

MEGI

Master Degree Program in
Statistics and Information Management

TIME SERIES INTERVENTION ANALYSIS OF PORTUGUESE AIRPORTS' PASSENGER THROUGHPUT (2004-2018)

Paulo Alexandre Caeiro Correia

Master Thesis

presented as partial requirement for obtaining a Master's Degree in Statistics and Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

**TIME SERIES INTERVENTION ANALYSIS OF PORTUGUESE AIRPORTS' PASSENGER
THROUGHPUT (2004-2018)**

by

Paulo Alexandre Caeiro Correia

Master Thesis presented as partial requirement for obtaining the Master's degree in Statistics and Information Management, with a specialization in Information Analysis and Management.

Supervised by

Jorge M. Mendes, PhD, NOVA Information Management School

November, 2024

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism, any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Lisbon, November 24, 2024

Paulo Correia

ABSTRACT

The evolution of passenger traffic is shaped by economics, demographics, operating environment and unexpected events. Understanding the dynamics of these anomalous events is essential for accurate forecasting and informed decision-making in civil aviation. This dissertation focuses on modelling a monthly time series with interventions, representing the number of passengers who embarked, disembarked, or were transferred at the main Portuguese airports between 2004 and 2018. The primary objective is to identify model specifications that deliver accurate forecasts, utilizing a recursive out-of-sample forecasting approach to assess performance against data from 2019. The findings demonstrate that incorporating interventions significantly improves the models' fit, both in-sample and out-of-sample, when compared to standard ARIMA models. However, the inclusion of additional regressors, such as outliers, interventions and breakpoints, did not consistently lead to better forecasting accuracy. Furthermore, the analysis reveals that monthly data aggregation is not ideal for detecting anomalies or assessing the impact of specific events in mature and competitive markets. In such contexts, the rapid responses of market participants often compensate for supply disruptions caused by the anomalous events, thereby diminishing their detectability in aggregated data.

KEYWORDS

ARIMA; Regression with ARIMA errors; Time series outliers; Intervention analysis.

Sustainable Development Goals (SDG):



TABLE OF CONTENTS

Statement of Integrity	ii
Abstract	iii
List of Figures.....	vii
List of Tables.....	viii
List of Abbreviations and Acronyms.....	ix
1. Introduction.....	1
2. Literature review	2
2.1. ARIMA Framework	2
2.1.1.Box-Jenkins iterative cycle	2
2.1.1.1. Seasonal ARIMA model.....	3
2.1.1.2. Stationarity.....	3
2.1.1.2.1. Augmented Dickey-Fuller test (ADF).....	4
2.1.1.2.2. Kwiatkowski-Phillips-Schmidt-Shin test (KPSS)	5
2.1.1.2.3. Hylleberg-Engle-Granger-Yoo test (HEGY)	5
2.1.1.2.4. Canova-Hansen test (CH)	6
2.1.1.3. ARIMA order and model selection.....	6
2.1.1.4. Residual diagnostics.....	7
2.1.1.4.1. Ljung-Box test.....	8
2.1.1.4.2. Lagrange multiplier test for conditional heteroskedasticity (ARCH test) ...	8
2.1.1.4.3. Jarque-Bera test	9
2.2. Outliers and interventions.....	9
2.2.1.Outliers and time series	9
2.2.2.Interventions	10
2.2.3.Classification of outliers	11
2.2.3.1. Additive outlier (AO)	13
2.2.3.2. Innovative outlier (IO).....	14
2.2.3.3. Level shift (LS) and Temporary level shift (TLS)	14
2.2.3.4. Seasonal level shift (SLS)	15
2.2.3.5. Temporary change (TC).....	15
2.2.3.6. Ramp shift outlier (RS)	15

2.2.4. Outlier detection	15
2.3. Intervention modelling	17
2.4. Dynamic regression	19
2.5. Forecasting and accuracy metrics	20
3. Methodology	22
4. Data description, data selection and data cleaning	24
4.1. Data source and description.....	24
4.2. Data selection and univariate time series generation	25
4.3. Data cleaning	25
5. Analysis of air passengers' throughput time series.....	27
5.1. Time series overview	27
5.2. Descriptive statistics.....	28
5.3. Seasonality and autocorrelation	29
5.4. Spectral density and periodogram with stationary data.....	33
5.5. Variance stability and data transformation	34
5.6. Decomposition	35
5.7. Stationarity	36
6. ARIMA modelling.....	39
6.1. Seasonal ARIMA Models.....	39
7. Selection of external regressors	42
7.1. Calendar effects.....	42
7.2. Outlier detection	46
7.2.1. Application of the Chen and Liu (1993) methodology	46
7.2.2. Outlier detection using decomposition and residual analysis	49
7.3. Structural breaks and modelling of the financial crisis period.....	52
7.4. Interventions	54
8. Regression with ARIMA errors	58
9. Linear regression model with Fourier terms and external regressors	60
10. Forecasting and accuracy	66
11. Discussion and conclusions	69
12. Future research	73
References.....	74
Appendix A: Data collection and screening	90
Appendix B: Database coverage (2016-2018).....	91
Appendix C: Coding of Easter and Leap-year effects	92

Appendix D: Residuals plots	93
Appendix E: Fitted and forecasts	97
Annex A: Traffic form	103

LIST OF FIGURES

Figure 2-1: Box and Jenkins iterative cycle for model building (Box et al., 2015, pp. 15–17)...	3
Figure 2-2: Outlier types.....	13
Figure 5-1: Passenger time series, rolling mean and line of polynomial regression.	27
Figure 5-2: Kernel density estimate by year.	28
Figure 5-3: Monthly distribution of passengers and seasonal variation.	30
Figure 5-4: Autocorrelation at lags 1 to 20.	31
Figure 5-5: Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of untransformed data.	32
Figure 5-6: Periodogram of first differenced, log-transformed series.....	34
Figure 5-7: Decomposition of log-transformed series into trend, seasonal, and residual components.....	35
Figure 5-8: Transformation and differencing of data: Original, log-transformation, first differencing, and seasonal differencing.	36
Figure 6-1: ACF and PACF of the transformed data.	39
Figure 6-2: Residual diagnostics for the <i>auto.arima</i> model.....	41
Figure 7-1: Residual diagnostics for model with calendar effects.	45
Figure 7-2: Figure 7-3: Detected outlier effect (Chen and Liu, 1993).	46
Figure 7-4: Residual diagnostics for model with calendar effects and outliers detected using Chen and Liu (1993) procedure.....	48
Figure 7-5: Residual diagnostics for model with calendar effects and outliers detected using decomposition and residual analysis.	51
Figure 7-6: Structural breakpoints in model residuals.....	53
Figure 7-7: Dates of interventions.	57
Figure 9-1: Residual diagnostics for log-linear model with Easter effect, Fourier terms and Intervention regressors.....	63

LIST OF TABLES

Table 2-1: Model identification using ACF and PACF	6
Table 2-2: Impacts of interventions	19
Table 5-1: Summary of passenger data statistics: 2004 – 2018.	29
Table 5-2: Autocorrelation and robust autocorrelation tests.....	33
Table 5-3: Results of stationarity and unit root tests for log-transformed and differenced series.	37
Table 5-4: Summary of Canova and Hansen test results for seasonal stability for different transformations.....	38
Table 6-1: Base model comparison.....	40
Table 7-1: Calendar effects models.....	43
Table 7-2: Summary of ARIMA (0,1,1)(0,1,1) ₁₂ model with calendar effects.	44
Table 7-3: Summary of ARIMA (0,1,1)(2,1,0) ₁₂ model with calendar effects and outliers detected using Chen and Liu (1993) procedure.....	47
Table 7-4: Summary of ARIMA (0,1,1)(0,1,1) ₁₂ model with calendar effects and outliers detected using decomposition and residual analysis.	50
Table 7-5: External regressors for structural breaks and financial crisis intervention.	54
Table 7-6: Interventions.	56
Table 8-1: Models with outliers, interventions and calendar effects.	58
Table 9-1: Estimated coefficients and significance levels for log-linear model with Fourier and intervention terms.	61
Table 9-2: Diagnostic tests for residuals and model adequacy.	64
Table 9-3: Estimated coefficients with Newey-West adjusted standard errors and significance levels for the log-linear model incorporating Fourier seasonal components and intervention terms.	65
Table 10-1: Model comparison of forecasting accuracy metrics.	67
Table 10-2: Pairwise comparison of forecasting models using Diebold-Mariano test (One-Step Ahead).	68

LIST OF ABBREVIATIONS AND ACRONYMS

ANAC	Autoridade Nacional da Aviação Civil
ANA	Aeroportos de Portugal, S.A.
ARIMA	Aeroportos de Portugal, S.A.
ARMA	Autoregressive–moving-average
EASA	European Union Aviation Safety Agency.
ICAO	International Civil Aviation Organization
UNFCCC	United Nations Framework Convention on Climate Change

1. INTRODUCTION

Air travel time series have long been one of the most interesting subjects to analyse with time series techniques ever since the release of the famous Box and Jenkins airline data comprising the number of international air passengers from 1949 to 1960 (McCoy & Stephens, 2004). With the intensification of global trade and globalization, most air passenger time series will show the presence of high seasonality, trend and a cyclical component making this type of dataset a superb candidate for experimenting and developing time series techniques.

Air traffic forecasting is of the utmost importance to most of the aviation activities. From providing justification to airport planning, airport development and investment decisions to city planning, forecasting of future demand supports all major decisions where air traffic is taken into consideration. Even with the availability of numerous alternative approaches, time series analysis techniques remain an essential toolset for forecasting activities.

These models are influenced by various external factors commonly referred to as *interventions* (S. C. Hillmer & Wei, 2006, p. 212) which are abnormal, known events that cause significant deviations in the series, typically manifesting as shifts or outliers.

Intervention analysis is a modelling approach for incorporating (estimating) the effects of interventions in time series analysis. An example of interventions in the air transport scenario are the several oil crises (1979-1980, 1990), the terrorist attacks of 2001, the Severe Acute Respiratory Syndrome (SARS) crisis in 2003 and the COVID-19 pandemics in 2020.

Interventions can affect the level, trend, or induce other response patterns in time series, potentially leading to distorted conclusions from modelling. When interventions are not accounted for, they can compromise the fit of ARIMA models and result in inaccurate forecasts. (S. C. Hillmer & Wei, 2006).

Likewise, Box, Jenkins, Reinsel and Ljung (2015, p. 492) emphasize the importance of recognizing the presence of outliers that distort the conclusions regarding the time series but caution against the overuse of “*outlier adjustment procedures*” that may lead to “*overspecification*” issues or compromise the interpretation of the results.

The advantages of interventions analyses, lie in its ability to detect and model changes in the “*in level, slope, or regime of a process*” (Yaffee & McGee, 2000, p. 345) particularly when the effects of an intervention are not immediately apparent. By modelling the response, intervention analysis enhances the understanding of the nature and dynamics of the intervention's effect.

2. LITERATURE REVIEW

ARIMA models are widely used to analyse and forecast air traffic indicators. Tsui et al., (2014) applied ARIMA models (with and without explanatory variables) to forecast the Hong Kong International Airport's passenger throughput¹ and concluded that the model with explanatory variables produced improved predictions despite both models were highly accurate. Payne and Taylor (2007) compared an ARIMA model with a dynamic regression model (referred to as an autoregressive-seasonal-trend model in their study) and concluded that the dynamic regression model outperformed the ARIMA model in both in-sample and out-of-sample forecasting accuracy. Li et al., (2017) proposed a seasonal ARIMA model to predict the passenger flow of an Asian airport and concluded that the model produced error ranges from 1% to 3%. Nieto et al., (2018) combined an ARIMA approach with other methods to forecast air transportation demand in the U.S. market.

2.1. ARIMA FRAMEWORK

In this section, the ARIMA framework is introduced as a methodology for time series modelling. The foundational Box-Jenkins iterative cycle is presented with its three stages: model identification, parameter estimation, and diagnostic checking.

2.1.1. BOX-JENKINS ITERATIVE CYCLE

A common strategy used to construct a ARIMA model is based on an iterative cycle (Box et al., 2015, pp. 15–17), as shown in Figure 2-1), aimed to obtain a model that can be used to forecast a time series. Starting from choosing a reasonable class of models, then by analysing the autocorrelation function, partial autocorrelation function, expanded autocorrelation function (EACF), Akaike criterion (AIC) or the Bayesian Information Criterion (BIC), it enables the identification of the underlying structure of the model. The order of the model is assessed also by taking into consideration the stationarity of the time series (Box et al., 2015, p. 180).

The parameter estimation of the specified model uses the maximization the conditional likelihood function or the least squares methods - conditional or unconditional - (Box et al., 2015, pp. 209–210) even though the MLE estimation is recommended for smaller samples (Box et al., 2015, p. 217). Subsequent diagnostic checks are performed namely by scrutinizing the residuals of the fitted model. If they exhibit characteristics of white noise, are not autocorrelated, are approximately normal and homoscedastic, the model is deemed as appropriate to represent the underlying structure of the data. Otherwise, the iterative cycle is reiterated. Modelling multiples seasonalities in ARIMA models, as far as we know, hasn't been extensively used, since ARIMA models are unsuitable do handle very distant autocorrelations (R. Hyndman, 2010).

¹ A concept referred to as “processed passenger” or “handled passengers”.

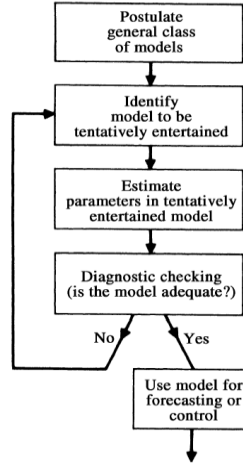


Figure 2-1: Box and Jenkins iterative cycle for model building (Box et al., 2015, pp. 15–17).

2.1.1.1. SEASONAL ARIMA MODEL

An ARIMA multiplicative model with a seasonal component can be expressed by:

$$\phi_p(B)\Phi_P(B^S)\nabla^d\nabla_s^D y_t = \theta_q(B)\Theta_Q(B^S)\varepsilon_t, \quad (2-1)$$

where ∇ is defined by $\nabla y_t = y_t - y_{t-1}$. Therefore, $\nabla y_t = y_t - y_{t-1} = (1 - B)y_t$. $\nabla^d\nabla_s^D$ denotes the differencing level for the seasonal and non-seasonal processes, meaning $\nabla_s = 1 - B^S$ and $\nabla_s^D y_t = (1 - B^S)^D y_t$. B is the backshift or lag operator with $B(y_t) = y_{t-1}$ implying that $B^m y_t = y_{t-m}$, and ε_t is a white noise process (e.g., Yaffee & McGee, 2000, p. 173). The moving average polynomial (MA) is defined by:

$$\theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q, \quad (2-2)$$

the autoregressive polynomial (AR) is defined by:

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p, \quad (2-3)$$

the seasonal MA is given by:

$$\Theta_Q(B^S) = 1 + \theta_1 B^S + \theta_2 B^{2S} + \dots + \theta_Q B^{QS}, \quad (2-4)$$

and the seasonal AR polynomial can be expressed by:

$$\Phi_P(B^S) = 1 - \phi_1 B^S - \phi_2 B^{2S} - \dots - \phi_P B^{PS}. \quad (2-5)$$

2.1.1.2. STATIONARITY

A stochastic process is weakly stationary (or covariance stationary) if *i*) it has the same expected value at all points ($E(x_t) = \mu, \forall t$) meaning μ doesn't depend on t ; *ii*) has the same

finite variance ($E(x_t)^2 < \infty, \forall t$), and *iii*) the covariance between the values at any two different time points depend only on the distance of these two values and not on the time difference, *i.e.*, $cov(x_{t_1}, x_{t_2}) = cov(x_{t_1+h}, x_{t_2+h}), \forall t_1, t_2, h$ (e.g., Mills, 2019, p. 33). Stationarity implies that the characteristic polynomial of degree p , $\mu(B)$, has to have a modulus value for all roots ($\Phi(B) = 0$) greater than 1 *i.e.*, the roots (can be complex) lie outside the unit circle being B a complex variable and not an operator (Woodward et al., 2017, p. 51).

Testing for stationarity can be done by testing for the presence of a unit root *i.e.*, if a given test provides statistical support to whether 1 is a solution to the characteristic equation of the modelled process, the time series is non-stationary² although differencing can reduce it to stationarity (Yaffee & McGee, 2000, p. 84).

Let an ARMA process expressed by:

$$\phi_p(B)\Phi_p(B^s)y_t = \theta_q(\lambda B)\Theta_q(\lambda B^s)\varepsilon_t, \quad (2-6)$$

where $\phi_p(B)$ and $\Phi_p(B^s)$ represent the non-seasonal and seasonal autoregressive components, respectively and $\theta_q(\lambda B)$ and $\Theta_q(\lambda B^s)$ represent the non-seasonal and seasonal moving average components, respectively.

The time series has a unit root if any λ , the eigenvalues of the characteristic equation, lies on or inside the unit circle.

2.1.1.2.1. AUGMENTED DICKEY-FULLER TEST (ADF)

The null hypothesis for the Dickey-Fuller test states that a unit root is present in an autoregressive model of a time series and that the process is not stationary. The non-augmented test can be used to test the presence of three types of unit root processes: *i*) unit root: $\nabla y_i = \delta y_{i-1} + u_i$; *ii*) unit root with drift: $\nabla y_i = a_0 + \delta y_{i-1} + u_i$ and *iii*) unit root with drift and a deterministic time trend: $\nabla y_i = a_0 + a_1 t + \delta y_{i-1} + u_i$.

The Augmented Dickey-Fuller test (Dickey & Fuller, 1979; Said & Dickey, 1984) enables the use of higher p orders addressing autocorrelation and allowing for the modelling of more complex dynamics, such as trends. The null hypothesis states that a unit root is present, hence the autoregressive model represents a process nonstationary. The regression equation for testing the trend stationarity is expressed by:

$$\nabla y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^p \delta_i \nabla y_{t-1} + \varepsilon_t, \quad (2-7)$$

Where α represents the intercept, ∇ is the first difference operator, p expresses the number of AR lags, δ is a parameter for the lagged differences and ε is assumed to be a white noise

² A linear time series is nonstationary in the presence of a unit root, but the reciprocal is not valid *i.e.*, the absence of a unit root does not assure stationarity.

process. To test for stationarity, equation (2-7) is used without the trend component, βt . In some unusual situations, the intercept can also be suppressed (Neusser, 2016, pp. 147–148). The ADF test is very sensitive to the lag length chosen (Said & Dickey, 1984) *apud* Das (2019, p. 320) and the order p can be chosen using an information criterion or the interactive procedure proposed by Ng and Perron (1995). The test statistic is expressed by:

$$DF_t = \frac{\hat{\gamma}}{SE(\hat{\gamma})}, \quad (2-8)$$

where $\hat{\gamma}$ is the estimated coefficient of y_{t-1} and $SE(\hat{\gamma})$ is its standard error. Hence the null hypothesis becomes $H_0: \gamma = 0$ (there is a unit root in the time series) against the alternative $H_1: \gamma < 0$ (there isn't a unit root in the time series). If the test statistic result is less than the tabulated critical value, then H_0 is rejected and there is statistical support to conclude that the time series does not contain a unit root and is trend-stationary.

2.1.1.2.2. KWIATKOWSKI-PHILLIPS-SCHMIDT-SHIN TEST (KPSS)

KPSS and ADF explore different aspects of non-stationarity and can be used in a complementary manner, by focusing on whether stationarity can be assumed, rather than testing for the presence of a unit root. In a KPSS test (Kwiatkowski et al., 1992), under the null hypothesis, a time series is stationary. The time series is decomposed in trend, stationary error and a random walk which is tested for the hypothesis that it has zero variance. The test can be used to infer level stationarity or trend stationarity. In case of testing for the level stationarity, with $H_0: \sigma_u^2 = 0$ against the $H_1: \sigma_u^2 \neq 0$ the test statistic is given by:

$$\hat{\eta}_\mu = \frac{\eta_\mu}{s^2(l)} = T^{-2} \frac{\sum S_t^2}{s^2(l)}, \quad (2-9)$$

where $S_t = \sum_{t=1}^T \varepsilon_t$ and $s^2(l)$ is an estimator for σ_ε^2 . The test statistic results can be compared to the critical values computed by Kwiatkowski et al., (1992). If the test statistic exceeds the critical value, the null hypothesis of stationarity is rejected, indicating that the series exhibits non-stationarity.

2.1.1.2.3. HYLLEBERG-ENGLE-GRANGER-YOO TEST (HEGY)

The test proposed by Hylleberg et al. (1990) as an extension to the Dickey and Fuller test was aimed at testing for seasonal unit roots on quarterly data. Expanding this concept, Beaulieu and Miron (1993) adapt the methodology to monthly time series.

The seasonal differencing operator for monthly data can be expressed as:

$$\nabla_{12} = (1 - B^{12}) = (1 - B)(1 + B)(1 + B^2)(1 + B^3 + \dots + B^{11}) \quad (2-10)$$

The auxiliary regression for the HEGY test in monthly data can be expressed as:

$$\psi(B)y_t = \pi_1 y_{1,t-1} + \pi_2 y_{2,t-1} + \dots + \pi_{11} y_{11,t-1} + \mu_t + \epsilon_t \quad (2-11)$$

where $y_{1,t}, y_{2,t}, \dots, y_{11,t}$ are transformed components of the series designed to isolate specific seasonal frequencies. The null hypotheses test the presence of unit roots at these frequencies are:

a) $H_0: \pi_1 = 0$ (tests for a non-seasonal unit root at frequency $\omega = 0$;

b) $H_0: \pi_2 = 0$ (tests for a non-seasonal unit root at frequency $\omega = \pi$, etc.

If the null hypothesis is rejected for a given π_i , it indicates the absence of a unit root at the corresponding frequency.

2.1.1.2.4. CANOVA-HANSEN TEST (CH)

The test proposed by Canova and Hansen (1995) has, under the null, no seasonal unit roots against the alternative hypothesis of the presence of seasonal roots. Unlike the HEGY test, which isolates specific seasonal frequencies, the Canova-Hansen test evaluates the joint presence of unit roots across all seasonal frequencies. The regression model used can be expressed by:

$$y_t = z + x_t\beta + d_t a + \epsilon_t \quad (2-12)$$

Where z is the intercept, x_t is deterministic seasonal component and d_t are seasonal dummy variables.

2.1.1.3. ARIMA ORDER AND MODEL SELECTION

The specification of a model is open to subsequent evaluation, diagnostic validation, and adjustments as needed. The ACF and PACF plots can be used to help to determine the SARIMA orders $(p, q)(P, Q)$ as discussed in Box et al., (2015a, p. 183).

Table 2-1: Model identification using ACF and PACF

Component	Order	ACF	PACF
AR	(p)	Exponential decay	Cuts off after lag p
MA	(q)	Cuts off after lag q	Exponential decay
ARMA	(p,q) _s	Exponential decay	Exponential decay
AR	(P)	Exponential decay at lag ks, k=1,2	Cuts off after lag P _s
MA	(Q)	Cuts off after lag Qs	Exponential decay at lag k _s , k=1,2,...
ARMA	(P,Q) _s	Exponential decay at lags ks	Exponential decay at lags k _s

Adapted from (Shumway & Stoffer, 2019, pp. 80, 114)

Even though sample ACF and PACF are useful tools in model identification, there are cases where the orders of an ARIMA model cannot be determined solely from the patterns observed in the ACF and PACF (Box et al., 2015, p. 190). In such instances, other tools can help to identify the model for the non-seasonal component, namely the extended sample autocorrelation function (EACF) introduced by Tsay and Tiao (1984).

A common approach for model selection is the use of goodness-of-fit measures such as the Akaike Information Criterion (Akaike, 1974), the Bayesian Information Criterion (Schwarz, 1978) or the Akaike Information Criterion Corrected for small sample sizes (Hurvich & Tsai, 1989). These criteria evaluate the trade-off between model complexity and goodness of fit, discouraging overfitting by penalizing less parsimonious models. The equations for these criteria are, as follows:

$$AIC = n \log\left(\frac{RSS}{n}\right) + 2k, \quad (2-13)$$

$$BIC = n \log\left(\frac{RSS}{n}\right) + k \log(n), \quad (2-14)$$

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}, \quad (2-15)$$

where RSS represents the residual sum of squares, k is the number of parameters included in the model and n is the number of available observations. The above estimators of prediction error provide an indication of the relative quality of a model and add a penalty to the least parsimonious models to discourage overfitting. For “reasonable” models, assuming that all the models use same sample and the same response variable, the best model is the one with the lowest AIC/BIC/AICc value, taking into consideration that BIC penalizes more the addition of more parameters. The assumption of normality is not required for the residuals for the use of AIC/BIC/AICc.

Burnham and Anderson (2004) propose a set of guidelines to compare multiple models by AIC, specifically by rescaling the AIC results of each model through differencing against the minimum AIC of the models in competition. Then, for each difference *i.e.*, $\Delta_i = AIC_i - AIC_{min}$, the above authors provide the following rule of thumb:

- i) If $\Delta_i < 2$, there is substantial support for the i -th model;
- ii) If $2 < \Delta_i < 4$, there is a strong support for the i -th model;
- iii) $4 < \Delta_i < 7$, there considerably support for the i -th model;
- iv) Beyond 10 there is essentially no support for the i -th model when compared against AIC_{min} .

2.1.1.4. RESIDUAL DIAGNOSTICS

In this stage, the residuals of the fitted model are checked for autocorrelation, homoscedasticity and normality.

2.1.1.4.1. LJUNG-BOX TEST

Ljung-Box test (1978) is a portmanteau type of test used to check for the absence of autocorrelation on the residuals. The null hypothesis assumes no significant autocorrelation up to lag h , against the alternative hypothesis of serial correlation up to an order h . The test statistics is:

$$Q = n(n+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n-k}, \quad (2-16)$$

where n is the number of observations, $\hat{\rho}$ is the sample autocorrelation at lag k and h represents the number of tested lags. Under H_0 the test statistic Q follows a χ^2 distribution with h degrees of freedom, assuming n is large.

The Ljung-Box test statistic Q becomes larger when the sample autocorrelations $\hat{\rho}_k$ increase. The corresponding p-value represents the probability of obtaining a value as large (or larger) than the observed one, under the assumption of uncorrelated residuals. Hence, a $p - value < \alpha$ provides statistical evidence against the null hypothesis, suggesting that the residuals are correlated.

2.1.1.4.2. LAGRANGE MULTIPLIER TEST FOR CONDITIONAL HETEROSKEDASTICITY (ARCH TEST)

The Lagrange multiplier test for conditional heteroscedasticity of Engle (1982) *apud* Tsay (2010, p. 114) evaluates whether the variance of the residuals depends on past information, pointing to ARCH effects. The squared residuals $\hat{a}_t^2$, which proxy the conditional variance, are tested for significant dependence on their lagged values using the regression:

$$\hat{a}_t^2 = \alpha_0 + \alpha_1 \hat{a}_{t-1}^2 + \alpha_m \hat{a}_{t-m}^2 + \varepsilon_t, \quad t = m+1, \dots, T, \quad (2-17)$$

where m denotes the number of lags to test, α_i are the coefficients to be estimated and ε_t is the error term. The null hypothesis posits that there are no ARCH effects, *i.e.*, $\alpha_1 = \alpha_2 = \dots = \alpha_m = 0$. The test statistic given by a F-statistic:

$$F = \frac{(SSR_0 - SSR_1)/m}{SSR_1/(T - 2m - 1)}, \quad (2-18)$$

where $SSR_0 = \sum_{t=m+1}^T (a_t^2 - \bar{w})^2$ with $\bar{w}$ is the mean of the squared residuals ($\bar{w} = \frac{1}{T} \sum_{t=1}^T a_t^2$) and SSR_1 is the sum of squared residuals from the fitted regression model (2-35). Under the null hypothesis, the F-statistic follows asymptotically a Chi-squared distribution with m degrees of freedom, such as $F \sim \chi_m^2$. The decision rule is to reject H_0 if $F > \chi_{m,\alpha}^2$, where $\chi_{m,\alpha}^2$ is the critical value of the chi-squared distribution for m degrees of freedom at significance level α .

2.1.1.4.3. JARQUE-BERA TEST

The Jarque-Bera test (Jarque & Bera, 1980) can be applied to the residuals of the fitted model to test for normality by assessing the skewness and kurtosis of the data when compared to those of the normal distribution. The null hypothesis states the data follows a normal distribution whereas the alternative hypothesis assumes the data does not follow a normal distribution.

The test statistic is expressed by:

$$JB = n \left(\frac{S^2}{6} + \frac{(K - 3)^2}{24} \right), \quad (2-19)$$

where S represents skewness and K represents kurtosis. Under the null hypothesis the test statistic follows a chi-squared distribution with 2 degrees of freedom.

2.2. OUTLIERS AND INTERVENTIONS

This section explores the role of outliers and intervention modelling in time series analysis, focusing on their identification, classification, and impact on the underlying data structure and modelling processes.

2.2.1. OUTLIERS AND TIME SERIES

An outlier is “an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism” (Atkinson & Hawkins, 1981) *apud* Aggarwal (2017, p. 1). Outliers may represent unwanted or erroneous data, in which case they can be deleted or replaced³ to improve data quality (Blázquez-García et al., 2021). Alternatively, outliers can be viewed as anomalous⁴ observations which contribute to the understanding of a phenomena. In fact, some activities rely exclusively on the analysis of the outliers themselves, *e.g.*, system intrusion, fraud detection, etc.

Outliers can significantly impact the process of model identification and model building by introducing bias in parameter estimation through the sample autocorrelations, and by distorting the power of the statistical tests based on those estimates. Furthermore, outliers can lead to unsuitable decomposition outcomes and the accuracy of confidence intervals for model parameters. Consequently, prediction quality is greatly affected (Chan, 1995; Chang et al., 1988; S. Hillmer, 1984; Ledolter, 1989). According to Chen and Liu (1993) “*whereas the usual time series model is designed to grasp the homogeneous memory pattern of a time*

³ Replacing outliers with regular values aims to ensure that the normality hypothesis cannot be rejected, and estimation isn't affected by abnormal values. Outlier corrections demand prior decisions about the patterns to be tested. The number of outliers to be treated should be kept to a minimum unless there are strong reasons to justify them (Kaiser & Maravall, 1999) and the series is large enough to compensate for the loss of degrees of freedom.

⁴ “Outlier”, “anomaly” and “discordant observations” are used interchangeably across the literature.

series, the presence of outlying observations or structural changes raises the question of efficiency and adequacy in fitting general auto-regressive moving average (ARMA)”.

In the context of outlier detection for univariate time series, an outlier is a point or an observation that deviates from a predicted or an expected value, $\hat{x}_t$, which, in its turn, can be computed by different methods, *e.g.*, the median absolute deviation (Mehrang, Helander, Pavel, Chieh, & Korhonen, 2015). Other techniques to compute $\hat{x}_t$ assume that outliers are the unlikely residuals after fitting a model or distribution, *e.g.*, (Reddy et al., 2017).

Another class of techniques involve the analysis of the residuals derived from a Seasonal-Trend decomposition using LOESS (STL) decomposition (Hochenbaum et al., 2017) or residuals from ARIMA models estimated iteratively after each identified outlier is removed (Akouemo & Povinelli, 2014) or by adding confidence intervals for the predictions (Zhang et al., 2012). There are also techniques that consider as outliers a certain number of neighbors that lie within a certain distance in value *e.g.*, Angiulli & Fassetti (2007).

A different approach involves the detection of points that, after removed, will result in an ordered histogram with lower error than the original representation (Muthukrishnan et al., 2004).

According to Blázquez-García et al., (2021) regarding the treatment of temporality, the methods to detect point outliers in a univariate time series can be classified as (i) methods that take into consideration the ordered nature of the observations and (ii) methods that ignore the time dependence. Despite the different analytical approaches, both types produce the same results if the whole time series is used in the analysis instead of segments of the time series.

Since the presence of outliers is usually not known at the start of an analysis, it is important to test for their presence and assess their significance. The presence of outliers at the end of the series are particularly difficult to handle. For example, a level shift occurring at the last point can be incorrectly modelled as an additive outlier, thus assigning its effects to the residual component. Therefore, outliers can influence the trend-cycle component or the residuals depending on the coding of the outlier, even though the seasonally adjusted series are unaffected (EUROSTAT, 2015, p. 23).

After identifying outliers, candidate models should jointly estimate the effects of the outliers and the model parameters to ensure robustness against their influence. An alternative approach is to use an estimation method that produce robust parameter estimates and, from there, identify and test the outliers (Becker et al., 2013, p. 243).

2.2.2. INTERVENTIONS

When the time of occurrence or prior information about an event that results in an outlying pattern are known, the effects can be captured by intervention analysis techniques. However, in many cases, the occurrence, causes, and duration of outliers are unknown, necessitating

the use of methods to detect both the location and impact of these unidentified outliers, which are not associated with any pre-identified events. In such situations, the detection of outliers relies on analysing the data itself, showing the close methodological relationship between outlier detection and intervention analysis. (Bruce & Martin, 1989; Chang et al., 1988; Chen & Liu, 1993; Tsay, 1986).

Campbell (1963) and Campbell & Stanley (1966) are credited with the first published research on quasi-experimental designs applied to time series. The term “intervention” was coined by Glass (1972) and Newbold et al., (1976) following the seminal work of Box and Tiao (1965). Tiao et al., (1975) published an influential paper in which the authors assessed the impact of some measures in the pollution levels of Los Angeles. Relevant research, also in the environment domain was presented by Box and Tiao (1975) and Reinsel et al. (1981).

Intervention analysis assumes the intervention changes the mean function, the trend or the autocovariance function of the time series (Cryer & Chan, 2008, p. 250) and that the system in *“which the input event and the impact response take place is assumed to be closed. Apart from the noise of the series itself, the only exogenous impact on the series is presumed to be that of the event or intervention. All other things are presumed to remain the same or to remain external to the system.”* (Yaffee & McGee, 2000, p. 267). In our context, interventions are particular cases of outliers that can be modelled as dynamic changes at some known points in a time series (Becker et al., 2013, p. 250). The main objective of these techniques is to incorporate external information into the process of modelling time series (Cryer and Chan, 2008: 249). Box and Tiao (1975) state that the objective is to *“frame a model for change which describes what is expected to occur given knowledge of the known intervention”*.

“Intervention analysis” and “interrupted time series” are often used interchangeably in the literature. However, the term “interrupted time series” is mostly used in contexts that utilize segmented regression techniques, particularly when seasonality does not play a significant role. Nevertheless, interrupted time series analysis methodology can be addressed by ARIMA models *e.g.* (Box & Tiao, 1975; Meyer et al., 1981) and by ordinary least-squares regression models *e.g.* (Linden, 2015; Velicer & McDonald, 1991) with the later regarded as more flexible (Box et al., 2015).

2.2.3. CLASSIFICATION OF OUTLIERS

Seminal work related to outlier classification/intervention analysis is attributed to Fox (1972), moving away from the assumption of *i.i.d.* in time series context. Fox proposed a typology of outliers affecting linear models and univariate time series, distinguishing between two types.

Type I outliers, later termed additive outliers (AO), refer to single observations that deviate significantly from their neighbouring observations (local outliers) or from the broader set of observations in the time series (global outliers). Type II outliers, subsequently designated as innovative outliers (IO), impact all observations following a specific point in time. These can be conceptualized as a sequence of additive outliers propagating through the series (Balke,

1993). To detect outliers in autoregressive models, Fox (1972) suggests the use of a list of likelihood ratio tests.

Later, Tsay (1988) proposed two new categories of outliers, the level shift (LS), representing a trend break or a change in the level of the series, and the temporary change (TC)⁵, for changes in the level of the time series that return exponentially to the previous level through the effect of a dampening factor (Battaglia & Orfei, 2005, p. 109).

The multivariate cases were addressed by Tsay et al. (2000). Other influential contributions to outlier modelling in time series were made by Tsay (1986) and Chen and Liu (1993, 2006) with all three references framing time series as an ARMA model with interventions added.

Another relevant approach was made by Tsay (1986) by proposing an iterative procedure to identify, detect and remove outliers in stationary and non-stationary ARMA series in order to specify a model. For the detection of the AO and IO outlier types, Chang et al., (1988) developed the likelihood ratio tests for both types. Chen and Liu (1993) introduced an interactive outlier detection method that contributed to improve the model parameter estimation in ARIMA models. This method is now widely used.

Chen and Liu (1993) emphasize that, for all the cases, except the IO case, the effect of the outlier does not depend on the model applied to the time series. This means that these outlier effects can be described without reference to the specific ARIMA structure of the series. However, for innovative outliers (IO), the impact depends explicitly on the model because IOs propagate through the series based on the dynamics of the ARIMA process. Also, different outlier types can produce similar effects on time series. For example, a TC with a parameter of 0,8 is equal to an IO for an AR(1) model with a coefficient of 0,8.

Figure 2-2) illustrates the different types of outliers and structural changes in a time series.

⁵ Some authors call this type of unusual joint behavior “subsequence outliers” (Blázquez-García et al., 2021).

