Analysis, Modeling, and Forecasting of a Time Series using a SARIMA model

Authors:
Pedro Martins, Nina Lichtenberger and Mariana de Saavedra Lourenço (Group 06)
Faculty of Sciences of University of Porto

August 4, 2025

Abstract

This report presents the analysis of the monthly passenger traffic at the San Francisco International Airport from 1999 until 2020. The analysis comprises an exploratory data analysis, SARIMA modeling, and the employment of different forecasting approaches. Based on residual diagnostics, an analysis of the Autocorrelation Function, and Information Criteria, three suitable SARIMA models are identified. Subsequently, a comparative analysis of the forecasting performance of these models is conducted.

1 Introduction

This report presents an analysis of the monthly passenger traffic at the San Francisco Airport. Accurate passenger forecasting is essential for effective operational planning, resource allocation, and decision making in the aviation sector. The report begins with an exploratory analysis, which is followed by a structured approach to finding the most suitable SARIMA model and the comparison of the forecasting performance of different SARIMA models.

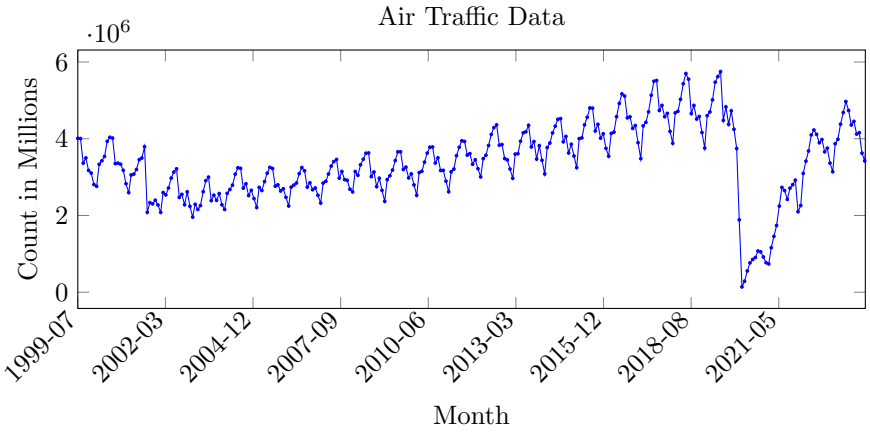
1.1 Data Description

The data source is the San Francisco International Airport Report on Monthly Passenger Traffic Statistics by Airline [2]. The data is self-reported by airlines and only available at a monthly level - we will be using the report from October 22, 2024. The data ranges from July 1999 until August 2024 and includes 302 observations. The original dataset consists of 12 columns with various information on airline, type of travel, etc. next to the passenger count. For our purpose, we will exclude all columns except for the passenger count. All passengers arriving at or departing from the San Francisco Airport will be counted and transit passengers will be counted twice, according to the convention of the San Francisco airport[2].

2 Exploratory Data Analysis

2.1 General Characteristics

Graph 1 displays the complete time series, ranging from 1999 to 2024. It demonstrates a general upward trend in passenger numbers over the years, indicating growth in air travel demand, beginning in 2004. Between 2001 and 2004, a sharp decline followed by a slight downward trend is evident. This may have been caused by 9/11. Moreover, seasonal patterns are evident, with regular peaks and troughs each year, reflecting fluctuations due to travel seasons. In 2020, there is a dramatic decline caused by the

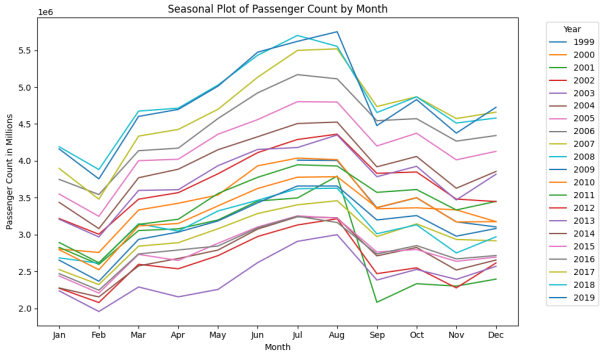


Graph 1: Time Plot of the Monthly Passenger Count from 1999 until 2024

COVID-19 pandemic, which severely impacted global travel. Following this dip, a gradual recovery is visible, with passenger counts rebounding toward pre-pandemic levels. Nonetheless, some fluctuations persist and the peaks before 2020 have not yet been reached again by August 2024.

Therefore, the plot highlights the necessity of segmenting the time series into pre- and post-pandemic. The effects of COVID were due to an extraordinary situation, and, considering that there were no comparable events in the past, we do not have enough information to accurately forecast the long-term impact and the rebound of the air travel industry. Therefore, this paper will be limited to modeling the time series before 2020 and forecasting how the passenger count would have developed on the condition that the COVID-19 pandemic had not happened. Analyzing the plot with the pre-pandemic observations, heteroskedasticity becomes evident. The variance of the plot is non-constant over time and depends on the mean, which is influenced by the trend. The upward trend does not only result in higher values, but also in higher variance. The data seems to be clean and without unusual values.

2.2 Seasonality and Trend



Graph 2: Seasonality Plot Showing the Seasonal Pattern for all Years

Graph 2 illustrates the passenger count over the years, broken down at a monthly level, revealing both the long-term trend and consistent seasonal patterns. Each line corresponds to a specific year, with passenger counts peaking in the summer months (June to August) and declining during the winter months (December to February). A noticeable dip occurs during the autumn months (September to November), likely reflecting reduced travel after the summer peak. Beginning in February, the number

of air passengers begins to increase again, until reaching the peak in summer. The plot also shows the steady increase in passenger traffic over the years.

3 Model Building

One of the most adequate models for time series with seasonality is the *Seasonal Autoregressive Moving Average* (SARIMA) model, which integrates the *Autoregressive* (AR) model and the *Moving Average* (MA) model with a seasonal component. The SARIMA model is defined by three main parameters: the order of the AR component (p), the degree of differencing (d), and the order of the MA component (q). In addition, it includes seasonal parameters: the order of the seasonal AR component (P), the degree of seasonal differencing (D), the order of the seasonal MA component (Q), and the length of the seasonal cycle (S), which represents the frequency of seasonality. The mathematical formulation of a SARIMA(p,d,q)$\times$(P, D, Q)$_S$ model can be described as follows:

$$\begin{aligned} \phi(B)\Phi(B^S)Z_t &= \theta(B)\Theta(B^S)\omega_t \\ Z_t &= (1-B^d)X_t \\ w_t &\sim \mathcal{N}(0,\sigma_w^2) \end{aligned}$$

where Z_t is the differenced time series up to order d from X_t, p is the order of the AR component, and q is the order of the MA component. w_t represents the residuals, which are assumed to follow a normal distribution. The polynomial operators are defined in such a way that the functions

$$\phi(z) = 1 - \sum_{n=1}^{p} \alpha_n z^n$$

and

$$\theta(z) = 1 - \sum_{n=1}^{q} \beta_n z^n$$

have roots with an absolute value exceeding 1 (outside the unit circle, $|z| > 1$) in order to attain stationarity and invertibility properties (this must also be true for the seasonal polynomial operators). If this is valid for $\phi(z)$ and $\Phi(z^S)$, then the time series process is stationary. Otherwise, it is non-stationary. If this is valid for $\theta(z)$, then it is invertible. If not, it is non-invertible. Without loss of generality, the characteristic polynomials for the seasonal components are expressed as:

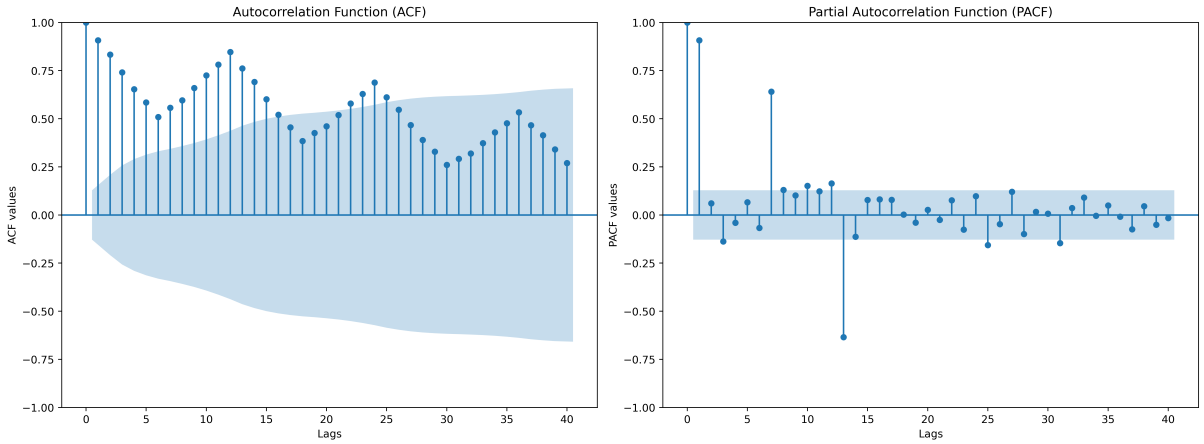
$$\Phi(z^S) = 1 - \sum_{n=1}^{P} \alpha_{n_S} z^{nS}$$

and

$$\Theta(z^S) = 1 - \sum_{n=1}^{Q} \beta_{n_S} z^{nS}$$

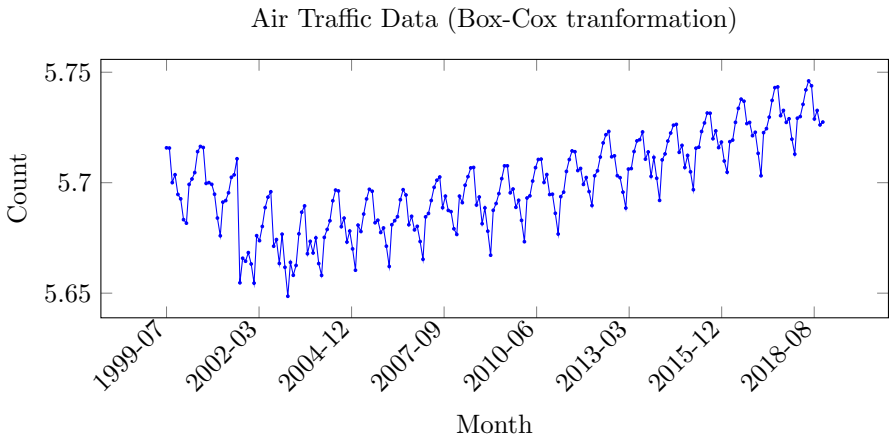
3.1 Data Preparation

In order to be able to evaluate the accuracy of our model and forecast, we first split the dataset into training and test data. The test dataset comprises 12 observations (one year), meaning we are using the data from 1999 until 2018 to predict the values for the year 2019. To determine which model is appropriate, we first plot the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF).



Graph 3: ACF and PACF before any transformation with the 95% confidence interval (Bartlett test)

Analyzing the ACF and PACF, we can see the seasonal pattern in the ACF, given that the correlation increases and decreases in yearly cycles. The ACF slowly decays to 0, while the PACF shows strongly significant values at lag 1, 6, and 12 and still significant, but weaker, correlations for lags 24 and 30. The significant PACF at lags 6, 12, 24, and 30 point to the seasonality of the data. The fact that the ACF slowly decays to 0 and the significant PACF values indicate that an AR component should be included in our model. We will check that after transforming our data.



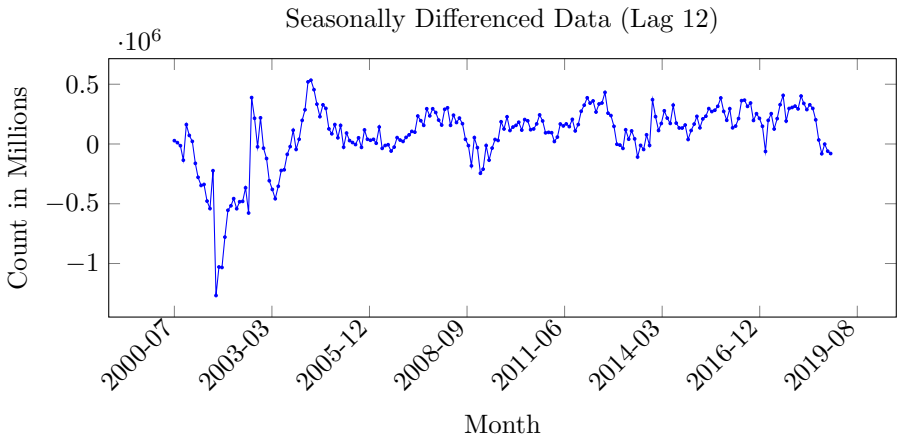
Graph 4: Monthly Passenger Count after Box-Cox transformation ($\lambda = -0.159$)

In order to model our data, the data has to be stationary, meaning it needs to have a constant variance as well as mean and that the autocovariance must only depend on lag h. As we could see in the EDA, our time series has neither a constant mean nor constant variance. Therefore, we will use a Box-Cox transformation to stabilize the variance with a lambda of -0.16. Mathematically, the Box-Cox tranformation can be defined as [1]:

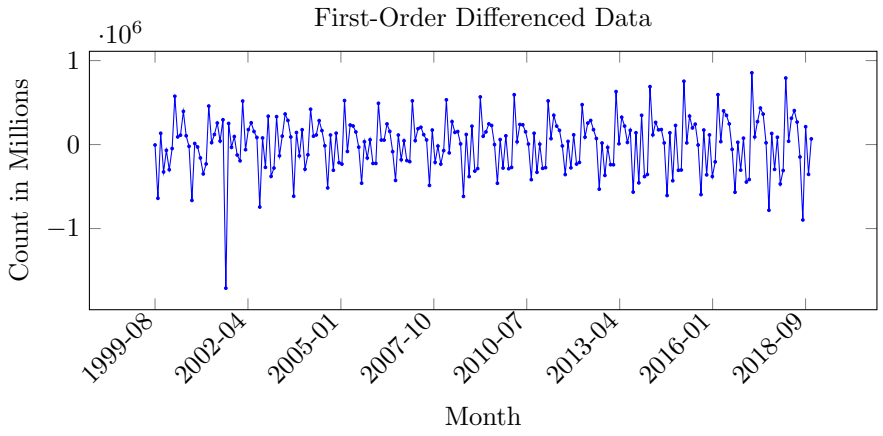
$$y_i^{(\lambda)} = \begin{cases} \frac{y_i^{\lambda}-1}{\lambda} & \text{if } \lambda \neq 0 \\ \ln y_i & \text{if } \lambda = 0 \end{cases}$$

Considering the strong seasonality of our data, we have to apply a seasonal difference filter to achieve a constant mean. According to the results of the Augmented Dickey-Fuller (ADF) test, the seasonally

differenced data is still not stationary. Therefore, we will apply the difference operator to the seasonally differenced data with a difference factor of 1. This leads to a stationary time series, as confirmed by the ADF test.



Graph 5: Seasonally Differenced Data with lag $h = 12$ (yearly cycles)

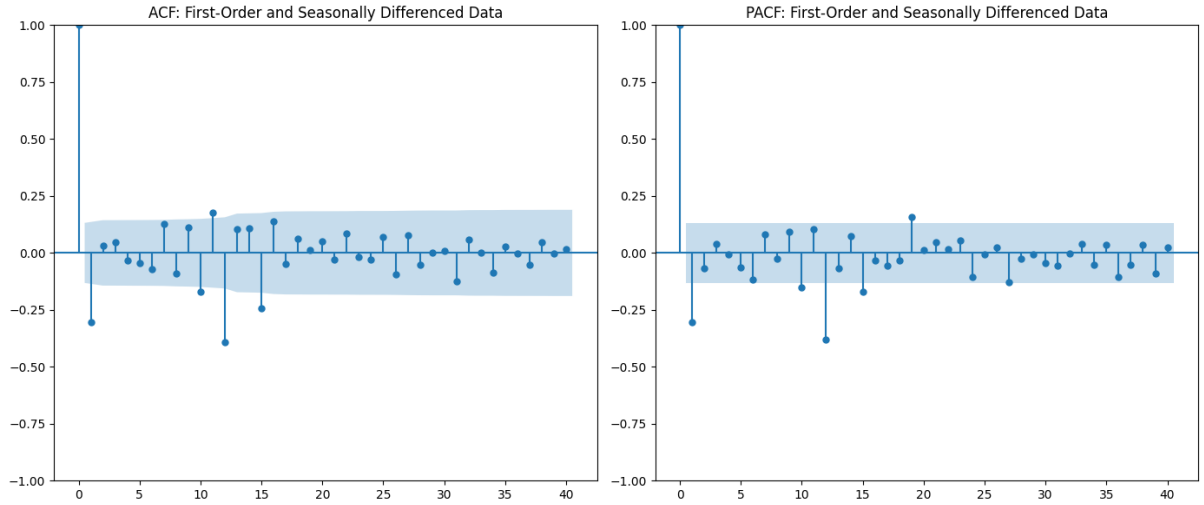


Graph 6: First-Order Differenced Data (Before Box-Cox transform) - Note that the mean appears relatively constant over the months

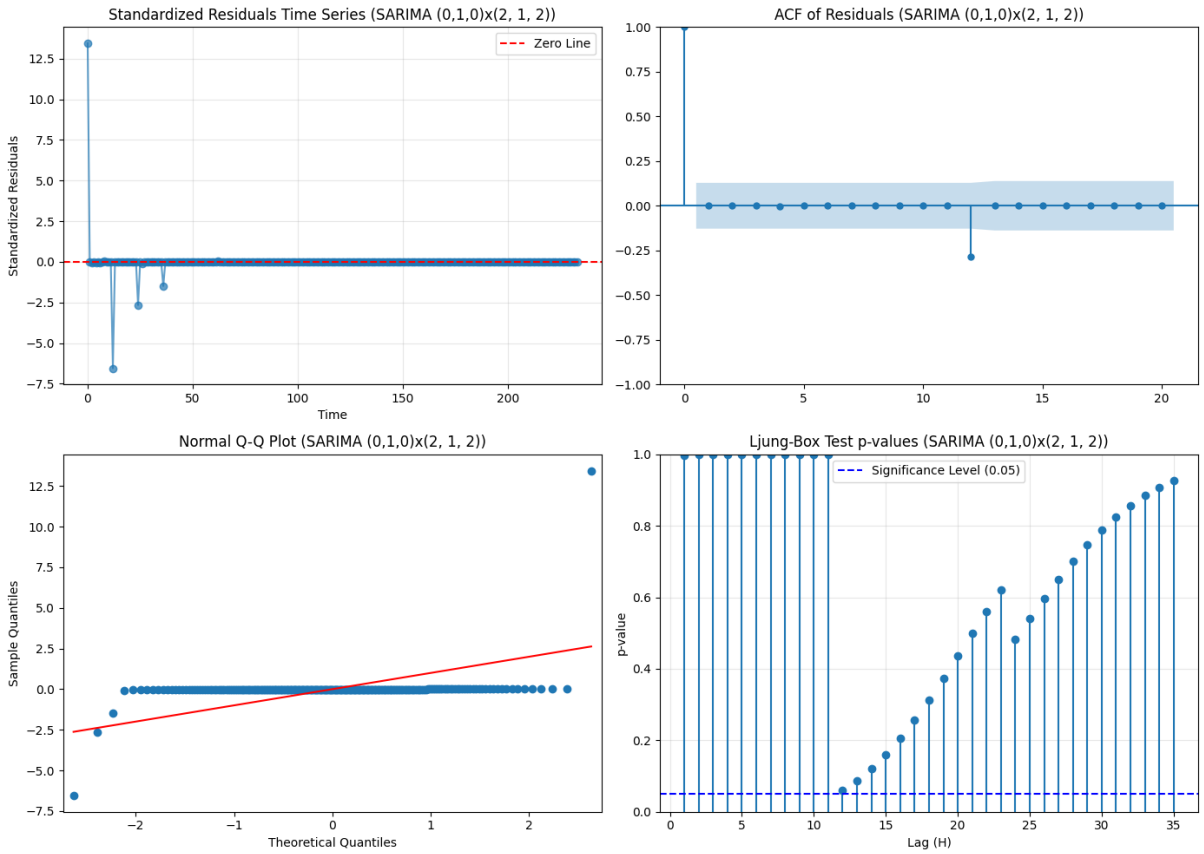
3.2 Approach for Finding the Best Model

Looking at the ACF and PACF of the differenced data, we can see a significant ACF for lag 1 and some lags close to 12. The same holds true for the PACF, which is significant for lag 1, lags close to 12, and very weakly significant for two higher lags. The ACF and PACF imply $q = 1$ and $p = 1$ or 0. Regarding the seasonal orders, it is probable that P as well as Q are necessary. As a starting point, we try SARIMA(0,1,0)$\times(2, 1, 2)_{12}$.

