARIMA(0,1,1) model show test statistics of 1.946723 and 2.346358 with p-values of 1.0 at lags 12 and 24, respectively. These high p-values indicate no significant autocorrelation in the residuals, suggesting that the residuals resemble white noise.

SARIMA(1,1,1)

At last, we will be fitting model SARIMA(1,1,1).

Residual Plot:



Ljung-Box Test:

Lag12:

test statistics: 7.197085 p-value: 0.844319

Lag24:

test statistics: 7.624028 p-value: 0.999392

The SARIMA(1,1,1)(1,1,1) model performs well, as indicated by the ACF and PACF plots showing residuals within the confidence bands, suggesting white noise. The Ljung-Box test results with p-values of 0.8443 (lag 12) and 0.9994 (lag 24) confirm no significant autocorrelation. Overall, the model adequately captures the data's structure, and no further adjustments are needed.

Comparison and Summary of Models

To determine the best time series model, we compare three different models based on their performance metrics, diagnostic tests, and overall suitability.

Model Overview and Comparison

Model	AIC	BIC	Ljung-Box (L1) p-value	Jarque-Bera p-value	Key Characteristics
ARIMA (12,1,12)	-379.926	<u>-310.871</u>	0.78	<u>0.02</u>	High-order ARMA structure; lowest AIC but poor residual normality
SARIMA (0,1,1)(0,1,1) [12]	-358.052	<u>-350.519</u>	<u>0.80</u>	0.10	Simple MA(1) with seasonal MA(12); no significant autocorrelation in residuals
SARIMA (1,1,1)(1,1,1) [12]	-353.151	-340.596	0.56	0.07	Additional AR(1) and seasonal AR(12); slightly worse AIC/BIC than ARIMA(12,1,12)

AIC/BIC Comparison

AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) measure model fit, with lower values indicating a better fit:

- ARIMA(12,1,12) has the lowest AIC (-379.926), suggesting the best fit. However, its BIC (-310.871) is significantly higher, indicating possible overfitting.
- SARIMA(0,1,1)(0,1,1)[12] has the lowest BIC (-350.519) and a relatively low AIC, making it a balanced choice.
- SARIMA(1,1,1)(1,1,1)[12] has higher AIC and BIC than the other models, indicating that adding AR(1) and seasonal AR(12) does not provide significant improvement.

Residual Analysis and Model Diagnostics

Ljung-Box Test (Autocorrelation Check)

- ARIMA(12,1,12): p-value = 0.78, indicating no significant autocorrelation in residuals.
- SARIMA(0,1,1)(0,1,1)[12]: p-value = 0.80, slightly higher than ARIMA(12,1,12), confirming white noise residuals.
- SARIMA(1,1,1)(1,1,1)[12]: p-value = 0.56, relatively lower, suggesting minor residual autocorrelation.

Conclusion: SARIMA(0,1,1)(0,1,1)[12] performs best in terms of residual independence.

Jarque-Bera Normality Test

- ARIMA(12,1,12): p-value = 0.02, indicating that residuals significantly deviate from a normal distribution, which may affect prediction reliability.
- SARIMA(0,1,1)(0,1,1)[12]: p-value = 0.10, failing to reject the normality assumption, meaning residuals are approximately normally distributed.
- SARIMA(1,1,1)(1,1,1)[12]: p-value = 0.07, slightly lower than SARIMA(0,1,1)(0,1,1)[12] but still better than ARIMA(12,1,12).

Conclusion: SARIMA(0,1,1)(0,1,1)[12] has the most normally distributed residuals, making it suitable for statistical modeling and inference.

Best Model Selection

Considering AIC/BIC, the Ljung-Box test, and residual normality, we recommend SARIMA(0,1,1)(0,1,1)[12] as the final model for the following reasons:

1. Lower complexity, avoiding overfitting compared to ARIMA(12,1,12).
2. Balanced AIC and the best BIC, providing a good trade-off between fit and generalization.
3. No significant autocorrelation in residuals based on the Ljung-Box test.
4. Residuals are closest to a normal distribution, as confirmed by the Jarque-Bera test.

Thus, SARIMA(0,1,1)(0,1,1)[12] is the optimal choice among the three models.

Forecasting

Now let's look at forecasting international airline passenger numbers for the next 12 months.