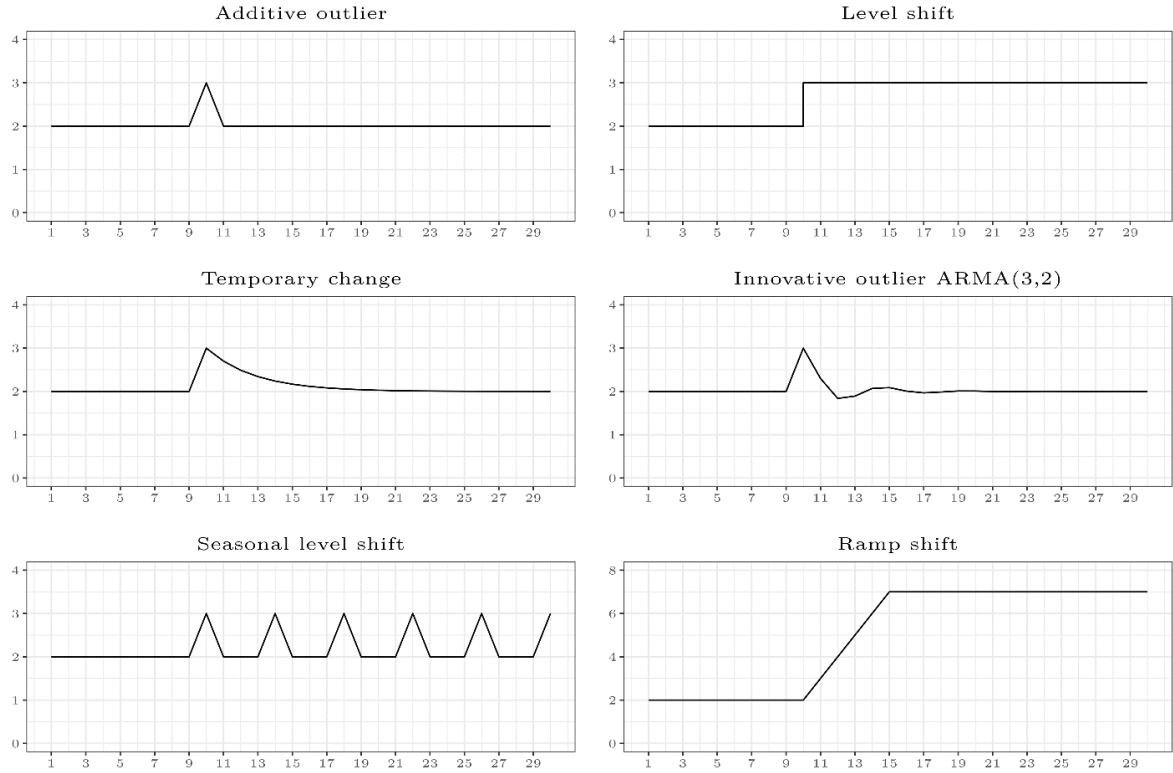


Figure 2-2: Outlier types.

2.2.3.1. ADDITIVE OUTLIER (AO)

An additive outlier models an exogenous change of a single observation, usually related to measurement errors or other external impulses without further effects in future observations. A time series affected by an AO at $t = T$ can be denoted by the equation

$$Y_t = \omega P_t^{(T)} + x_t \quad (2-20)$$

where x_t is underlying time series, assumed to follow an ARIMA model, ω represents the magnitude of the outlier and $P_t^{(T)}$ is an indicator variable that has the value 1 if $t = T$ and 0 otherwise.

An additive outlier can also be represented by i) $x_t = \omega P_t^{(T)} + \Psi(B)a_t$ (Garcia-Ferrer, 2003, p. 154) where $\Psi(B)$ is a polynomial that describes the ARIMA process and a_t represents white noise; and ii) $\pi(B)(z_t - \omega_A P_t^{(T)}) = a_t$ (Pena, 2001), where the ARIMA model $\pi(B)$ governs the adjusted time series, $z_t - \omega_A P_t^{(T)}$.

The transfer function for an additive outlier in moment T is $v(B) = 1$, i.e., the outlier only affects the T^{th} observation. An additive outlier produces a negative bias on the estimated parameter and creates lower autocorrelation coefficients than the coefficients obtained on the same series not contaminated by the additive outlier (Pena, 2001).

2.2.3.2. INNOVATIVE OUTLIER (IO)

An innovative outlier (also called innovational outlier) is usually related to incidents connected to internal causes. The effects of an IO depend on the stationarity of the time series. In the presence of stationarity, the effect tends to fade after a few observations while in the presence of non-stationarity, all observations after the intervention tend to be affected.

The outlier affects the observations $Y_t, Y_{t+j}, j = 1, 2, \dots$, in accordance with a given model, a transfer function $v(b)$ describing the dynamic relationship between the outliers the observed series, expressed as $v(b) = \frac{\omega(B)}{\delta(B)}$. The model adds a pulse effect to the residuals of the original process:

$$Y_t = \psi(B)(\omega P_t^{(T)} + a_t), \quad (2-21)$$

where Y_t is the observed time series, $\psi(B) = \frac{\theta(B)}{\nabla^d \phi(B)}$, is the transfer function derived from the ARIMA model, ω is the magnitude of the outlier and $P_t^{(T)}$ is an indicator variable with zero value in all lags different than T at which point becomes 1. The behaviour of the innovational outlier depends on the ARIMA process because after T , the observations Y_t are affected by the model ψ , i.e., $z_t = y_t + \omega \psi_j$, with $t = T + j$. It can be seen from equation (2-21) that the innovation of a time series affected an IO at $t = T$ is $a_t = e_t + \omega P_t^{(T)}$ where e_t is the series without the IO and $a_t = \Psi(B)\omega P_t^{(T)} + x_t$.

According to (Pena, 2001, pp. 141–142) innovative outliers have less severe effects on the time series than additive outliers and, for large sample sizes, the impact of IOs on the parameter estimates are often negligible.

2.2.3.3. LEVEL SHIFT (LS) AND TEMPORARY LEVEL SHIFT (TLS)

The LS outlier produces an immediate and permanent change on the mean level of the time series on $t = T$ that continues until the end of series, thus making a stationary process non-stationary, unless addressed. It severely affects the identification and the estimation of the observed series (Becker et al., 2013, p. 248) and its effects are larger for stationary than for non-stationary processes (Garcia-Ferrer, 2003, p. 145). A time series affected by a LS is given by a step function that has a value of 0 before T and 1 when $t > T$ i.e., $y_t = \omega S_t^{(T)} + x_t$, where $S_t^{(T)} = \frac{1}{(1-B)} P_t^{(T)}$. This means that by taking a difference to the step function, a pulse function is obtained. As such, the transfer function for a LS is given by $v(B) = \frac{1}{1-B}$.

The innovation of a time series affected a LS at $t = k$ are expressed by $a_t = \pi(B)(z_t - \omega S_t^{(T)})$. The LS can be modeled as a sequence of AO with the same effect during and after T :

$$z_t = \begin{cases} y_t & t < T \\ y_t + \omega & t \geq T \end{cases} \quad (2-22)$$

A temporary level shift implies that:

$$z_t = \begin{cases} y_t & t < T_b \\ y_t + \omega & T_e \geq t \geq T_b \\ y_t & t > T_e \end{cases} \quad (2-23)$$

where T_b denotes the first period of the level shift and T_e represents the last period of the level shift.

2.2.3.4. SEASONAL LEVEL SHIFT (SLS)

A seasonal level shift (SLS) introduces a change in a single seasonal period (*e.g.*, a specific month or quarter) and often results in a permanent change in the level of the series that recurs every seasonal cycle. The transfer function for an SLS is given by $v(B) = \frac{1}{(1-B^s)}$ where s represents the seasonal period. Kaiser and Maravall (1999) present three basic specifications for modelling SLS and show that certain innovative outlier (IO) effects can be equivalently represented by one of these SLS specifications. This equivalence is particularly valid when the time series can be expressed as $\nabla \nabla_s Y_t = \theta(B) a_t$, where the IO takes the form $IO^{(T)} = \omega \frac{\theta(B)}{\nabla \nabla_s} I_t^{(T)}$.

2.2.3.5. TEMPORARY CHANGE (TC)

Also called “transitory change”, produces an immediate effect, ω , at moment T and then converges to 0 by an exponential decay defined by the decay parameter δ which takes values between 0 and 1. The transfer function is expressed by $v(B) = \frac{1}{1-\delta B}$, with $0 < \delta < 1$. If δ is close to zero, the TC behaves like an AO and if δ approaches 1 then the intervention transposes to a LS. Chen and Liu (1993) recommend $\delta = 0.7$ as a default value for identifying temporary changes.

2.2.3.6. RAMP SHIFT OUTLIER (RS)

Other types of outliers can be modelled by using the previous specifications. A ramp shift (Chen & Tiao, 1990) is a change in the trend in an ARIMA $(p, 1, q)$ model starting at $t = T$ until the end of the series. A time series affected by a RS is given by $y_t = \omega R_t^{(T)} + x_t$ where $R_t^{(T)} = (1-B)^{-1} S_t^{(T)}$ is a ramp function. The innovations affected by RS at $t = k$ are expressed by $a_t = e_t + \pi(B) \omega R_t^{(T)}$, where $\pi(B) = \theta(B)^{-1} \phi(B) (1-B)^d$.

2.2.4. OUTLIER DETECTION

Until the contribution of Chen and Liu (1993), earlier procedures for detecting outliers (*e.g.*, Chang et al., 1988; S. Hillmer, 1984; Tsay, 1988) faced notable limitations. These methods were susceptible to masking effects and exhibited inconsistencies

in the classification and location of outliers during different stages of the iterative detection process. Adding to these drawbacks, the previous procedures were not conceived to estimate jointly the outlier effects and the parameters of the specified model.

Chen and Liu (1993), following Tsay (1986), extended the procedure proposed by Chang et al., (1988) of an iterative procedure to detect IO and AO outliers, and included TC and (S)LS outliers as well, as to estimate simultaneously the effect of the outliers and the parameters of the model.

The sequence consists in 4 iterative stages (Box et al., 2015a, pp. 491–492; Chen & Liu, 1993, Morettin & Toloi, 2018, p. 318-321) where the outliers are identified sequentially:

- i. Estimate a model assuming there are no outliers present and collect the residuals by

$$\hat{e}_t = \frac{\hat{\phi}(B)}{\hat{\theta}(B)} Z_t, \quad (2-24)$$

where the variance of the white noise processes (a_t) is estimated by

$$\hat{\sigma}_a^2 = \frac{1}{N} \sum \hat{e}_t^2; \quad (2-25)$$

- ii. Compute $\hat{\lambda}_T = \max_t \max_j \{|\lambda_{j,t}|\}$, where T is the time point where observation is maximum and i is the type of outlier. If $\hat{\lambda}_T > C$, where C is a constant between 3 and 4, then the series has an outlier at time t of the type i and an estimated effect of $\hat{\omega}_{i,T}$. Letting $\pi(B) = \frac{\phi(B)}{\theta(B)}$ and $e_t = \pi(B)Y_t$, we get:

$$\text{For an additive outlier (AO)} \quad e_t = \omega \pi(B) I_t^{(T)} + a_t, \quad (2-26)$$

$$\text{for an innovative outlier (IO)} \quad e_t = \omega I_t^{(T)} + a_t, \quad (2-27)$$

$$\text{for a level shift (LS)} \quad e_t = \omega \frac{\pi(B)}{1-B} I_t^{(T)} + a_t \quad (2-28)$$

$$\text{for a temporary change (TC)} \quad e_t = \omega \frac{\pi(B)}{1-\delta B} I_t^{(T)} + a_t, \quad (2-29)$$

$$\text{for a seasonal level shift (SLS)} \quad e_t = \omega \frac{\pi(B)}{1-B^S} I_t^{(T)} + a_t \quad (2-30)$$

where S denotes the seasonal period.

The impact ω of an outlier (estimated by least squares) when T is known and all the parameters of an ARIMA model were previously estimated. Following the least square estimate of the effect of an outlier at time $t = t_1$ is expressed by:

$$\hat{\omega}_{AO}(t) = \frac{\sum_t^n \hat{e}_t x_{2t}}{\sum_t^n x_{2t}^2}, \quad (2-31)$$

$$\hat{\omega}_{IO}(t) = \hat{e}_t, \quad (2-32)$$

$$\hat{\omega}_{LS}(t) = \frac{\sum_{t=T}^n \hat{e}_t x_{3t}}{\sum_{t=T}^n x_{3t}^2}, \quad (2-33)$$

$$\hat{\omega}_{TC}(t) = \frac{\sum_{t=T}^n \hat{e}_t x_{4t}}{\sum_{t=T}^n x_{4t}^2}, \quad (2-34)$$

The series is then adjusted by removing the effect of the outlier $\tilde{Z}_t = Z_t - \hat{\omega}_{i,T} I_t^{(T)}$ and new errors are computed, and the variance is recalculated;

- iii. Recalculate $\hat{\lambda}_{i,t}$, $i = 1, 2, \dots, 5$, $t = 1, \dots, N$ and repeat the previous step until all the outliers are identified while maintaining the first model for estimation;
- iv. Estimate simultaneously ω_k for the i types of outliers through the model (Equation (2-36)). Steps (ii) to (iv) are repeated until all the outliers are identified, and their impacts are estimated.

2.3. INTERVENTION MODELLING

The modelling of an intervention effect in a time series requires specifying several elements: the type of intervention (*e.g.*, pulse, step, ramp), the timing of the intervention, the order of the transfer function numerator, and the order of the transfer function denominator. Two generic approaches are widely employed to evaluate the impacts of a given intervention. The first approach involves estimating the parameters of the transfer function model, evaluating their statistical significance (*e.g.* Newbold et al., 1976) and then model the residuals (Yaffee & McGee, 2000, p. 282). The second approach, proposed by Box and Tiao (1976) comprising three generic steps *i)* Estimate an appropriate stochastic model for the section of the time series prior to the intervention; *ii)* Forecast for the periods after the intervention with the previous model and compare with the actual values to establish the impact of the intervention; *iii)* Compute the functions of forecasting errors and conclude about the changes on the model.

In the context of the first approach, an intervention model can be expressed as:

$$F_t = Z_t + \frac{\omega(B)}{\delta(B)} I_t^{(T)}, \quad (2-35)$$

where F_t presents the observed time series, Z_t denotes the underlying time series unaffected by the intervention, and $I_t^{(T)}$ is the intervention variable that captures the effect of the intervention starting at time T . The term $\frac{\omega(B)}{\delta(B)}$, also expressed by $v(b)$ is the transfer function. Here $\omega(B)$ models the immediate impact of the intervention, and $\delta(B)$ captures the decay

pattern of this effect over time (*e.g.*, Wei, 2006, p. 215). The intervention effect is represented by ω that diminishes geometrically at a rate determined by δ . Binary variables I_t are used, where observations affected by the intervention are coded as "1," and unaffected observations are coded as "0."

After dropping the constant to address nonstationary, for multiple outliers/interventions following Abraham (1980), Box et al., p. (2015, p. 482), Box & Tiao (1975), Ismail et al. (2009), Montgomery & Weatherby (1980), the intervention model can be re-written as (Box et al., 2015, p. 491):

$$F_t = \sum_{i=1}^k \omega_i v_i(B) I_t^{(T_i)} + \frac{\theta(B)}{\phi(B)} a_t, \quad (2-36)$$

where F_t represents the observed time series, ω_i denotes the magnitude of the i -th outlier or intervention and $v_i(B)$ is the transfer function that models its dynamic pattern. $I_t^{(T_i)}$ is an indicator function representing the presence of the i -th outlier or intervention at time T_i . The ARIMA process $\frac{\theta(B)}{\phi(B)} a_t$ accounts for the residual structure of the series, with $\theta(B)$ and $\phi(B)$ as polynomials in the backshift operator B describing the autoregressive and moving average components, respectively, and a_t representing white noise. This model allows for the representation of multiple interventions or outliers, each contributing dynamically to the observed series based on its transfer function $v_i(B) = \frac{\omega_i(B)}{\delta_i(B)}$ where $\omega_i(B)$ models the immediate impact of the i -th outlier or intervention, and $\delta_i(B)$ captures the decay pattern of this effect over time. $I_t^{(T_i)}$ indicator functions allows that the timing of each outlier or intervention is explicitly accounted for.

The assessment of intervention impacts in time series can be defined based on the nature and duration of the intervention's effect. Most commonly, interventions fall into one of two categories: pulse functions and step functions, which have been extensively documented (Yaffee & McGee, 2000, Chapter 8; Wei, 2006, Chapter 10; Box et al., 2015, Chapter 13). Using the notation $I_t^{(T)}$ to represent the intervention series, where T denotes the intervention time index, different functional forms can be defined to model the intervention.

A Pulse series $P_t^{(T)}$ is used for an intervention that occurs in a single time t_0 and lasts for one single time point, that is:

$$I_t^{(T)} = P_t^{(T)} = \begin{cases} 1 & \text{if } t = T \\ 0 & \text{if } t \neq T \end{cases} \quad (2-37)$$

A Step series $S_t^{(T)}$ implies that the intervention continues to affect the time series after (T) introducing a step towards a new level:

$$I_t^{(T)} = S_t^{(T)} = \begin{cases} 1 & \text{if } t \geq T \\ 0 & \text{if } t < T \end{cases} \quad (2-38)$$

Additionally, a Ramp series $R_t^{(T)}$ that can be used to model interventions that introduce a gradual recovery or trajectory, often associated with changes in trend:

$$f(I_t) = I_t^{(T)} = R_t^{(T)} = \begin{cases} t - T + 1 & \text{if } t \geq T \\ 0 & \text{if } t < T \end{cases} \quad (2-39)$$

The indicator variable used to model the period of intervention can be formulated either as a step function or as a pulse function, *i.e.*, $P_t^{(T)} = S_t^{(T)} - S_{t-1}^{(T)}$ or, using the backshift notation, $P_t^{(T)} = (1 - B)S_t^{(T)}$.

Following an intervention, the time series may either stabilize at a new equilibrium state, representing a permanent change in level, or return to its pre-intervention state, indicating a temporary impact. When these outcomes are further classified by the dynamics of the interventions (abrupt or gradual), a matrix of four potential impact patterns emerges (Table 2-2)

Table 2-2: Impacts of interventions

	Permanent change	Temporary change
Gradual start	gradual-permanent	gradual-temporary
Abupt start	abrupt-permanent	abrupt-temporary

Adapted from (McDowall et al., 2011, p. 8)

McDowall et al., (2011, p. 8) argue that the gradual-temporary intervention lacks theoretical significance. They emphasize that each of the remaining types (abrupt-permanent, abrupt-temporary, and gradual-permanent) requires distinct nested sub-models of $v_i(B)I_t^{(Ti)}$. Moreover, a robust theoretical framework is essential to justify the nature and dynamics of the intervention being modelled.

2.4. DYNAMIC REGRESSION

Dynamic regression with ARIMA errors extends the linear regression framework by incorporating time series characteristics into the error term, which is modelled as an ARIMA process. A general dynamic regression model with ARIMA errors can be expressed as:

$$y_t = \alpha + \beta_1 x_{1,t} + \dots + \beta_k x_{k,t} + N_t \quad (2-40)$$

Where α is the intercept, $\beta_1, \dots, \beta_k$ are the regression coefficients for the regressors (outliers and interventions) and N_t represents the residuals, which are modelled as an ARIMA process:

$$\phi_p(B)\Phi_P(B^S)\nabla^d\nabla_S^D N_t = \theta_q(B)\Theta_Q(B^S)e_t, \quad (2-41)$$

where

B is the backshift operator,

$\Phi_p(B)$ is the non-seasonal autoregressive polynomial of order p ,

$\Phi_p(B^S)$ is the Seasonal autoregressive polynomial of order P ,

$\theta_q(B)$ is the non-seasonal moving average polynomial of order q ,

$\theta_Q(B^S)$ is the seasonal moving average polynomial of order Q ,

∇^d is the non-seasonal differencing operator,

∇_s^D is the seasonal differencing operator and

N_t is the error term from the regression model.

Solving (2-41) and (2-40) for y_t and adding the need to differentiate all regressors to the same order of integration of N_t , following Pankratz (1991, p. 119) we get:

$$\nabla^d \nabla_s^D y_t = \alpha + \beta_1 \nabla^d \nabla_s^D x_{1,t} + \dots + \beta_k \nabla^d \nabla_s^D x_{k,t} + \frac{\theta_q(B)\theta_Q(B^S)}{\nabla^d \nabla_s^D \Phi_p(B)\Phi_P(B^S)} e_t, \quad (2-42)$$

This means that every variable of the regression equation must have the same order of differentiation that is applied to N_t unless the variables are cointegrated, since N_t is going to be stationary in this case.

The parameters are estimated by maximum likelihood, namely by the conditional likelihood function described by Box et al., (1976).

2.5. FORECASTING AND ACCURACY METRICS

The models are compared using a recursive out-of-sample forecasting approach. In this method, the models forecast one-steps ahead, and the resulting forecast errors are recorded. After each forecast, the models are re-estimated, incorporating the newly available "known" values into the training sample. The updated models then generate another one-step forecast, repeating the process iteratively until the end of the sample is reached.

The accuracy metrics we used for in-sample and out-of-sample evaluation are the following:

Mean error (ME) to which provides the average of the residuals and indicates the direction of bias of the forecasts:

$$ME = \frac{1}{n} \sum_{t=1}^n (\hat{y}_t - y_t). \quad (2-43)$$

Root mean squared error (RMSE), which measures the average magnitude of forecast errors, with larger penalties for significant deviations:

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{n}}. \quad (2-44)$$

Mean absolute error (MAE), representing the mean of absolute deviations.

$$MAE = \frac{1}{n} \sum_{t=1}^n |\hat{y}_t - y_t|. \quad (2-45)$$

Mean percentage error (MPE), yielding the average percentage error of the forecasts.

$$MPE = \frac{1}{n} \sum_{t=1}^n \frac{\hat{y}_t - y_t}{y_t} 100. \quad (2-46)$$

Mean absolute percentage error (MAPE), which quantifies forecast accuracy as a percentage.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{\hat{y}_t - y_t}{y_t} \right| 100. \quad (2-47)$$

Mean absolute scaled error (MAE), which scales errors relative to a naïve forecasting model.

$$MAE = \frac{\frac{\sum_{t=1}^n |\hat{y}_t - y_t|}{n}}{\frac{1}{n-1} \sum_{t=2}^n |\hat{y}_t - y_{t-1}|}. \quad (2-48)$$

Where $\hat{y}_t$ is the predicted value, y_t is the actual value and n is the number of observations.

The Diebold-Mariano test (Diebold & Mariano, 1995) is used to compare the out-of-sample forecast accuracy of the selected models. By computing the forecast errors over a recursive forecast horizon, the DM test evaluates whether differences in accuracy are statistically significant. The test relies on the loss differential, d_t , which represents the difference between the loss functions of the two models for each forecast period t .

$$d_t = (g(e_{1,t}) - g(e_{2,t})), \quad (2-49)$$

Where $e_{1,t}$ and $e_{2,t}$ denote the forecast errors from models 1 and 2, respectively and $g(\cdot)$ is the loss function applied to the forecast errors.

3. METHODOLOGY

This study uses a single-case study framework combined with time series analysis, adopting a quantitative approach. Building on Myers et al., (2011), case studies are particularly suitable be used *“for the exploration, classification and hypothesis development stages of the knowledge building process”*.

The research follows three generic phases: description, exploration, and explanation. The unit of analysis is the number of passengers that were processed by the main Portuguese airports. Due to the specificity of the data series, the generalization potential of the findings is limited.

The dissertation follows with a detailed description of the data, outlining its sources, coverage and limitations. This is followed by the processes of data selection and cleaning. After generating the time series, descriptive statistics are discussed to summarize the main characteristics of the data.

The analysis then focuses on identifying and examining the seasonal pattern, performing data transformation, and decomposition of the time series to complete the description of its key features.

Three seasonal ARIMA models without external regressors are initially estimated. Two models are manually specified based on an analysis of the autocorrelation function (ACF), partial autocorrelation function (PACF), and extended autocorrelation function (EACF). The third model is automatically selected using the *auto.arima* function from the forecast R package.

Following the estimation of these initial models, calendar effects, outliers, structural breaks, and interventions are evaluated using the significance of coefficients and information criteria. For each scenario, multiple models are tested to assess the inclusion of external regressors and their potential role in subsequent analyses.

All automatically selected models are estimated with consistent parameter settings: *stepwise = FALSE, approximation = FALSE, lambda = 0, and biasadj = TRUE*.

The assumption of data independence inherent in linear regression is violated in the context of time series data due to its autoregressive nature. To address the serial dependence, regressions with ARIMA errors are estimated following the framework described by R. J. Hyndman & Athanasopoulos (2018). The selection of external regressors involves evaluating multicollinearity using the Variance Inflation Factor (VIF) for each set of candidate variables. Additionally, informed judgment is applied to ensure the suitability and relevance of the chosen regressors for each analysis.

As an alternative, a linear regression model (*tslm*) with harmonic components is implemented to capture cyclical patterns in the data, explicitly modelling seasonal and periodic fluctuations

and a selection of regressors. This approach was adopted because achieving a good fit with lagged and external regressors proved unsuccessful.

Competing models are evaluated during the model selection phase using criteria such as information measures, residual diagnostics (including tests for autocorrelation, heteroskedasticity, normality, and stationarity), and parsimony to guide the selection of the final models.

In the final section before discussion and conclusions, forecasting is performed both in-sample and out-of-sample to evaluate model accuracy and predictive performance employing a recursive, one-step-ahead forecasting method, as outlined by R. J. Hyndman & Athanasopoulos (2018).

4. DATA DESCRIPTION, DATA SELECTION AND DATA CLEANING

The complete dataset is distributed across two databases, one containing detailed records of flights that departed from or arrived at Portuguese airports since the 1990s and a secondary database that account for less than 0.1% of all the passenger processed (Appendix A).

4.1. DATA SOURCE AND DESCRIPTION

Most flight reporting forms (“Formulários de Tráfego”; see Annex A) are submitted by airlines or the ground handling companies responsible for providing assistance and are finalized by airport management services, with one form submitted per flight movement.

The flight report data, which encompasses all flights landing at or departing from an airport, is submitted monthly to the Autoridade Nacional da Aviação Civil (ANAC). This independent regulatory agency is responsible for creating regulations, supervising compliance, and enforcing civil aviation rules. ANAC compiles and stores this data in two separate databases: one for the main airports (20 as of 2018) and another for small aerodromes (19 as of 2018), with the latter containing only a limited subset of the variables reported by the main airports. As of 2018, ANA–Aeroportos de Portugal, SA. provides data regarding the air traffic of the airports localized in Lisboa, Porto, Faro, Funchal, Ponta Delgada, Terceira, Horta, Porto Santo, Santa Maria, Flores and Beja and Sata Aerodromos, SA. provides data regarding the air traffic of the airports localized at Pico, Graciosa, Corvo and São Jorge. The remaining airports are managed by other entities, mostly municipalities

All major airports provide high-quality data, tested through regular statistical audits conducted by ANAC to ensure completeness, accuracy, and validity. Due to the high automation of airport processes, the occurrence of a flight without the generation of the corresponding flight form is considered unlikely.

In contrast, reporting errors are more common at small airports, primarily due to less robust data quality processes, limited information technology resources, and manual data entry errors. These small aerodromes, which are mostly managed by municipalities, host aviation activities such as pilot training flights, aerial work, and search and rescue operations. These activities are typically non-commercial and account for only a residual number of passengers processed.

Due to the less demanding reporting requirements for the smaller database—where the identification of the origin and destination is not required—these aerodromes are able to report their movements in a simplified form but fail to fully capture their activities according to the standards of the main database.

By comparing the total number of passengers processed in both databases, most passengers are reported in the main database (Appendix B). For instance, in 2018, the main database

accounted for approximately 99.92% of all passengers processed by national airports. Of these, Lisbon alone was responsible for processing around half of all passengers.

Analyzing data from the main database from 2004 to 2018, the three airports with the highest passenger volumes consistently accounted for more than 80% of the total across all years in the series, ranging from a minimum of 81.8% in 2005 to a maximum of 87.4% in 2017.

The main database granularity of data is at individual flight level and time granularity is set at one minute. The main database contains 40 unique variables that are exported to an aggregated database with 70 variables with some of the variable being linear transformations of others. Other interesting variables and traffic indicators could be easily generated by including external information. For instance, from these variables we could compute the Vicenty distances (Vincenty, 1975) flown by adding latitude and longitude data for each airport. CO₂ estimations per flight could also be performed by relating the already existing variables to the tables published by ICAO and UNFCCC.

4.2. DATA SELECTION AND UNIVARIATE TIME SERIES GENERATION

A univariate time series was computed, representing the number of processed passengers at the airports, irrespective of their origin. A key characteristic of this variable is that it does not reflect the actual number of passengers who travelled by air, due to the well-known issue of double counting. For counting purposes, a passenger traveling between two domestic airports should ideally be counted only once. However, both the origin and destination airports report the same passenger, with the origin recording them as embarking and the destination as disembarking. This double counting often leads to misreporting in the media, as the total number of passengers processed by airports is not equivalent to the actual number of passengers who travelled. Nonetheless, both perspectives are valid: the airport's perspective reflects the utilization of its infrastructure, while the airline's perspective represents the actual *number of passengers*.

The variable of interest "Passengers Processed" is created from the sum of the variables that relate to passengers embarked and passenger disembarked, *i.e.*:

Passengers = embarked local + embarked exempt + embarked in transfer + disembarked direct + disembarked local + disembarked exempt + disembarked in transfer.

4.3. DATA CLEANING

We started by analysing the daily time series of all the airports for discordant observations to minimize the masking and swamping effect of outliers. Anomalies were found on several data points. There was an abnormal value recorded on a single flight from 16 February 2009 with origin in Corvo and destination Lajes reporting 3143 passengers. Seeing we were unable to find the true value, the discordant value was rejected, becoming a missing value problem. For its imputation, seeing we are dealing with a missing single value among millions of

observations, in a missing completely at random mechanism⁶ (MCAR), we used a hot deck method that has the advantage of being computationally simple.

Hot deck procedures are used to impute missing values, edit data or outlier removal where missing data is replaced by similar records from the same dataset. The candidate records to provide imputation values are designated as “donors” and the records to be replaced by the donors are designated “recipients” (Ford, 1983). One possible hot deck method is the nearest neighbour imputation that uses the item value of the nearest record to fill in the missing information. Considering that air traffic has a weekly seasonal pattern and flights within the Autonomous Regions of Azores and Madeira are mostly scheduled flights⁷, we chose for the donor the number of passengers from a flight scheduled for the following week at approximately the same time and for the same origin, destination, and license plate, resulting in 4 passengers instead of 3143. The donor record is not unusual as its values are within the typical distribution of values from that month and airport.

In some smaller airports we observed that the series had some days without arrivals and departures due to weather constraints or due to the absence of demand. Even though an aggregated monthly time series is unaffected by the presence of a few daily gaps in a few airports, the daily time series were “padded” by using the *padr* R package. Consequently, the value “0” was imputed to the days without any flight record in given airport. Additionally, some airports did not operate for the entire period from 2004 to 2018, meaning their start dates are later than the beginning of 2004. These periods without data were coded as *missing*. The screening process is depicted in Appendix A: Data collection and screening.

After extraction, the dataset was cleaned by harmonizing categorical variables and replacing Latin characters to avoid conflicts during the R coding phases. At this stage the dataset was aggregated to a monthly level.

⁶ Following Richards, Little and Rubin (1989, pp. 13–17) taxonomy, in a MCAR missingness mechanism, the causes of the missing values are not related to the data, *i.e.* the probability of a missing value is the same across all values.

⁷ A scheduled air service is “an air service open to use by the general public and operated according to a published timetable or with such a regular frequency that it constitutes an easily recognizable systematic series of flights” (ICAO & International Civil Aviation Organization, 2016, p. 194).

5. ANALYSIS OF AIR PASSENGERS' THROUGHPUT TIME SERIES

In this section, we provide a description of the monthly air passenger throughput time series, focusing on its key structural characteristics.

5.1. TIME SERIES OVERVIEW

The following figure presents the time series, along with the rolling mean, standard deviation bands, and a polynomial regression line.

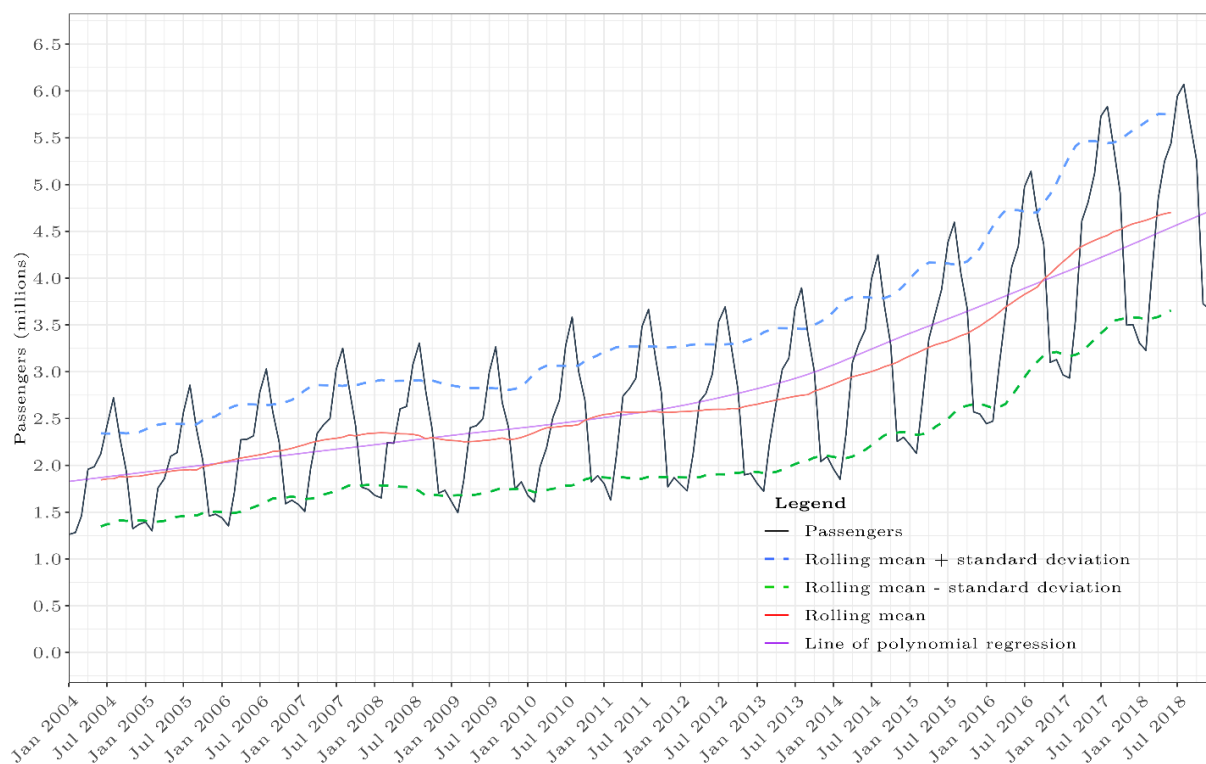


Figure 5-1: Passenger time series, rolling mean and line of polynomial regression.

This figure reveals a pronounced seasonal pattern, with the rolling mean (displayed in red) capturing the central tendency and showing a general upward trend, indicative of growing passenger figures over time.

The rolling mean plus and minus one standard deviation bands, shown as blue and green dashed lines, respectively, offer a measure of variability around the mean. These bands widen over time, suggesting an increase in variability, implying that growth in air travel demand is accompanied by larger fluctuations in monthly passenger volumes. The polynomial regression line (in purple) further emphasizes the long-term upward trend in passenger volumes.

This pattern suggests that subsequent modelling approaches should account for the trend, seasonality and heteroscedasticity to accurately capture the underlying dynamics of the time series.

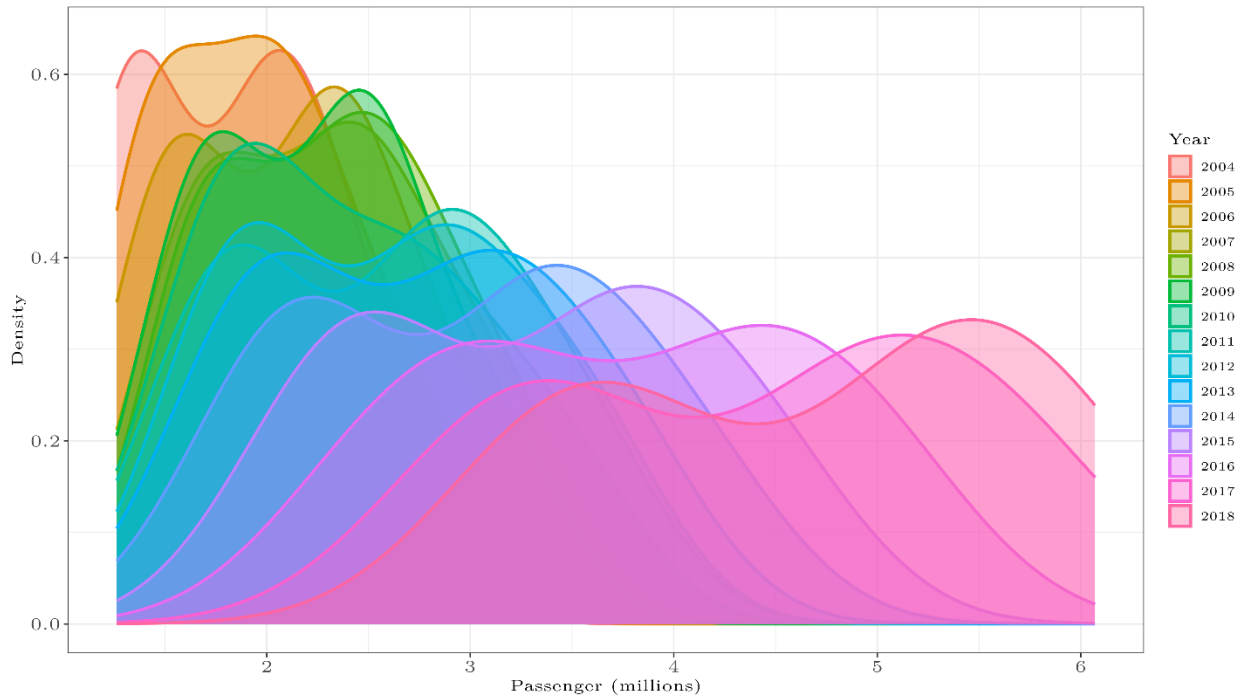


Figure 5-2: Kernel density estimate by year.