The ACF in graph 8 shows that the residuals of the SARIMA(0,1,0)$\times$(2,1,2) still have a significant correlation at lag 12. Moreover, the summary in table 1 shows that the seasonal MA components of our model are not significant. Therefore, we should increase the seasonal AR coefficients in our model and reduce the MA order. According to the Ljung-Box Test displayed in graph 8, the residuals are, however, uncorrelated, speaking in favor of the model. The Q-Q plot shows that the residuals are not normally distributed. This is likely to be caused by the trend break in our data in the early 2000s.



Graph 7: ACF and PACF for the Seasonally and First-Order Differenced Data



Graph 8: Residual Diagnostics of SARIMA$(0,1,0)\times(2,1,2)$

Based on the findings, we adjusted the P and Q - we also tried different values for p and q. For the sake of brevity, we will not go into detail about the diagnostics of all models we have tried. We tried values from 0 to 2 for p, 0 to 3 for q, 1 to 11 for P, and 1 to 4 for Q. Our approach was to first check if all coefficients were significant. If they were not, we fitted another model without the insignificant

Table 1: Summary of Model Parameters for SARIMA(0,1,0)×(2,1,2)

Parameter	**Coef.**	**Std Err**	z	$P > \|z\|$
ar.S.L12	−0.54	0.02	−22.37	0.00
ar.S.L24	−0.23	0.02	−9.59	0.00
ma.S.L12	0.02	0.07	0.28	0.78
ma.S.L24	−0.12	0.09	−1.28	0.20
AIC		-1689.649		
BIC		-1673.258		

coefficients. Then, we analyzed if the residuals still showed a significant ACF and how significant it was. Lastly, we analyzed if the Ljung-Box Test indicated a correlation of the residuals. The Ljung-Box Test looks at the accumulation of autocorrelation rather than individual correlations seen in the ACF. We also analyzed the Q-Q Plot, but the plot behaved similarly for all models, wherefore this was not a decision criterion. The goal was to achieve ACF values as low as possible (though we were unable to completely remove the autocorrelation at lag 12) and uncorrelated residual values for all lags according to the Ljung-Box Test. If a model fulfilled all of these criteria, we included it in our shortlist for appropriate models. Among these appropriate models, we will use the forecast performance and the Alkaike (AIC) as well as the Bayesian Information Criterion (BIC) to choose the best model.

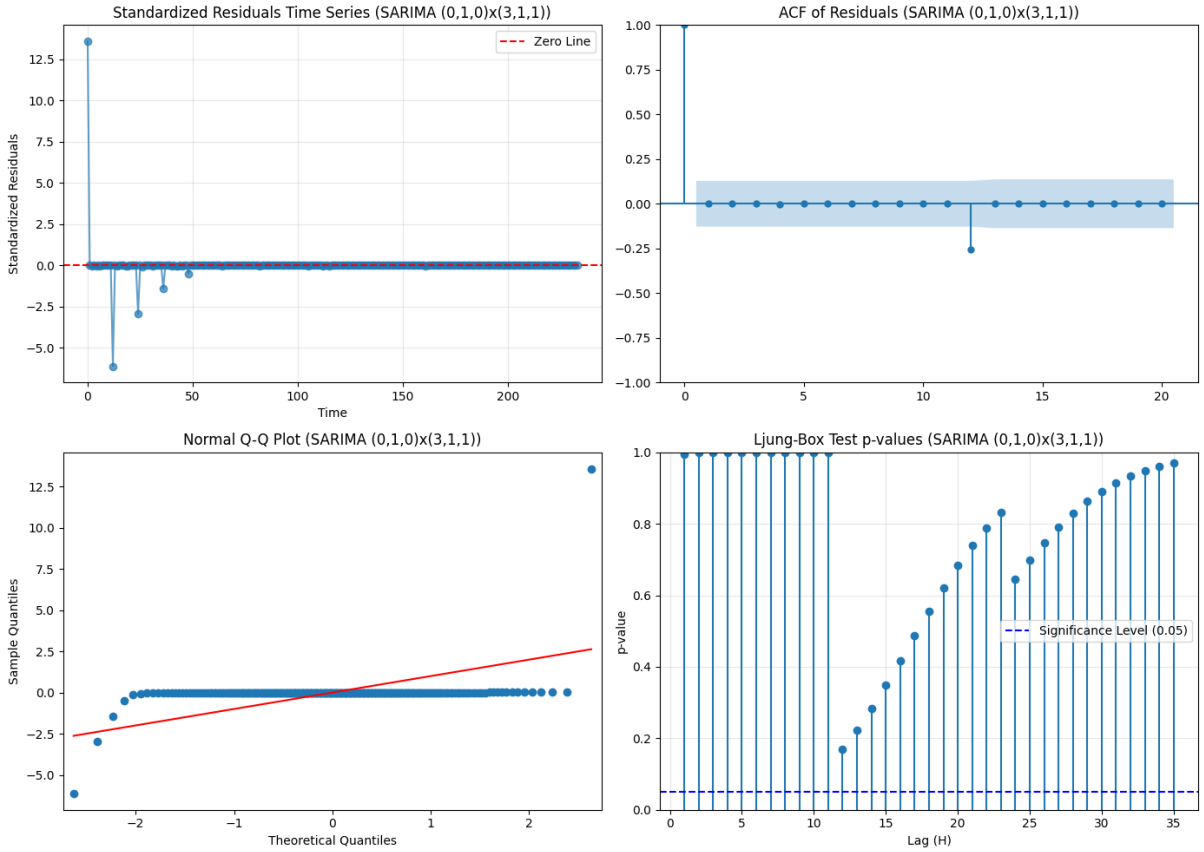
Lastly, we built a function similar to the auto.arima function in R to be able to compare how our manually fit model compares to an automatically found model. This function does not consider any residual diagnostics and simply finds the optimal model based on AIC.