Three time series models were employed:

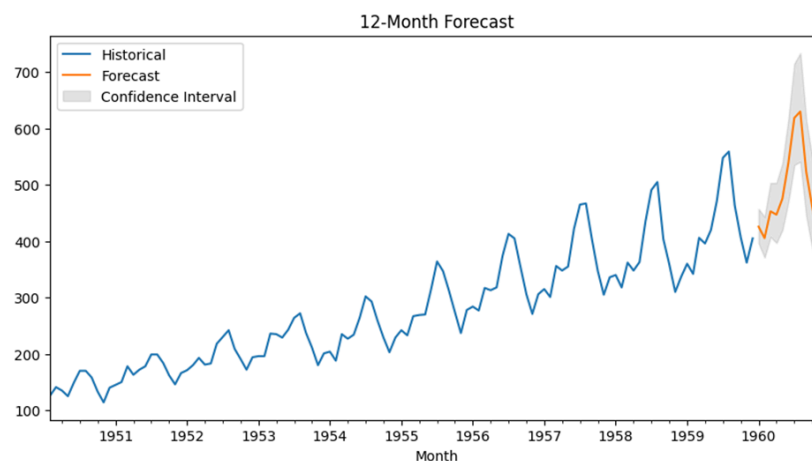
- $ARIMA(12,1,12)$
- Classic $SARIMA(0,1,1)(0,1,1)[12]$
- $SARIMA(1,1,1)(1,1,1)[12]$

The models were fitted on a log-transformed and appropriately differenced version of the historical data, ensuring stationarity and capturing both non-seasonal and seasonal patterns. The forecasts were generated in the log scale and then converted back to the original scale for interpretation. Confidence intervals were included in each forecast plot to convey the uncertainty of the predictions.

Forecasting Visualization and Interpretation

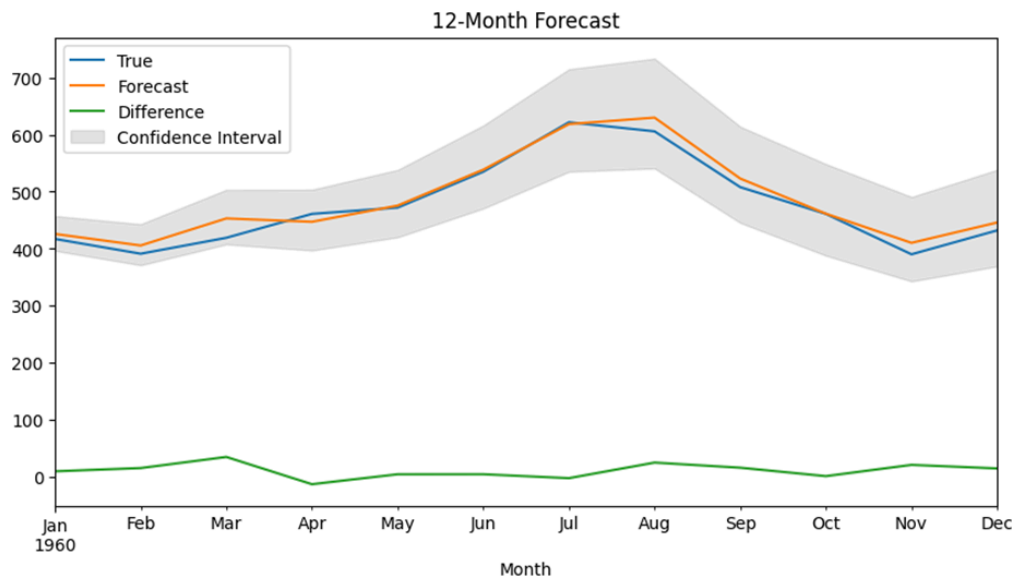
ARIMA(12,1,12)

Visualization:



The forecast plot for the $ARIMA(12,1,12)$ model overlays historical passenger data with the predicted values for the next 12 months. The confidence intervals—typically shown as a shaded band around the forecast—provide a visual measure of forecast uncertainty. The forecast line

follows the upward trend observed in the historical data, and the intervals appear moderately tight, suggesting reasonable certainty in the short-term predictions.



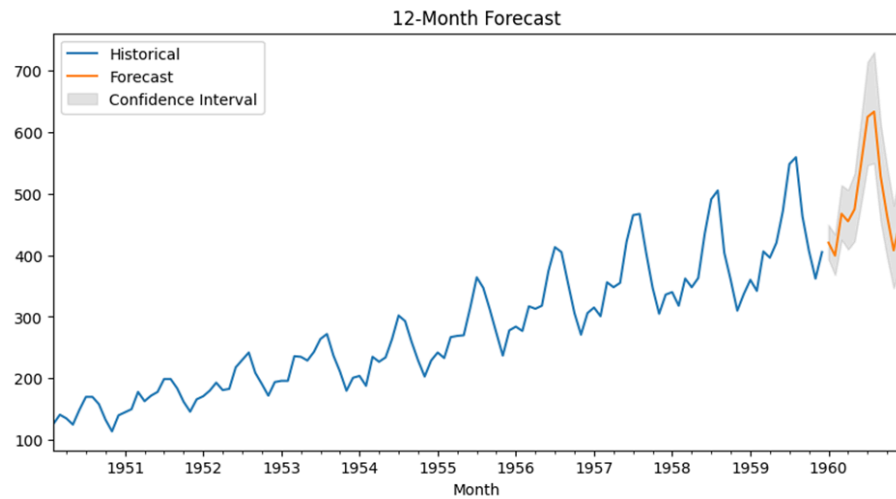
Quantitative Evaluation:

- RMSE: 16.02
- MAE: 12.93
- MAPE: 2.86%

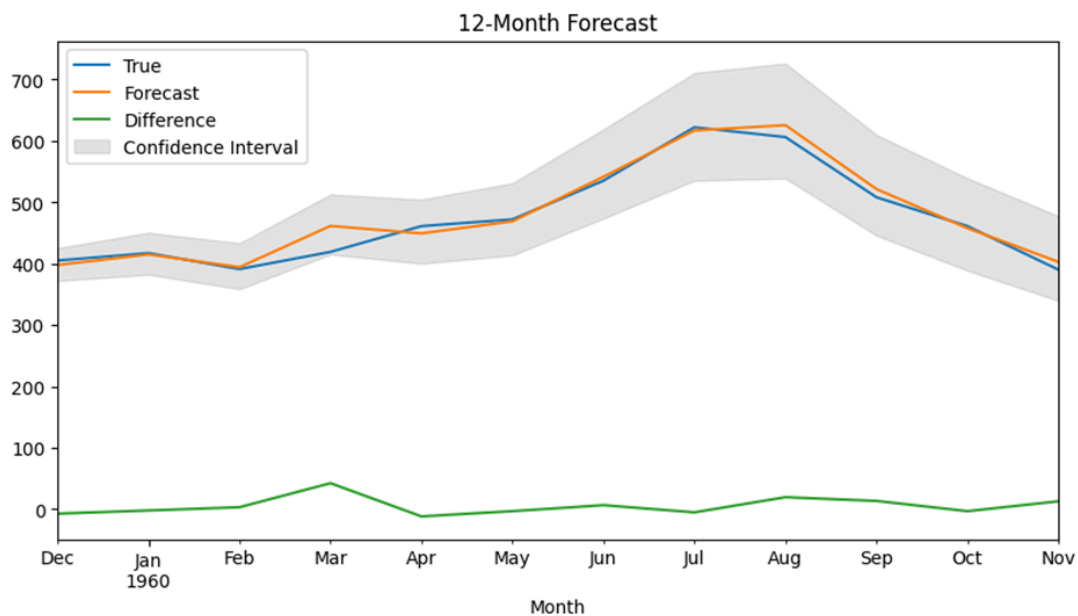
These metrics indicate that while the model captures the general trend, there is still room for improvement in terms of absolute and relative errors.

SARIMA(0,1,1)(0,1,1)[12]

Visualization:



The forecast produced by the Classic SARIMA model is visually compelling. Its forecast plot closely follows the historical trend, and the confidence intervals are slightly narrower compared to the ARIMA(12,1,12) model. This suggests that the model is more certain about its predictions, especially in handling the seasonal components of the data.



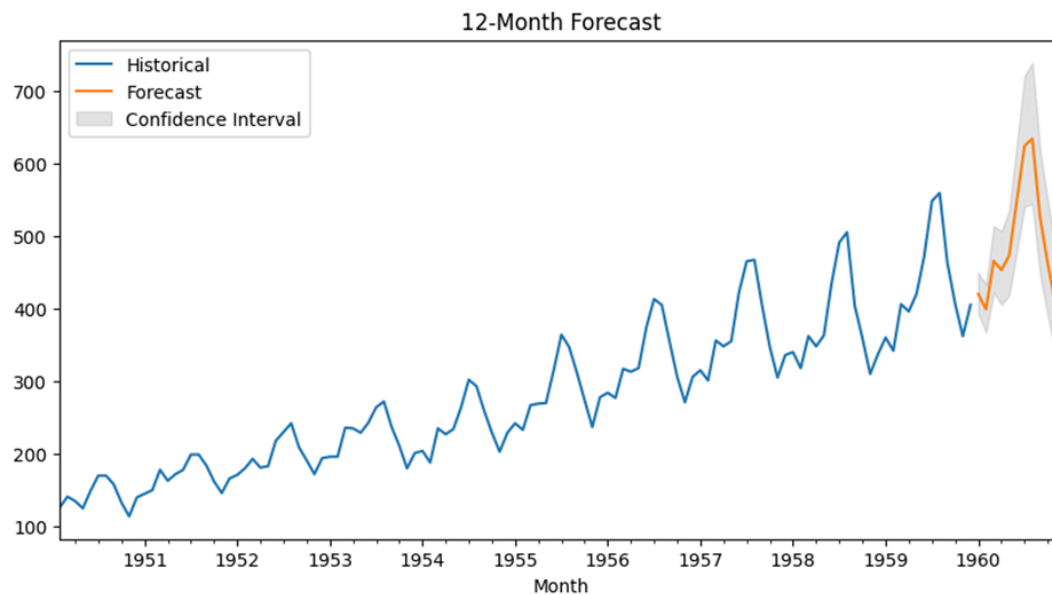
Quantitative Evaluation:

- RMSE: 15.30
- MAE: 10.89
- MAPE: 2.36%

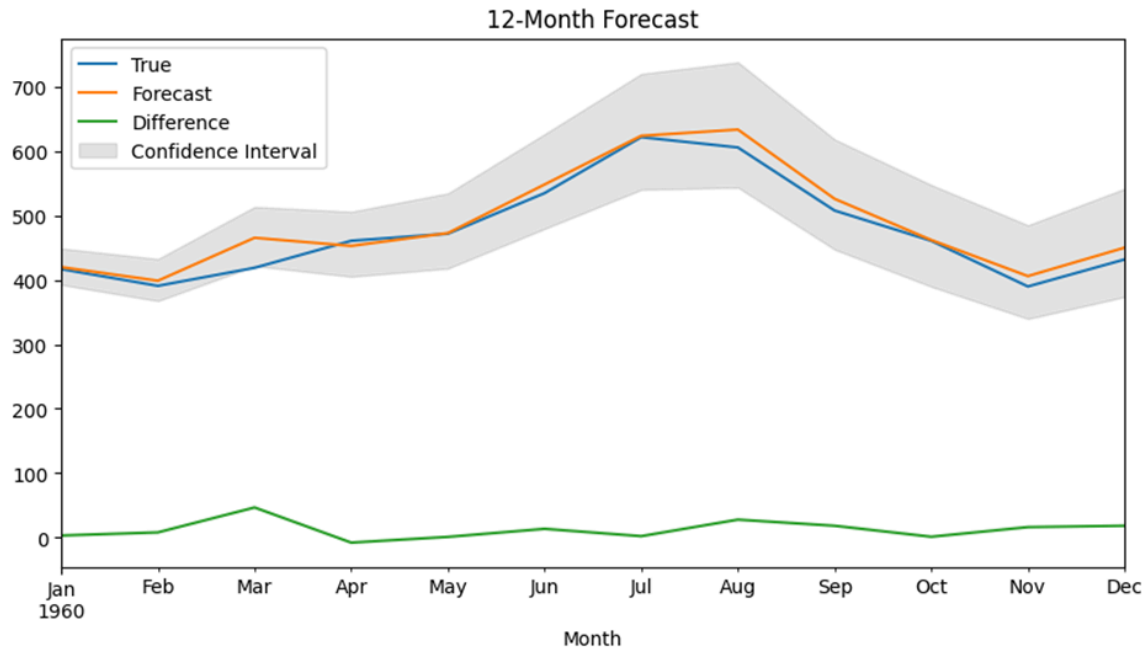
With the lowest RMSE, MAE, and MAPE among the three models, the SARIMA(0,1,1)(0,1,1)[12] model demonstrates superior performance. This indicates that it more effectively captures both the seasonal pattern and the overall trend in passenger numbers.

SARIMA(1,1,1)(1,1,1)[12]

Visualization:



The SARIMA(1,1,1)(1,1,1)[12] model, which incorporates both autoregressive and moving average components in its non-seasonal and seasonal parts, produces a forecast that follows the trend of the historical data but with wider confidence intervals. The increased width of these intervals suggests greater uncertainty in the model's forecasts, which could be due to the model's increased complexity and potential overfitting.



Quantitative Evaluation:

- RMSE: 18.70
- MAE: 13.64
- MAPE: 2.96%

These metrics are higher than those from the other two models, indicating that while this model is more flexible, it does not provide as accurate predictions as the classic SARIMA specification for this dataset.

Evaluation of Forecast Accuracy

The evaluation of forecast accuracy was performed using three standard metrics:

- **Root Mean Squared Error (RMSE):** This metric penalizes larger errors more than smaller ones, and it reflects the model's ability to predict accurately on the same scale as the original data. The Classic SARIMA model showed the lowest RMSE (15.30), suggesting it has the smallest average error magnitude.