The density plot illustrates a clear rightward shift over time, signifying a steady increase in passenger volumes. The peak density values transition from lower passenger counts in the earlier years (2004-2008) to significantly higher counts in the later years, especially from 2015 onward. Additionally, the distribution's spread widens in recent years, reflecting more pronounced seasonal peaks as travel demand continues to grow.

5.2. DESCRIPTIVE STATISTICS

The mean number of passengers rose from approximately 1.84 million in 2004 to over 4.7 million by 2018. A similar upward trend is seen in the median values, emphasising the general growth in passenger numbers across the years. In conjunction with the rising mean, there is an increase in the standard deviation, which grew from 497,226 in 2004 to over 1 million by 2018. As shown before, this increase in dispersion reflects greater fluctuations in passenger numbers, suggesting that while the overall trend is upward, the magnitude of changes from month to month became more pronounced.

Table 5-1: Summary of passenger data statistics: 2004 – 2018.

Year	Mean	Median	std dev	Minimum	Maximum	Skewness	Kurtosis	Quartile 25	Quartile 75
2004	1 844 315	1 947 807	497 226	1 262 284	2 721 645	0,234	-1,197	1 357 414	2 163 143
2005	1 946 191	1 957 241	495 845	1 300 387	2 856 501	0,336	-0,953	1 473 257	2 200 937
2006	2 101 528	2 262 469	548 843	1 352 597	3 029 547	0,132	-1,187	1 618 135	2 379 293
2007	2 281 252	2 385 241	574 112	1 508 044	3 252 270	0,190	-1,132	1 760 009	2 581 087
2008	2 336 363	2 314 791	566 999	1 652 357	3 307 508	0,197	-1,138	1 725 626	2 668 572
2009	2 264 787	2 403 082	562 260	1 494 329	3 265 952	0,238	-1,024	1 804 459	2 542 100
2010	2 412 625	2 350 737	652 144	1 608 911	3 585 474	0,401	-1,066	1 873 381	2 772 654
2011	2 565 519	2 759 295	706 255	1 628 365	3 666 159	0,061	-1,352	1 852 427	2 992 092
2012	2 598 075	2 726 641	694 714	1 727 495	3 692 885	0,151	-1,336	1 909 741	3 043 783
2013	2 724 726	2 831 228	740 142	1 723 212	3 896 225	0,099	-1,314	2 079 742	3 205 693
2014	2 980 949	3 193 158	815 376	1 851 777	4 250 146	0,024	-1,348	2 288 608	3 511 983
2015	3 307 846	3 480 677	854 462	2 127 054	4 597 115	0,005	-1,386	2 565 243	3 922 850
2016	3 783 180	3 863 713	942 642	2 444 773	5 141 260	-0,042	-1,354	3 086 790	4 437 358
2017	4 403 984	4 712 038	1 059 194	2 932 456	5 833 507	-0,120	-1,473	3 502 839	5 185 897
2018	4 702 322	5 051 316	1 047 181	3 228 658	6 070 780	-0,181	-1,524	3 711 725	5 492 064
All years	2 816 911	2 587 237	1 102 461	1 262 284	6 070 780	0,956	0,405	1 952 736	3 360 072

Regarding skewness, early years show slight positive skewness. However, by 2016, the skewness shifts to negative, indicating that the distribution becomes more symmetric. Moreover, the kurtosis values across all years are negative, indicating a platykurtic distribution. This reflects a flatter distribution compared to a normal distribution, with fewer extreme values in the tails. This means that while variability increases, the frequency of outliers remains relatively low compared to a normal distribution. As a result, the dataset demonstrates a greater level of complexity than the widely studied *AirPassengers* dataset. When comparing the entropy, which reflects the spectral density and randomness of the data, the *AirPassengers* dataset presents an entropy value of 0.29, while our dataset exhibits a higher entropy value of 0.392. This difference indicates a greater degree of randomness and a less structured pattern in our dataset, making it more challenging to forecast.

5.3. SEASONALITY AND AUTOCORRELATION

The figure below illustrates the monthly distribution of air passenger volumes across the observed period, displaying both variability within each month and the pronounced seasonality, with passenger volumes significantly increasing during the summer months, peaking in July and August. The median passenger count, as well as the interquartile range, expands in these peak months, indicating a higher volume and variability of air travel during summer. In contrast, the winter months, especially January and February, show lower medians and narrower interquartile ranges, reflecting reduced travel demand and more stable passenger volumes during this period. The presence of outliers in most months suggests occasional deviations from typical seasonal patterns.

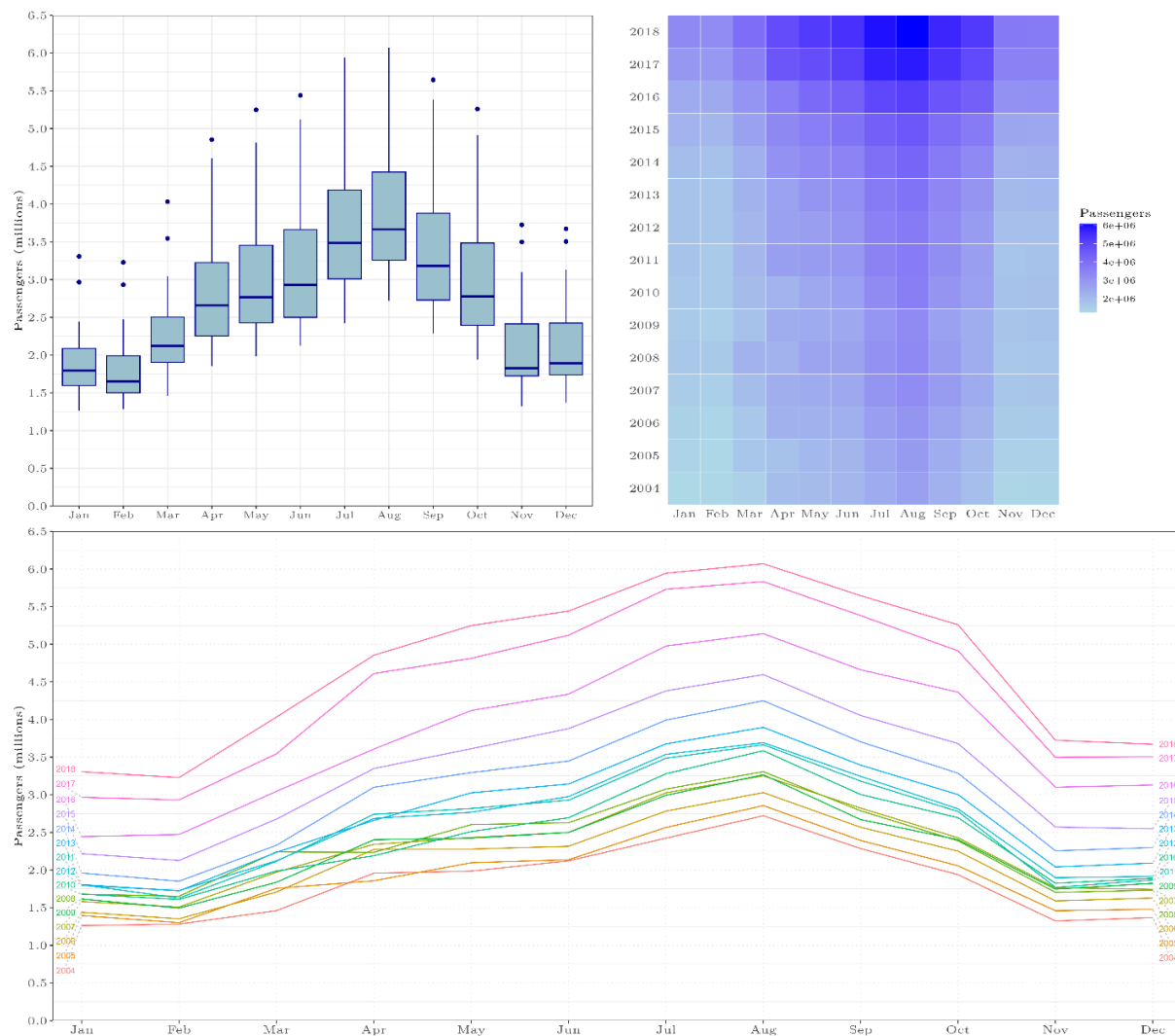


Figure 5-3: Monthly distribution of passengers and seasonal variation.

While air passenger traffic continues to peak during the summer months of July and August, there has been a noticeable increase in passenger volume throughout the entire year over time. The seasonal plot also shows a clear upward trend across all months, indicating sustained growth in passenger volumes across the entire calendar.

The heatmap points to a shift in the seasonal patterns. In the earlier years, particularly from 2004 to around 2014, the distribution of higher values was more evenly spread throughout the year. Starting around 2015, there is a distinct change in this seasonality. The summer season (June to September) begins to show a much stronger concentration of higher values, with a notable expansion of this high-value period extending into more months. By the time we reach the years 2016 to 2018, the intense activity is evident from June through September, indicating that the high-value season is not only more pronounced but also longer in duration. This elongation of the summer season, with sustained high values, suggests a change in the seasonality pattern. Results from SEATS (Seasonal Extraction in ARIMA Time Series) method (Bee Dagum & Bianconcini, 2016) reveal, for instance, a seasonality index of 2005:08=1,447

whereas for 2018:08 it had declined to 1,298⁸. This shift suggests a broadening of travel demand, likely influenced by off-season travel, lower air fares and the liberalization of the airline market, which have made flights more accessible year-round.

An analysis of the autocorrelation structure at lags up to 20 (Figure 5-4) indicates that the most prominent correlations occur at lags 1, 11, 12, and 13. The significant correlations at lags 11, 12, and 13 highlights a pronounced yearly seasonality, while the high correlation at lag 1 it's a strong indication of an autoregressive component of order 1.

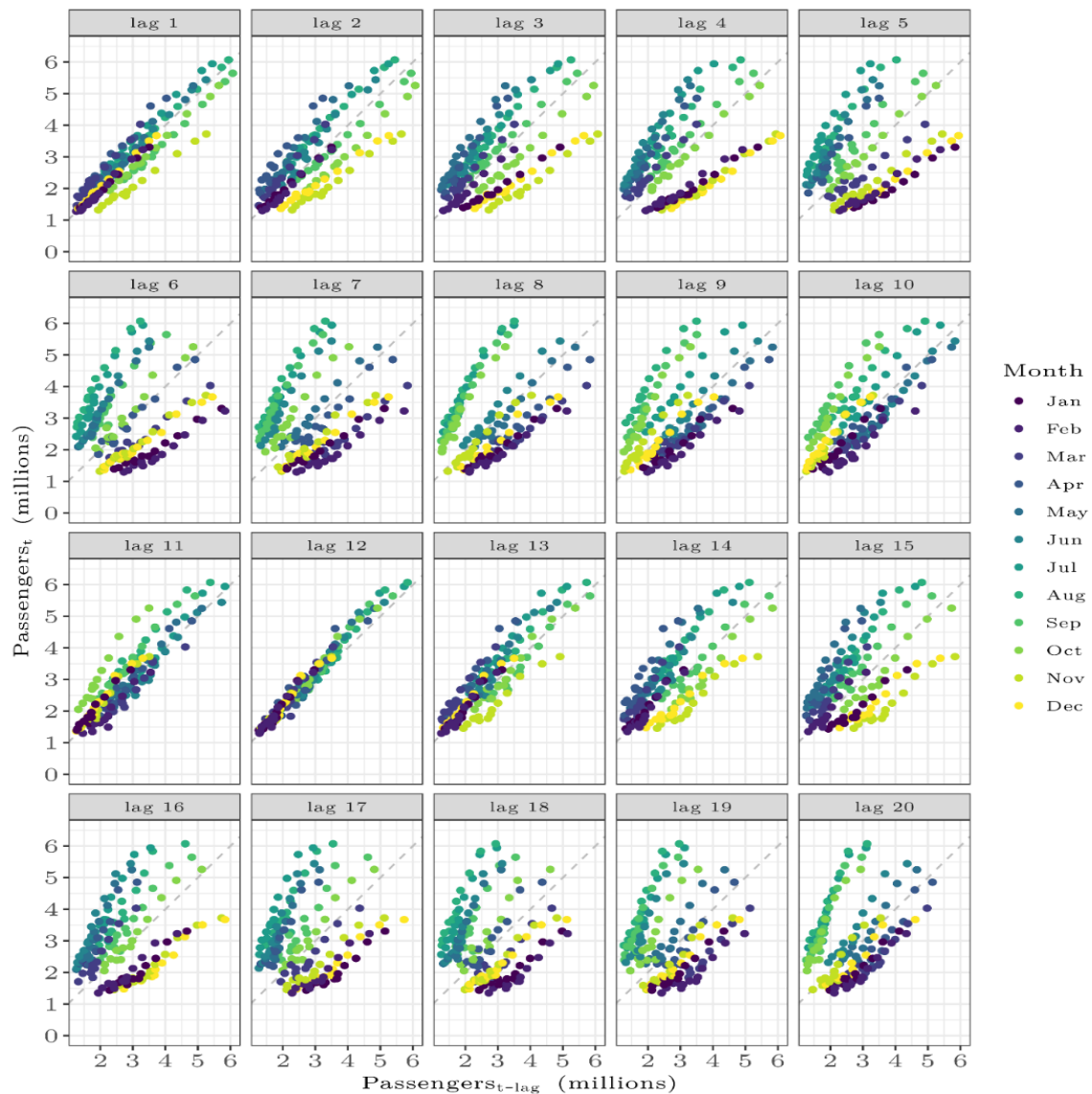


Figure 5-4: Autocorrelation at lags 1 to 20.

⁸ The estimate of the seasonal component using the X-11 method are similar: 2005:08=1,45; 2018:08=1,314.

Building on the prior analysis that pointed to the presence of yearly seasonality, the Autocorrelation Function (ACF) plot further reinforces the presence of this annual pattern.

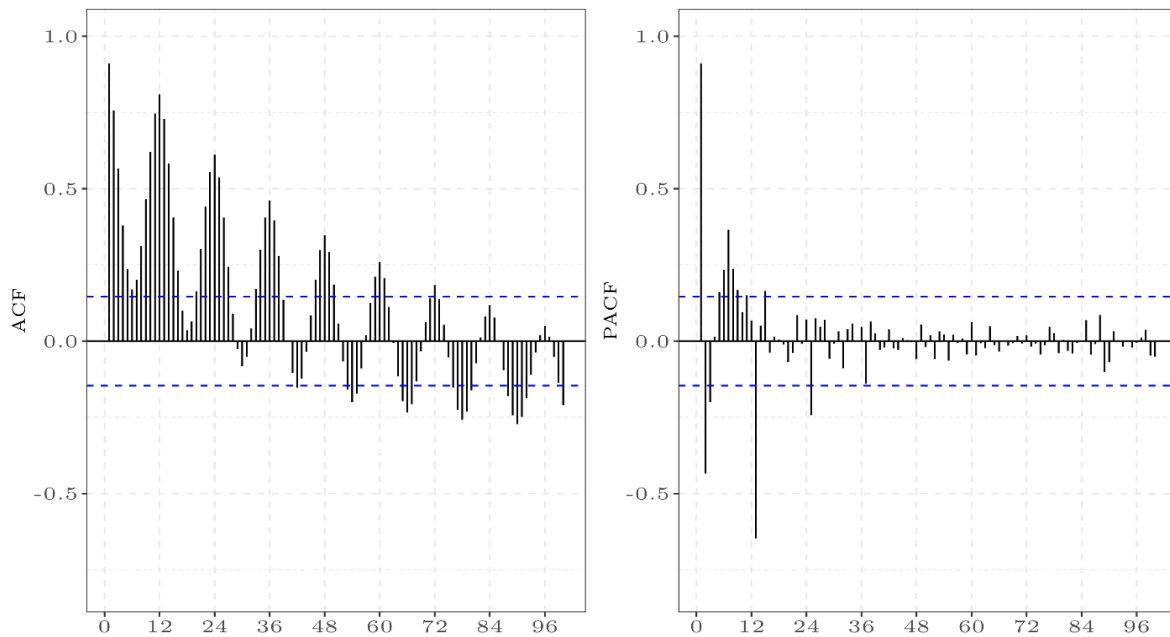


Figure 5-5: Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of untransformed data.

There are regular spikes at intervals that are multiples of 12, indicating a seasonal component with a period of 12. Additionally, the fact that the ACF starts with a strong positive autocorrelation near lag 0, which decays slowly over time, indicates the presence of a trend and the non-stationarity of the data, which was already evident from the plot of the series. The combination of the slow decay in the ACF and a spike at lag 1 in the PACF is also a strong indication of a trend in the series.

Finally, we applied a robust procedure for testing the significance of autocorrelation in univariate data, as outlined by Dalla, Giraitis, and Phillips (2022). The test was done using the first-order difference and seasonal difference with a lag of 12 of the log-transformed time series. In this case, a notable discrepancy was observed between the significant lags indicated by the standard t-test and the robust $\tilde{t}$ -statistic (Table 5-2).

Table 5-2: Autocorrelation and robust autocorrelation tests.

Lag	AC	Stand. CB(95%)	Robust CB(95%)	t	p-value (t)	t-tilde	p-value (t-tilde)
1	-0,527	(-0.152, 0.152)	(-0.290, 0.290)	-6,810	0,000	-3,556	0,000
2	0,070	(-0.152, 0.152)	(-0.149, 0.149)	0,907	0,364	0,924	0,355
3	-0,027	(-0.152, 0.152)	(-0.103, 0.103)	-0,353	0,724	-0,520	0,603
4	0,048	(-0.152, 0.152)	(-0.082, 0.082)	0,624	0,533	1,159	0,247
5	-0,059	(-0.152, 0.152)	(-0.089, 0.089)	-0,756	0,450	-1,283	0,199
6	0,043	(-0.152, 0.152)	(-0.091, 0.091)	0,555	0,579	0,920	0,358
7	-0,028	(-0.152, 0.152)	(-0.090, 0.090)	-0,357	0,721	-0,601	0,548
8	0,035	(-0.152, 0.152)	(-0.071, 0.071)	0,455	0,649	0,976	0,329
9	-0,003	(-0.152, 0.152)	(-0.095, 0.095)	-0,039	0,969	-0,062	0,951
10	-0,043	(-0.152, 0.152)	(-0.115, 0.115)	-0,558	0,577	-0,737	0,461
11	0,319	(-0.152, 0.152)	(-0.214, 0.214)	4,126	0,000	2,931	0,003
12	-0,559	(-0.152, 0.152)	(-0.349, 0.349)	-7,228	0,000	-3,141	0,002
13	0,319	(-0.152, 0.152)	(-0.251, 0.251)	4,119	0,000	2,487	0,013
14	-0,028	(-0.152, 0.152)	(-0.136, 0.136)	-0,367	0,714	-0,409	0,682
15	-0,002	(-0.152, 0.152)	(-0.080, 0.080)	-0,026	0,979	-0,050	0,960
16	-0,068	(-0.152, 0.152)	(-0.073, 0.073)	-0,884	0,376	-1,849	0,064
17	0,085	(-0.152, 0.152)	(-0.078, 0.078)	1,099	0,272	2,137	0,033
18	-0,067	(-0.152, 0.152)	(-0.084, 0.084)	-0,869	0,385	-1,561	0,119
19	0,042	(-0.152, 0.152)	(-0.102, 0.102)	0,540	0,589	0,803	0,422
20	-0,056	(-0.152, 0.152)	(-0.071, 0.071)	-0,728	0,467	-1,563	0,118
21	0,028	(-0.152, 0.152)	(-0.078, 0.078)	0,358	0,720	0,700	0,484
22	-0,024	(-0.152, 0.152)	(-0.104, 0.104)	-0,304	0,761	-0,445	0,656
23	0,001	(-0.152, 0.152)	(-0.168, 0.168)	0,015	0,988	0,013	0,989
24	0,024	(-0.152, 0.152)	(-0.288, 0.288)	0,315	0,753	0,166	0,868

While the cumulative Ljung-Box test at lag 17 is still significant ($p\text{-value} < 0.001$), the PACF at lag 17 is not and a visual inspection of the original time series supports the absence of autocorrelation at lag=17 (Figure 5-4). This suggests that the autocorrelation at lag 17 may be isolated, rather than part of a broader autocorrelation structure requiring modelling through the ARIMA framework.

5.4. SPECTRAL DENSITY AND PERIODOGRAM WITH STATIONARY DATA

Given a time series the spectral density can be estimated by replacing $\gamma(\cdot)$ by the sample autocovariance $\hat{\gamma}(\cdot)$ in the equation:

$$f(v) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i v h} \quad (5-1)$$

where $\gamma(h)$ represents the autocovariance at lag h , v is the frequency at which the spectral density is evaluated and $vh = k/n$ are Fourier frequencies where $k = 0, 1, 2 \dots \lfloor n/2 \rfloor$ for a time series of length n .

Spectral analysis enables the identification of sine and cosine functions of frequencies (*i.e.*, the number of cycles per unit of time) with the time series values. Larger correlations mean an underlying stronger periodicity of that period. The periodogram is “*a measure of the relative power of a time series as a function of frequency*” (Schuster, 1898) *apud* (Hernandez, 1999) *i.e.*, it enables to identify the dominant frequencies of a time series.

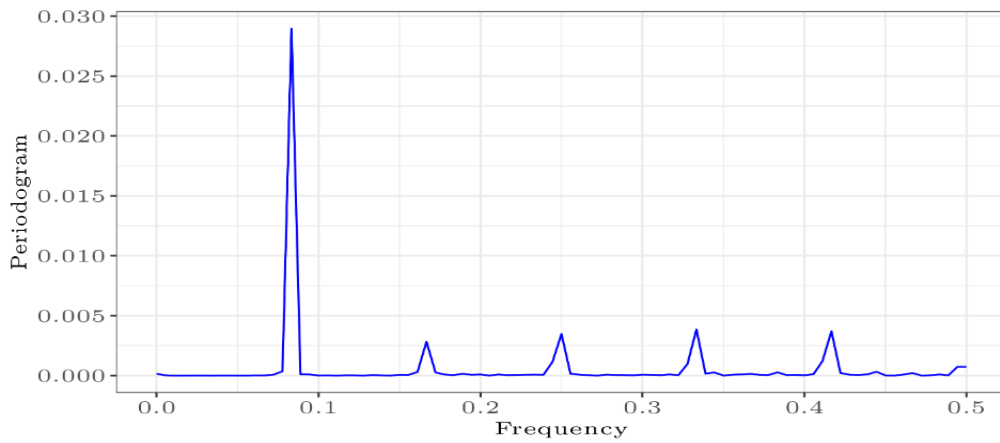


Figure 5-6: Periodogram of first differenced, log-transformed series.

For the periodogram presented in (Figure 5-6), we used the log-transformed and first-differenced series to emphasize the seasonal component by removing the trend. We then calculated the Fast Fourier Transform (FFT) of the detrended series, where the frequency values were normalized by the length of the original series and the power spectrum was calculated as the squared magnitude of the FFT results. To locate the dominant period, we identified the largest estimates of the periodogram (P) corresponding to the frequencies (f) at which the spectral density was estimated, up to $N/2$. The inverse of the largest harmonic frequency indicates the dominant period. Since $\max(P) \approx 0,029$ and $f \approx 0,0833$ then $1/f = 12$. As anticipated from the visual inspection, the periodogram confirms that it takes 12 months to complete a cycle. This also means that most of the variance of the passenger time series is described by long periods, *i.e.*, low frequencies.

5.5. VARIANCE STABILITY AND DATA TRANSFORMATION

In order to establish the need for a transformation in the data, we started by addressing the variance stability. Time series often exhibit heteroscedasticity which can lead to biased parameter estimates and poor model selection (e.g., Box et al., 2015, p. 197). A widely used method to stabilize variance is the Box-Cox transformation, a power transformation, where the parameter, λ determines the appropriate degree of transformation.

Using the Guerrero method, the optimal Box-Cox parameter was estimated as $\lambda = 0,11$. Since this value is close to zero, a logarithmic transformation was deemed suitable and preferable after comparison. Our choice is supported by the observation that when a time series exhibits increased dispersion associated with its level (where higher values correspond to greater variation, indicating heteroscedasticity) a logarithmic transformation stabilizes the variance, resulting in a series with more consistent variability. Furthermore, when a series exhibits an exponential trend, applying a log transformation converts it into a linear time trend, which can then be modeled as a stationary process by taking first differences (Cryer & Chan, 2008, pp. 98–99). This transformation also enhances interpretability, as changes in the logged series can be explained in relative terms.

5.6. DECOMPOSITION

For the decomposition of the log-transformed time series, we applied the MSTL function from the *forecast* package, which is particularly useful when dealing with complex seasonal patterns, as it adapts to various periodicities present in the data. Additionally, we tested the decomposition using the *seasonal* package, which uses the X-13 Arima-Seats method for seasonal adjustment. Both approaches yielded similar results

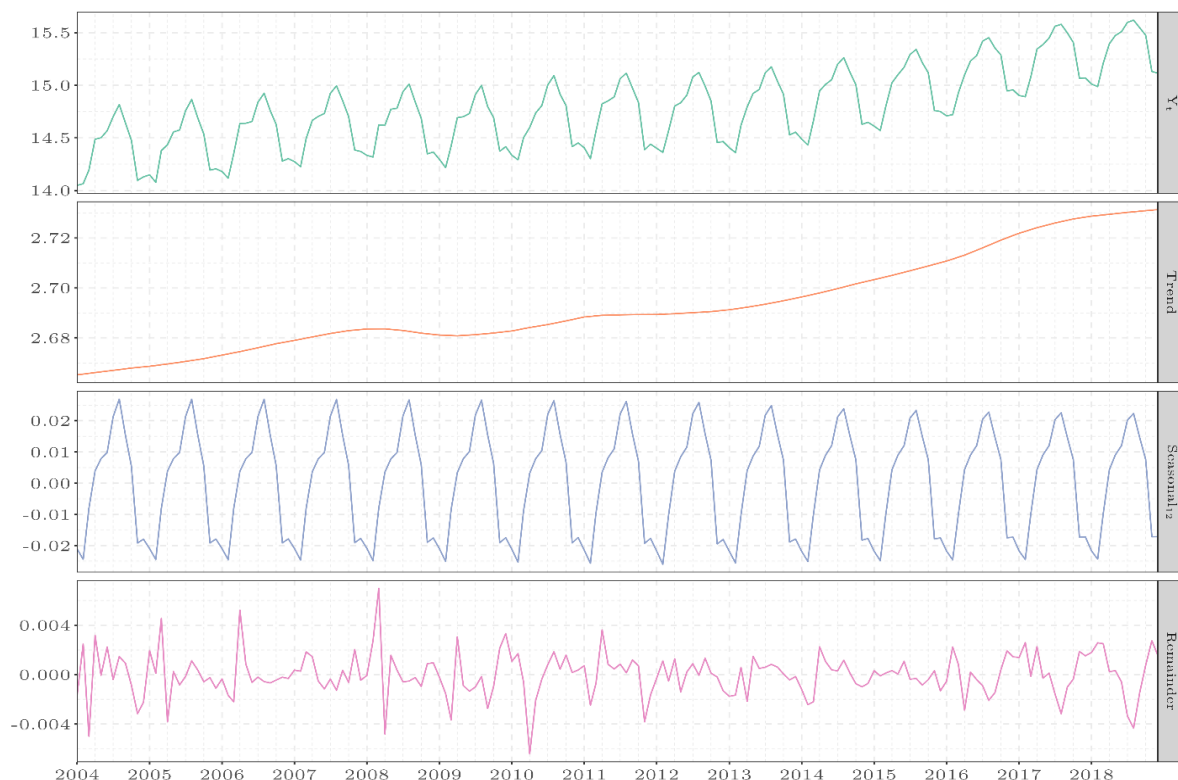


Figure 5-7: Decomposition of log-transformed series into trend, seasonal, and residual components.

The trend in the log-transformed series shows a long-term upward movement. This indicates a consistent growth in passenger numbers over time and the seasonal component in the decomposition displays a regular, recurring pattern, as expected. The remainder exhibits some spikes, which might indicate outliers not accounted for by the decomposition model.

5.7. STATIONARITY

A visual inspection of the time series plots for air passengers reveals the typical features of non-stationarity, with a clear upward trend and seasonal fluctuations (Figure 5-8).

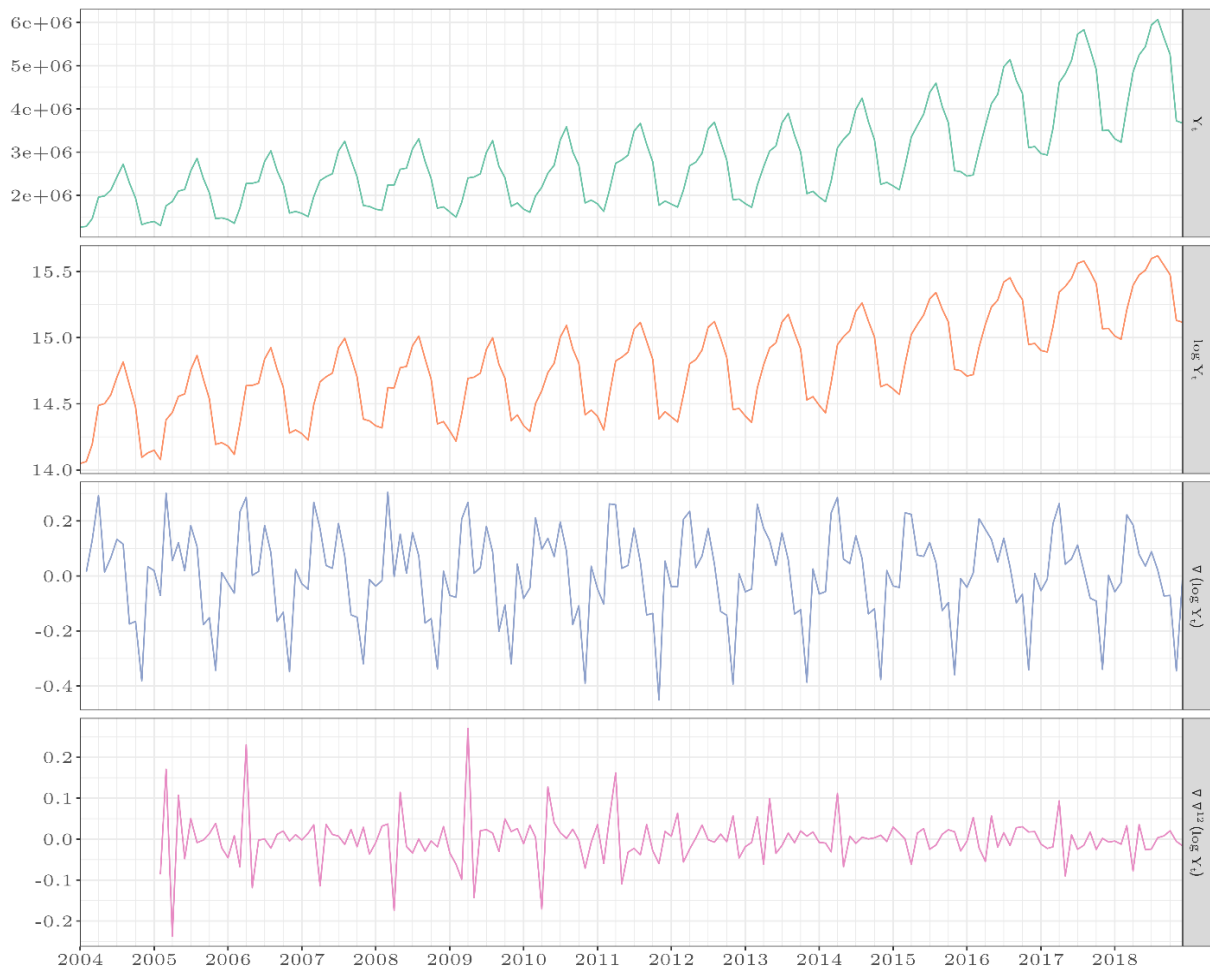


Figure 5-8: Transformation and differencing of data: Original, log-transformation, first differencing, and seasonal differencing.

After applying the logarithmic transformation to the original series, the reduction in the amplitude of the seasonal component suggests that the transformation handled most of the heteroscedasticity. However, the trend persisted, indicating the need for non-seasonal differencing. The first differencing of the log-transformed series successfully removed the trend, improving stationarity while preserving the seasonal structure, which is typical of air passenger series. To address the seasonality related to recurring fluctuations, we then applied a seasonal difference, resulting in the final chart that resembles white noise.

The visual inspection suggests that non-seasonal and seasonal differencing of the log-transformed series is sufficient to achieve stationarity.

To complement the visual inspection, we conducted formal tests to determine the appropriate orders of integration (Table 5-3).

Table 5-3: Results of stationarity and unit root tests for log-transformed and differenced series.

		Log-transformed	First differenced (log)	Seasonally differenced (log)	Regular and seasonal differenced (log)
KPSS (level) - (stationary around a constant mean) ($\alpha = 0,05$)	statistic	2,384	0,014	0,994	0,037
	p-value	<0,01 (**)	<0,1 (ns)	<0,01 (**)	<0,1 (ns)
KPSS (trend) - (stationary around a deterministic linear trend) ($\alpha = 0,05$)	statistic	0,132	0,011	0,283	0,036
	p-value	0,08 (ns)	<0,1 (ns)	<0,01 (**)	<0,1 (ns)
Augmented Dickey-Fuller (ADF) ($\alpha = 0,05$)	statistic	-8,263	-12,317	-2,140	-7,053
	p-value	<0,01 (**)	<0,01 (**)	0,52 (ns)	<0,01 (**)
Phillips-Perron (PP)($\alpha = 0,05$)	statistic	-59,584	-118,630	-92,651	-221,340
	p-value	<0,01 (**)	<0,01 (**)	<0,01 (**)	<0,01 (**)

Legend: $p < 0,001$ (***); $0,001 \leq p < 0,01$ (**); $0,01 \leq p < 0,05$ (*); $p \geq 0,05$ (ns)

The KPSS test evaluates whether a series is stationary under two assumptions: stationarity around a constant mean (KPSS Level) and stationarity around a deterministic linear trend (KPSS Trend). As expected, the log-transformed series has a notable trend a seasonality resulting the rejection of the null hypothesis of stationarity around a constant mean. Regarding trend, the KPSS Trend test fails to reject the null hypothesis of stationarity around a deterministic trend, suggesting that after accounting for the trend, it has constant variance and no unit root.

After applying first differencing to the log-transformed series, the KPSS Level test produced a statistic of 0,014 with a p-value of less than 0,1. This result fails to reject the null hypothesis, indicating that the series has achieved stationarity following the differencing process. This conclusion is further supported by the ADF test, confirming stationarity. The KPSS Trend test also fails to reject stationarity reinforcing the conclusion that the differenced series is stationary.

For the seasonally differenced series, the KPSS Level test produced a statistic of 0,994 with a p-value of less than 0,01, rejecting the null hypothesis of stationarity around a constant mean, as anticipated. Additionally, the ADF test fails to reject the null hypothesis of the presence of a unit root at the 95% confidence level, as indicated by a p-value of 0,518. This result indicates that seasonal differencing alone is insufficient to achieve stationarity in the series.

Finally, after applying both regular and seasonal differencing, the KPSS Level test results in a statistic of 0,037 with a p-value of less than 0.1, failing to reject the null hypothesis, indicating stationarity around a constant mean. Similarly, the ADF test returns a statistic of -7,053 with

a p-value of less than 0.01, rejecting the null hypothesis that the time series has a unit root. The KPSS Trend test also fails to reject stationarity.

Testing for the need of seasonal differencing for the log-transformed series with the function *nsdiffs* from the R package *Forecast*, produced mixed results. The Osborn, Chui, Smith, and Birchenhall (OCSB) test (Osborn et al., 1988), designed to detect seasonal unit roots, returned a statistic of -2,51 leading to the rejection of the null hypothesis of stationarity of the seasonal component. However, the Hylleberg-Engle-Granger-Yoo (HEGY) test (Hylleberg et al., 1990) suggested the presence of a seasonal unit root, indicating the potential need for seasonal differencing.

Table 5-4: Summary of Canova and Hansen test results for seasonal stability for different transformations.