3.3 Best Models based on Residual and Coefficient Analysis: SARIMA(0,1,0)×(3,1,1)$_{12}$ and SARIMA(0,1,1)×(3,1,1)$_{12}$

Table 2: Statistical Summary of Model Parameters for SARIMA(0,1,0)×(3,1,1)

Parameter	**Coef.**	**Std Err**	z	$P > \|z\|$
ar.S.L12	−0.65	0.09	−7.40	0.00
ar.S.L24	−0.35	0.06	−6.27	0.00
ma.S.L12	−0.19	0.02	−8.05	0.00
ma.S.L24	0.30	0.12	2.57	0.01
AIC		-1661.351		
BIC		-1645.250		

Based on the manual inspection of residual diagnostics and the significance of coefficients, SARIMA(0,1,0)×$(3,1,1)_{12}$ and SARIMA(0,1,1)×$(3,1,1)_{12}$ were found to be the most suitable models. As can be seen in graph 9 (and 10) and table 2 (and 3), all of the coefficients are significant, and the Ljung-Box Test shows that the residuals are uncorrelated for all lags. The Ljung-Box Test p-values are well beyond the significance level of 0.05. Regarding the ACF of the residuals, the only significant correlation is at lag 12. In comparison to the other models, the correlation at lag 12 is, however, relatively small. Based on the diagnostic plots, the fit of the two models is approximately the same. The



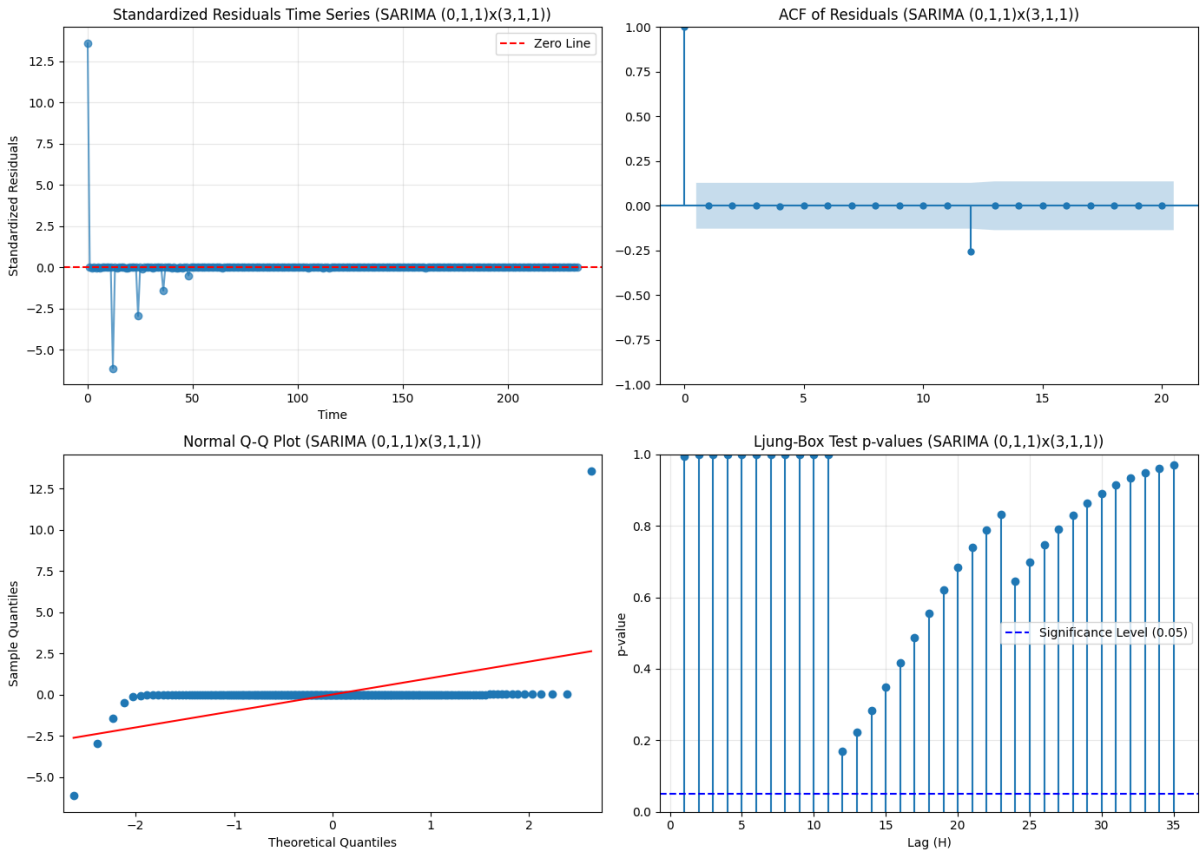
Graph 9: Residual Diagnostics of SARIMA(0,1,0)×(3,1,1)

Table 3: Statistical Summary of Model Parameters for SARIMA(0,1,1)×(3,1,1)

Parameter	**Coef.**	**Std Err**	***z***	$P > \|z\|$
ma.L1	−0.32	0.08	−4.01	0.00
ar.S.L12	−0.65	0.07	−9.05	0.00
ar.S.L24	−0.35	0.05	−7.43	0.00
ar.S.L36	−0.19	0.02	−8.87	0.00
ma.S.L12	0.30	0.10	2.97	0.00
AIC		-1683.090		
BIC		-1663.768		

SARIMA(0,1,1)×$(3,1,1)_{12}$ has a slightly lower AIC and BIC score with -1683 and -1664, respectively. It does, however, also have a higher number of parameters, making it more prone to overfitting. The SARIMA(0,1,0)×$(3,1,1)_{12}$ has an AIC of -1661 and BIC of -1645 and also has, although it has one parameter less, a high number of coefficients. The forecast on the test dataset will show if the models overfit.

The Q-Q plot of the residuals compares the quantiles of the residuals to the theoretical quantiles of a normal distribution (indicated by the red line). The plot demonstrates a substantial deviation



Graph 10: Residual Diagnostics of SARIMA$(0,1,1)\times(3,1,1)$

from normality for both models. Specifically the tails of the distribution exhibit significant divergence, with extreme residuals at both ends deviating substantially from the expected values under normality. This indicates the presence of heavy tails in the residual distribution. The clustering of points near the center of the plot suggests that residuals around the mean are closer to normality. However, the overall departure from the red line implies that the model residuals are not normally distributed. This is likely due to a structural trend break trend at the beginning of the time series. From 2004 until 2020, we could see a strong upward trend in number of passengers. Conversely, from 1999 until 2004 a downward trend is evident. Considering that increasing trend makes up the majority of the sample, our models will be fit to this trend, leading to high residuals at the beginning of the time series.

To resolve this issue, we might use a log transformation instead of a Box-Cox transformation to stabilize the variance of the data. A log transformation may resolve the issue more effectively than a Box-Cox transformation because it is simpler to apply and interpret, particularly when dealing with time series data exhibiting structural breaks and heavy tails. Unlike the Box-Cox transformation, which requires estimating an optimal lambda parameter to stabilize variance, the log transformation directly addresses variance stabilization in a straightforward manner. This is especially useful when the data already shows a strong multiplicative structure, as is common in time series with exponential growth or decline, such as the passenger data in this case. The log transformation inherently compresses large values, reducing the influence of extreme observations and mitigating the divergence caused by the early downward trend in the series. Moreover, it avoids potential issues with numerical instability or overfitting that can arise from Box-Cox's reliance on parameter estimation, making it a more robust choice for addressing the non-normally distributed residuals in this context.