- **Mean Absolute Error (MAE):** MAE gives a clear indication of the average magnitude of errors in the same units as the data. Again, the Classic SARIMA model recorded the lowest MAE (10.89).
- **Mean Absolute Percentage Error (MAPE):** Expressed as a percentage, MAPE provides an intuitive understanding of forecast accuracy relative to the actual values. The SARIMA(0,1,1)(0,1,1)[12] model achieved the lowest MAPE (2.36%), further confirming its relative superiority.

Overall, the evaluation metrics rank the models in terms of forecast accuracy as follows:

1. Classic SARIMA(0,1,1)(0,1,1)[12] – **Best performance with the lowest errors.**
2. ARIMA(12,1,12) – **Moderate performance with slightly higher errors.**
3. SARIMA(1,1,1)(1,1,1)[12] – Highest error metrics, suggesting increased complexity may not have translated into better forecasts for this dataset.

Discussion and Conclusion

Interpretation of Results

The analysis of the International Airline Passengers dataset involved testing different time series models to identify the most accurate forecasting approach. Three models - ARIMA (12,1,12), SARIMA (1,1,1)(1,1,1,12), and SARIMAX (0,1,1)(0,1,1,12) - were evaluated based on their predictive performance.

Trends and Patterns Identified

- The dataset exhibits a strong seasonal pattern with a yearly cycle, indicating that passenger numbers rise and fall predictably over time.
- A clear upward trend suggests increasing air travel demand over the years.
- The best-performing model effectively captured both trend and seasonality while minimizing forecast errors.

Forecasting Implications

The SARIMAX (0,1,1)x(0,1,1,12) model provides reliable forecasts with low error rates. This suggests that it can be used to project future passenger numbers with confidence, aiding airlines and policymakers in resource planning and decision-making.

Comparison

Model	RMSE	MAE	MAPE
ARIMA (12,1,12)	16.02	12.93	2.86%
SARIMA (1,1,1)(1,1,1,12)	18.70	13.64	2.96%
SARIMAX (0,1,1)x(0,1,1,12)	<u>15.30</u>	<u>10.89</u>	<u>2.36%</u>

- **ARIMA (12,1,12):** This model performed reasonably well, capturing short-term dependencies but struggled with seasonality.
- **SARIMA (1,1,1)(1,1,1,12):** The inclusion of seasonal components improved the understanding of yearly trends, but the model exhibited higher forecasting errors.
- **SARIMAX (0,1,1)x(0,1,1,12):** The best-performing model, with the lowest RMSE (15.30), MAE (10.89), and MAPE (2.36%). This model effectively captured both trend and seasonality.

Limitations of the Analysis

- **Exogenous Variables Not Considered:** Economic conditions, fuel prices, or global disruptions (e.g., pandemics) were not included.
- **Fixed Seasonal Effects:** The model assumes that seasonal patterns remain constant over time, which may not always hold.
- **Heteroskedasticity in Residuals:** Some variation in forecast errors suggests that external shocks may influence predictions.

Suggestions for Future Work

- **Hybrid Models:** Combining SARIMAX with machine learning approaches (e.g., LSTM, XGBoost) could further enhance accuracy.
- **Incorporating External Factors:** Adding economic indicators, airline pricing, or tourism trends could refine forecasts.
- **Re-evaluating Seasonality Over Time:** Checking for evolving seasonal trends could improve long-term predictions.

Conclusion

The SARIMAX (0,1,1)x(0,1,1,12) model emerged as the most effective forecasting approach for international airline passengers. With the lowest forecast errors, it provides a reliable tool for anticipating passenger demand. Future work could enhance accuracy by incorporating external variables and hybrid modeling techniques.

Appendix

References

The following references were used as key resources for time series modeling, particularly the ARIMA methodology:

1. **Box, G. E. P., & Jenkins, G. M. (1976).** *Time Series Analysis: Forecasting and Control*. Holden-Day.
 - This foundational book introduced the Box-Jenkins methodology for ARIMA modeling, widely used in time series forecasting.
2. **Anderson, O. D. (1976).** *Time Series Analysis and Forecasting: The Box-Jenkins Approach*. Butterworth-Heinemann.
 - This work provides a practical guide to applying the Box-Jenkins method, covering model identification, estimation, and diagnostic checking.
3. **O'Donovan, G. (1983).** *Short-Term Forecasting: An Introduction to the Box-Jenkins Approach*. Wiley.
 - A comprehensive introduction to ARIMA modeling for short-term forecasting, particularly useful for seasonal data such as airline passenger trends