Month	Log-transformed		First differenced		Regular and Seasonal	
	Statistic	p-value	Statistic	p-value	Statistic	p-value
Jan	1,0212	0,0003***	0,3472	0,1056	0,2092	0,2839
Feb	0,9994	0,0004***	0,2073	0,2858	0,1367	0,4959
Mar	0,9859	0,0005***	0,1326	0,5071	0,1364	0,4974
Apr	1,0234	0,0003***	0,1237	0,5444	0,0513	0,9417
May	1,0552	0,0002***	0,1074	0,6214	0,0394	0,9856
Jun	1,0697	0,0001***	0,3418	0,1097	0,093	0,7027
Jul	1,0779	0,0001***	0,6973	0,0075**	0,4012	0,0717.
Aug	1,0756	0,0001***	0,9755	0,0005***	0,0582	0,9056
Sep	1,0677	0,0002***	0,8807	0,0014**	0,1394	0,4856
Oct	1,0689	0,0001***	0,8528	0,0019**	0,0521	0,9381
Nov	1,0275	0,0003***	0,156	0,4227	0,0905	0,7166
Dec	1,0384	0,0002***	0,3259	0,1226	0,0707	0,8329
Joint	1,8648	0,381	2,0733	0,3118	1,5286	0,4926

***: p-value < 0,001; **: p-value < 0,01; *: p-value < 0,05; . : p-value < 0,1

In the log-transformed series, every month exhibits seasonal instability. This suggests that the seasonal patterns are not stable over time, with all months showing changes in their seasonal effects. Despite this, the joint test remains non-significant implying that the overall seasonal structure, when considered collectively, does not exhibit significant deviations from stability. After applying first differencing, seasonal instability persists in July, August, September, and October. However, when both regular and seasonal differencing are applied, the seasonal instability is largely reduced, implying that none of the months show significant seasonal instability, suggesting that applying both regular and seasonal differencing successfully addresses the instability in both the trend and seasonal components of the time series produce stable seasonal patterns throughout the year.

6. ARIMA MODELLING

6.1. SEASONAL ARIMA MODELS

As previously discussed, to achieve weak stationarity the log-transformed series required to be differenced to remove the trend and seasonally differenced to eliminate the recurring seasonal patterns. To identify a set of candidate orders for a SARIMA model we'll begin by analysing patterns in the PACF and ACF.

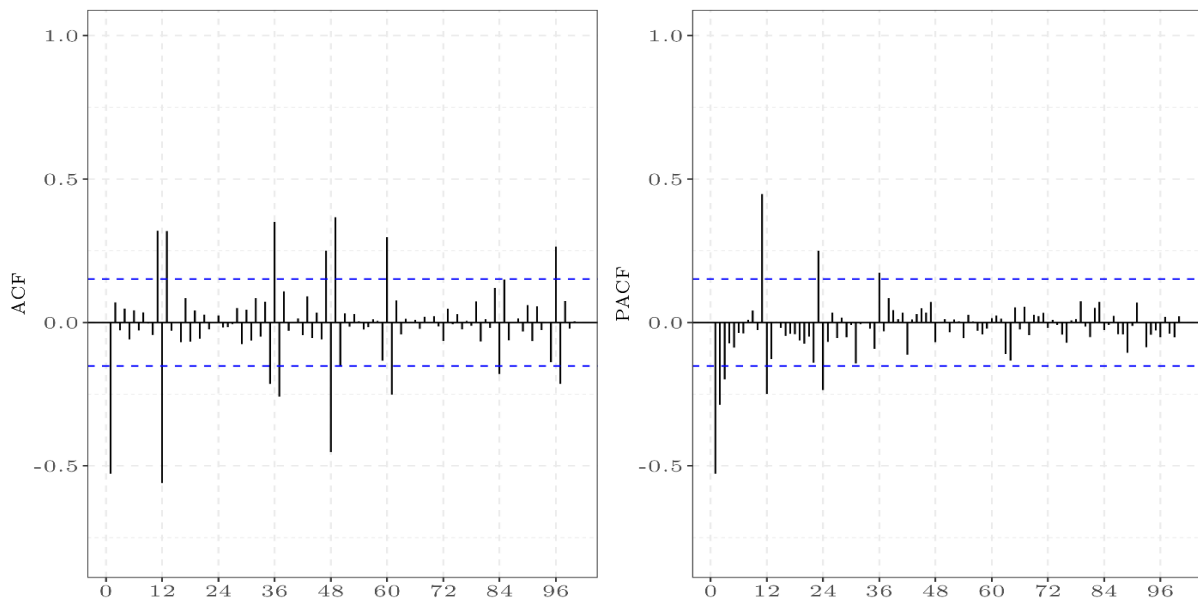


Figure 6-1: ACF and PACF of the transformed data.

In the ACF plot, significant spikes are observable at lags that are multiple of 12, confirming the annual seasonal component. Moreover, the pattern of alternating spikes at these seasonal lags suggests the presence of both seasonal autoregressive (AR) and seasonal moving average (MA) components. The gradual decay of the ACF values over time highlights the need for first differencing. The PACF plot exhibits a sharp drop after lag 1, suggesting a non-seasonal autoregressive term of order 1 and the presence of significant spikes at seasonal lags points to a seasonal AR component of order 1. Additionally, the persistent seasonal spikes in both the ACF and PACF implied that a seasonal difference is needed.

We then applied the *eacf* function from the TSA R package for identifying candidate ARMA models by visually analysing the patterns of significant autocorrelations. The EACF analysis is constrained to identifying only non-seasonal orders and as such, its recommendations did not align completely with the seasonal ARIMA structures tested, even for the non-seasonal components. For the log-transformed, first differenced and seasonal differenced series, the pattern observed in the EACF matrix pointed to an ARMA(0,2) configuration, due to the presence of a complete rectangular pattern of non-significant values at AR=0 and MA=2. Other

potential models, such as ARMA(1,2) and ARMA(2,1) are also plausible under the EACF analysis.

Following this analysis, two candidate models were subject to testing: ARIMA(1,1,0)(1,1,1)₁₂ and ARIMA(1,1,1)(1,1,1)₁₂. Additionally, an automatic model selection procedure using the *auto.arima* function from the forecast R package was applied, following the Hyndman and Khandakar (2008) algorithm, which selects a model by minimizing the AIC. This approach yielded the model ARIMA(0,1,1)(2,1,1)₁₂ as the best candidate.

These 3 models were fitted to the data and evaluated using standard model selection criteria, and residual diagnostics were conducted to verify that the models capture the underlying structure of the series.

Table 6-1: Base model comparison.

Model id	Arima order	AICc	AIC	BIC	Ljung-Box	ARCH Test	Aug. Dickey-Fuller	Jarque-Bera	Param.	Coeff.	SE	Sig
Prep 1	(1,1,0)	-597,28	-597,53	-585,06	$\chi^2 = 15,956$	$\chi^2 = 10,247$	DF = -6,4504	$\chi^2 = 125,39$	ar1	-0,4207	0,0739	***
	(1,1,1) ₁₂				df = 12	df = 12	Lag order = 5	df = 2	sar1	-0,3242	0,1088	**
					p-value =	p-value =	p-value < 0,01	p-value = 2,2e	sma1	-0,4796	0,1062	***
Prep 2	(1,1,1)	-604,64	-605,01	-589,42	$\chi^2 = 8,4812$	$\chi^2 = 11,17$	DF = -5,5713	$\chi^2 = 104,48$	ar1	0,0269	0,1529	(ns)
	(1,1,1) ₁₂				df = 12	df = 12	Lag order = 5	Lag order = 5	ma1	-0,5723	0,1234	***
					p-value =	p-value =	p-value < 0,01	p-value = 2,2e	sar1	-0,2975	0,1206	*
auto arima	(0,1,1)	-624,2	-624,57	-608,98	$\chi^2 = 9,7043$	$\chi^2 = 6,5242$	DF = -5,4594	$\chi^2 = 61,62$	ma1	-0,501	0,0777	***
	(2,1,1) ₁₂				df = 12	df = 12	Lag order = 5	df = 2	sar1	-1,2026	0,0978	***
					p-value =	p-value =	p-value < 0,01	p-value =	sar2	-0,6407	0,0626	***
					0,6419	0,8874		4,163e-14	sma1	0,5303	0,118	***

Legend: $p < 0,001$ (***); $0,001 \leq p < 0,01$ (**); $0,01 \leq p < 0,05$ (*); $p \geq 0,05$ (ns)

The comparison of the three models indicates that the *auto.arima* model outperforms the others based on the selected information criteria (AIC, AICc, and BIC), achieving the better fit of the three models. Residual diagnostics further confirm that the residuals are uncorrelated, as evidenced by the non-significant Ljung-Box test results. Moreover, the residuals of this model kept the ACF spikes to a minimum, suggesting that most of the autocorrelation structure has been captured. However, the Jarque-Bera test for all 3 models reveals significant deviations from normality in the residuals, implying the need for other approaches. The residuals for Models id Prep 1 and Prep 2 are provided in the Appendices (Figure A - 7, Figure A - 8) for reference and comparison.

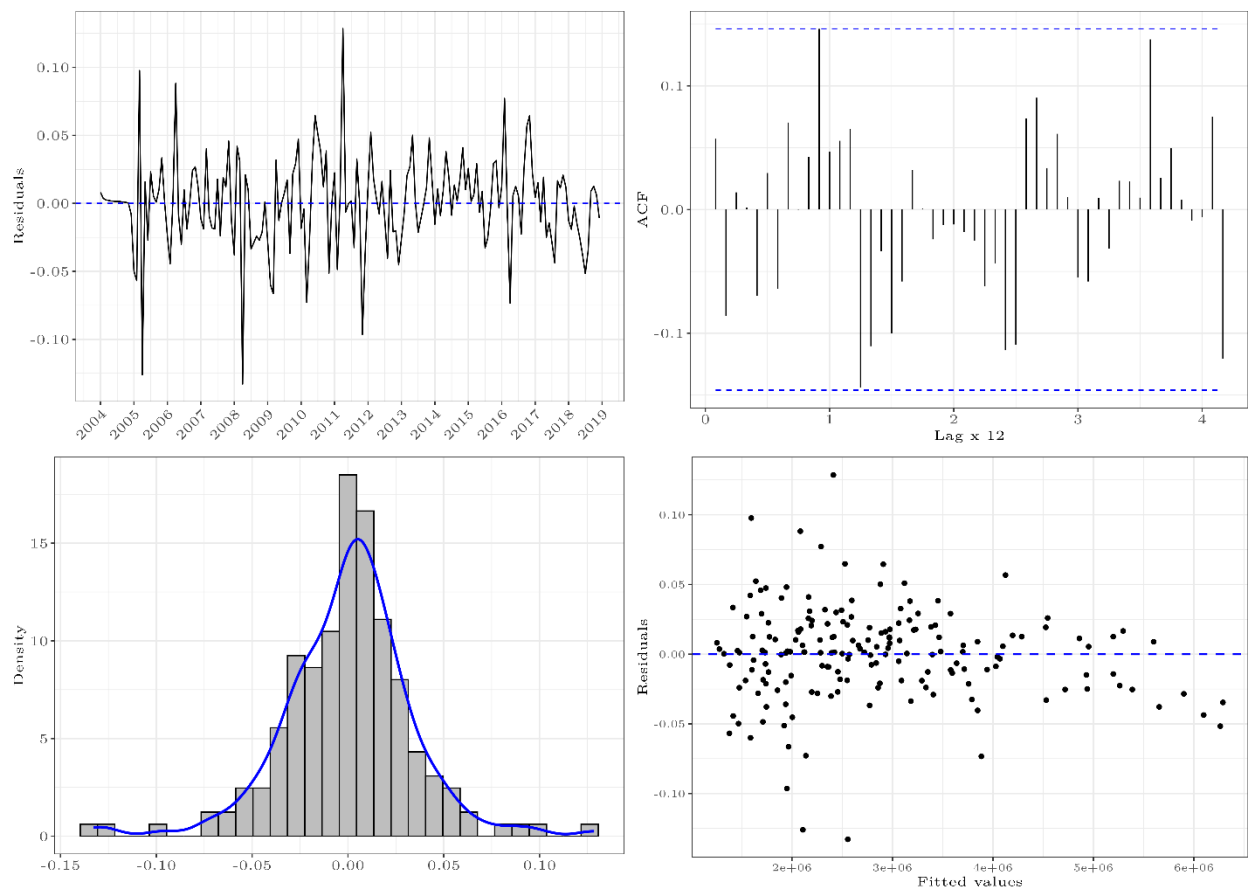


Figure 6-2: Residual diagnostics for the *auto.arima* model.

Further inspection of the residuals for the $ARIMA(0,1,1)(2,1,1)_{12}$ model reveals a slight funnel shape in the Residuals vs. Fitted plot, indicating potential non-constant variance.

7. SELECTION OF EXTERNAL REGRESSORS

This chapter focuses on the selection of external regressors: calendar, outliers, structural breaks and interventions.

7.1. CALENDAR EFFECTS

Incorporating calendar effects (Easter, working days⁹, and leap years¹⁰) into time series models is needed for capturing the systematic variations in the data related to air traffic, (e.g., Beria and Laurino, 2016).

Due to its coding structure, the month preceding Easter is typically assigned a value of -0,5, while the month containing Easter is coded as 0,5 (Appendix C). This structure results in a coefficient that represents the cumulative effect of Easter distributed across these two months. This indicates that the coefficient captures the cumulative impact distributed between the month before and the month during Easter, rather than reflecting the effect in a single period.

To determine the most significant calendar effects to include in the models to assess, a set of candidate models were estimated, each incorporating distinct calendar effects such as Easter, Leap Year, and Working Days (Table 7-1). The selection criterion was based on minimizing the AICc values, with the models exhibiting the lowest AICc for each calendar effect being selected.

⁹ Accounts for the variation in the number of working days across months.

¹⁰ Captures the effect of the additional day that occurs in leap years in February.

Table 7-1: Calendar effects models.

Model id	Arima order	AICc	AIC	BIC	Ljung-Box	ARCH Test	Aug. Dickey-Fuller	Jarque-Bera	Parameter	Coeff	SE	Sig.
1	(0,1,1)				$\chi^2 = 8,1361$	$\chi^2 = 27,3$	DF = -5,8216	$\chi^2 = 2,9531$	ma1	-0,4260	0,0821	***
	(0,1,1) ₁	-678,71	-678,95	-666,48	df = 12	df = 12	Lag order = 5	df = 2	sma1	-0,4810	0,0766	***
	2				p-value = 0,7744	p-value = 0,00699	p-value < 0.01	p-value = 0,2284	xreg	0,0804	0,0073	***
									ma1	-0,4421	0,0785	***
2	(0,1,1)				$\chi^2 = 7,8307$	$\chi^2 = 31,294$	DF = -5,6722	$\chi^2 = 6,9469$	sar1	-1,1748	0,1699	***
	(2,1,1) ₁	-665,07	-665,59	-646,88	df = 12	df = 12	Lag order = 5	df = 2	sar2	-0,4740	0,0806	***
	2				p-value = 0,7982	p-value = 0,001776	p-value < 0.01	p-value = 0,03101	sma1	0,6655	0,1925	***
									xreg	0,0732	0,0093	***
3	(0,1,1)				$\chi^2 = 7,8318$	$\chi^2 = 31,477$	DF = -5,6644	$\chi^2 = 6,6036$	ma1	-0,4407	0,0788	***
	(2,1,1) ₁	-664,72	-665,25	-646,54	df = 12	df = 12	Lag order = 5	df = 2	sar1	-1,2073	0,1725	***
	2				p-value = 0,7981	p-value = 0,001665	p-value < 0.01	p-value = 0,03682	sar2	-0,4779	0,0801	***
									sma1	0,7012	0,1972	***
4	(0,1,1)				$\chi^2 = 7,3421$	$\chi^2 = 28,203$	DF = -5,6265	$\chi^2 = 5,4898$	xreg	0,0722	0,0091	***
	(2,1,1) ₁	-659,72	-660,24	-641,53	df = 12	df = 12	Lag order = 5	df = 2	ma1	-0,4497	0,0781	***
	2				p-value = 0,8342	p-value = 0,005166	p-value < 0.01	p-value = 0,06426	sar1	-1,2595	0,1852	***
									sar2	-0,4868	0,0814	***
5	(0,1,1)				$\chi^2 = 7,0579$	$\chi^2 = 17,395$	DF = -5,5445	$\chi^2 = 6,0764$	sma1	0,7550	0,2154	***
	(2,1,1) ₁	-646,63	-647,16	-628,45	df = 12	df = 12	Lag order = 5	df = 2	xreg	0,0657	0,0087	***
	2				p-value = 0,8538	p-value = 0,1353	p-value = < 0.01	p-value = 0,04792	ma1	-0,4620	0,0799	***
									sar1	-1,1900	0,1115	***
6	(0,1,1)				$\chi^2 = 6,5358$	$\chi^2 = 19,874$	DF = -5,5584	$\chi^2 = 3,8816$	sar2	-0,5735	0,0704	***
	(2,1,1) ₁	-652,34	-652,86	-634,15	df = 12	df = 12	Lag order = 5	df = 2	sma1	0,5722	0,1279	***
	2				p-value = 0,8867	p-value = 0,0695	p-value < 0.01	p-value = 0,1436	xreg	-0,0069	0,0013	***
									ma1	-0,4580	0,0792	***
7	(0,1,1)				$\chi^2 = 11,984$	$\chi^2 = 7,007$	DF = -5,5405	$\chi^2 = 58,912$	sar1	-1,1845	0,1145	***
	(2,1,1) ₁	-638,04	-638,56	-619,86	df = 12	df = 12	Lag order = 5	df = 2	sar2	-0,5590	0,0711	***
	2				p-value = 0,4469	p-value = 0,8572	p-value < 0.01	p-value = 1,612e-13	sma1	0,5915	0,1301	***
									xreg	0,0076	0,0013	***
8	(0,1,1)				$\chi^2 = 9,793$	$\chi^2 = 6,6748$	DF = -5,4414	$\chi^2 = 54,88$	ma1	-0,4859	0,0740	***
	(2,1,1) ₁	-622,71	-623,02	-604,53	df = 12	df = 12	Lag order = 5	df = 2	sar1	-1,2601	0,0820	***
	2				p-value = 0,6341	p-value = 0,8783	p-value < 0.01	p-value = 1,21e-12	sar2	-0,6882	0,0539	***
									sma1	0,6409	0,1059	***
9	(0,1,1)				$\chi^2 = 9,7979$	$\chi^2 = 7,7572$	DF = -5,6267	$\chi^2 = 57,343$	xreg	0,0452	0,0105	***
	(2,1,1) ₁	-629,92	-630,45	-611,74	df = 12	df = 12	Lag order = 5	df = 2	ma1	-0,5059	0,0769	***
	2				p-value = 0,6337	p-value = 0,8038	p-value < 0.01	p-value = 3,534e-13	sar1	-1,2010	0,0995	***
									sar2	-0,6346	0,0637	***
									sma1	0,5399	0,1198	***
									xreg	0,0024	0,0029	(ns)
									ma1	-0,4653	0,0835	***
									sar1	-1,2199	0,0954	***
									sar2	-0,6548	0,0625	***
									sma1	0,5174	0,1171	***
									xreg	-0,0015	0,0005	***

Legend: $p < 0,001$ (***); $0,001 \leq p < 0,01$ (**); $0,01 \leq p < 0,05$ (*); $p \geq 0,05$ (ns)

Despite some models not passing the Jarque-Bera test for normality (e.g., models 2, 3, and 4 with p-values of 0,031, 0,0368, and 0,0643, respectively), the inclusion of Easter as an external regressor was justified due to its significant impact on the series, as shown by the low p-values of the xreg coefficients. Similarly, the leap year effect was tested in model 7, revealing high statistical significance even though the residuals strongly deviate from normality (p-value < 0,0001 for the Jarque-Bera test). The effect of working days was also explored in model 9, which demonstrated a statistically significant impact (p-value = 0,0043), although the residuals did not satisfy the assumption of normality. As for the Ljung-Box test results, none of the models show significant autocorrelation in the residuals. Overall, these series of models, demonstrates that incorporating calendar effects can enhance the model's explanatory power, even if they lead to some deviation from the ideal assumption of normality of the residuals.

After identifying the most relevant calendar effects, a comprehensive model was estimated using the *auto.arima* function from the *forecast* R package, incorporating the selected external regressors: Easter and Leap Year. Following additional testing, the Working Days (or weekend) variable was excluded from the model due to its minimal impact on model fit and the negligible change in AIC and BIC values upon its removal, despite its statistically significant coefficient¹¹. The resulting model exhibited the best fit among the tested configurations, with the residual diagnostics indicating adequate statistical properties (Table 7-2).

Table 7-2: Summary of ARIMA (0,1,1)(0,1,1)₁₂ model with calendar effects.

Model id	Arima order	AICc	AIC	BIC	Ljung-Box	ARCH Test	Aug. Dickey-Fuller	Jarque-Bera	Param.	Coeff.	SE	Sig
10	(0,1,1) (0,1,1) ₁₂	-691,17	-691,54	-675,95	$\chi^2 = 6,2317$	$\chi^2 = 28,07$	DF = -5,8191	$\chi^2 = 4,4195$	ma1	-0,383	0,0824	***
					df = 12	df = 12	Lag order = 5	df = 2	sma1	-0,455	0,0753	***
					p-value =	p-value =	p-value < 0.01	p-value =	Easter	0,0811	0,0067	***
					0,904	0,0054		0,1097	Leap year	0,0441	0,0111	***

Legend: $p < 0,001$ (***); $0,001 \leq p < 0,01$ (**); $0,01 \leq p < 0,05$ (*); $p \geq 0,05$ (ns)

This model exhibits the lowest BIC among all models so far, suggesting a better fit compared to the models without external regressors and the model with individual calendar effects, at the cost of a greater level of complexity. All the estimated parameters are statistically significant.

¹¹ Although the distinction between weekends and working days is relevant in the context of air traffic for daily or weekly data, its influence becomes less pronounced when using monthly aggregated data due to the smoothing effect of the aggregation.

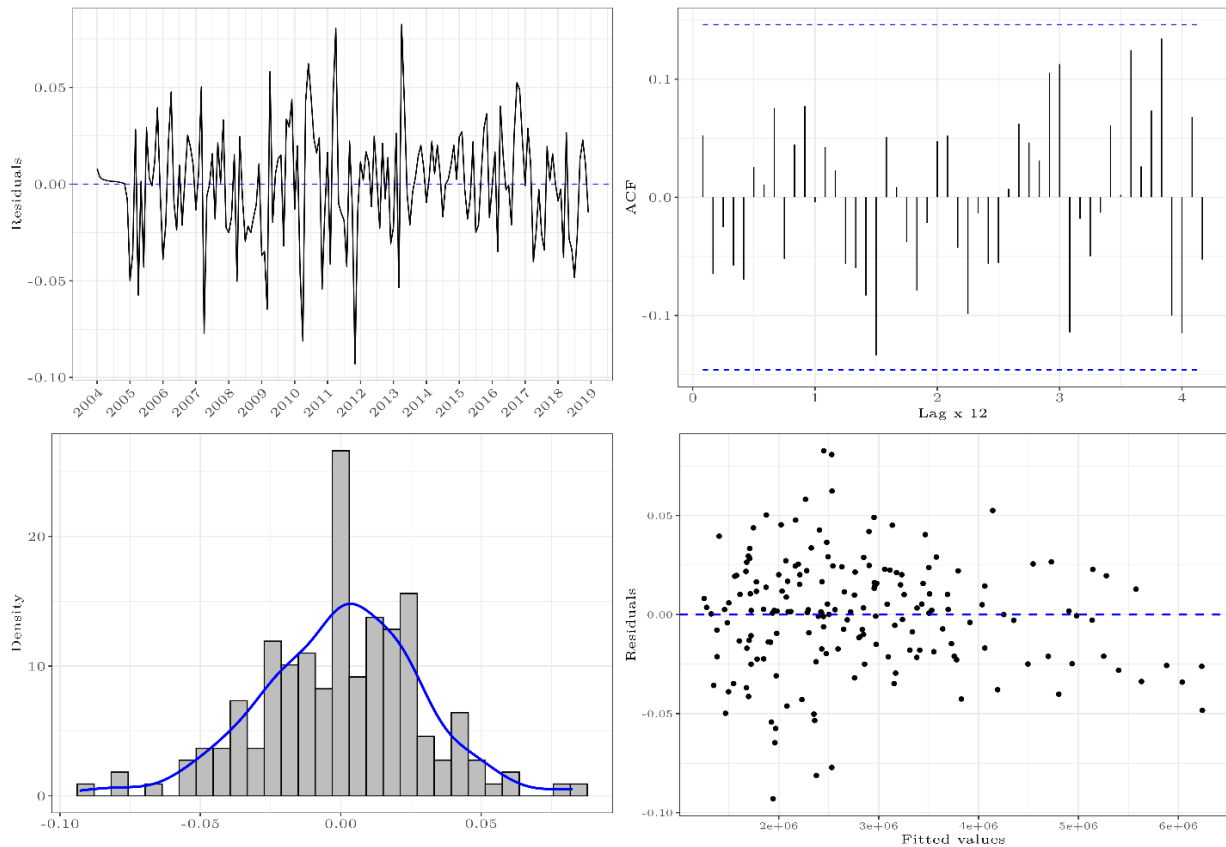


Figure 7-1: Residual diagnostics for model with calendar effects.

Furthermore, the diagnostic statistics confirm that the residuals are well-behaved. The non-significant result of the Jarque-Bera test (Chi-squared=4,4195, p-value=0,1097) suggests that the residuals approximate a normal distribution. Similarly, the Ljung-Box test (Chi-squared=6,2317, p-value=0,904) indicates that there is no remaining autocorrelation in the residuals.

When considering the effects of the 2 calendar regressors, both are statistically significant at a high confidence level ($p < 0,001$). The Easter effect (coefficient=0,0811) indicates a notable positive deviation in passenger numbers around Easter. The Leap Year effect (coefficient=0,0441) also exhibits a significant impact, reflecting variations in passenger volume due to the additional day in February on leap years. Additionally, the ARCH test (Chi-squared=28,07, p-value=0,0054) indicates some heteroskedasticity in the residuals, and the ADF test (DF=-5.8191, p-value<0.01) provides statistical evidence for stationarity.

Overall, the model accounts for important calendar-related patterns, making it a robust candidate for modelling passenger evolution over time.

7.2. OUTLIER DETECTION

In this section, we tested for the presence of outliers using two distinct approaches.

7.2.1. APPLICATION OF THE CHEN AND LIU (1993) METHODOLOGY

We applied the methodology described in Chapter 2.2.4, using the R package *tsoutliers* to identify additive outliers (AO), innovation outliers (IO), level shifts (LS), temporary changes (TC), and seasonal level shifts (SLS). The Chen and Liu (1993) method identified a single temporary change (TC) in April 2010. With the Easter external regressor retained in the model, the timing of the outlier suggests a potential overlap, indicating that part of the deviation attributed to the outlier may reflect the Easter holiday period. The effect of the identified outlier is shown in Figure 7-2).

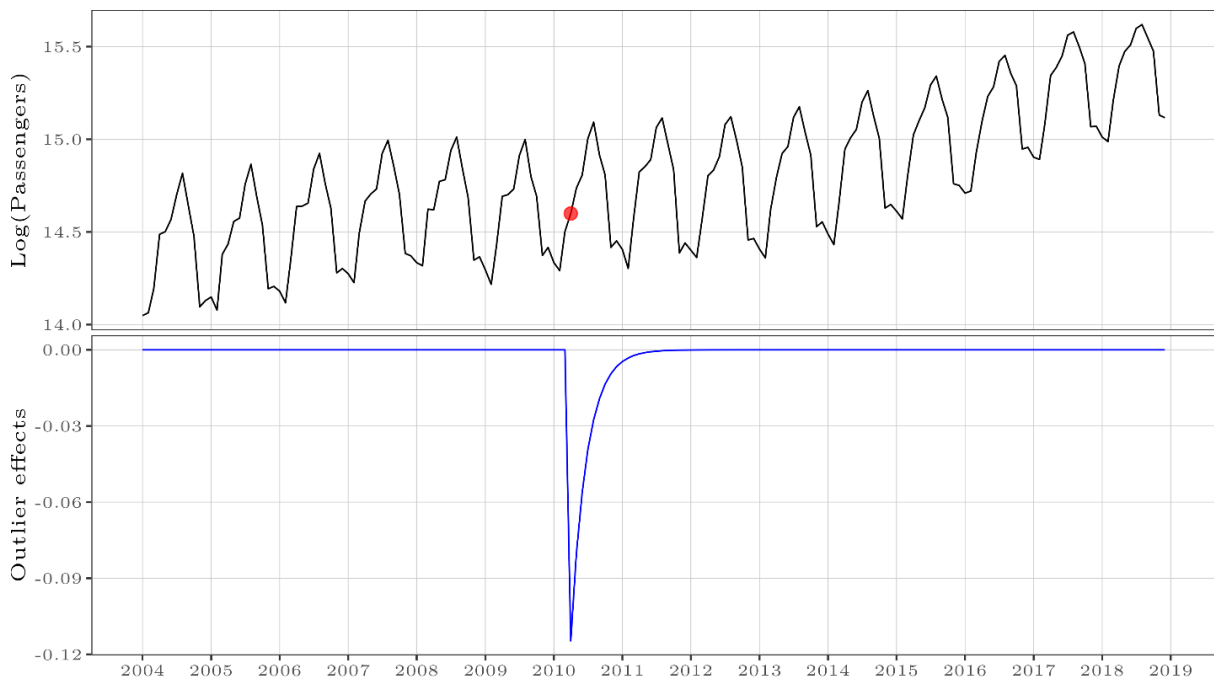


Figure 7-2: Figure 7-3: Detected outlier effect (Chen and Liu, 1993).

The outlier effect is the estimate of the deviation caused by an unknown abnormal event. This means that, if the series is log-transformed, a coefficient of -0.11 implies that the observed value at that point is approximately $(e^{-0.11} - 1) \times 100 \approx -10.42\%$ lower than the expected value based on the model without the outlier or intervention.

The model that achieved the best information criteria, incorporating the external regressors for calendar effects and the identified outlier, was an $\text{ARIMA}(0,1,1)(2,1,0)_{12}$.

Table 7-3: Summary of ARIMA (0,1,1)(2,1,0)₁₂ model with calendar effects and outliers detected using Chen and Liu (1993) procedure.

Model id	Arima order	AICc	AIC	BIC	Ljung-Box	ARCH Test	Aug. Dickey-Fuller	Jarque-Bera	Param.	Coeff.	SE	Sig
11	(0,1,1) (2,1,0) ₁₂	-712,6	-713,31	-691,48	$\chi^2 = 6,2252$ df = 12 p-value = 0,9043	$\chi^2 = 18,807$ df = 12 p-value = 0,0933	DF = - 4,9138 Lag order = 5 p-value <0.01	$\chi^2 = 5,425$ df = 2 p-value = 0,06637	ma1	-0,4057	0,0724	***
									sar1	-0,3741	0,0801	***
									sar2	-0,2351	0,0814	***
									Easter	0,0779	0,0063	***
									Leap year	0,0461	0,0099	***
									TC (2010:04)	-0,1147	0,0222	***

Legend: $p < 0,001$ (***); $0,001 \leq p < 0,01$ (**); $0,01 \leq p < 0,05$ (*); $p \geq 0,05$ (ns)

All parameters of resulting model are significant at conventional levels (p-value<0.05). The Jarque-Bera test yielded a non-significant result (p-value=0,0637), indicating that the residuals do not significantly deviate from the assumption of normality, with a skewness value of -0,375. The Ljung-Box test for autocorrelation at lag 12 returned a non-significant p-value (p=0.9043), suggesting that no significant autocorrelation remains in the residuals.

The ARCH test for conditional heteroscedasticity returned a non-significant result (p-value=0.0933), indicating no strong evidence of heteroscedasticity in the residuals. Lastly, the Augmented Dickey-Fuller (ADF) test rejected the null hypothesis of a unit root (p-value<0.01). These results, combined with a visual inspection of various residual diagnostics (Figure 7-4) support the adequacy of the model's specification.

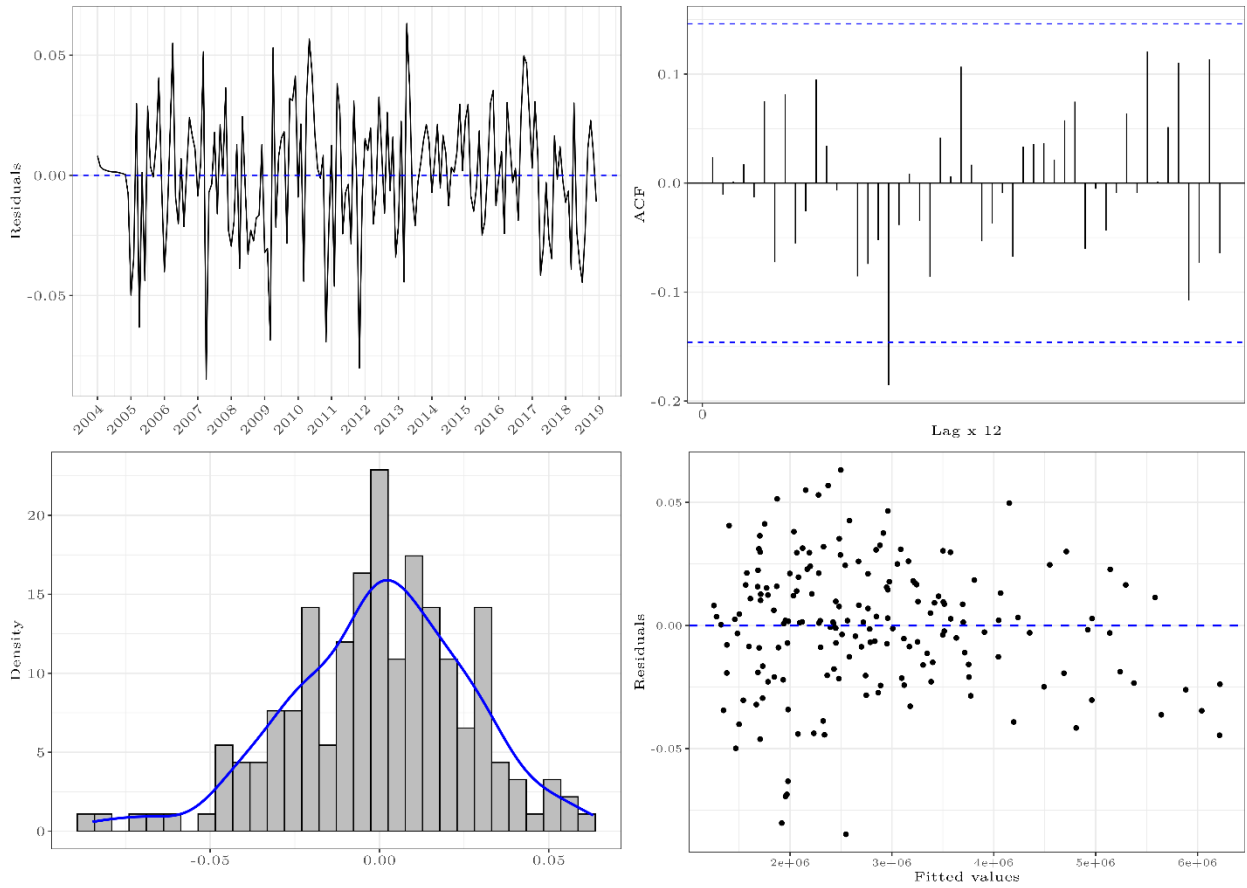


Figure 7-4: Residual diagnostics for model with calendar effects and outliers detected using Chen and Liu (1993) procedure.

Following equation (2-36) the estimated model can be expressed as follows:

$$\log(Y_t) = -\frac{0,1147}{1-0,7B} I_t^{(76)} + 0,0779 \text{Easter}_t + 0,0461 \text{Leap Year}_t + \frac{(1 - 0,4057 B)(1 - 0,3741B - 0,2351B^2)}{(1 - B)(1 - B^{12})} a_t \quad (7-1)$$

where:

Y_t : Represents the log-transformed number of passengers at time t ;

$I_t^{(76)}$ is a variable representing an TC, such that:

$$I_t^{(76)} = \begin{cases} \delta^k & \text{for } k = 0, 1, 2, \dots \\ 0 & \text{otherwise} \end{cases}$$

where δ represents the decay factor to model the effect reduction over time and k denotes the number of time periods that have elapsed since the occurrence of the initial outlier event. We assume that $\delta = 0,7$ which indicates that the initial impact decays by 30% each period following the intervention. The temporary change effect at time t is $I_t^{(\text{TC index})} = \hat{\beta} \times 0,7^k$ and

$\frac{1}{1-0,7B}$ represents the decay factor with a coefficient of 0,7 to modulate the impact of the outlier over time;

$Easter_t$ is an external regressor such that:

$$Easter_t = \begin{cases} -0,5 & \text{if } t = \text{Month before Easter (e.g., March 2004)} \\ 0,5 & \text{if } t = \text{Easter month (e.g., April 2004)} \\ -0,214 & \text{if } t = \text{Easter impact in March 2015} \\ 0,214 & \text{if } t = \text{Easter month in 2015 (April)} \\ -0,071 & \text{if } t = \text{Easter impact in March 2010} \\ 0,071 & \text{if } t = \text{Easter month in 2010 (April)} \\ -0,357 & \text{if } t = \text{Easter impact in March 2018} \\ 0,357 & \text{if } t = \text{Easter month in 2018 (April)} \\ 0 & \text{otherwise} \end{cases}$$

$Leap Year_t$ is an external regressor such that:

$$Leap_{year}_t = \begin{cases} 0,75 & \text{if } t = \text{February in Leap Years (e.g., February 2004)} \\ -0,25 & \text{if } t = \text{February in non-Leap Years (e.g., February 2005)} \\ 0 & \text{otherwise} \end{cases}$$

and $\frac{(1 - 0,4057 B)(1 - 0,3741B - 0,2351B^2)}{(1 - B)(1 - B^{12})} a_t$ represents the ARIMA structure of the residuals where $(1 - 0,4057 B)$ represents the moving average component, $(1 - 0,3741B - 0,2351B^2)$ represents the seasonal autoregressive component with a seasonal period of 12 and $(1 - B)(1 - B^{12})$ represents regular and seasonal differencing in the ARIMA model.