After log transforming the dataset, the two models were fitted to the data again. The results showed, however, a worse performance than with the Box-Cox transformed data. While the residuals did still not display a normal distribution, the ACF of the residuals at lag 12 became more pronounced and the Ljung-Box Test also showed significant correlations, which were not present with the Box-Cox transformed data. Furthermore, the AIC and BIC values increased. Hence, we will continue to work with the Box-Cox transformed data.

Moreover, we attempted to remove the significant ACF of the residuals at lag 12 by increasing D to 2. This did, however, neither remove the ACF at lag 12, nor improve the other diagnostics. Despite trying values ranging from 1 to 4 for Q ($P = 0$ due to insignificance of the MA coefficient in the seasonal component, $p = 0$, q ranging from 0 to 1), all of the models showed a significant correlation of residuals beginning at lag 12 according to the Ljung-Box Test and the ACF at lag 12 was higher than for the models with $D = 1$. The residuals remained non-normally distributed.

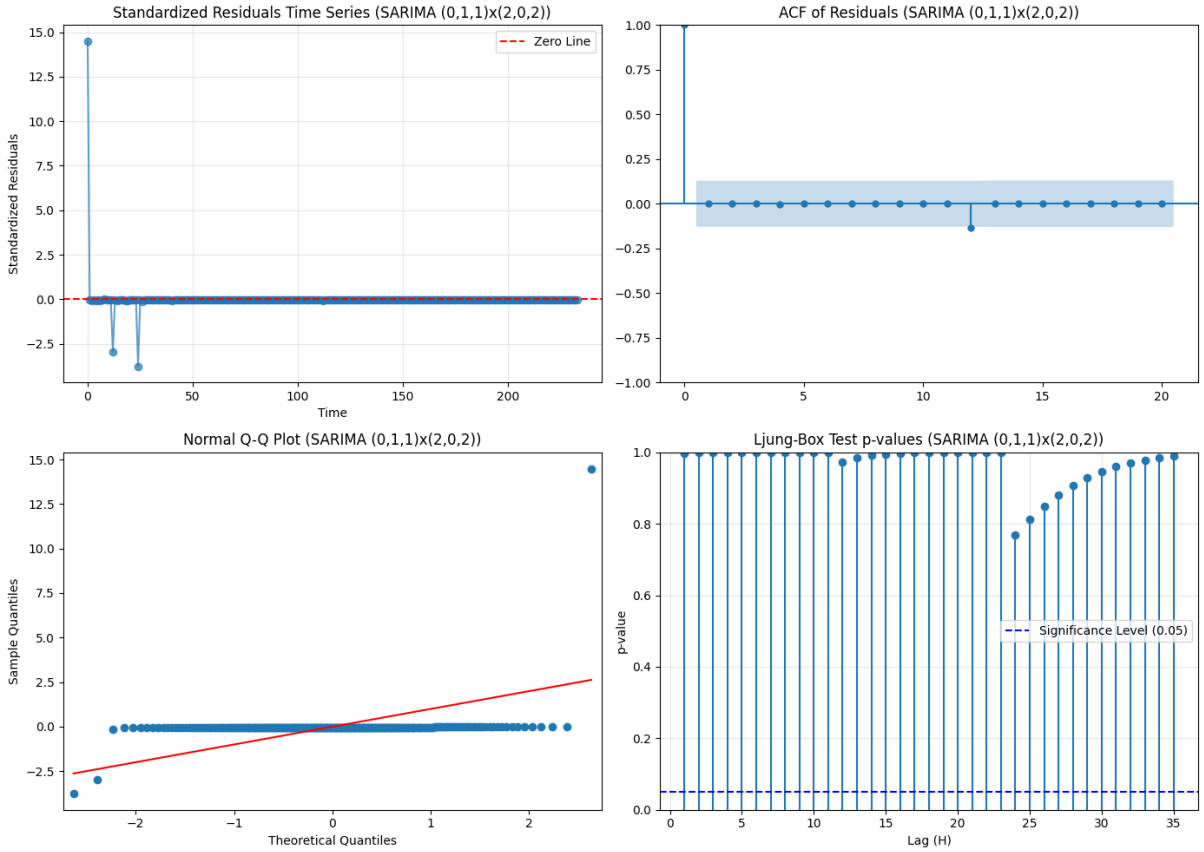
In conclusion, none of the adjustments were able to resolve the significant ACF at lag 12 and the non-normal distribution of the residuals. Consequently, SARIMA$(0,1,0)\times(3,1,1)_{12}$ and SARIMA$(0,1,1)\times(3,1,1)_{12}$ with Box-Cox transformed data will be used for the forecast.

3.4 Best Model based on AIC: SARIMA$(0,1,1)\times(2,0,2)_{12}$

After manually finding the best models, we built a function which loops through all possible combinations of p, q, P, and Q between 0 and 4 (inclusive) and d as well as D between 0 and 2 (inclusive) to find the best SARIMA model based on AIC. The model with the lowest AIC was SARIMA$(0,1,1)\times(2,0,2)_{12}$ with a score of -1840. Although that model was found without residual diagnostics, the diagnostics displayed in graph 11 confirm the model's adequacy.The ACF plot shows that there is no significant autocorrelation between the residuals except for lag 12. When comparing to the previous models, the peak at lag 12 is less pronounced and barely significant. The Ljung-Box test results indicate that the p-values for all lags are well above the significance level of 0.05. Table 4 does, however, show that one of the seasonal MA coefficients is not significant. Nonetheless, we will keep the coefficient considering that the model was found automatically using the AIC criteria. It, therefore, serves the purpose of comparing how the two approaches for finding a suitable model can lead to an out- or underperformance.

Table 4: Summary of Model Parameters for SARIMA$(0,1,1)\times(2,0,2)$

Parameter	**Coef.**	**Std Err**	z	$P > \|z\|$
ma.L1	−0.24	0.05	−5.21	0.00
ar.S.L12	0.40	0.09	4.72	0.00
ar.S.L24	0.55	0.09	6.31	0.00
ma.S.L12	−0.08	0.15	−0.55	0.58
ma.S.L24	−0.48	0.10	−4.70	0.00
AIC		-1839.59		
BIC		-1819.60		



Graph 11: Residual Diagnostics of SARIMA$(0,1,1)\times(2,0,2)$

4 Forecast Performance on the Test Dataset

To compare the performance of our models, the passenger count for 2019 was forecasted based on the data from 1999 until 2018. The objective of forecasting is to make predictions about the future with the greatest possible degree of accuracy [3]. The three models used for forecasting were the ones described in Section 3.3 and 3.4. Furthermore, different forecasting approaches were employed: fixed, recursive, and rolling scheme.