Given that that $t = 76$ was subsequently coded as an Additive Outlier (AO) intervention associated with the Volcano event, we retained the Temporary Change (TC) intervention suggested by this method to evaluate how well the model performs when an event is treated either as an unknown outlier or as a predefined intervention.

7.2.2. OUTLIER DETECTION USING DECOMPOSITION AND RESIDUAL ANALYSIS

Following the procedure proposed by Hyndman (2021), we applied a detection method based on the analysis of the residuals obtained from decomposition. Observations are classified as outliers if they are either below $Q1 - 3 \times IQR$ or above $Q3 + 3 \times IQR$, where $Q1$ and $Q3$ denotes the first and third quartiles, respectively, and IQR is the interquartile range. This procedure was executed using the *tsoutliers* function from the *forecast* R package.

The procedure initially detected outliers at $t = \{15, 16, 51, 52, 76, 176\}$. However, through model testing and refinement, the outlier at $t = 176$ was found to be statistically non-significant and was subsequently excluded from Model 12. The ARIMA (0,1,1)(0,1,1)₁₂ model incorporating calendar effects and the remaining detected outliers is summarized below.

Table 7-4: Summary of ARIMA (0,1,1)(0,1,1)₁₂ model with calendar effects and outliers detected using decomposition and residual analysis.

Model id	Arima order	AICc	AIC	BIC	Ljung-Box	ARCH Test	Aug. Dickey-Fuller	Jarque-Bera	Param.	Coeff.	SE	Sig
12	(0,1,1) (0,1,1) ₁₂	-741,3	-742,7	-711,5	$\chi^2 = 11.373$ df = 12 p-value = 0.4972	$\chi^2 = 5.1771$ df = 12 p-value = 0.9518	DF = -5,3671 Lag order = 5 p-value < 0.01	$\chi^2 = 19.149$ df = 2 p-value = 6.948e-05	ma1	-0,3129	0,0807	***
									sma1	-0,2926	0,0930	***
									Easter	0,0462	0,0069	***
									Leap year	0,0464	0,0088	***
									Outlier t=15	0,0514	0,0189	***
									Outlier t=16	-0,0681	0,0187	***
									Outlier t=51	0,0714	0,0188	***
									Outlier t=52	-0,0747	0,0187	***
									Outlier t=76	-0,1181	0,0169	***

Legend: $p < 0,001$ (***); $0,001 \leq p < 0,01$ (**); $0,01 \leq p < 0,05$ (*); $p \geq 0,05$ (ns)

Following the general form from Equation (2-36), the estimated model can be written as:

$$\log(Y_t) = 0,0462 \text{Easter}_t + 0,0464 \text{LeapYear}_t + 0,0514 I_t^{(15)} - 0,0681 I_t^{(16)} + 0,0714 I_t^{(51)} - 0,0747 I_t^{(52)} - 0,1181 I_t^{(76)} + \frac{1 - 0,3129B}{1 - 0,2926B^{12}} a_t \quad (7-2)$$

Where:

Y_t represents the dependent variable at time t ;

Easter_t is defined as in Equation (2-45);

LeapYear_t is an indicator variable representing the effect of leap years in February;

$I_t^{(15)}, I_t^{(16)}, I_t^{(51)}, I_t^{(52)}, I_t^{(76)}$ are additive outliers coded as dummy variables with zero all months except at the index;

$\frac{1 - 0,3129B}{1 - 0,2926B^{12}} a_t$ represents the ARIMA component of the residuals.

Given log-transformed series ($\lambda = 0$) each coefficient reflects an approximate percentage change in the passenger numbers relative to a baseline month without an intervention. A positive coefficient suggests an increase in passenger traffic attributable to the intervention, whereas a negative coefficient signifies a decrease, under the assumption that the log-transformation captures proportional changes.

Due to the coding structure of the Easter dummy variable, where the month preceding Easter is typically coded as -0,5 and the month of Easter as 0,5, the coefficient of 0,0462 represents the combined effect spread across these two months. This indicates that the coefficient captures the cumulative impact distributed between the month before and the month during Easter, rather than reflecting the effect in a single period.

During the month in which Easter occurs, there is an estimated increase in passenger numbers of approximately $(e^{0,0462 \times 0,5} - 1) \times 100 \approx 2,34\%$, relative to the months outside the Easter period. In the preceding month, the magnitude of the overall Easter effect, compared to a

typical non-Easter month is lower by $(e^{0,0462 \times (-0,5)} - 1) \times 100 \approx -2,28\%$ meaning that that a portion of the anticipated rise in passenger demand is shifted forward into the month when Easter takes place, reflecting a temporal redistribution of the overall Easter effect.

Using the same methodology, the Leap year coefficient (0,0464) implies an increase of $(e^{0,0464} - 1) \times 100 \approx 4,75\%$ in the number of passengers in February compared to non-leap years. Similarly, the coefficient for the Outlier at index $t = 51$ ($\beta = 0,0714$) reflects an increase of approximately 7,4% passengers going through the Portuguese airports compared to the expected levels in the absence of that particular outlier. In absolute terms, the impact of this event can be estimated by $Y_t \times (e^\beta - 1)$. Considering that May 2010 recorded 2.511.505 passengers, the total impact of the volcanic event, in this model, amounts to an estimated decrease of 314.834 passengers.

A visual inspection of the residuals corroborates the resemblance to white noise (Figure 7-5).

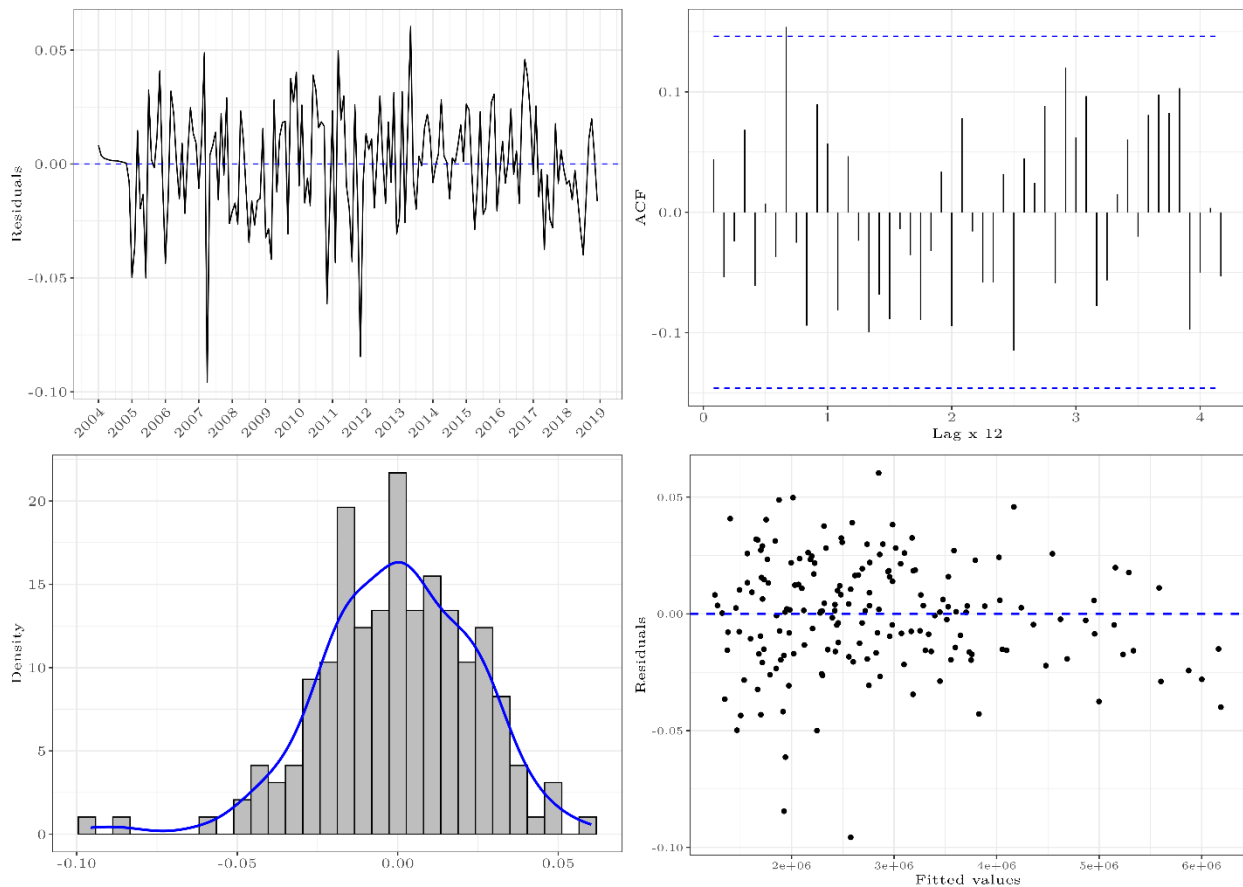


Figure 7-5: Residual diagnostics for model with calendar effects and outliers detected using decomposition and residual analysis.

The integration of the calendar effects into the model, combined with the identified outliers, revealed overlaps where some detected outliers' effects were already accounted for by the calendar effect variables. To assess the presence of multicollinearity among the independent variables, we ran a linear regression using the residuals from the model as the dependent

variable and the external regressor matrix as the predictors and calculated the Variance Inflation Factor (VIF) for each predictor. All VIFs were well below 5, confirming minimal multicollinearity.

7.3. STRUCTURAL BREAKS AND MODELLING OF THE FINANCIAL CRISIS PERIOD

The description of a time series benefits from the analysis of potential discontinuities or breakpoints and inhomogeneities caused by shifts or other changes in trend or in variance (e.g., Tsay, 1988). Breakpoints are considered as boundaries between two or more segments of data (Topál et al., 2016). Empirical research has shown that statistically significant changes in the mean or in the coefficients in a linear regression match the time of historical events that might have produced these breaks (e.g., Zeileis et al., 2003). Breakpoints can be gradual or abrupt where the latter implies the mean, trend and variance of the time series change after a certain point in time in such a way that the time series needs to be studied on both sides of a breakpoint given that this breakpoint is not as simple data artefact. We conducted the structural break analysis on the residuals derived from an automatic model selection procedure using the *auto.arima* function from the *forecast* package in R. This analysis was performed without the inclusion of external regressors, focusing on identifying significant structural changes in the residuals that may not have been captured by the ARIMA model. To achieve this, we utilized the breakpoints function from the *strucchange* package.

The main objective was to use these breakpoints to model the financial crisis that began toward the end of 2008 and continued until approximately 2013. Identifying precise dates for these structural changes proved challenging, as many airline tickets are purchased well in advance, and inbound tourism may not directly align with the financial crisis timeline. Additionally, this extended period was influenced by external factors, such as the Arab Spring, which began in late 2010 and had a counterbalancing effect by increasing tourism.

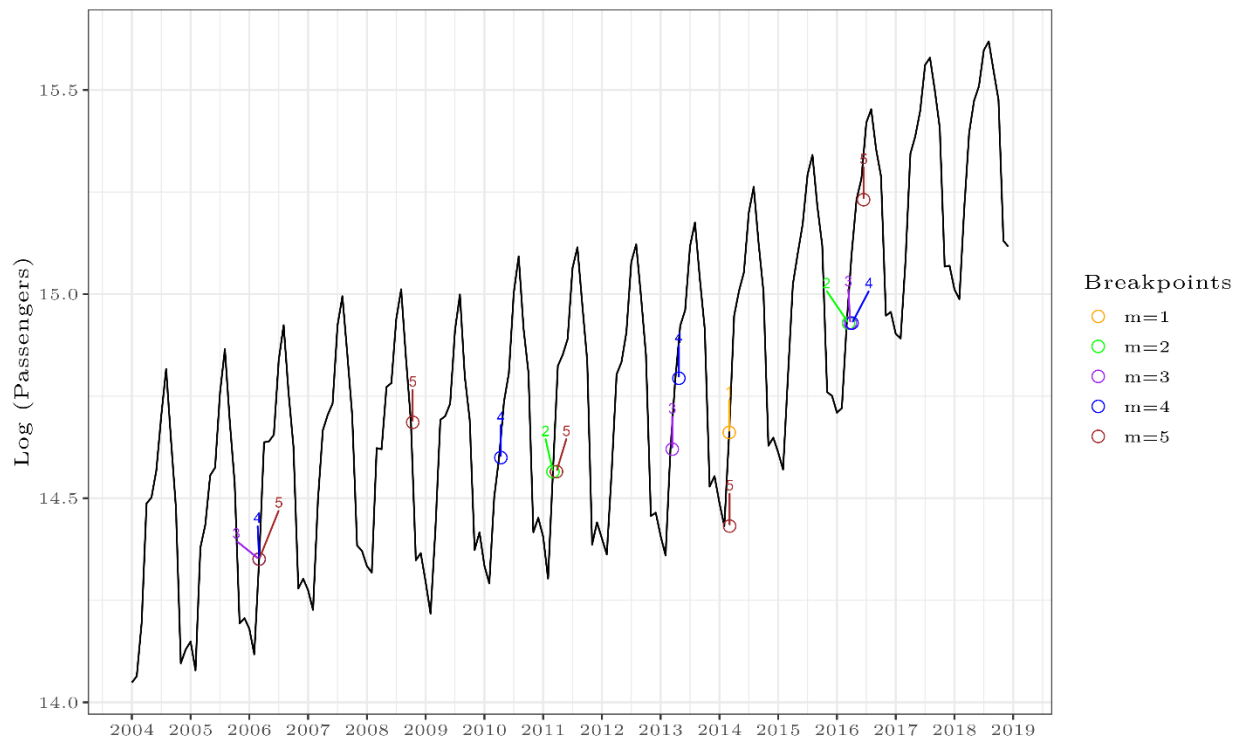


Figure 7-6: Structural breakpoints in model residuals.

By examining the output from the *strucchange* package, the BIC initially decreases as additional breakpoints are included, indicating an improved model fit. However, this trend plateaus around 2 to 3 breakpoints, suggesting that the marginal benefits of adding more breakpoints diminish beyond this point. Similarly, the RSS (Residual Sum of Squares) decreases progressively with each additional breakpoint, but the rate of decrease slows after 2 to 3 breakpoints. This shows that selecting 2 or 3 breakpoints achieves a balance between identifying structural changes and avoiding overfitting.

The breakpoints and resulting segments were subsequently tested for statistical significance in several configurations, focusing on both level and trend changes. Additionally, we tested other external regressor candidates created through observation of the series, leading to the development of the Financial Crisis H1 to H8 regressors (Table 7-5)

Table 7-5: External regressors for structural breaks and financial crisis intervention.

External regressor	Start	End	Effect type
Financial crisis H1	2008/10	2010/06	Level
Financial crisis H2	2008/09	2010/05	Level
Financial crisis H3	2008/09	2010/06	Level
Financial crisis H4	2008/09	2010/07	Level
Financial crisis H5	2008/11	2010/09	Level
Financial crisis H6	2009/01	2013/03	Level
Financial crisis H7	2008/11	2010/09	Trend
Financial crisis H8	2008/08	2010/07	Trend
Structural break (2016/05)	2016/05	2018/12	Level
Structural break (2016/05)	2016/05	2018/12	Trend
Structural break (2008/10 to 2011/03)	2008/10	2011/03	Level
Structural break (2008/10 to 2013/04)	2008/10	2013/04	Level
Structural break (2009/03 to 2013/09)	2009/03	2013/09	Level
Structural break (2009/03 to 2013/03)	2009/03	2013/03	Level

We concluded that the trend component was already adequately captured by the ARIMA model features, rendering additional trend segments non-significant. We also found that some candidates were significant when tested individually but became non-significant in the presence of other intervention variables due to potential collinearity or overlapping effects. Among the numerous configurations evaluated, a promising candidate was Financial Crisis H6 (2009:01 to 2013:03) which consistently demonstrated significance in most tests for the selection of external regressors and interventions. As a consequence, we retained the Financial Crisis H6 segment, which became the main candidate to model the crisis period despite the inherent ambiguity of this period due to the presence of other potential interventions that are equally challenging to model accurately.

7.4. INTERVENTIONS

We classified the interventions of Table 7-6) into seven groups:

1. Strikes and labour events: Includes disruptions such as airline and ground handling strikes;
2. Sports events: Events that influence passenger numbers due to increased travel demand;
3. Accidents: Events that have the potential to alter the willingness to travel;
4. Bankruptcies: Financial collapses of airlines that impact short-term route availability;

5. Religious Events: Events that may trigger surges in passenger traffic due to large-scale gatherings
6. Natural Disasters: Disruptions caused by natural phenomena that can severely impact air traffic.
7. Terrorism and other criminal acts: Disruptions arising from acts of terrorism or criminal activities that significantly impact air travel.

Testing various models revealed that most of the interventions were statistically non-significant when included individually or in combination. However, the Eyjafjallajökull volcanic eruption and the general strike stood out as highly significant, indicating their substantial impact on the monthly passenger totals.

Events like the bankruptcies of major airlines (e.g., Monarch in October 2017 and Air Berlin in November 2017) and the Germanwings crash in March 2015 illustrate that, in the European context, disruptions are often mitigated swiftly due to the open market structure where all European airlines can operate flights to any destination within the EU. This flexibility ensures that passenger demand is quickly met by other carriers stepping in to fill the gap.

A contrasting scenario is observed in the case of airport slots and slot allocation, as these are more complex to manage due to their scarcity. The limited number of available slots can create significant challenges and delays, affecting the features of the time series by imposing constraints on trends. This scarcity can hinder the rapid recovery of disrupted services, leading to more noticeable shifts in monthly passenger data when significant events occur.

Table 7-6: Interventions.

Intervention	Months	Intervention Classification	Expected Type of Intervention
General Strike	2011-11	Strikes and Labour Events	Additive Outlier
UEFA Champions League 2013-2014 Final	2014-05	Sport Events	Additive Outlier
Germanwings crash	2015-03	Accidents	Additive Outlier
Strikes (e.g., ground handling, airline personnel)	2016-08	Strikes and Labour Events	Additive Outlier
Strikes (e.g., ground handling, airline personnel)	2016-10	Strikes and Labour Events	Additive Outlier
Monarch Airlines bankruptcy	2017-10	Bankruptcies	Additive Outlier / Temporary Change
Air Berlin bankruptcy	2017-11	Bankruptcies	Additive Outlier
Aga Khan meeting	2018-07	Religious Events	Additive Outlier
Primera Air bankruptcy	2018-10	Bankruptcies	Additive Outlier
Eyjafjallajökull Volcano Eruption	2010-04 to 2010-05	Natural Disasters	Additive Outlier
Brussels terrorist attack	2016-03	Terrorism and other criminal acts	Additive Outlier / Temporary Change
Pope's visit	2017-05	Religious Events	Additive Outlier
TAP Air Portugal strikes	2017-09	Strikes and Labour Events	Additive Outlier
Ryanair strikes	2018-08	Strikes and Labour Events	Additive Outlier
Ryanair strikes	2018-09	Strikes and Labour Events	Additive Outlier
Second Boeing 737 Max crash	2019-03	Accidents	Additive Outlier / Temporary Change
Strikes (e.g., ground handling, airline personnel)	2019-04	Strikes and Labour Events	Additive Outlier
UEFA Champions League 2018-2019 Final (Increased passenger traffic through Portugal due to logistical reasons)	2019-05 to 2019-06	Sport Events	Additive Outlier / Temporary Change

The 2019 interventions are also presented in the previous table because they are intended to be included in the exogenous regressor matrix to evaluate the predictive accuracy of the models on holdout data beyond the main study period, which concludes in 2018.

It is important to note that the temporal aggregation of the time series plays a crucial role in determining which types of interventions materially impact its features. Minor disruptions in air traffic may not be evaluated as significant within a monthly aggregation due to airlines' rescheduling practices and rapid recovery measures driven by fierce competition. This monthly aggregation tends to smooth out short-term fluctuations, allowing only sustained interventions to manifest as observable changes in the series.

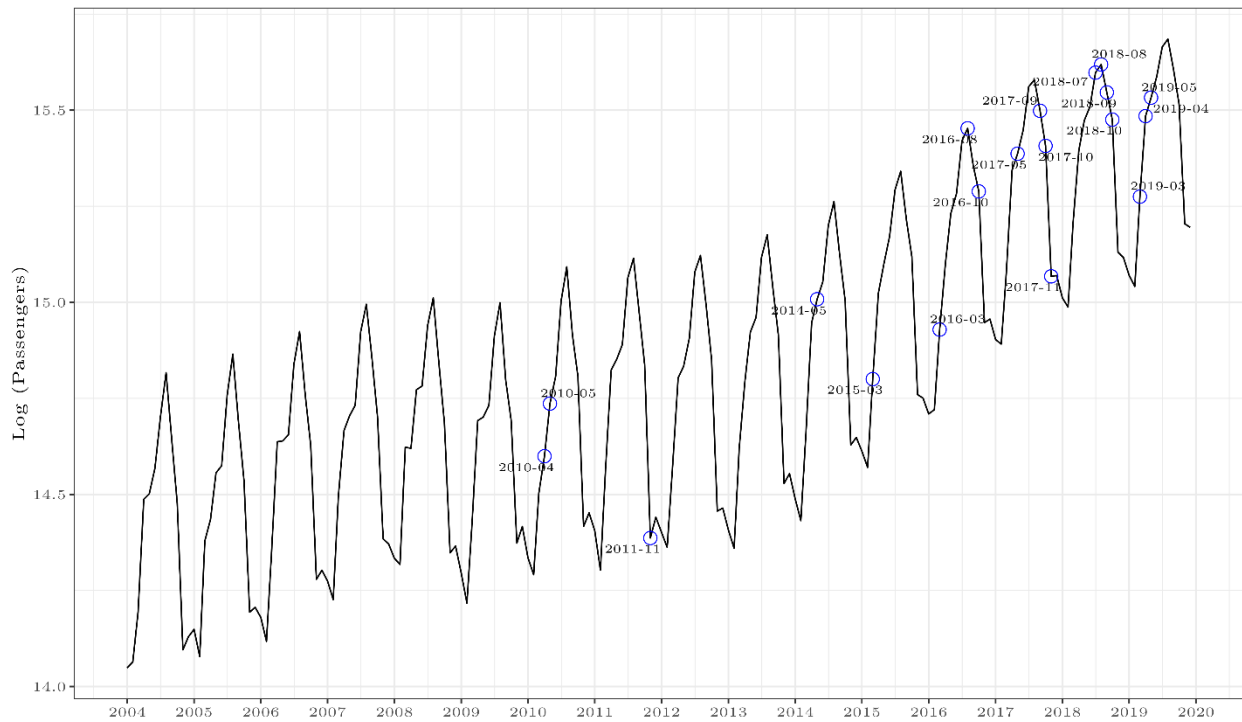


Figure 7-7: Dates of interventions.

To illustrate this effect, visual markers were added to the time series plot (Figure 7-7). As shown, localized and low adherence strikes, such as the one in September 2017, which were confirmed to be non-significant, do not reveal a visible impact on the series. Conversely, the two-day disruption caused by the volcanic eruption is evident, highlighting how only substantial and lasting events create noticeable changes in a monthly aggregated time series.

8. REGRESSION WITH ARIMA ERRORS

Using the full set of potential external regressors, multiple candidate models were tested within a regression framework with ARIMA errors (Table 8-1).

Table 8-1: Models with outliers, interventions and calendar effects.

Model id	Arima order	AICc	AIC	BIC	Ljung-Box	ARCH Test	Aug. Dickey-Fuller	Jarque-Bera	Param.	Coeff.	SE	Sig
13	(0, 1, 1) (0, 1, 1) ₁₂	-721,25	-722,16	-697,21	$\chi^2 = 10,873$	$\chi^2 = 12,558$	DF = -5,1552 Lag order = 5 p-value < 0,01	$\chi^2 = 0,96376$ df = 2 p-value = 0,6176	mal	-0,453	0,071	***
					smal	-0,4142			0,079	***		
					Easter	0,0829			0,006	***		
					Leap year	0,0419			0,01	***		
					Financial crisis H6	-0,0432			0,014	**		
					General Strike Nov 2011	-0,0579			0,019	**		
					Volcano Apr 2010	-0,0747			0,016	***		
					ar1	0,9898			0,009	***		
					mal	-0,4035			0,073	***		
					sar1	-0,342			0,083	***		
14	(1, 0, 1) (1, 1, 0) ₁₂	-733,54	-735,55	-698,07	$\chi^2 = 10,525$ p-value = 0,57	$\chi^2 = 6,5382$ p-value = 0,8866	DF = -5,1225 p-value < 0,01	$\chi^2 = 1,4726$ p-value = 0,4789	Easter	0,0612	0,007	***
					Leap year	0,0462			0,009	***		
					Financial crisis H6	-0,0311			0,014	*		
					General Strike Nov 2011	-0,0569			0,017	**		
					Volcano Apr 2010	-0,0816			0,015	***		
					AO Apr 2005	-0,0569			0,021	***		
					AO Mar 2008	0,0667			0,02	**		
					AO Apr 2008	-0,0481			0,019	*		
					mal	-0,4555			0,072	***		
					smal	-0,3805			0,082	***		
15	(0, 1, 1) (0, 1, 1) ₁₂	-717,69	-718,6	-693,66	$\chi^2 = 7,8441$ df = 12 p-value = 0,7972	$\chi^2 = 15,194$ df = 12 p-value = 0,231	DF = -5,2052 Lag order = 5 p-value < 0,01	$\chi^2 = 1,0539$ df = 2 p-value = 0,5904	Easter	0,0795	0,006	***
					Leap year	0,0419			0,01	***		
					General Strike Nov 2011	-0,0582			0,019	**		
					Volcano Apr 2010	-0,0736			0,016	***		
					Level 2008:10 to 2011:03	-0,0333			0,014	*		
					ar1	0,9899			0,009	***		
					mal	-0,4654			0,067	***		
					sar1	-0,2856			0,083	***		
					Easter	0,0779			0,006	***		
					Leap year	0,0423			0,01	***		
17	(1, 0, 1) (1, 1, 0) ₁₂	-719,96	-721,36	-690,12	$\chi^2 = 15,945$ df = 12 p-value = 0,1938	$\chi^2 = 14,918$ df = 12 p-value = 0,2459	DF = -4,8111 Lag order = 5 p-value < 0,01	$\chi^2 = 3,7126$ df = 2 p-value = 0,1562	General Strike Nov 2011	-0,0566	0,018	**
					Volcano Apr 2010	-0,0738			0,016	***		
					TC Apr 2005	-0,0456			0,022	*		
					Financial crisis H6	-0,0402			0,014	**		
					mal	-0,4255			0,073	***		
					smal	-0,3613			0,085	***		
					Easter	0,0801			0,006	***		
					Leap year	0,0414			0,01	***		
					General Strike Nov 2011	-0,0566			0,018	**		
					Volcano Apr 2010	-0,0668			0,016	***		
18	(0, 1, 1) (0, 1, 1) ₁₂	-724,97	-726,12	-698,06	$\chi^2 = 10,057$ df = 12 p-value = 0,6109	$\chi^2 = 11,829$ df = 12 p-value = 0,4595	DF = -4,925 Lag order = 5 p-value < 0,01	$\chi^2 = 2,9126$ df = 2 p-value = 0,2331	Financial crisis H6	-0,0427	0,014	**
					AO Apr 2009	0,0488			0,02	*		

Legend: $p < 0,001$ (***); $0,001 \leq p < 0,01$ (**); $0,01 \leq p < 0,05$ (*); $p \geq 0,05$ (ns)

All parameters are statistically significant, with the t-values exceeding the critical threshold of 1.96, which means that the parameter estimates are greater than twice their standard errors, consistent with a 97.5% confidence level under a two-tailed test. The diagnostic tests indicate that the models exhibit well-behaved residuals, supporting their adequacy for capturing the dynamics of the series. The non-significant results from the Jarque-Bera test suggest that the residuals approximate normality, while the ARCH and Ljung-Box tests confirm the absence of conditional heteroscedasticity and autocorrelation, respectively. Finally, the Augmented

Dickey-Fuller tests suggest stationarity leaving residuals that closely approximate white noise. In the previous table, Model 14 ranks highest across all criteria, followed by Model 18.

9. LINEAR REGRESSION MODEL WITH FOURIER TERMS AND EXTERNAL REGRESSORS

This model integrates a linear trend, with added Fourier terms for capturing cyclical seasonality and external regressors without relying on the autoregressive dynamics. The model extends the standard linear form to include cyclical seasonality and specific event-based adjustments, and can be represented as follows:

$$Y_t = \beta_0 + \beta_1 trend_t + \sum_{k=1}^K \left(\alpha_k \sin \frac{2\pi kt}{m} + \gamma_k \cos \left(\frac{2\pi kt}{m} \right) \right) + \sum_i \delta_i Intervention_i + \epsilon_t \quad (9-1)$$

Y_t represents the response variable time t ,

β_0 is the intercept,

$\beta_1 trend_t$ models the linear trend over time,

α_k and γ_k are coefficients for the Fourier terms for $k = \{1,2,3\}$, representing sine and cosine pairs, respectively, that capture cyclical seasonality with periodicity m ,

δ_i represents the influence of external regressors on the response variable and

ϵ_t the error term.

The estimation of this model, with the most relevant regressors, is as follows:

$$\log(Y_t) = \beta_0 + trend_t - \sum_{k=1}^3 \left(\alpha_k \sin \left(\frac{2\pi kt}{12} \right) + \gamma_k \cos \left(\frac{2\pi kt}{12} \right) \right) + \delta_1 Level_{2009:03 \text{ to } 2013:09} + \delta_2 Level_{2016:05 \text{ to } 2018:12} + \delta_3 Strikes + \delta_4 Easter + \epsilon_t \quad (9-2)$$

Where,

Y_t : Passenger count at time t ,

$trend_t$ represents the underlying time trend component at time t , defined as $trend_t = t$,

$\sin 2\pi t/12$ and $\cos 2\pi t/12$ represent the annual cycle,

$\sin 4\pi t/12$ and $\cos 4\pi t/12$ represent the semi-annual cycle,

$\sin 6\pi t/12$ and $\cos 6\pi t/12$ represent the quarter cycle,

$$Level_{2009:03 \text{ to } 2013:09} = \begin{cases} 1 & \text{if } t \in [63, 117] \\ 0 & \text{otherwise} \end{cases},$$

$$Level_{2016:05 \text{ to } 2018:12} = \begin{cases} 1 & \text{if } t \in [149, 180] \\ 0 & \text{otherwise} \end{cases},$$

$$Strikes = \begin{cases} 1 & \text{if } t = 95 \\ 0 & \text{otherwise} \end{cases} \text{ and}$$

$$Easter = \begin{cases} -0,5 & \text{if } t = \text{Month before Easter (e.g., March 2004)} \\ 0,5 & \text{if } t = \text{Easter month (e.g., April 2004)} \\ -0,214 & \text{if } t = \text{Easter impact in March 2015} \\ 0,214 & \text{if } t = \text{Easter month in 2015 (April)} \\ -0,071 & \text{if } t = \text{Easter impact in March 2010} \\ 0,071 & \text{if } t = \text{Easter month in 2010 (April)} \\ -0,357 & \text{if } t = \text{Easter impact in March 2018} \\ 0,357 & \text{if } t = \text{Easter month in 2018 (April)} \\ 0 & \text{otherwise;} \end{cases}$$

The significance levels are shown in Table 9-1).

Table 9-1: Estimated coefficients and significance levels for log-linear model with Fourier and intervention terms.

Regressor (Model 19)	Est.	SE	t-value	p-value	Sig.
Intercept	14,3905	0,0102	1416,97	< 2e-16	***
Trend	0,0041	0,0001	34,034	< 2e-16	***
Fourier S ₁ -12	-0,1872	0,0064	-29,035	< 2e-16	***
Fourier C ₁ -12	-0,2719	0,0065	-41,906	< 2e-16	***
Fourier S ₂ -12	-0,0001	0,0065	-0,021	0.9837	(ns)
Fourier C ₂ -12	-0,0594	0,0064	-9,222	< 2e-16	***
Fourier S ₃ -12	0,016	0,0065	2,461	0.0149	*
Fourier C ₃ -12	0,0455	0,0064	7,062	4.14e-11	***
Level _{2009:03 to 2013:09}	-0,0589	0,0109	-5,403	2.22e-7	***
Easter	0,0694	0,0244	2,842	0.0050	**
Level _{2016:05 to 2018:12}	0,1965	0,0173	11,345	< 2e-16	***
Strikes	-0,1495	0,0626	-2,388	0.0180	*

Legend: $p < 0,001$ (***); $0,001 \leq p < 0,01$ (**); $0,01 \leq p < 0,05$ (*); $p \geq 0,05$ (ns)

The model is specified as a linear regression of a log-linear form and includes three Fourier terms (resulting in six seasonal components) to capture periodicity with varying harmonic frequencies, that representing seasonal patterns at intervals of 12, 6, and 4 months. The model has strong explanatory power with an $R^2=0,975$ (adjusted $R^2=0,974$) meaning it explains approximately 97.5% of the variance in the log-transformed passenger numbers with most predictors showing statistical significance.

The positive trend coefficient indicates a steady, underlying growth in passenger counts over time. this coefficient translates to a monthly increase of approximately $(e^{0,0041} - 1) \times 100 \approx 0,41\%$. For a baseline of 3.000.000 passengers, this corresponds to an additional 12.300 passengers per month, on average.

The independent variable $\text{Level}_{2009:03 \text{ to } 2013}$ is coded as a level shift to capture the effects of the financial crisis, with a significant negative coefficient ($-0,0589$; $p < 0,001$), indicating a reduction 5,72% reduction in passenger numbers during this period in relation to the previous periods. For the baseline, this reduction equates to around 171.600 fewer passengers each month.

Conversely, $\text{Level}_{2016:05 \text{ to } 2018:12}$, is coded as a level shift to capture air traffic growth from May 2016 up to December 2018, with a significant positive coefficient ($0,1965$; $p < 0,001$) point to an approximate 21,7% upward shift in passenger counts from May 2016 onward, relative to the preceding periods. This level shift suggests a structural change had a lasting effect, elevating passenger counts over the period. The starting date was chosen based on breakpoints detected previously in the residuals, while the end date aligns with forecasts of economic deceleration for 2019 (e.g. World Bank, 2018).

The dummy variable *Strikes* represents the general strike of November 2011, is included as an intervention term to capture the temporary disruption in travel demand caused by this event. The coefficient ($-0,1495$; $p = 0,018$) indicates a 13,88% reduction in passenger numbers during this period. For a baseline of 3.000.000 passengers, this reduction equates to approximately 416.400 fewer passengers.

The Easter coefficient of 0.0694 indicates a 7,19% increase in passenger numbers for each Easter effect. In terms of passenger number this equates to an expected increase of 215.700 passengers for a baseline of 3,000,000 passengers when the Easter effect is accounted in fully in one month.

Regarding model diagnostics and residual analysis (Figure 9-1, Table 9-2), the residual time plot (Figure 9-1, top-left panel) reveals no trend but there are some clustered deviations not captured by the model. This observation is consistent with the results of the Ljung-Box test, which returns a significant p-value ($p < 0,001$) indicating strong evidence of residual autocorrelation at the selected lags. Also, the ARCH Test yields a significant p-value ($p < 0,01$), suggesting the presence of conditional heteroscedasticity in the residuals, which is consistent with the funnel shape in the residuals versus fitted values plot (bottom-right panel), indicating increasing variability. Finally, the ACF plot (top-right panel) still shows significant autocorrelation at many lags.

This means the model fails to address autocorrelated noise even though the Fourier terms capture the seasonal cycles.

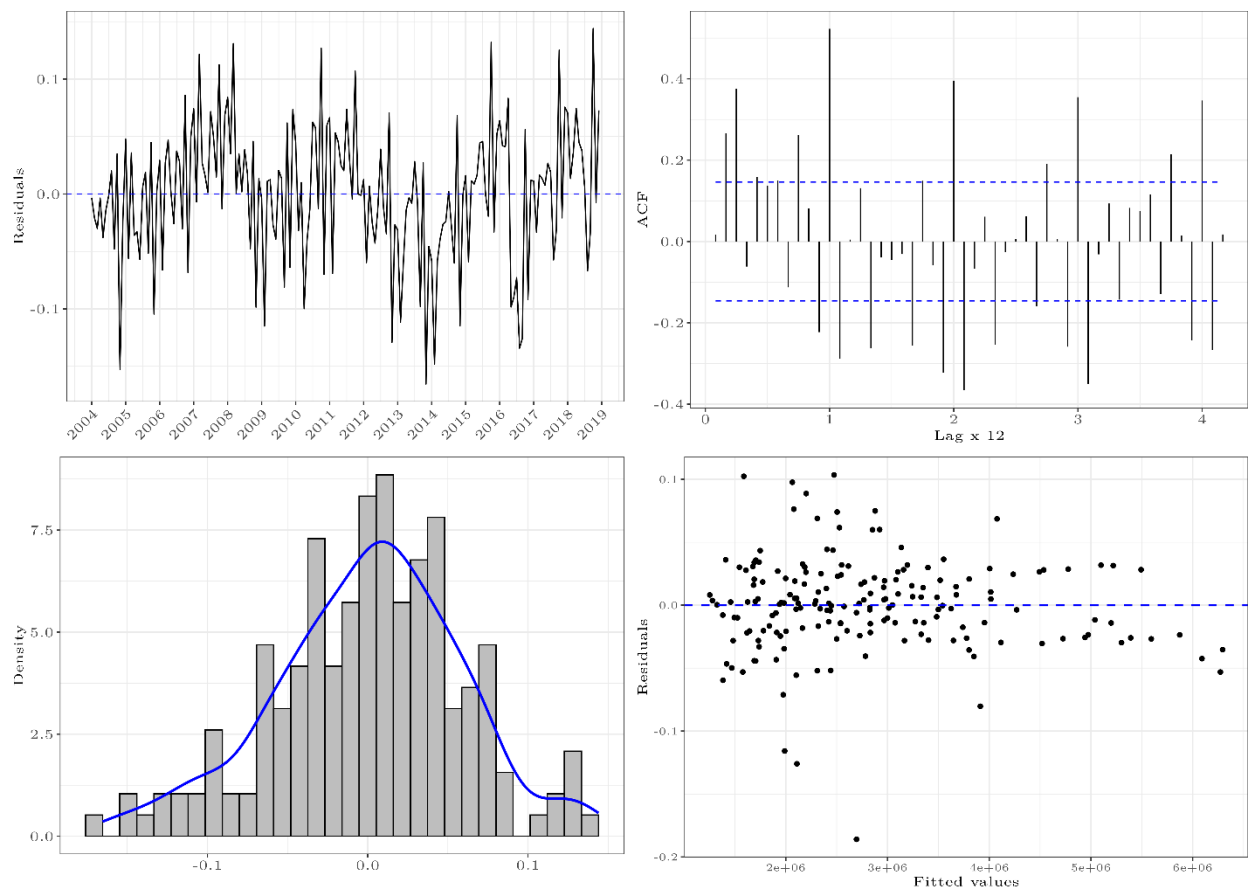


Figure 9-1: Residual diagnostics for log-linear model with Easter effect, Fourier terms and Intervention regressors.

The histogram of residuals (bottom-left panel) shows that the residuals are approximately centred around zero and the density curve suggests a near-normal distribution yet there may still be deviations from normality at the tails. This aligns with the Jarque-Bera test result ($p\text{-value}=0,38$), which indicates no strong evidence against normality. On the other hand, the ADF results, with a test statistic of $-3,87$ and a significant $p\text{-value}$ ($p<0,05$), indicate that the residuals are stationary.

Table 9-2: Diagnostic tests for residuals and model adequacy.

Statistic (Model 19)	Value/Sig.
Multiple R-Squared	0,9750
Adjusted R-Squared	0,9740
Residual Standard Error	0,0609
Box-Ljung Test (lag = 12)	$\chi^2=132,13$; $p < 0,001$ (***)
Augmented Dickey-Fuller Test	$DF=-3,8731$; $p=0,0169$ (*)
Breusch-Pagan Test	$BP=38,893$; $p<0,001$ (***)
ARCH Test	$\chi^2=30,94$; $p=0,002$ (**)
Jarque-Bera Test	$\chi^2=1,9429$; $p=0,3785$ (ns)
Ramsey RESET Test (2nd degree)	$RESET=4,4965$; $p=0,03544$ (*)

Legend: $p < 0,001$ (***) ; $0,001 \leq p < 0,01$ (**); $0,01 \leq p < 0,05$ (*); $p \geq 0,05$ (ns)

The Ramsey RESET Test was applied to assess whether the current model structure adequately captured the underlying dynamics of passenger numbers. The test yielded a result of 4,497 with a p-value of 0,0354, providing statistically significant evidence of model misspecification *i.e.*, there may be missing key functional components or relevant explanatory variables that could improve the fit. As a result, several configurations with quadratic terms for the trend were tested to model an accelerating growth rate of passengers.

These adjustments allowed the model to capture potential nonlinear evolution patterns. Consequently, the RESET test improved to a non-significant level, suggesting a better alignment with the data's underlying structure. However, while the RESET test results improved, the residual diagnostics showed heteroscedasticity, autocorrelation and deviations from normality. Based on these findings, we concluded that while addition of a quadratic term improved the model's specification, it compromised the residual behaviour quality.

Considering the residual diagnostics' importance, we applied the Newey-West adjusted standard errors to provide robust estimates that are resilient to heteroscedasticity and autocorrelation. This adjustment ensures that the significance tests for the coefficients remain valid.

Table 9-3: Estimated coefficients with Newey-West adjusted standard errors and significance levels for the log-linear model incorporating Fourier seasonal components and intervention terms.

Regressor (Model 19)	Est.	SE (N-W)	t-value	p-value	Sig.
Intercept	14,3905	0,0124	1.162,83	$< 2.2\text{e-}16$	***
Trend	0,0041	0,0002	22,58	$< 2.2\text{e-}16$	***
Fourier S1–12	-0,1872	0,0062	-30,30	$< 2.2\text{e-}16$	***
Fourier C1–12	-0,2719	0,0061	-44,78	$< 2.2\text{e-}16$	***
Fourier S2–12	-0,0001	0,0051	-0,03	0,9791	(ns)
Fourier C2–12	-0,0594	0,0039	-15,14	$< 2.2\text{e-}16$	***
Fourier S3–12	0,016	0,0041	3,91	0,0001	**
Fourier C3–12	0,0455	0,0046	9,87	$< 2.2\text{e-}16$	***
Level 2009:03 to 2013:09	-0,0589	0,0143	-4,12	5.98e-05	***
Easter	0,0694	0,017	4,09	6.67e-05	**
Level 2016:05 to 2018:12	0,1965	0,0324	6,06	8.67e-09	***
Strikes	-0,1495	0,0147	-10,16	$< 2.2\text{e-}16$	***

Legend: $p < 0,001$ (***); $0,001 \leq p < 0,01$ (**); $0,01 \leq p < 0,05$ (*); $p \geq 0,05$ (ns)

Therefore, the original model was retained despite its limitations in accounting for autoregressive structures, as it provided a balanced trade-off between model specification and residual stability.

10. FORECASTING AND ACCURACY

The evaluation of the models based on forecasting accuracy metrics shows that Model 12 and Model 14 are the two best-performing models in terms of both AICc and all forecasting error metrics, both in-sample and out-of-sample.

Model 12, specified as a regression with ARIMA (0,1,1)(0,1,1) errors, achieves the lowest in-sample forecasting errors, with an RMSE of 67.706 passengers, an MAE of 51.010 passengers, and a MAPE of 1.87%. This performance reflects the model's ability to capture the dynamics of the training data with precision, despite employing a more parsimonious structure with fewer external regressors. Model 12 incorporates the following external regressors: Easter, Leap Year, and five Additive Outliers detected by analyzing the remainder of the time series decomposition (March 2005, April 2005, March 2008, April 2008, April 2010).

Model 14, a regression with ARIMA (1,0,1)(1,1,0) errors, is also a viable model, achieving an out-of-sample RMSE of 101.748 passengers, an MAE of 74.010 passengers, and a MAPE of 1.5%. This model incorporates a broader set of external regressors, including Easter, Leap Year, Financial Crisis H6, strikes (2011:11 and 2019:04 in the out-of-sample period), the 2010 volcanic disruption at 2010:04, and three Additive Outliers (April 2005, March 2008, April 2008). However, despite the additional regressors, Model 14 does not achieve significantly better forecast accuracy compared to Model 12, as both models are statistically equivalent based on the Diebold-Mariano test ($p = 0.701$).

Given that Model 12 and Model 14 do not exhibit statistically significant differences in forecasting accuracy, Model 12 is preferable due to its simpler structure and superior RMSE performance, aligning with the principle of model parsimony.

Table 10-1: Model comparison of forecasting accuracy metrics.

Prep. Model 1	In-sample	Out-of-sample
Mean Error (ME)	-2.222	-211.559
Root Mean Squared Error (RMSE)	103.765	281.856
Mean Absolute Error (MAE)	74.400	218.575
Mean Percentage Error (MPE)	-0.11%	-3.80%
Mean Absolute Percentage Error (MAPE)	2.68%	3.96%
Mean Absolute Scaled Error (MASE)	0.22	0.64
Autocorrelation of Errors (ACF1)	0.07	n/a
Prep. Model 2	In-sample	Out-of-sample
Mean Error (ME)	-2.054	-189.755
Root Mean Squared Error (RMSE)	102.499	257.179
Mean Absolute Error (MAE)	71.170	197.661
Mean Percentage Error (MPE)	-0.07%	-3.39%
Mean Absolute Percentage Error (MAPE)	2.56%	3.57%
Mean Absolute Scaled Error (MASE)	0.21	0.58
Autocorrelation of Errors (ACF1)	0.18	n/a
Auto Arima (no external regressors)	In-sample	Out-of-sample
Mean Error (ME)	-3.389	-197.920
Root Mean Squared Error (RMSE)	92.806	262.205
Mean Absolute Error (MAE)	66.047	212.213
Mean Percentage Error (MPE)	-0.10%	-3.68%
Mean Absolute Percentage Error (MAPE)	2.43%	3.99%
Mean Absolute Scaled Error (MASE)	0.19	0.62
Autocorrelation of Errors (ACF1)	0.20	n/a
Model 12	In-sample	Out-of-sample
Mean Error (ME)	-3.659	-26.376
Root Mean Squared Error (RMSE)	67.706	91.950
Mean Absolute Error (MAE)	51.010	67.219
Mean Percentage Error (MPE)	-0.10%	-0.35%
Mean Absolute Percentage Error (MAPE)	1.87%	1.27%
Mean Absolute Scaled Error (MASE)	0.15	0.20
Autocorrelation of Errors (ACF1)	0.15	n/a
Model 10	In-sample	Out-of-sample
Mean Error (ME)	-3.334	-88.644
Root Mean Squared Error (RMSE)	81.822	164.262
Mean Absolute Error (MAE)	60.719	131.960
Mean Percentage Error (MPE)	-10.09%	-144.99%
Mean Absolute Percentage Error (MAPE)	218.94%	246.11%
Mean Absolute Scaled Error (MASE)	0.18	0.39
Autocorrelation of Errors (ACF1)	0.16	n/a
Model 13	In-sample	Out-of-sample
Mean Error (ME)	-3.371	-49.072
Root Mean Squared Error (RMSE)	76.366	145.097
Mean Absolute Error (MAE)	56.638	112.890
Mean Percentage Error (MPE)	-0.09%	-0.72%
Mean Absolute Percentage Error (MAPE)	2.02%	2.12%
Mean Absolute Scaled Error (MASE)	0.17	0.33
Autocorrelation of Errors (ACF1)	0.16	n/a

Model 14	In-sample	Out-of-sample
Mean Error (ME)	2.655	27.781
Root Mean Squared Error (RMSE)	68.952	101.748
Mean Absolute Error (MAE)	53.982	74.010
Mean Percentage Error (MPE)	0.17%	0.48%
Mean Absolute Percentage Error (MAPE)	1.98%	1.50%
Mean Absolute Scaled Error (MASE)	0.16	0.22
Autocorrelation of Errors (ACF1)	0.14	n/a
Model 15	In-sample	Out-of-sample
Mean Error (ME)	-3.832	-32.562
Root Mean Squared Error (RMSE)	76.096	143.893
Mean Absolute Error (MAE)	55.915	110.976
Mean Percentage Error (MPE)	-0.10%	-0.42%
Mean Absolute Percentage Error (MAPE)	2.00%	2.10%
Mean Absolute Scaled Error (MASE)	0.16	0.32
Autocorrelation of Errors (ACF1)	0.16	n/a
Model 11	In-sample	Out-of-sample
Mean Error (ME)	-3.533	-88.326
Root Mean Squared Error (RMSE)	75.738	167.218
Mean Absolute Error (MAE)	56.575	135.891
Mean Percentage Error (MPE)	-9.00%	-147.00%
Mean Absolute Percentage Error (MAPE)	205.00%	259.00%
Mean Absolute Scaled Error (MASE)	0.17	0.40
Autocorrelation of Errors (ACF1)	0.15	n/a
Model 17	In-sample	Out-of-sample
Mean Error (ME)	2.465	42.063
Root Mean Squared Error (RMSE)	73.196	115.120
Mean Absolute Error (MAE)	57.048	77.627
Mean Percentage Error (MPE)	0.17%	0.96%
Mean Absolute Percentage Error (MAPE)	2.08%	1.60%
Mean Absolute Scaled Error (MASE)	0.17	0.23
Autocorrelation of Errors (ACF1)	0.13	n/a
Model 18	In-sample	Out-of-sample
Mean Error (ME)	-3.648	-27.420
Root Mean Squared Error (RMSE)	74.361	134.541
Mean Absolute Error (MAE)	55.718	102.427
Mean Percentage Error (MPE)	-0.10%	-0.33%
Mean Absolute Percentage Error (MAPE)	1.99%	1.94%
Mean Absolute Scaled Error (MASE)	0.16	0.30
Autocorrelation of Errors (ACF1)	0.15	n/a
Model 19	In-sample	Out-of-sample
Mean Error (ME)	-2.196	1.109.798
Root Mean Squared Error (RMSE)	181.314	1.152.750
Mean Absolute Error (MAE)	129.075	1.109.798
Mean Percentage Error (MPE)	-0.35%	18.26%
Mean Absolute Percentage Error (MAPE)	4.61%	22.39%
Mean Absolute Scaled Error (MASE)	0.38	3.24
Autocorrelation of Errors (ACF1)	0.12	n/a

In order to evaluate the accuracy of forecasts produced by competing models, Diebold-Mariano tests were conducted between all pairs of models using one-step-ahead forecast errors, *i.e.*, for each month in the out-of-sample period, models were refitted using all available data up to the current point, generating a forecast solely for the next observation.

In Table 10-2) the smallest p-values indicate significant differences in forecasting performance. For example, Model 12 consistently outperforms others, reflected in their statistically significant differences (p-values<0,05) against less accurate forecasts. Notably, the

linear regression model (Model 19) displayed significant differences against most other models, suggesting it consistently performs worse when compared to alternative Models.

Table 10-2: Pairwise comparison of forecasting models using Diebold-Mariano test (One-Step Ahead).

	Auto	Model	Prep.	Prep.	Model	Model	Model	Model	Model	Model	Model
Model 19	0,000										
Prep. Model 1	0,366	0,000									
Prep. Model 2	0,784	0,000	0,015								
Model 13	0,027	0,000	0,043	0,054							
Model 15	0,037	0,000	0,053	0,068	0,883						
Model 17	0,047	0,000	0,059	0,074	0,277	0,148					
Model 18	0,032	0,000	0,047	0,060	0,234	0,049	0,310				
Model 14	0,030	0,000	0,043	0,052	0,103	0,028	0,230	0,074			
Model 12	0,014	0,000	0,027	0,030	0,008	0,010	0,476	0,014	0,701		
Model 10	0,038	0,000	0,053	0,071	0,030	0,052	0,096	0,027	0,030	0,005	
Model 11	0,050	0,000	0,070	0,097	0,019	0,013	0,060	0,005	0,019	0,003	0,711

Based on this, Model 12 and Model 14 emerge as the producers of the most accurate one-step-ahead forecasts. Between the one-step-ahead forecasts of these two models, a p-value of 0,701 indicates that there is no statistical evidence to reject the null hypothesis, hence supporting the conclusion that neither of the two models offer a clear advantage in terms of out-of-sample predictive performance.

11. DISCUSSION AND CONCLUSIONS

Airports are at the heart of air transportation, serving as pivotal nodes for passenger traffic. The modelling of passenger flows from an airport perspective has historically been central to air transport planning, capacity management, and policymaking. Up until February 2020, the evolution of passenger traffic at airports followed a relatively stable and predictable trajectory, shaped by many exogenous forces, (Strand, 1999, Graham, 2006). This apparent stability, however, does not imply an absence of disruptions over time. Major events such as the 9/11 terrorist attacks, the SARS outbreak, the global financial crisis, and oil price shocks all exerted measurable impacts on passenger flows.

Nonetheless, these events were generally moderate in intensity or short-lived, allowing the fundamental dynamics of passenger demand to persist and enabling conventional time series modelling techniques to remain reliable.

Our research demonstrates that airport passenger throughput possesses features that make it particularly well-suited for ARIMA modelling. These features include consistent seasonal patterns, calendar effects, a sustained long-term growth trend, and short-term fluctuations influenced by economic factors, population growth, and occasional disruptions. In most of the tested models, the differencing components effectively addressed the data's non-stationarity.

Even ARIMA specifications with no external regressors—or with minimal inclusion of external variables—produced good fits to the data. This robustness can be attributed to the inherent characteristics of airport passenger throughput, including stable and predictable seasonal patterns that evolve gradually over time, as well as a long-term trend that ARIMA models effectively captured despite the presence of structural breakpoints. While the explicit modelling these breakpoints did not result in significant improvements in modelling, the incorporation of intervention variables further enhanced model's accuracy.

Unsurprisingly, given the similarities with the well-known *AirPassengers* dataset, our model selection process revealed that, as additional external regressors were incorporated into the ARIMA models, the underlying structure of the models often conformed to the traditional airline model specification, ARIMA (0,1,1)(0,1,1). This finding highlights the ability of the airline model to remain a reliable baseline in the airline passenger contexts.

Regarding the selection of the interventions, we observed that calendar effects were significant across all models, which is expected given the nature of air travel demand. However, Easter effects were not fully captured as intended. This limitation was evidenced by the improved fitting of certain models when AO were explicitly incorporated for specific Easter periods. Even when applying the Chen and Liu (1993) outlier detection method some observations in March and April were flagged as outliers, suggesting that the Easter effect coding was insufficient to fully account for the specific impacts of Easter in those cases. These

findings suggest that Easter's impact may vary significantly from year to year, extending beyond standard calendar adjustments.

As for the crisis segment coding, we found potential limitations in attributing the observed effect exclusively to the financial crisis due to the presence of other concurrent structural events that may have influenced passenger numbers during the same period. For instance, the Arab Spring, which began in late 2010 and continued into 2011, likely redirected tourism flows towards Portuguese destinations as travellers sought alternatives to regions affected by political instability. Additionally, the already pessimistic outlook for the Portuguese economy prior to the global financial crisis makes the temporal delimitation of this segment more difficult, as some slowdown effects may have started earlier. Consequently, the assumption that the estimated coefficient captures only the financial crisis's impact may be violated indicating the coefficient could be biased, as it may inadvertently absorb the effects of other phenomena that coincide temporally with the crisis period.

Similarly, the terrorist attacks in Tunisia in June 2015 may have further enhanced the attractiveness of Portuguese destinations in the following years by deterring tourists from visiting competing Mediterranean markets. This shift in tourist flows, however, was not explicitly accounted for in the models due to lack of data.

To adequately isolate the impact of these two interventions, the inclusion of additional data, such as variables reflecting market dynamics (e.g., airline entry or re-entry, city-pair variations, hotel occupancy rates, and the number of nights stayed), would have been ideal. Without such data, the interpretation of the intervention coefficients becomes uncertain, reflecting a mix of multiple contemporaneous factors rather than a singular cause.

Another limitation of this study is the phenomenon of low-cost carriers (LCCs) within the European airline market, which was left unmodeled. Following the implementation of the Third Package of Air Transport liberalization measures in 1992, the European market underwent significant shifts, introducing new competitive dynamics. Our calculations show that in 2004, only 3% of passengers processed at Portuguese airports were transported by the two main low-cost carriers. By 2010, this figure had risen sharply to 23%, before declining slightly to 19% in 2019. A preliminary analysis reveals a pattern of rapid growth followed by stagnation, as the low-cost market segment consolidated among a few dominant carriers while legacy airlines adapted to the evolving competition.

Additionally, it is well established that lower time granularity increases the likelihood of other events influencing an intervention study (Ramsay et. Al., 2003). In this context, the interventions, predominantly coded as AO, were not consistently significant, as the temporal aggregation process smoothed out much of their impact, reducing their detectability. Since the data is monthly, significant interventions must result in substantial and prolonged effects to show a visible disruption. The effect of smaller or short-lived events might be distributed across the monthly timeframe often smoothed out as flights are typically rescheduled quickly,

limiting the visible impact in aggregated monthly data. Consequently, some disruptions may have been masked, showing a trade-off between data aggregation and the ability to detect effects.

The breakpoints we tested were non-significant in most cases due to the ARIMA process effectively capturing the underlying dynamics of the time series. The ARIMA model's differencing components removed much of the trend and seasonal variability, which often diminishes the apparent impact of structural changes or temporary disruptions. Additionally, the AR and MA terms modelled much of the short-term dependencies and residual correlations, further smoothing the series and reducing the statistical significance of most breakpoints and segments. This suggests that while breakpoints may exist in the time series, their influence is rendered negligible within the ARIMA framework.

Regarding the forecasting performance, most of the tested models delivered highly accurate 12-month forecasts (Appendix E) and exhibited some generalizability, provided that the conditions remained consistent with those observed before March 2020.

ARIMA models without intervention regressors, such as “Auto Arima Model”, “Prep. Model 1” and “Prep. Model 2” performed reasonably well in the training data but their performance degraded substantially when applied to the out-of-sample data.

The Linear regression with Fourier components mode (Model 19) proved to be the weakest of all models. Although it captured some periodic trends in the training phase, with a MAPE of 22,39%, it performed poorly with the out-of-sample data.

The model with the best out-of-sample accuracy, as indicated by its optimal balance of low RMSE, MAE, and MAPE values, was Model 12, a regression with ARIMA (0,1,1)(0,1,1) errors. This model achieved the lowest in-sample MAPE, and a Diebold-Mariano test comparing its performance with the second best model, indicates that the difference in predictive performance between the best two models is not statistically significant.

The second best model in terms of out-of-sample accuracy was Model 14, a regression with ARIMA (1,0,1)(1,1,0) errors. This model deviates from the traditional ARIMA differencing approach ($d = 1$) used in other models to address the trend. The absence of differencing to remove the trend in this case can be attributed to the combined effect of the external regressors and the autoregressive (AR) term, which together captured the underlying dynamics of the data and rendered the residuals stationary.

Finally, a significant limitation of these models is the implicit assumption that airport passenger throughput can grow without bounds. However, the growth of air passenger traffic is constrained by the physical capacity of airports and the limitations of associated infrastructures. This issue is particularly acute when airports, such as Lisbon, approach the limit of operational capacity, creating bottlenecks that constrain growth and limit passenger throughput.

12. FUTURE RESEARCH

Future research directions emerge by investigating higher-frequency datasets and delving into the cases of specific airports.

Increasing the temporal resolution by using daily or weekly data would facilitate the identification of more granular patterns and disruptions that remain obscured in monthly aggregates.

Another promising line of research could focus on investigating the specificities of certain airports, particularly those located in autonomous regions. The airports in the Azores, for example, present interesting operational characteristics due to the prevalence of circular routes, where flights serve multiple islands in a single operation. This networked configuration could be studied using network modelling techniques to analyse the air connectivity between islands.

Furthermore, the autonomous regions frequently experience weather-related disruptions, with entire days of limited or cancelled flights due to adverse conditions. Future research could use the flight timestamps to explore whether the frequency or severity of weather-related delays has increased over time, contributing to the broader discussion on climate variability. Quantifying these disruptions would allow to estimate the cumulative impacts of weather-related delays and cancellations, including passenger time losses, airline operational costs, and broader economic effects.

However, transitioning to higher-frequency data or flight-level analyses introduces significant methodological challenges. Firstly, there is a critical need for robust methodological approaches to manage the variability and complexity inherent in high-frequency data.

Secondly, access to detailed flight-level data is often restricted due to confidentiality concerns and represents a significant obstacle, particularly for routes served by a limited number of airlines. In addition, the inconsistent availability and reliability of flight cancellation data further complicate the analysis.

REFERENCES

- Abraham, B. (1980). *Intervention Analysis and Multiple Time Series*. Biometrika. <https://doi.org/10.2307/2335318>
- Aggarwal, C. C. (2017). *Outlier Analysis*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-47578-3>
- Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19(6). <https://doi.org/10.1109/TAC.1974.1100705>
- Akouemo, H. N., & Povinelli, R. J. (2014). Time series outlier detection and imputation. *IEEE Power and Energy Society General Meeting*, 2014-October(October). <https://doi.org/10.1109/PESGM.2014.6939802>
- Angiulli, F., & Fasseti, F. (2007). Detecting distance-based outliers in streams of data. *International Conference on Information and Knowledge Management, Proceedings*. <https://doi.org/10.1145/1321440.1321552>
- Atkinson, A. C., & Hawkins, D. M. (1981). Identification of Outliers. *Biometrics*, 37(4). <https://doi.org/10.2307/2530182>
- Balke, N. S. (1993). Detecting Level Shifts In Time Series. *Journal of Business and Economic Statistics*. <https://doi.org/10.1080/07350015.1993.10509934>
- Battaglia, F., & Orfei, L. (2005). Outlier detection and estimation in nonlinear time series. *Journal of Time Series Analysis*. <https://doi.org/10.1111/j.1467-9892.2005.00392.x>
- Becker, C., Fried, R., & Kuhnt, S. (2013). Robustness and complex data structures: Festschrift in honour of Ursula Gather. In *Robustness and Complex Data Structures: Festschrift in Honour of Ursula Gather*. <https://doi.org/10.1007/978-3-642-35494-6>
- Bee Dagum, E., & Bianconcini, S. (2016). *Seasonal Adjustment Methods and Real Time Trend-Cycle Estimation*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-31822-6>
- Beria, P., & Laurino, A. (2016). Determinants of daily fluctuations in air passenger volumes. The effect of events and holidays on Milan Malpensa airport. *Journal of Air Transport Management*, 53, 73–84. <https://doi.org/10.1016/j.jairtraman.2016.01.005>
- Blázquez-García, A., Conde, A., Mori, U., & Lozano, J. A. (2021). A Review on Outlier/Anomaly Detection in Time Series Data. *ACM Computing Surveys*, 54(3). <https://doi.org/10.1145/3444690>
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). Wiley. ISBN: 9781119202101

- Box, G. E. P., & Tiao, G. C. (1965). A Change in Level of a Non-Stationary Time Series. *Biometrika*, 52(1/2), 181. <https://doi.org/10.2307/2333823>
- Box, G. E. P., & Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*. <https://doi.org/10.1080/01621459.1975.10480264>
- Box, G. E. P., & Tiao, G. C. (1976). Comparison of Forecast and Actuality. *Applied Statistics*, 25(3), 195. <https://doi.org/10.2307/2347226>
- Bruce, A. G., & Martin, R. D. (1989). Leave-K-Out Diagnostics for Time Series. *Journal of the Royal Statistical Society: Series B (Methodological)*, 51(3), 363–401. <https://doi.org/10.1111/j.2517-6161.1989.tb01435.x>
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. In *Sociological Methods and Research* (Vol. 33, Issue 2). <https://doi.org/10.1177/0049124104268644>
- Campbell, D. (1963). From description to experimentation: interpreting trends as quasi-experiments. In C. W. Harris (Ed.), *Problems in Measuring Change*. Madison.
- Campbell, D., & Stanley, J. (1966). *Experimental and Quasi-Experimental Designs for Research*. (H. M. Company, Ed.).
- Canova, F., & Hansen, B. E. (1995). Are seasonal patterns constant over time? A test for seasonal stability. *Journal of Business and Economic Statistics*, 13(3). <https://doi.org/10.1080/07350015.1995.10524598>
- Chan, W. S. (1995). Understanding the effect of time series outliers on sample autocorrelations. *Test*, 4(1). <https://doi.org/10.1007/BF02563108>
- Chang, I., Tiao, G. C., & Chen, C. (1988). Estimation of time series parameters in the presence of outliers. *Technometrics*. <https://doi.org/10.1080/00401706.1988.10488367>
- Chen, C., & Liu, L.-M. (1993). Joint Estimation of Model Parameters and Outlier Effects in Time Series. *Journal of the American Statistical Association*. <https://doi.org/10.2307/2290724>
- Chen, C., & Tiao, G. C. (1990). Random level-shift time series models, ARIMA approximations, and level-shift detection. *Journal of Business and Economic Statistics*. <https://doi.org/10.1080/07350015.1990.10509779>
- Cryer, J. D., & Chan, K.-S. (2008). *Time Series Analysis with applications in R* (Second edi). Springer Science.
- Dalla, V., Giraitis, L., & Phillips, P. C. B. (2022). Robust tests for white noise and cross-correlation. *Econometric Theory*, 38(5). <https://doi.org/10.1017/S0266466620000341>

- Das, P. (2019). Econometrics in Theory and Practice. In Econometrics in Theory and Practice. <https://doi.org/10.1007/978-981-32-9019-8>
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, 74(366), 427. <https://doi.org/10.2307/2286348>
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3). <https://doi.org/10.1080/07350015.1995.10524599>
- Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4). <https://doi.org/10.2307/1912773>
- EUROSTAT. (2015). ESS guidelines on seasonal adjustment (2015th ed.). Publications Office of the European Union. <https://doi.org/10.2785/317290>
- Ford, B. L. (1983). An overview of hot-deck procedures. In *Incomplete data in sample surveys*, Volume 2, Part IV.
- Fox, A. J. (1972). Outliers in Time Series. *Journal of the Royal Statistical Society: Series B (Methodological)*. <https://doi.org/10.1111/j.2517-6161.1972.tb00912.x>
- Garcia-Ferrer, A. (2003a). A Course in Time Series Analysis, Daniel Peña, George C. Tiao and Ruey S. Tsay (.... *International Journal of Forecasting*.
- Garcia-Ferrer, A. (2003b). A Course in Time Series Analysis, Daniel Peña, George C. Tiao and Ruey S. Tsay (.... *International Journal of Forecasting*.
- Glass, G. V. (1972). Estimating the Effects of Intervention Into a Non-stationary Time-Series. *American Educational Research Journal*, 9(3), 463–477. <https://doi.org/10.3102/00028312009003463>
- Graham, A. (2006). Have the major forces driving leisure airline traffic changed? *Journal of Air Transport Management*, 12(1). <https://doi.org/10.1016/j.jairtraman.2005.09.002>
- Hernandez, G. (1999). Time series, periodograms, and significance. *Journal of Geophysical Research: Space Physics*, 104(A5). <https://doi.org/10.1029/1999ja900026>
- Hillmer, S. (1984). Monitoring and adjusting forecasts in the presence of additive outliers. *Journal of Forecasting*, 3(2). <https://doi.org/10.1002/for.3980030208>
- Hillmer, S. C., & Wei, W. W. S. (2006). Time Series Analysis: Univariate and Multivariate Methods. *Journal of the American Statistical Association*, 86(413), 245. <https://doi.org/10.2307/2289741>

- Hochenbaum, J., Vallis, O. S., & Kejariwal, A. (2017). Automatic Anomaly Detection in the Cloud Via Statistical Learning.
- Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2). <https://doi.org/10.1093/biomet/76.2.297>
- Hylleberg, S., Engle, R. F., Granger, C. W. J., & Yoo, B. S. (1990). Seasonal integration and cointegration. *Journal of Econometrics*, 44(1–2). [https://doi.org/10.1016/0304-4076\(90\)90080-D](https://doi.org/10.1016/0304-4076(90)90080-D)
- Hyndman, R. (2010). Forecasting with long seasonal periods. <https://robjhyndman.com/hyndsight/longseasonality/> Accessed on 01 May 2024
- Hyndman, R. J. (2021). Detecting time series outliers. <https://robjhyndman.com/hyndsight/tsoutliers/> Accessed on 01 June 2023
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. Principles of Optimal Design. <https://doi.org/10.1017/9781316451038.010>
- Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(3). <https://doi.org/10.18637/jss.v027.i03>
- International Civil Aviation Organization. (2016). *Manual on the Regulation of International Air Transport - Third edition*. International Civil Aviation Organization (ICAO) Journal, Doc 9626, 218.
- Ismail, Z., Suhartono, Yahaya, A., & Efendi, R. (2009). Intervention model for analyzing the impact of terrorism to tourism industry. *Journal of Mathematics and Statistics*. <https://doi.org/10.3844/jmssp.2009.322.329>
- Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6(3). [https://doi.org/10.1016/0165-1765\(80\)90024-5](https://doi.org/10.1016/0165-1765(80)90024-5)
- Joseph Beaulieu, J., & Miron, J. A. (1993). Seasonal unit roots in aggregate U.S. data. *Journal of Econometrics*, 55(1–2). [https://doi.org/10.1016/0304-4076\(93\)90018-Z](https://doi.org/10.1016/0304-4076(93)90018-Z)
- Kaiser, R., & Maravall, A. (1999). *Seasonal Outliers in Time Series*. Statistics and Econometrics Series 15.
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54(1–3). [https://doi.org/10.1016/0304-4076\(92\)90104-y](https://doi.org/10.1016/0304-4076(92)90104-y)

- Ledolter, J. (1989). The effect of additive outliers on the forecasts from ARIMA models. *International Journal of Forecasting*, 5(2). [https://doi.org/10.1016/0169-2070\(89\)90090-3](https://doi.org/10.1016/0169-2070(89)90090-3)
- Li, Z., Bi, J., & Li, Z. (2017). Passenger Flow Forecasting Research for Airport Terminal Based on SARIMA Time Series Model. *IOP Conference Series: Earth and Environmental Science*, 100(1). <https://doi.org/10.1088/1755-1315/100/1/012146>
- Linden, A. (2015). Conducting interrupted time-series analysis for single- and multiple-group comparisons. *Stata Journal*.
- Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2). <https://doi.org/10.1093/biomet/65.2.297>
- McCoy, E. J., & Stephens, D. A. (2004). Bayesian time series analysis of periodic behaviour and spectral structure. *International Journal of Forecasting*, 20(4), 713–730. <https://doi.org/10.1016/j.ijforecast.2004.02.001>
- McDowall, D., McCleary, R., Meidinger, E., & Hay, R. (2011). Interrupted Time Series Analysis. In *Interrupted Time Series Analysis*. Oxford University Press, USA. <https://doi.org/10.4135/9781412984607>
- Meyer, S. M., McDowall, D., McCleary, R., Meidinger, E. E., & Hay, R. (1981). Interrupted Time-Series Analysis. *Journal of Policy Analysis and Management*. <https://doi.org/10.2307/3324157>
- Mills, T. C. (2019). Applied Time Series Analysis: A Practical Guide to Modeling and Forecasting. In *Applied Time Series Analysis: A Practical Guide to Modeling and Forecasting*. <https://doi.org/10.1016/C2016-0-03956-6>
- Montgomery, D. C., & Weatherby, G. (1980). Modeling and forecasting time series using transfer function and intervention methods. *AIIE Transactions*. <https://doi.org/10.1080/05695558008974521>
- Morettin, P., & Toloj, C. (2018). *Análise de Séries Temporais: Modelos Lineares Univariados* (Third edition). Edgar Blucher.
- Muthukrishnan, S., Shah, R., & Vitter, J. S. (2004). Mining deviants in time series data streams. *Proceedings of the International Conference on Scientific and Statistical Database Management, SSDBM*, 16. <https://doi.org/10.1109/ssdm.2004.1311192>
- Myers, M., Avison, D., Benbasat, I., Goldstein, D. K., & Mead, M. (2011). The Case Research Strategy in Studies of Information Systems. In *Qualitative Research in Information Systems*. <https://doi.org/10.4135/9781849209687.n5>
- Neusser, C. (2016). *Time series econometrics*. Springer International Publishing Switzerland.