In the fixed scheme, the training dataset remains unchanged as forecasts are made for all future points. The model is estimated only once, meaning each forecast is constructed using the same parameters. This approach is fast and convenient because it requires only one estimation. It does, however, not allow for parameter updating, which can be problematic in the presence of structural breaks or changes in model stability over time. The recursive scheme updates the training dataset sequentially by incorporating the actual values of previously forecasted points as new inputs. The model is re-estimated whenever a new observation is added to the estimation sample. This method has the advantage of incorporating as much information as possible into the model's estimation, making it beneficial if the model remains stable over time. If the data, however, exhibits structural breaks, the model's stability can be compromised. Lastly, the rolling scheme uses a sliding window approach, where the training dataset is updated by dropping the earliest data point and adding the most recent one. The model is re-estimated within each rolling sample, making it more robust against structural breaks in the data. While this scheme avoids the stability issues that may affect the recursive scheme, it does not utilize all available data, which can sometimes limit its predictive power. While the fixed scheme forecasts were computed for every model,

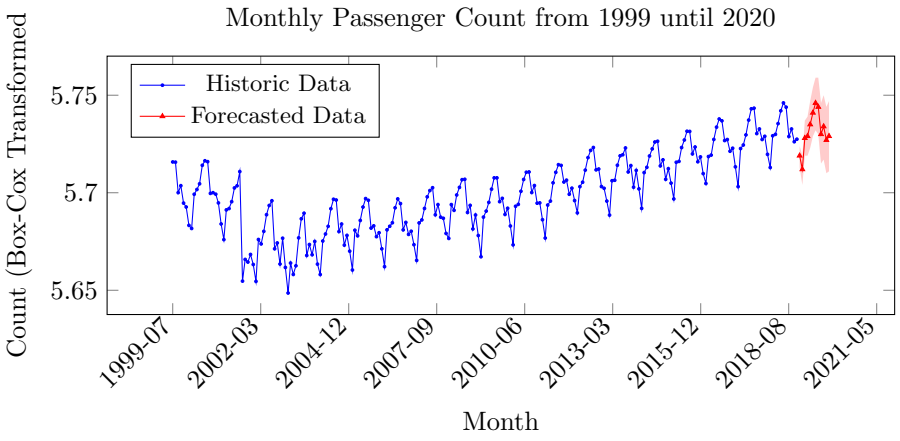
the recursive and rolling scheme forecasts were computed only for the SARIMA$(0,1,0)\times(3,1,1)_{12}$ since it was the best-performing model with the fixed scheme.

Metric	$(0,1,0)\times(3,1,1)$	$(0,1,1)\times(3,1,1)$	$(0,1,0)\times(2,0,2)$	**Rolling Scheme**	**Recursive Scheme**
Absolute Errors					
MAE	92,015	97,772	113,629	95,665	92,006
RMSE	121,600	130,875	150,522	117,610	121,588
Relative Errors					
MAE	1.921 %	2.041 %	2.372 %	1.997 %	1.921 %
RMSE	2.538 %	2.732 %	3.142 %	2.455 %	2.538 %
MAPE	1.952 %	2.074 %	2.410 %	2.075 %	1.952 %
Indices					
Theil's U	0.277	0.298	0.343	0.363	0.277

Table 5: Forecast Performance of the different SARIMA Models (Fixed and Rolling Scheme computed for (0,1,0)x(3,1,1))

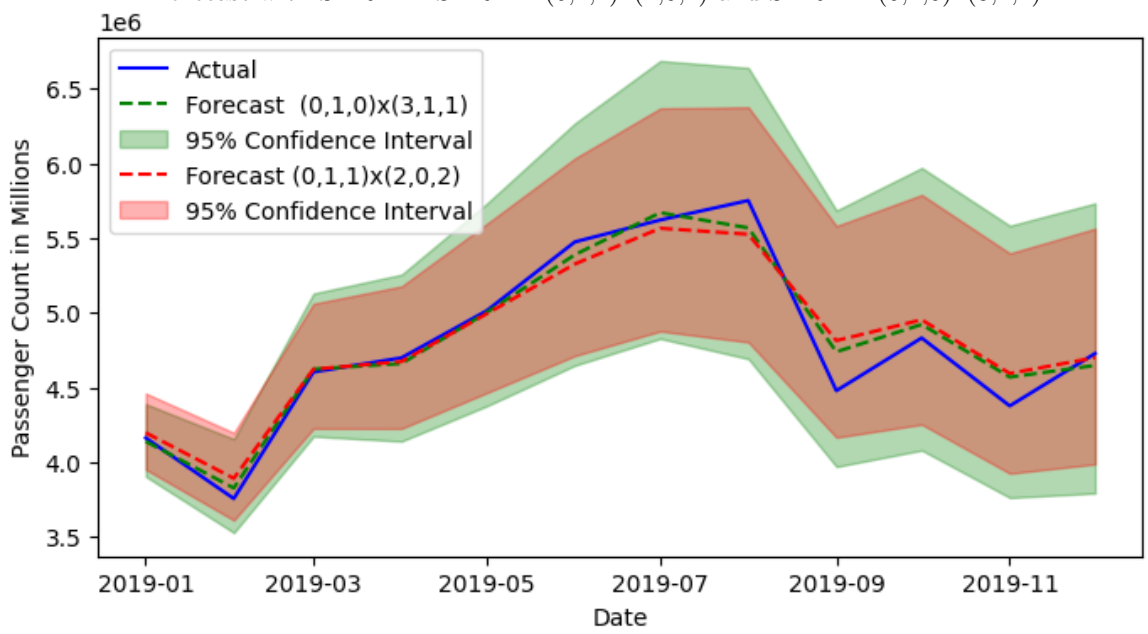
the forecasts produced by the SARIMA$(0,1,0)\times(3,1,1)_{12}$ and SARIMA$(0,1,1)\times(3,1,1)_{12}$ are very similar. They only deviate slightly from each other, with the SARIMA$(0,1,0)\times(3,1,1)_{12}$ having lower errors. As can be seen in table 5, it has a Root Mean Squared Error (RMSE) of 121,600 in comparison to 130,875. Considering that the SARIMA$(0,1,0)\times(3,1,1)_{12}$ also has less parameters than SARIMA$(0,1,1)\times(3,1,1)_{12}$, this is clearly the more favorable model. The AIC and BIC of the SARIMA$(0,1,0)\times(3,1,1)_{12}$ are lower than the AIC and BIC of the SARIMA$(0,1,1)\times(3,1,1)_{12}$.

Finally, when comparing the SARIMA$(0,1,0)\times(3,1,1)_{12}$ and SARIMA$(0,1,1)\times(2,0,2)_{12}$ models, they have the same number of parameters but the AIC and BIC of the SARIMA$(0,1,1)\times(2,0,2)_{12}$ model are the lowest ones. However, the SARIMA$(0,1,1)\times(2,0,2)_{12}$ has a significantly lower forecasting accuracy than the other two models with a RMSE of 150,522. Therefore, the Information Criteria did not prove to be a valuable comparison metric for the forecasting quality in our case.