- Newbold, P., Glass, G. V., Willson, V. L., & Gottman, J. M. (1976). Design and Analysis of Time-Series Experiments. *Journal of the American Statistical Association*, 71(353), 247. <https://doi.org/10.2307/2285785>
- Ng, S., & Perron, P. (1995). Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association*, 90(429). <https://doi.org/10.1080/01621459.1995.10476510>
- Nieto, M. R., & Carmona-Benítez, R. B. (2018). ARIMA + GARCH + Bootstrap forecasting method applied to the airline industry. *Journal of Air Transport Management*, 71. <https://doi.org/10.1016/j.jairtraman.2018.05.007>
- Osborn, D. R., Chui, A. P. L., Smith, J. P., & Birchenhall, C. R. (1988). SEASONALITY AND THE ORDER OF INTEGRATION FOR CONSUMPTION. *Oxford Bulletin of Economics and Statistics*, 50(4). <https://doi.org/10.1111/j.1468-0084.1988.mp50004002.x>
- Pankratz, A. E. (1991). *Forecasting with Dynamic Regression Models*. John Wiley & Sons, Inc.
- Payne, J. E., & Taylor, J. P. (2007). Modelling and forecasting airport passengers: a case study for an introductory forecasting course. *International Journal of Information and Operations Management Education*, 2(2). <https://doi.org/10.1504/ijiome.2007.015282>
- Pena, D. (2001). Outliers, Influential Observations, and Missing Data. In D. Pena, G. Tiao, & R. Tsay (Eds.), *A course in time series analysis*. John Wiley & Sons, INC.
- Ramsay, C. R., Matowe, L., Grilli, R., Grimshaw, J. M., & Thomas, R. E. (2003). Interrupted time series designs in health technology assessment: Lessons from two systematic reviews of behavior change strategies. In *International Journal of Technology Assessment in Health Care*. <https://doi.org/10.1017/S0266462303000576>
- Reddy, A., Ordway-West, M., Lee, M., Dugan, M., Whitney, J., Kahana, R., Ford, B., Muedsam, J., Henslee, A., & Rao, M. (2017). Using Gaussian mixture models to detect outliers in seasonal univariate network traffic. *Proceedings - 2017 IEEE Symposium on Security and Privacy Workshops, SPW 2017, 2017-December*. <https://doi.org/10.1109/SPW.2017.9>
- Reinsel, G., Tiao, G. C., Wang, M. N., Lewis, R., & Nychka, D. (1981). Statistical analysis of stratospheric ozone data for the detection of trends. *Atmospheric Environment* (1967), 15(9). [https://doi.org/10.1016/0004-6981\(81\)90140-2](https://doi.org/10.1016/0004-6981(81)90140-2)
- Richards, L. E., Little, R. J. A., & Rubin, D. B. (1989). Statistical Analysis with Missing Data. *Journal of Marketing Research*. <https://doi.org/10.2307/3172915>
- Said, S. E., & Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3). <https://doi.org/10.1093/biomet/71.3.599>

- Schuster, A. (1898). On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. *Journal of Geophysical Research*, 3(1). <https://doi.org/10.1029/tm003i001p00013>
- Schwarz, G. (1978). "Estimating the Dimension of a Model." *The Annals of Statistics*, 6(2). <https://doi.org/10.2307/2958889>
- Shumway, R. H., & Stoffer, D. S. (2019). Time Series: A Data Analysis Approach Using R. In *Time Series: A Data Analysis Approach Using R*. <https://doi.org/10.1201/9780429273285>
- Strand, S. (1999). Airport-specific traffic forecasts: The resultant of local and non-local forces. *Journal of Transport Geography*, 7(1), 17–29. [https://doi.org/10.1016/S0966-6923\(98\)00036-2](https://doi.org/10.1016/S0966-6923(98)00036-2)
- Tiao, G. C., Box, G. E. P., & Box, G. E. P. (1975). Analysis of los angeles photochemical smog data: A statistical overview. *Journal of the Air Pollution Control Association*, 25(3). <https://doi.org/10.1080/00022470.1975.10470082>
- Topál, D., Matyasovszky, I., Kern, Z., & Hatvani, I. G. (2016). Detecting breakpoints in artificially modified- and real-life time series using three state-of-the-art methods. *Open Geosciences*, 8(1), 78–98. <https://doi.org/10.1515/geo-2016-0009>
- Tsay, R. S. (1986). Time series model specification in the presence of outliers. *Journal of the American Statistical Association*. <https://doi.org/10.1080/01621459.1986.10478250>
- Tsay, R. S. (1988). Outliers, level shifts, and variance changes in time series. *Journal of Forecasting*. <https://doi.org/10.1002/for.3980070102>
- Tsay, R. S. (2010). *Analysis of Financial Time Series, Third Edition*, Willey Series in Probability and Statistics. In *Health Services Management Research* (Vol. 4, Issue 3).
- Tsay, R. S., Peña, D., & Pankratz, A. E. (2000). Outliers in multivariate time series. *Biometrika*, 87(4). <https://doi.org/10.1093/biomet/87.4.789>
- Tsay, R. S., & Tiao, G. C. (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models. *Journal of the American Statistical Association*. <https://doi.org/10.1080/01621459.1984.10477068>
- Tsui, W. H. K., Ozer Balli, H., Gilbey, A., & Gow, H. (2014). Forecasting of Hong Kong airport's passenger throughput. *Tourism Management*, 42. <https://doi.org/10.1016/j.tourman.2013.10.008>
- Velicer, W. F., & McDonald, R. P. (1991). Cross-Sectional Time Series Designs: A General Transformation Approach. *Multivariate Behavioral Research*. https://doi.org/10.1207/s15327906mbr2602_3

- Vincenty, T. (1975). Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Survey Review*, 23(176).
<https://doi.org/10.1179/sre.1975.23.176.88>
- Wei, W. W. S. (2006). Time Series Analysis: Univariate and Multivariate Methods Second Edition. In Pearson Education, Inc. <https://doi.org/10.1016/j.ijforecast.2004.02.001>
- Woodward, W. A., Gray, H. L., & Elliott, A. C. (2017). Applied time series analysis with R, second edition. In Applied Time Series Analysis with R, Second Edition. <https://doi.org/10.1201/9781315161143>
- World Bank. (2018). Global Economic Prospects, June 2018: The Turning of the Tide? In Global Economic Prospects, June 2018: The Turning of the Tide? (Issue January). <https://doi.org/10.1596/978-1-4648-1257-6>
- Yaffee, R., & McGee, M. (2000). Introduction to Time Series Analysis and Forecasting: With Applications of SAS and SPSS. With Applications of SAS and SPSS. ACADEMIC PRESS, INC.
- Zeileis, A., Kleiber, C., Walter, K., & Hornik, K. (2003). Testing and dating of structural changes in practice. *Computational Statistics and Data Analysis*, 44(1–2).
[https://doi.org/10.1016/S0167-9473\(03\)00030-6](https://doi.org/10.1016/S0167-9473(03)00030-6)
- Zhang, Y., Hamm, N. A. S., Meratnia, N., Stein, A., van de Voort, M., & Havinga, P. J. M. (2012). Statistics-based outlier detection for wireless sensor networks. *International Journal of Geographical Information Science*, 26(8).
<https://doi.org/10.1080/13658816.2012.654493>
- Abraham, B. (1980). Intervention Analysis and Multiple Time Series. *Biometrika*.
<https://doi.org/10.2307/2335318>
- Aggarwal, C. C. (2017). *Outlier Analysis*. Springer International Publishing.
<https://doi.org/10.1007/978-3-319-47578-3>
- Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19(6). <https://doi.org/10.1109/TAC.1974.1100705>
- Akouemo, H. N., & Povinelli, R. J. (2014). Time series outlier detection and imputation. *IEEE Power and Energy Society General Meeting, 2014-October*(October).
<https://doi.org/10.1109/PESGM.2014.6939802>
- Angiulli, F., & Fasseti, F. (2007). Detecting distance-based outliers in streams of data. *International Conference on Information and Knowledge Management, Proceedings*.
<https://doi.org/10.1145/1321440.1321552>
- Atkinson, A. C., & Hawkins, D. M. (1981). Identification of Outliers. *Biometrics*, 37(4).
<https://doi.org/10.2307/2530182>

- Balke, N. S. (1993). Detecting Level Shifts In Time Series. *Journal of Business and Economic Statistics*. <https://doi.org/10.1080/07350015.1993.10509934>
- Battaglia, F., & Orfei, L. (2005). Outlier detection and estimation in nonlinear time series. *Journal of Time Series Analysis*. <https://doi.org/10.1111/j.1467-9892.2005.00392.x>
- Becker, C., Fried, R., & Kuhnt, S. (2013). Robustness and complex data structures: Festschrift in honour of Ursula Gather. In *Robustness and Complex Data Structures: Festschrift in Honour of Ursula Gather*. <https://doi.org/10.1007/978-3-642-35494-6>
- Bee Dagum, E., & Bianconcini, S. (2016). *Seasonal Adjustment Methods and Real Time Trend-Cycle Estimation*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-31822-6>
- Beria, P., & Laurino, A. (2016). Determinants of daily fluctuations in air passenger volumes. The effect of events and holidays on Milan Malpensa airport. *Journal of Air Transport Management*, 53, 73–84. <https://doi.org/10.1016/j.jairtraman.2016.01.005>
- Blázquez-García, A., Conde, A., Mori, U., & Lozano, J. A. (2021). A Review on Outlier/Anomaly Detection in Time Series Data. *ACM Computing Surveys*, 54(3). <https://doi.org/10.1145/3444690>
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time Series Analysis: Forecasting and Control, 5th Edition, by. In *Journal of Time Series Analysis*. <https://doi.org/10.1111/jtsa.12194>
- Box, G. E. P., & Tiao, G. C. (1965). A Change in Level of a Non-Stationary Time Series. *Biometrika*, 52(1/2), 181. <https://doi.org/10.2307/2333823>
- Box, G. E. P., & Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*. <https://doi.org/10.1080/01621459.1975.10480264>
- Box, G. E. P., & Tiao, G. C. (1976). Comparison of Forecast and Actuality. *Applied Statistics*, 25(3), 195. <https://doi.org/10.2307/2347226>
- Bruce, A. G., & Martin, R. D. (1989). Leave- K -Out Diagnostics for Time Series. *Journal of the Royal Statistical Society: Series B (Methodological)*, 51(3), 363–401. <https://doi.org/10.1111/j.2517-6161.1989.tb01435.x>
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. In *Sociological Methods and Research* (Vol. 33, Issue 2). <https://doi.org/10.1177/0049124104268644>
- Campbell, D. (1963). From description to experimentation: interpreting trends as quasi-experiments. In C. W. Harris (Ed.), *Problems in Measuring Change*. Madison.

- Campbell, D., & Stanley, J. (1966). *Experimental and Quasi-Experimental Designs for Research*. (H. M. Company, Ed.).
- Canqva, F., & Hansen, B. E. (1995). Are seasonal patterns constant over time? A test for seasonal stability. *Journal of Business and Economic Statistics*, 13(3). <https://doi.org/10.1080/07350015.1995.10524598>
- Chan, W. S. (1995). Understanding the effect of time series outliers on sample autocorrelations. *Test*, 4(1). <https://doi.org/10.1007/BF02563108>
- Chang, I., Tiao, G. C., & Chen, C. (1988a). Estimation of time series parameters in the presence of outliers. *Technometrics*. <https://doi.org/10.1080/00401706.1988.10488367>
- Chang, I., Tiao, G. C., & Chen, C. (1988b). Estimation of time series parameters in the presence of outliers. *Technometrics*. <https://doi.org/10.1080/00401706.1988.10488367>
- Chang, I., Tiao, G. C., & Chen, C. (1988c). Estimation of time series parameters in the presence of outliers. *Technometrics*. <https://doi.org/10.1080/00401706.1988.10488367>
- Chen, C., & Liu, L.-M. (1993a). Joint Estimation of Model Parameters and Outlier Effects in Time Series. *Journal of the American Statistical Association*. <https://doi.org/10.2307/2290724>
- Chen, C., & Liu, L.-M. (1993b). Joint Estimation of Model Parameters and Outlier Effects in Time Series. *Journal of the American Statistical Association*. <https://doi.org/10.2307/2290724>
- Chen, C., & Liu, L.-M. (1993c). Joint Estimation of Model Parameters and Outlier Effects in Time Series. *Journal of the American Statistical Association*. <https://doi.org/10.2307/2290724>
- Chen, C., & Liu, L.-M. (2006). Joint Estimation of Model Parameters and Outlier Effects in Time Series. *Journal of the American Statistical Association*. <https://doi.org/10.2307/2290724>
- Chen, C., & Tiao, G. C. (1990). Random level-shift time series models, ARIMA approximations, and level-shift detection. *Journal of Business and Economic Statistics*. <https://doi.org/10.1080/07350015.1990.10509779>
- Cryer, J. D., & Chan, K.-S. (2008). *Time Series Analysis with applications in R* (Second edi). Springer Science.
- Dalla, V., Giraitis, L., & Phillips, P. C. B. (2022). ROBUST TESTS FOR WHITE NOISE AND CROSS-CORRELATION. *Econometric Theory*, 38(5). <https://doi.org/10.1017/S0266466620000341>
- Das, P. (2019). Econometrics in Theory and Practice. In *Econometrics in Theory and Practice*. <https://doi.org/10.1007/978-981-32-9019-8>

- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, 74(366), 427. <https://doi.org/10.2307/2286348>
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3). <https://doi.org/10.1080/07350015.1995.10524599>
- Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4). <https://doi.org/10.2307/1912773>
- EUROSTAT. (2015). *ESS guidelines on seasonal adjustment* (2015th ed.). Publications Office of the European Union. <https://doi.org/10.2785/317290>
- Ford, B. L. (1983). An overview of hot-deck procedures. In *Incomplete data in sample surveys, Volume 2, Part IV*.
- Fox, A. J. (1972). Outliers in Time Series. *Journal of the Royal Statistical Society: Series B (Methodological)*. <https://doi.org/10.1111/j.2517-6161.1972.tb00912.x>
- Garcia-Ferrer, A. (2003a). A Course in Time Series Analysis, Daniel Peña, George C. Tiao and Ruey S. Tsay (.... *International Journal of Forecasting*.
- Garcia-Ferrer, A. (2003b). A Course in Time Series Analysis, Daniel Peña, George C. Tiao and Ruey S. Tsay (.... *International Journal of Forecasting*.
- Glass, G. V. (1972). Estimating the Effects of Intervention Into a Non-stationary Time-Series. *American Educational Research Journal*, 9(3), 463–477. <https://doi.org/10.3102/00028312009003463>
- Graham, A. (2006). Have the major forces driving leisure airline traffic changed? *Journal of Air Transport Management*, 12(1). <https://doi.org/10.1016/j.jairtraman.2005.09.002>
- Hernandez, G. (1999). Time series, periodograms, and significance. *Journal of Geophysical Research: Space Physics*, 104(A5). <https://doi.org/10.1029/1999ja900026>
- Hillmer, S. (1984). Monitoring and adjusting forecasts in the presence of additive outliers. *Journal of Forecasting*, 3(2). <https://doi.org/10.1002/for.3980030208>
- Hillmer, S. C., & Wei, W. W. S. (2006). Time Series Analysis: Univariate and Multivariate Methods. *Journal of the American Statistical Association*, 86(413), 245. <https://doi.org/10.2307/2289741>
- Hochenbaum, J., Vallis, O. S., & Kejariwal, A. (2017). *Automatic Anomaly Detection in the Cloud Via Statistical Learning*.

- Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2). <https://doi.org/10.1093/biomet/76.2.297>
- Hylleberg, S., Engle, R. F., Granger, C. W. J., & Yoo, B. S. (1990). Seasonal integration and cointegration. *Journal of Econometrics*, 44(1–2). [https://doi.org/10.1016/0304-4076\(90\)90080-D](https://doi.org/10.1016/0304-4076(90)90080-D)
- Hyndman, R. (2010). *Forecasting with long seasonal periods*. <https://robjhyndman.com/hyndsight/longseasonality/>
- Hyndman, R. J. (2021). *Detecting time series outliers*. <https://robjhyndman.com/hyndsight/tsoutliers/>
- Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice. *Principles of Optimal Design*. <https://doi.org/10.1017/9781316451038.010>
- Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(3). <https://doi.org/10.18637/jss.v027.i03>
- ICAO, & International Civil Aviation Organization. (2016). Manual on the Regulation of International Air Transport - Third edition. *International Civil Aviation Organization (ICAO) Journal, Doc 9626*, 218.
- Ismail, Z., Suhartono, Yahaya, A., & Efendi, R. (2009). Intervention model for analyzing the impact of terrorism to tourism industry. *Journal of Mathematics and Statistics*. <https://doi.org/10.3844/jmssp.2009.322.329>
- Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6(3). [https://doi.org/10.1016/0165-1765\(80\)90024-5](https://doi.org/10.1016/0165-1765(80)90024-5)
- Joseph Beaulieu, J., & Miron, J. A. (1993). Seasonal unit roots in aggregate U.S. data. *Journal of Econometrics*, 55(1–2). [https://doi.org/10.1016/0304-4076\(93\)90018-Z](https://doi.org/10.1016/0304-4076(93)90018-Z)
- Kaiser, R., & Maravall, A. (1999). Seasonal Outliers in Time Series. *Statistics and Econometrics Series 15*.
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54(1–3). [https://doi.org/10.1016/0304-4076\(92\)90104-y](https://doi.org/10.1016/0304-4076(92)90104-y)
- Ledolter, J. (1989). The effect of additive outliers on the forecasts from ARIMA models. *International Journal of Forecasting*, 5(2). [https://doi.org/10.1016/0169-2070\(89\)90090-3](https://doi.org/10.1016/0169-2070(89)90090-3)

- Li, Z., Bi, J., & Li, Z. (2017). Passenger Flow Forecasting Research for Airport Terminal Based on SARIMA Time Series Model. *IOP Conference Series: Earth and Environmental Science*, 100(1). <https://doi.org/10.1088/1755-1315/100/1/012146>
- Linden, A. (2015). Conducting interrupted time-series analysis for single- and multiple-group comparisons. *Stata Journal*.
- Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2). <https://doi.org/10.1093/biomet/65.2.297>
- McCoy, E. J., & Stephens, D. A. (2004). Bayesian time series analysis of periodic behaviour and spectral structure. *International Journal of Forecasting*, 20(4), 713–730. <https://doi.org/10.1016/j.ijforecast.2004.02.001>
- McDowall, D., McCleary, R., Meidinger, E., & Hay, R. (2011). Interrupted Time Series Analysis. In *Interrupted Time Series Analysis*. Oxford University Press, USA. <https://doi.org/10.4135/9781412984607>
- Meyer, S. M., McDowall, D., McCleary, R., Meidinger, E. E., & Hay, R. (1981). Interrupted Time-Series Analysis. *Journal of Policy Analysis and Management*. <https://doi.org/10.2307/3324157>
- Mills, T. C. (2019). Applied Time Series Analysis: A Practical Guide to Modeling and Forecasting. In *Applied Time Series Analysis: A Practical Guide to Modeling and Forecasting*. <https://doi.org/10.1016/C2016-0-03956-6>
- Montgomery, D. C., & Weatherby, G. (1980). Modeling and forecasting time series using transfer function and intervention methods. *AIIE Transactions*. <https://doi.org/10.1080/05695558008974521>
- Morettin, P., & Toloi, C. (2018). *Análise de Séries Temporais: Modelos Lineares Univariados* (Third). Edgar Blucher.
- Muthukrishnan, S., Shah, R., & Vitter, J. S. (2004). Mining deviants in time series data streams. *Proceedings of the International Conference on Scientific and Statistical Database Management, SSDBM*, 16. <https://doi.org/10.1109/ssdm.2004.1311192>
- Myers, M., Avison, D., Benbasat, I., Goldstein, D. K., & Mead, M. (2011). The Case Research Strategy in Studies of Information Systems. In *Qualitative Research in Information Systems*. <https://doi.org/10.4135/9781849209687.n5>
- Neusser, C. (2016). *Time series econometrics*. Springer International Publishing Switzerland.
- Newbold, P., Glass, G. V., Willson, V. L., & Gottman, J. M. (1976). Design and Analysis of Time-Series Experiments. *Journal of the American Statistical Association*, 71(353), 247. <https://doi.org/10.2307/2285785>

- Ng, S., & Perron, P. (1995). Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association*, 90(429). <https://doi.org/10.1080/01621459.1995.10476510>
- Nieto, M. R., & Carmona-Benítez, R. B. (2018). ARIMA + GARCH + Bootstrap forecasting method applied to the airline industry. *Journal of Air Transport Management*, 71. <https://doi.org/10.1016/j.jairtraman.2018.05.007>
- Osborn, D. R., Chui, A. P. L., Smith, J. P., & Birchenhall, C. R. (1988). SEASONALITY AND THE ORDER OF INTEGRATION FOR CONSUMPTION. *Oxford Bulletin of Economics and Statistics*, 50(4). <https://doi.org/10.1111/j.1468-0084.1988.mp50004002.x>
- Pankratz, A. E. (1991). *Forecasting with Dynamic Regression Models*. John Wiley & Sons, Inc.
- Payne, J. E., & Taylor, J. P. (2007). Modelling and forecasting airport passengers: a case study for an introductory forecasting course. *International Journal of Information and Operations Management Education*, 2(2). <https://doi.org/10.1504/ijiome.2007.015282>
- Pena, D. (2001). Outliers, Influential Observations, and Missing Data. In D. Pena, G. Tiao, & R. Tsay (Eds.), *A course in time series analysis*. John Wiley & Sons, INC.
- Ramsay, C. R., Matowe, L., Grilli, R., Grimshaw, J. M., & Thomas, R. E. (2003). Interrupted time series designs in health technology assessment: Lessons from two systematic reviews of behavior change strategies. In *International Journal of Technology Assessment in Health Care*. <https://doi.org/10.1017/S0266462303000576>
- Reddy, A., Ordway-West, M., Lee, M., Dugan, M., Whitney, J., Kahana, R., Ford, B., Muedsam, J., Henslee, A., & Rao, M. (2017). Using Gaussian mixture models to detect outliers in seasonal univariate network traffic. *Proceedings - 2017 IEEE Symposium on Security and Privacy Workshops, SPW 2017, 2017-December*. <https://doi.org/10.1109/SPW.2017.9>
- Reinsel, G., Tiao, G. C., Wang, M. N., Lewis, R., & Nychka, D. (1981). Statistical analysis of stratospheric ozone data for the detection of trends. *Atmospheric Environment* (1967), 15(9). [https://doi.org/10.1016/0004-6981\(81\)90140-2](https://doi.org/10.1016/0004-6981(81)90140-2)
- Richards, L. E., Little, R. J. A., & Rubin, D. B. (1989). Statistical Analysis with Missing Data. *Journal of Marketing Research*. <https://doi.org/10.2307/3172915>
- Said, S. E., & Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3). <https://doi.org/10.1093/biomet/71.3.599>
- Schuster, A. (1898). On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. *Journal of Geophysical Research*, 3(1). <https://doi.org/10.1029/tm003i001p00013>

- Schwarz, G. (1978). "Estimating the Dimension of a Model." *The Annals of Statistics*, 6(2). <https://doi.org/10.2307/2958889>
- Shumway, R. H., & Stoffer, D. S. (2019). Time Series: A Data Analysis Approach Using R. In *Time Series: A Data Analysis Approach Using R*. <https://doi.org/10.1201/9780429273285>
- Strand, S. (1999). Airport-specific traffic forecasts: The resultant of local and non-local forces. *Journal of Transport Geography*, 7(1), 17–29. [https://doi.org/10.1016/S0966-6923\(98\)00036-2](https://doi.org/10.1016/S0966-6923(98)00036-2)
- Tiao, G. C., Box, G. E. P., & Box, G. E. P. (1975). Analysis of los angeles photochemical smog data: A statistical overview. *Journal of the Air Pollution Control Association*, 25(3). <https://doi.org/10.1080/00022470.1975.10470082>
- Topál, D., Matyasovszky, I., Kern, Z., & Hatvani, I. G. (2016). Detecting breakpoints in artificially modified- and real-life time series using three state-of-the-art methods. *Open Geosciences*, 8(1), 78–98. <https://doi.org/10.1515/geo-2016-0009>
- Tsay, R. S. (1986). Time series model specification in the presence of outliers. *Journal of the American Statistical Association*. <https://doi.org/10.1080/01621459.1986.10478250>
- Tsay, R. S. (1988a). Outliers, level shifts, and variance changes in time series. *Journal of Forecasting*. <https://doi.org/10.1002/for.3980070102>
- Tsay, R. S. (1988b). Outliers, level shifts, and variance changes in time series. *Journal of Forecasting*. <https://doi.org/10.1002/for.3980070102>
- Tsay, R. S. (2010). Analysis of Financial Time Series, Third Edition, Wiley Series in Probability and Statistics. In *Health Services Management Research* (Vol. 4, Issue 3).
- Tsay, R. S., Peña, D., & Pankratz, A. E. (2000). Outliers in multivariate time series. *Biometrika*, 87(4). <https://doi.org/10.1093/biomet/87.4.789>
- Tsay, R. S., & Tiao, G. C. (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models. *Journal of the American Statistical Association*. <https://doi.org/10.1080/01621459.1984.10477068>
- Tsui, W. H. K., Ozer Balli, H., Gilbey, A., & Gow, H. (2014). Forecasting of Hong Kong airport's passenger throughput. *Tourism Management*, 42. <https://doi.org/10.1016/j.tourman.2013.10.008>
- Velicer, W. F., & McDonald, R. P. (1991). Cross-Sectional Time Series Designs: A General Transformation Approach. *Multivariate Behavioral Research*. https://doi.org/10.1207/s15327906mbr2602_3

- Vincenty, T. (1975). Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Survey Review*, 23(176).
<https://doi.org/10.1179/sre.1975.23.176.88>
- Wei, W. W. S. (2006). Time Series Analysis: Univariate and Multivariate Methods Second Edition. In *Pearson Education, Inc.* <https://doi.org/10.1016/j.ijforecast.2004.02.001>
- Woodward, W. A., Gray, H. L., & Elliott, A. C. (2017). Applied time series analysis with R, second edition. In *Applied Time Series Analysis with R, Second Edition*.
<https://doi.org/10.1201/9781315161143>
- World Bank. (2018). Global Economic Prospects, June 2018: The Turning of the Tide? In *Global Economic Prospects, June 2018: The Turning of the Tide?* (Issue January).
<https://doi.org/10.1596/978-1-4648-1257-6>
- Yaffee, R. (2000). Introduction to Time Series Analysis and Forecasting (with applications of SAS and SPSS). *Time*. https://doi.org/10.1142/9789812833204_0024
- Yaffee, R., & McGee, M. (2000). *Introduction to Time Series Analysis and Forecasting: With Applications of SAS and SPSS. With Applications of SAS and SPSS*. ACADEMIC PRESS, INC.
- Zeileis, A., Kleiber, C., Walter, K., & Hornik, K. (2003). Testing and dating of structural changes in practice. *Computational Statistics and Data Analysis*, 44(1–2).
[https://doi.org/10.1016/S0167-9473\(03\)00030-6](https://doi.org/10.1016/S0167-9473(03)00030-6)
- Zhang, Y., Hamm, N. A. S., Meratnia, N., Stein, A., van de Voort, M., & Havinga, P. J. M. (2012). Statistics-based outlier detection for wireless sensor networks. *International Journal of Geographical Information Science*, 26(8).
<https://doi.org/10.1080/13658816.2012.654493>

APPENDIX A: DATA COLLECTION AND SCREENING

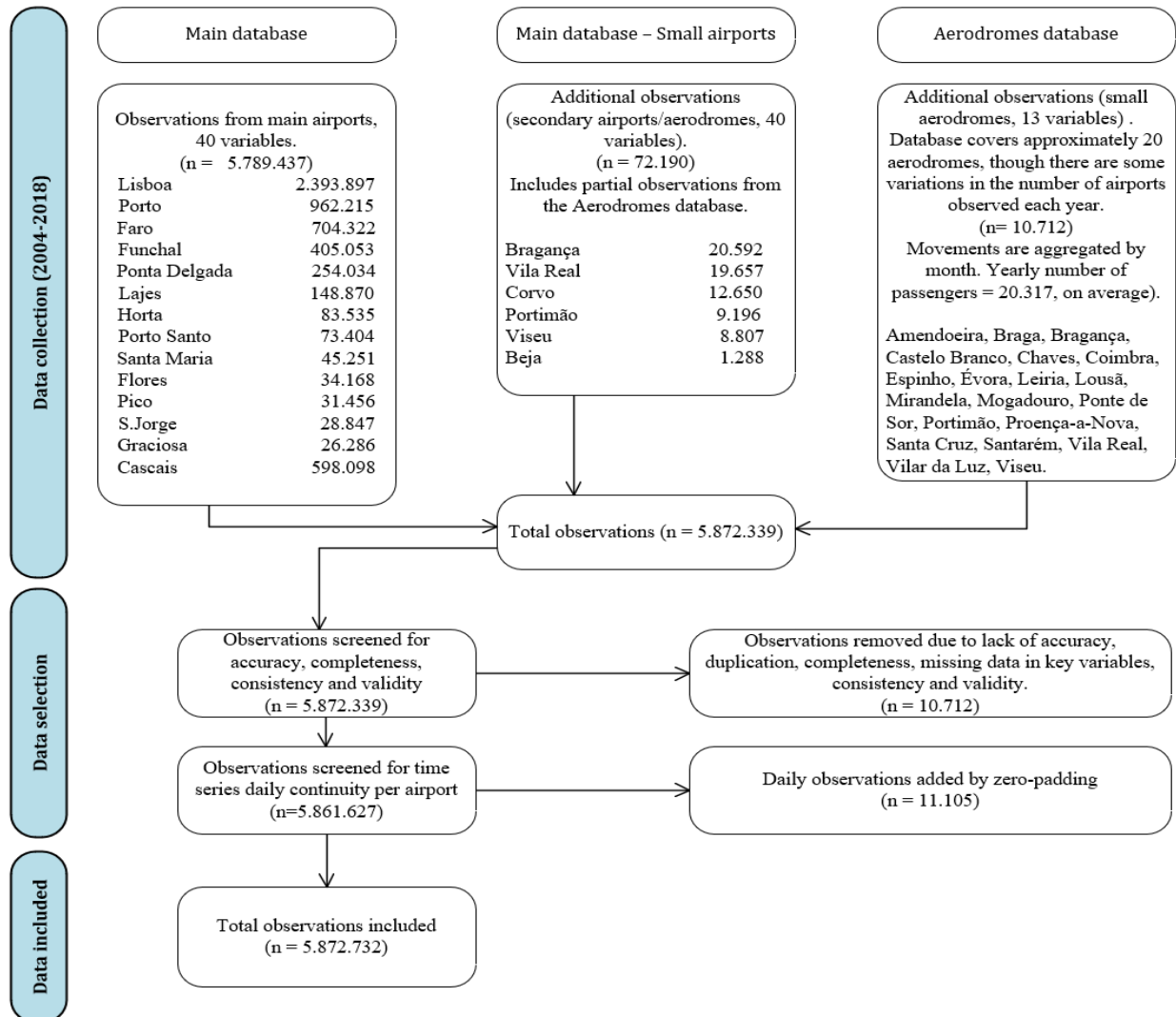


Figure A - 1: Data collection and screening.

APPENDIX B: DATABASE COVERAGE (2016-2018)

Table A - 1: Database coverage by airport (2016-2018).

Database	Aerodrome	2016	% over total	2017	% over total	2018	% over total
Main database	Lisboa	22.462.128	49,4%	26.675.708	50,4%	29.045.759	51,4%
Main database	Porto	9.379.599	20,6%	10.789.013	20,4%	11.943.800	21,1%
Main database	Faro	7.632.089	16,8%	8.727.732	16,5%	8.687.406	15,4%
Main database	Funchal	2.973.840	6,5%	3.204.438	6,1%	3.182.480	5,6%
Main database	Ponta Delgada	1.518.765	3,3%	1.853.164	3,5%	1.909.459	3,4%
Main database	Lajes	631.016	1,4%	740.961	1,4%	756.852	1,3%
Main database	Horta	221.640	0,5%	238.235	0,5%	247.249	0,4%
Main database	Porto Santo	159.225	0,4%	178.278	0,3%	167.141	0,3%
Main database	Pico	123.149	0,3%	121.560	0,2%	136.221	0,2%
Main database	Santa Maria	99.227	0,2%	93.316	0,2%	97.775	0,2%
Main database	São Jorge	63.986	0,1%	70.950	0,1%	75.553	0,1%
Main database	Flores	55.435	0,1%	63.605	0,1%	69.067	0,1%
Main database	Graciosa	47.403	0,1%	51.527	0,1%	53.767	0,1%
Main database	Cascais	9.064	0,0%	10.488	0,0%	13.071	0,0%
Main database	Viseu	5.190	0,0%	8.865	0,0%	9.649	0,0%
Main database	Corvo	7.112	0,0%	7.656	0,0%	9.261	0,0%
Main database	Portimão	1.870	0,0%	2.673	0,0%	6.762	0,0%
Main database	Vila Real	4.070	0,0%	5.285	0,0%	6.542	0,0%
Main database	Beja	393	0,0%	1.202	0,0%	5.816	0,0%
Main database	Bragança	2.961	0,0%	3.153	0,0%	4.065	0,0%
Small aerodrome database	Viseu	8.183	0,0%	11.492	0,0%	13.966	0,0%
Small aerodrome database	Vila Real	4.894	0,0%	6.745	0,0%	7.729	0,0%
Small aerodrome database	Portimão	4.210	0,0%	6.066	0,0%	7.609	0,0%
Small aerodrome database	Leiria	-	0,0%	716	0,0%	5.196	0,0%
Small aerodrome database	Bragança	3.229	0,0%	3.686	0,0%	4.337	0,0%
Small aerodrome database	Vilar da Luz	204	0,0%	5.183	0,0%	3.605	0,0%
Small aerodrome database	Chaves	1.275	0,0%	1.585	0,0%	1.135	0,0%
Small aerodrome database	Coimbra	2.014	0,0%	2.094	0,0%	1.111	0,0%
Small aerodrome database	Castelo Branco	2.011	0,0%	5.115	0,0%	605	0,0%
Small aerodrome database	Braga	496	0,0%	314	0,0%	280	0,0%
Small aerodrome database	Évora	283	0,0%	280	0,0%	160	0,0%
Small aerodrome database	Santarém	166	0,0%	190	0,0%	100	0,0%
Small aerodrome database	Espinho	504	0,0%	-	0,0%	-	0,0%
Small aerodrome database	Lousã	-	0,0%	-	0,0%	-	0,0%
Small aerodrome database	Mirandela	-	0,0%	-	0,0%	-	0,0%
Small aerodrome database	Mogadouro	-	0,0%	18	0,0%	-	0,0%
Small aerodrome database	Ponte de Sor	-	0,0%	-	0,0%	-	0,0%
Small aerodrome database	Proença-a-Nova	5	0,0%	16	0,0%	-	0,0%
Small aerodrome database	Santa Cruz	-	0,0%	-	0,0%	-	0,0%
Total (both databases)		45.425.636		52.891.309		56.473.528	
Main database total		45.398.162		52.847.809		56.427.695	
Main database coverage		99,94%		99,92%		99,92%	

Note: Four aerodromes (Bragança, Portimão, Vila Real and Viseu) report the flights statistics to the two databases.

APPENDIX C: CODING OF EASTER AND LEAP-YEAR EFFECTS

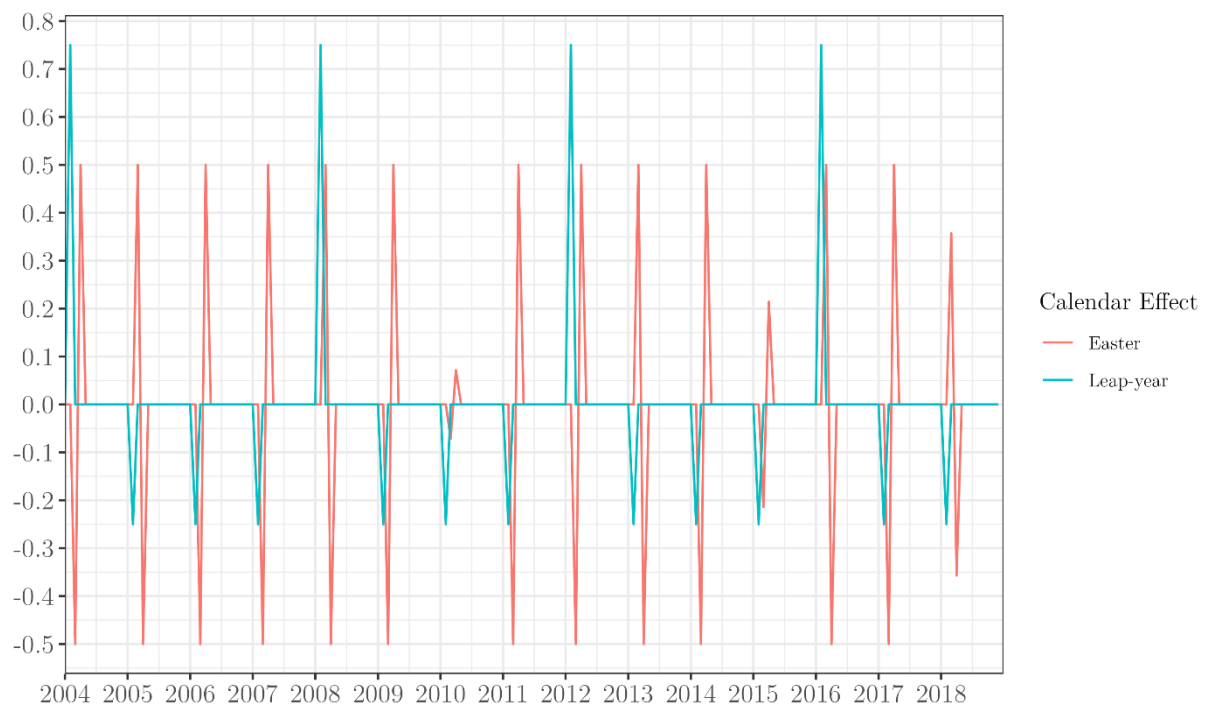


Figure A - 2: Easter and leap-year effects. The Easter effect includes the six days leading up to Easter Sunday.