Graph 12: Monthly Passenger Count including forecasted values for 2019 with the 95 % confidence interval with SARIMA$(0,1,0)\times(3,1,1)$

As demonstrated in Graphs 12 and 13, a more detailed investigation of the produced forecasts reveals



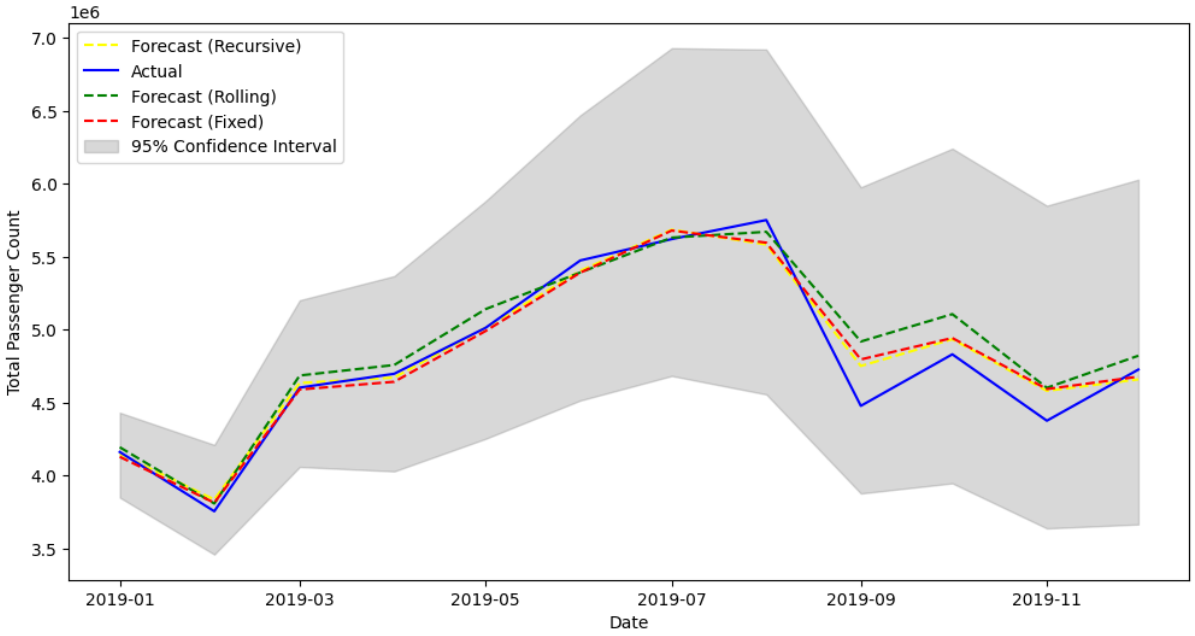
Graph 13: Actual Passenger Count vs Forecast with SARIMA(0,1,1)x(2,0,2) and SARIMA(0,1,0)x(3,1,1) including 95% Confidence Intervals

that both SARIMA(0,1,1)x(2,0,2) and SARIMA(0,1,0)x(3,1,1) models exhibit a decline in forecast accuracy towards the end of the forecasted period. This decline is anticipated, as the uncertainty surrounding the near future diminishes, leading to an accumulation of forecasting errors over time. Over the past four months, the models have consistently overestimated the number of passengers, with the summer peak forecast to occur a month earlier than expected. Additionally, the SARIMA$(0,1,1)\times(2,0,2)_{12}$ model marginally overestimates passenger volume at the beginning of the year and underestimates it during the summer months. However, it is important to note that the observed values of passenger volume fall within the 95th confidence intervals for both models. It is also noteworthy that the confidence interval for the SARIMA(0,1,0)x(3,1,1) model is more extensive, indicating a higher degree of uncertainty or variability in the model's predictions.

In consideration of the superior performance exhibited by the SARIMA$(0,1,0)\times(3,1,1)_{12}$ model in comparison to the other models, the SARIMA$(0,1,0)\times(3,1,1)_{12}$ model was utilised to derive rolling and recursive forecasts (see Graph 14). The recursive approach is advantageous due to its capacity to maintain the precision of the fixed model whilst concurrently offering adaptability for dynamic forecasting. This makes it an attractive option when adapting predictions to new data over time. A comparison of the rolling scheme with the fixed implementation reveals a slight reduction in accuracy. This slight decline in performance can be attributed to the inherent limitations of rolling schemes, which propagate errors forward with each forecasting iteration. Nevertheless, the rolling scheme remains a viable alternative for scenarios where adaptive predictions are essential.

5 Conclusion

This report has provided significant insights into the modelling of air passenger volume with a SARIMA methodology. The EDA revealed a pronounced seasonality and upward trend in the time series. The findings of this analysis have guided the subsequent selection of a SARIMA methodology, deemed to



Graph 14: SARIMA(0,1,0)x(3,1,1) Forecasts with the fixed, rolling, and recursive (expanding) scheme

be the most suitable approach for the given context. The dataset was transformed into a stationary time series using a Box-Cox transformation, as well as seasonal and first-order differencing. A manual analysis of the residual diagnostics and coefficients was then conducted, leading to the proposal of SARIMA$(0,1,0)\times(3,1,1)_{12}$ and SARIMA$(0,1,1)\times(3,1,1)_{12}$ as the most suitable models. The forecasting results of these models were then compared to those of the SARIMA$(0,1,1)\times(2,0,2)_{12}$ model, which was found automatically and has the lowest possible AIC. Finally, the three models were employed for forecasting. The SARIMA$(0,1,0)\times(3,1,1)_{12}$ model produced the most accurate forecasts, both with a fixed scheme and a recursive scheme.

The primary strengths of the proposed model are evident in its capacity to forecast passenger volume with a reasonable degree of accuracy under typical conditions. This is attributed to the stability of seasonal patterns and trends, which exhibit minimal disruptions. Consequently, the observed values invariably fell within the 95th percentile confidence intervals across all models, and the error metrics remained modest, given the magnitude of passenger volume (Relative Root Mean Square Error ranged from 2.4% to 3.2%).

However, the model's limitations become evident in the presence of unexpected events, such as the emergence of the novel (SARS-CoV-2). For instance, the pandemic caused a complete disruption of the upward trend, making it impossible to model that portion of the data accurately. Additionally, a slight downward trend at the beginning of the dataset contributed to a non-normal distribution of residuals in the models. Furthermore, a substantial degree of autocorrelation was observed in all model residuals at lag 12.

References

[1] George EP Box and David R Cox. An analysis of transformations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 26(2):211–243, 1964.

[2] San Francisco Government. Air traffic passenger statistics, 2024. URL `https://data.sfgov.org/Transportation/Air-Traffic-Passenger-Statistics/rkru-6vcg/`. Last accessed 22 September 2024.

[3] RJ Hyndman. *Forecasting: principles and practice*. OTexts, 2018.