APPENDIX D: RESIDUALS PLOTS

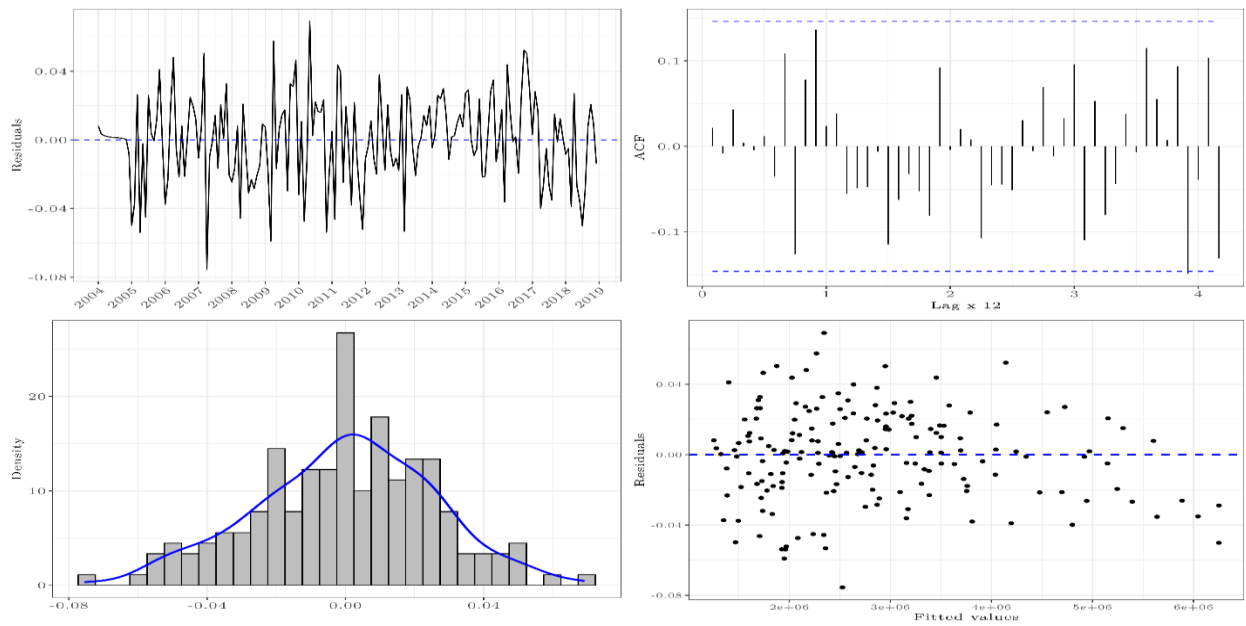


Figure A - 3: Fitted Residuals from Model 13.

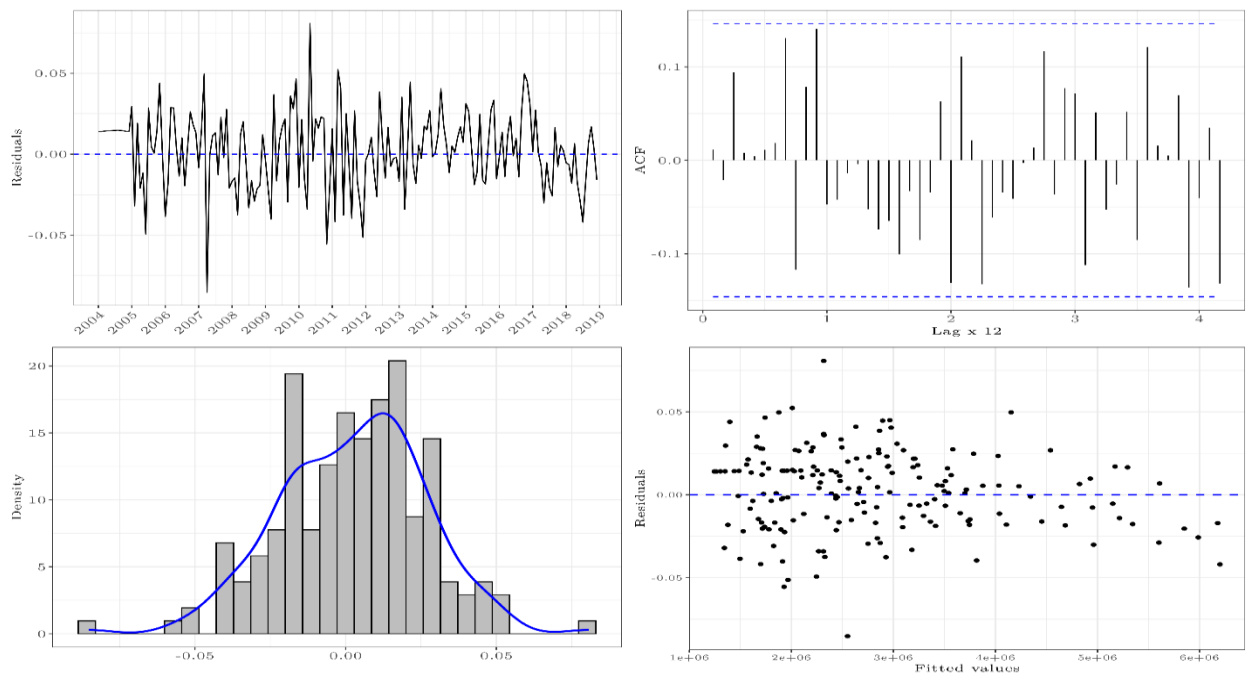


Figure A - 4: Residuals from Model 14.

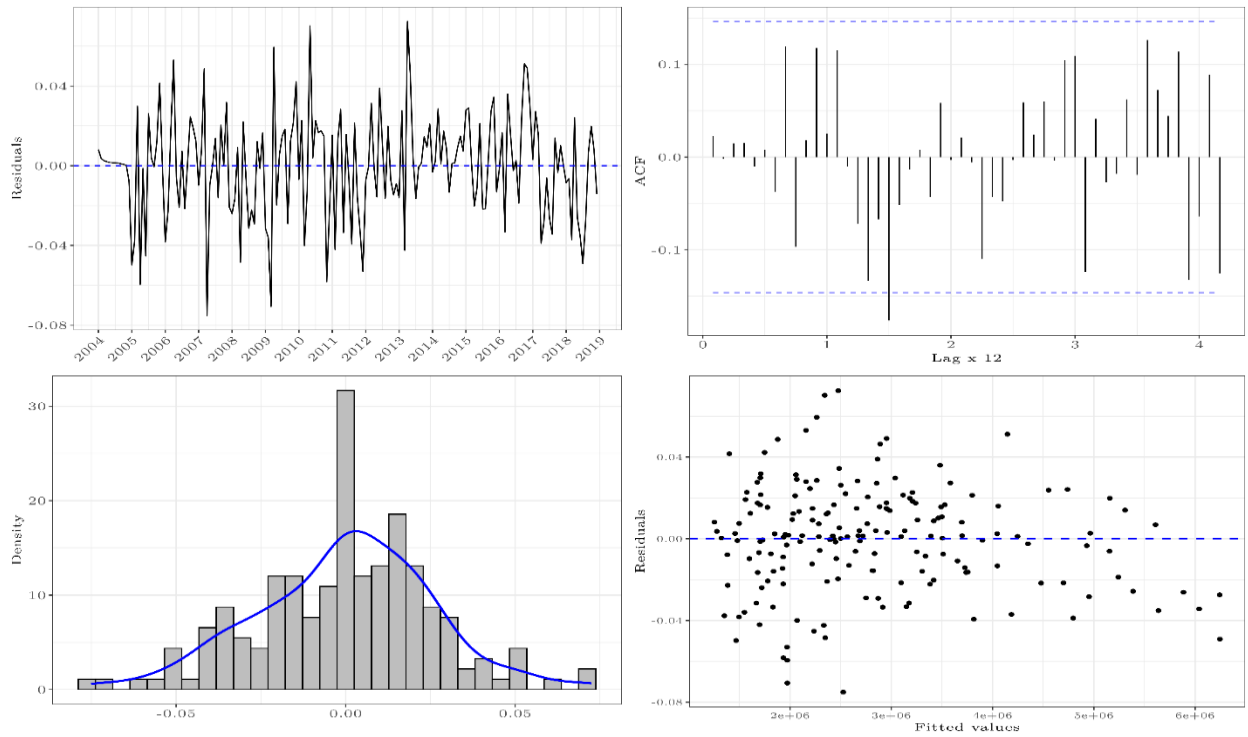


Figure A - 5: Residuals from Model 15.

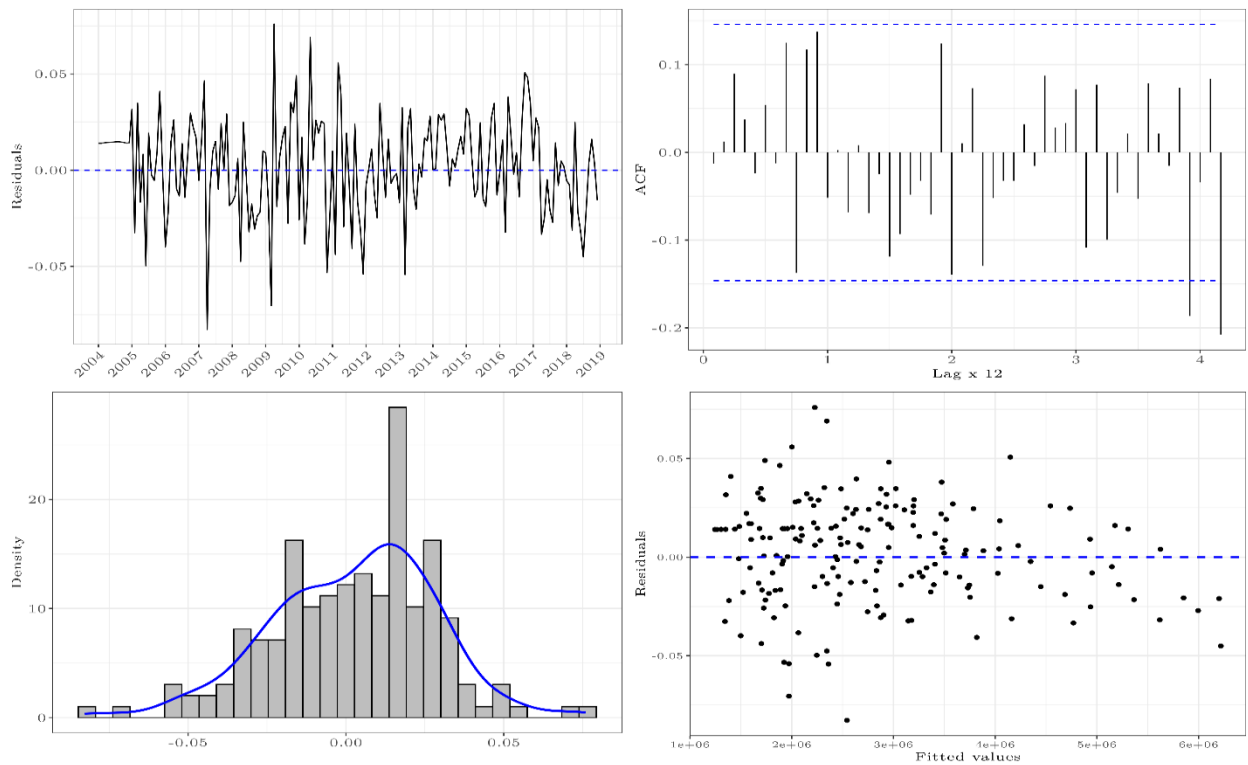


Figure A - 6: Residuals from Model 17.

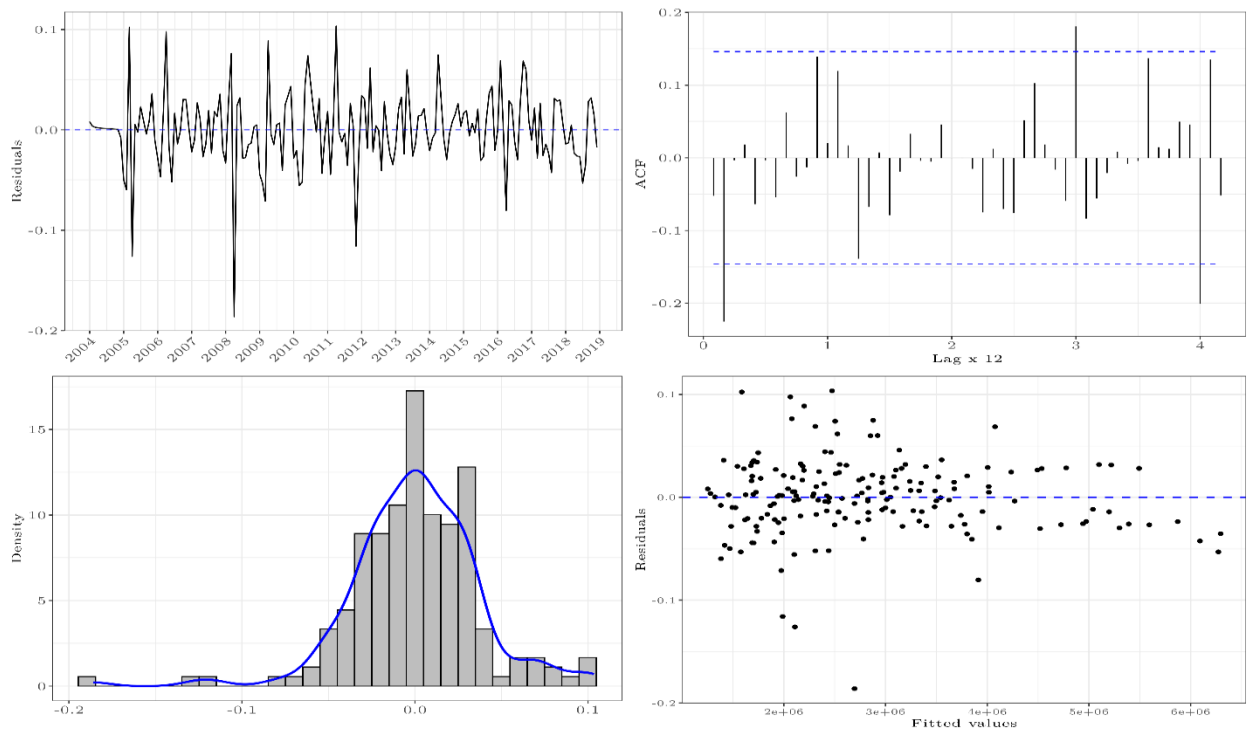


Figure A - 7: Residuals plots for Model 1 (Orders selected form visual inspection).

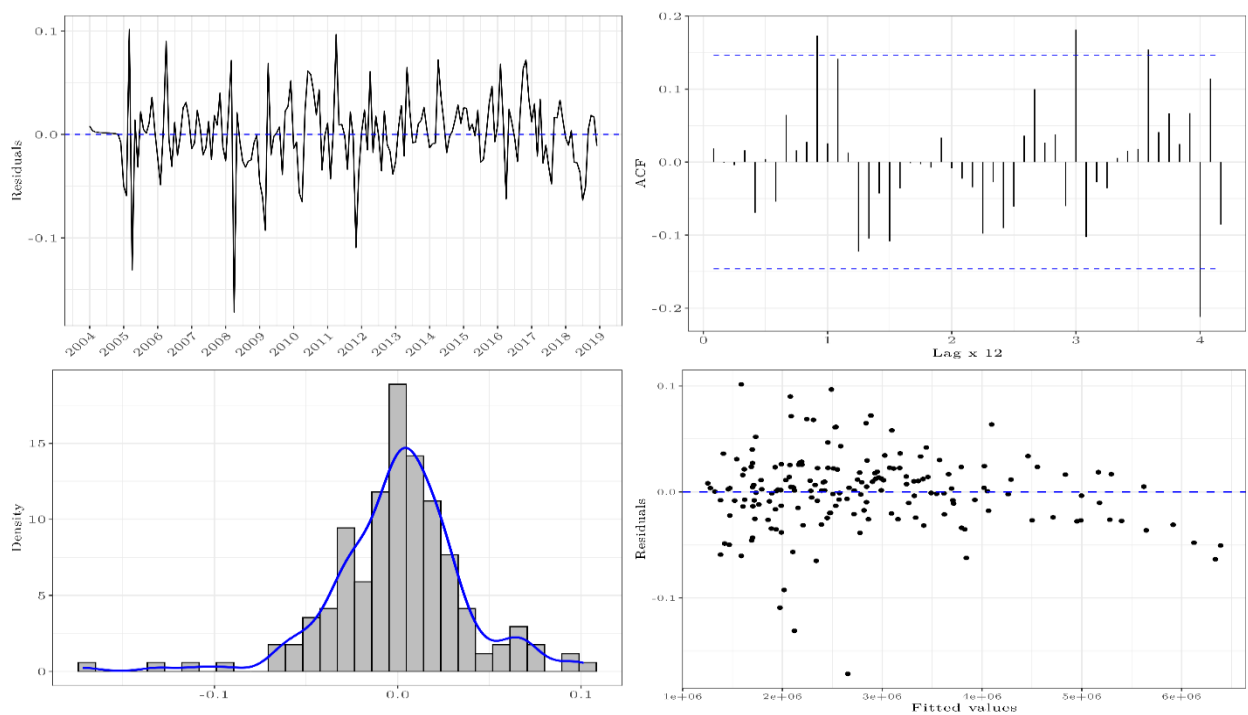


Figure A - 8: Residuals plots for Model 2 (Orders selected form visual inspection).

APPENDIX E: FITTED AND FORECASTS

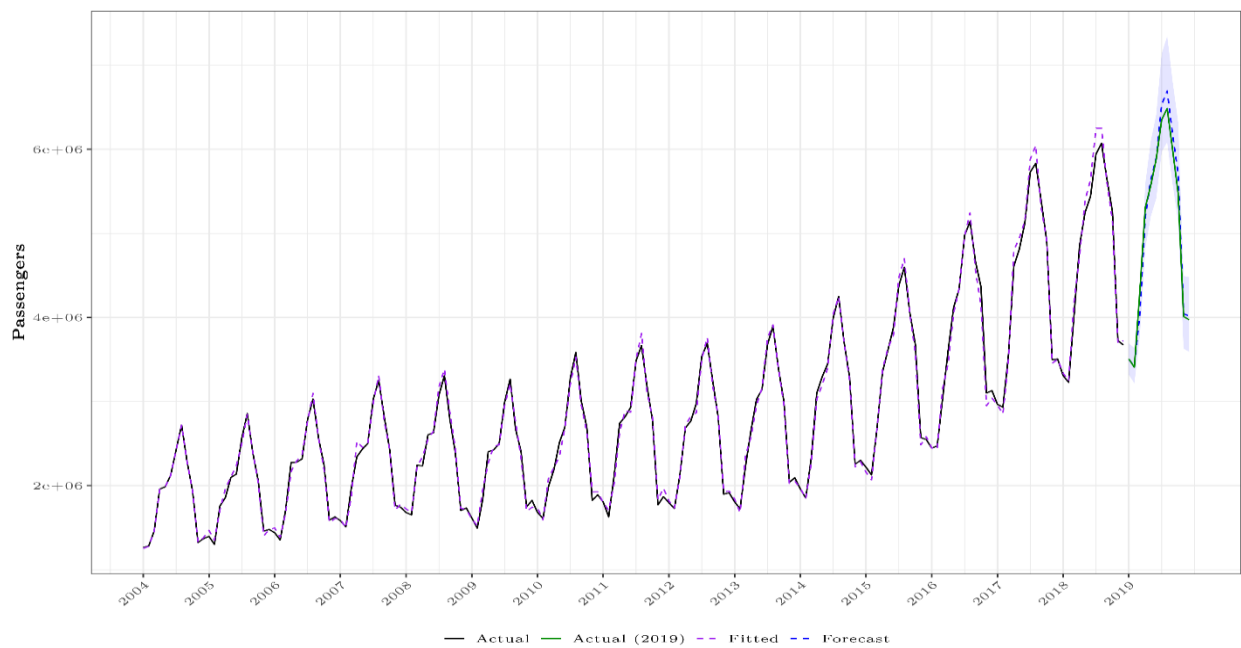


Figure A - 9: Fitted and forecasts from Model 13.

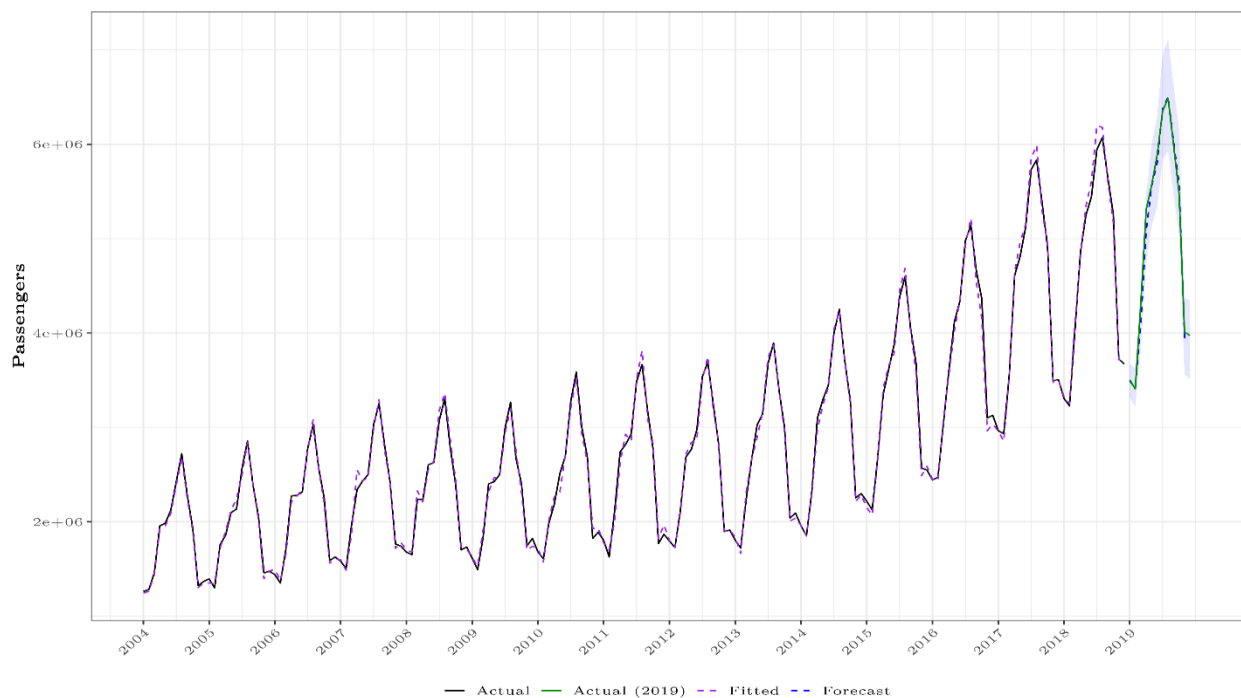


Figure A - 10: Fitted and forecasts from Model 14.

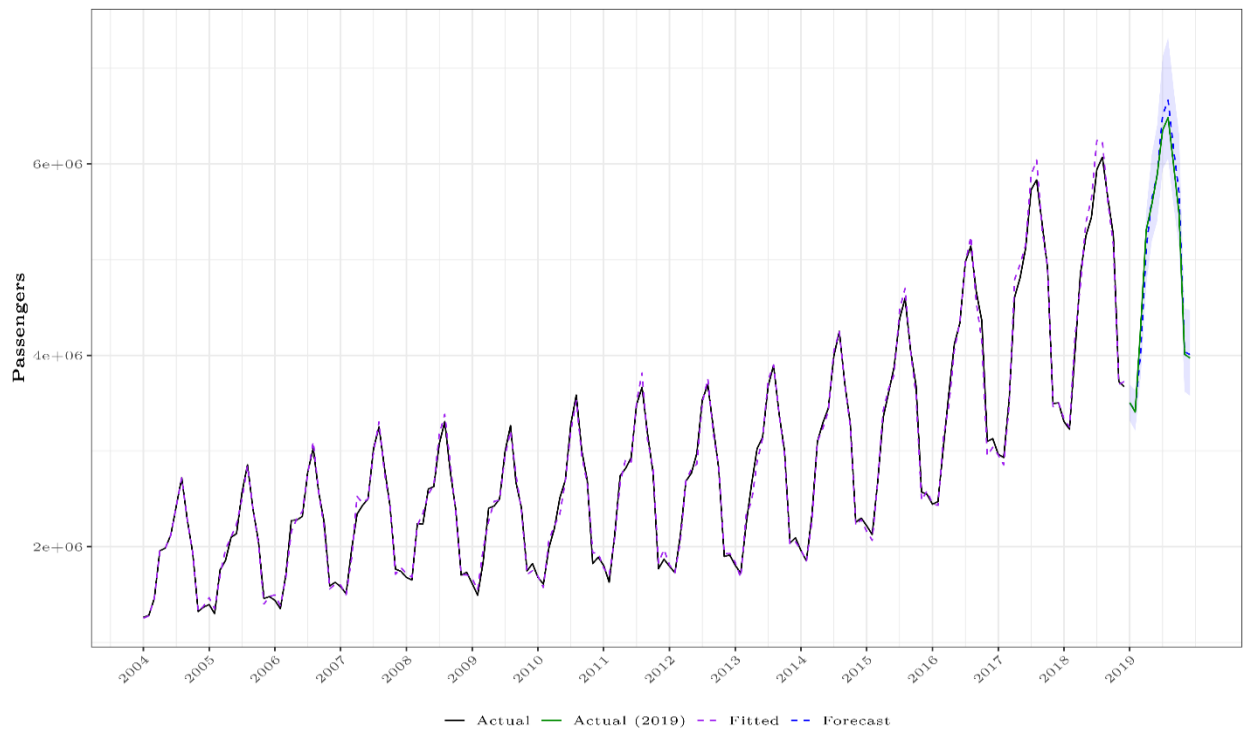


Figure A - 11: Fitted and forecasts from Model 15.

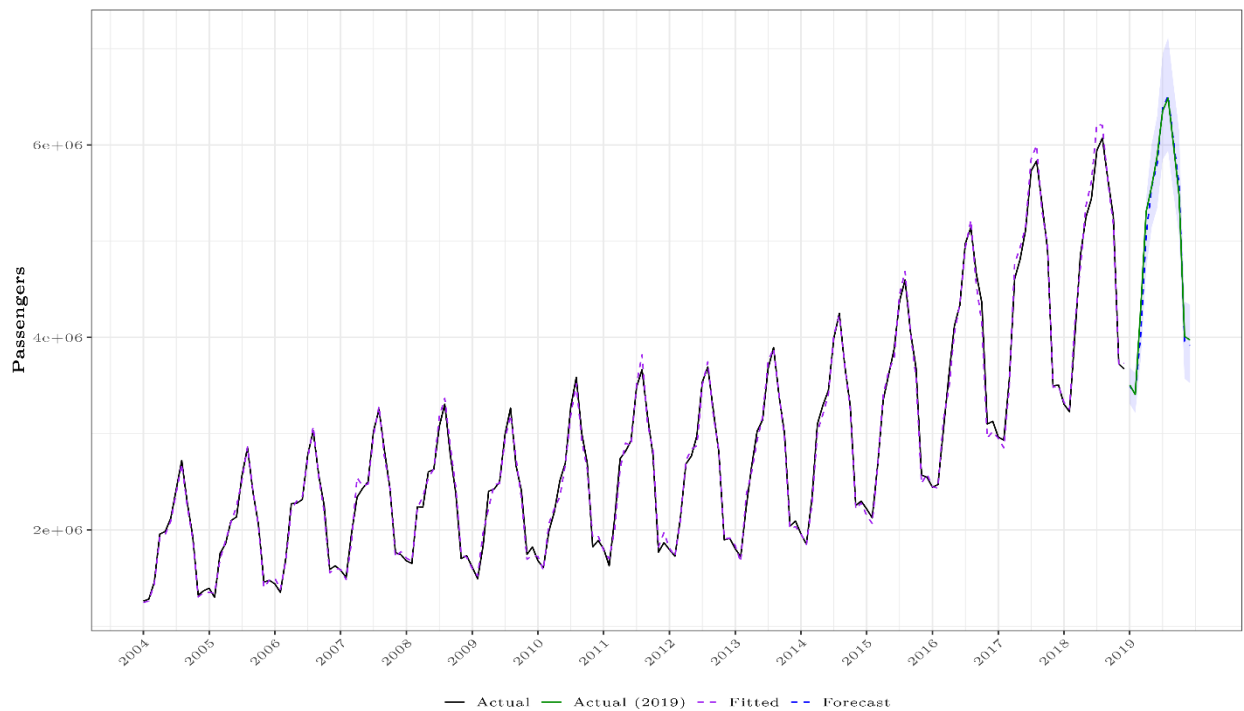


Figure A - 12: Fitted and forecasts from Model 17.

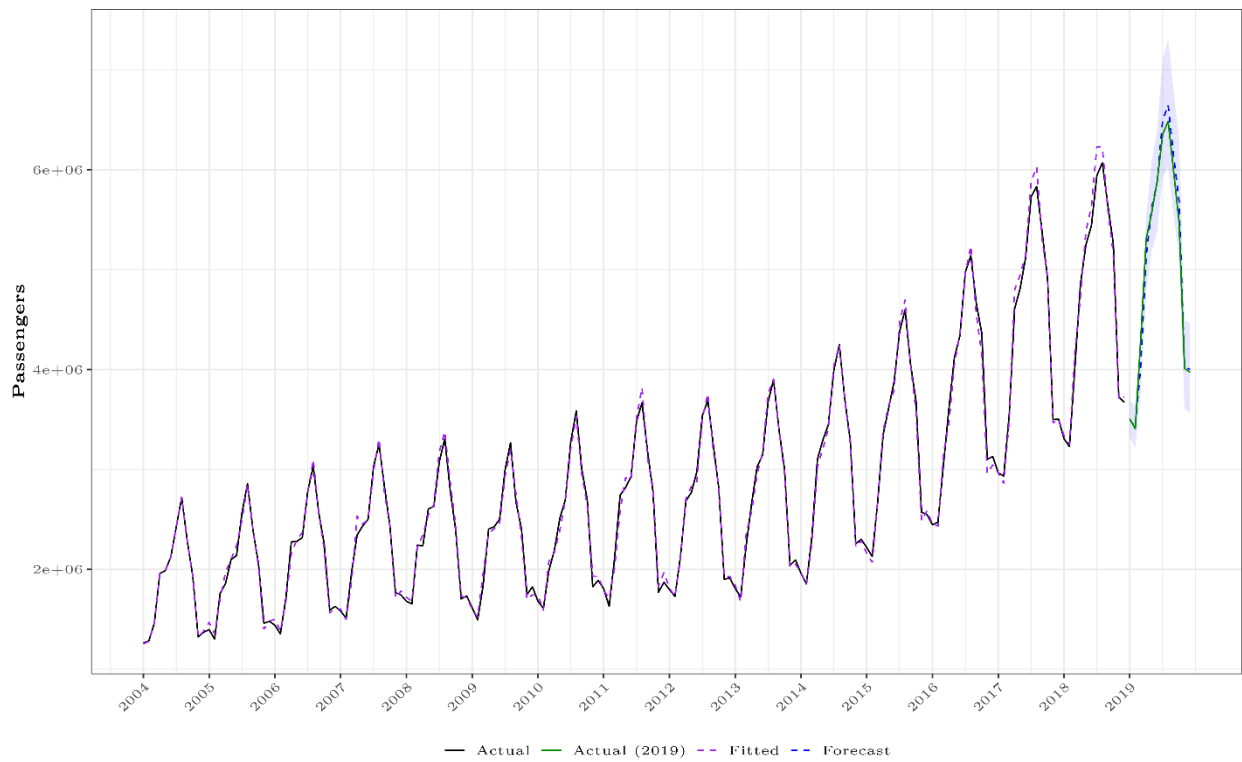


Figure A - 13: Fitted and forecasts from Model 18.

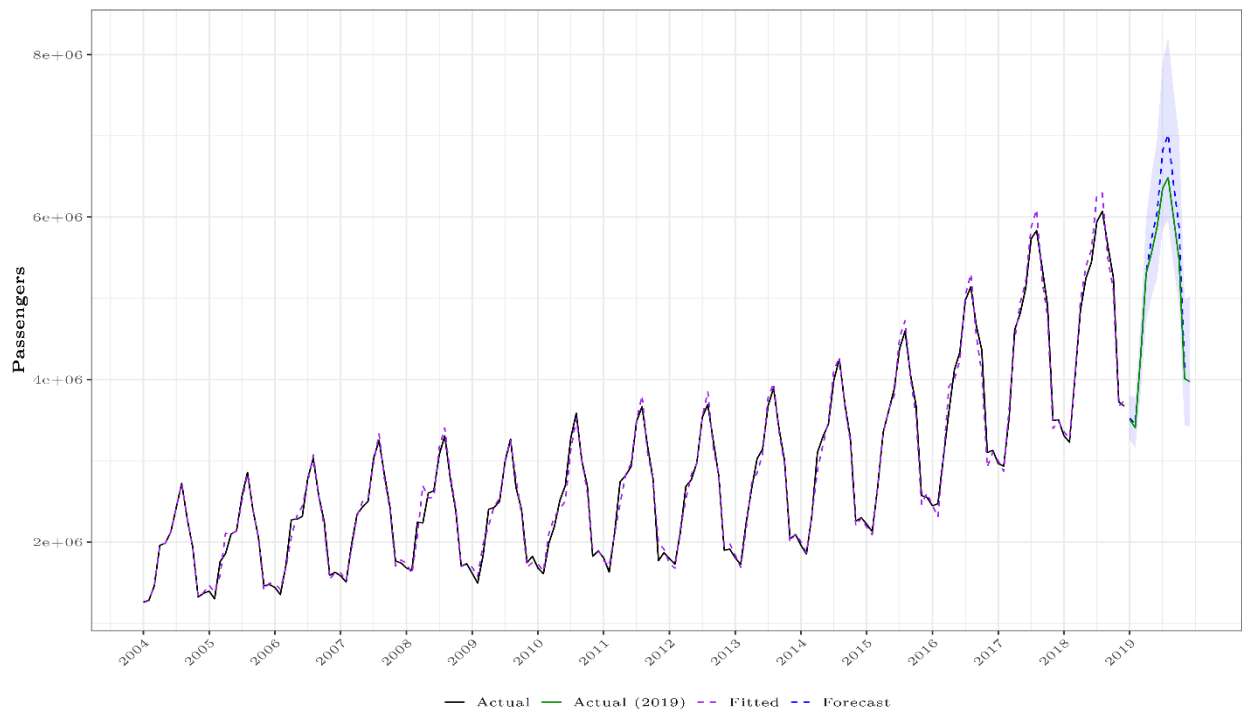


Figure A - 14: Fitted and forecasts from tentative Model 1.

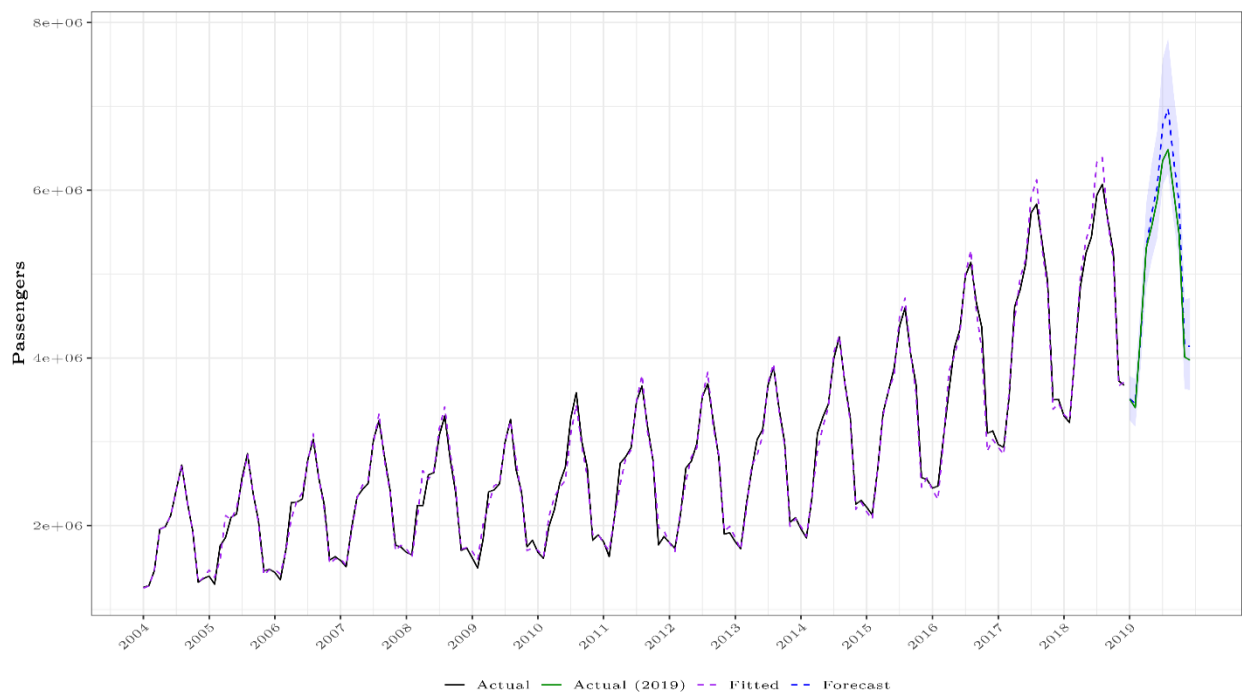


Figure A - 15: Fitted and forecasts from tentative Model 2.

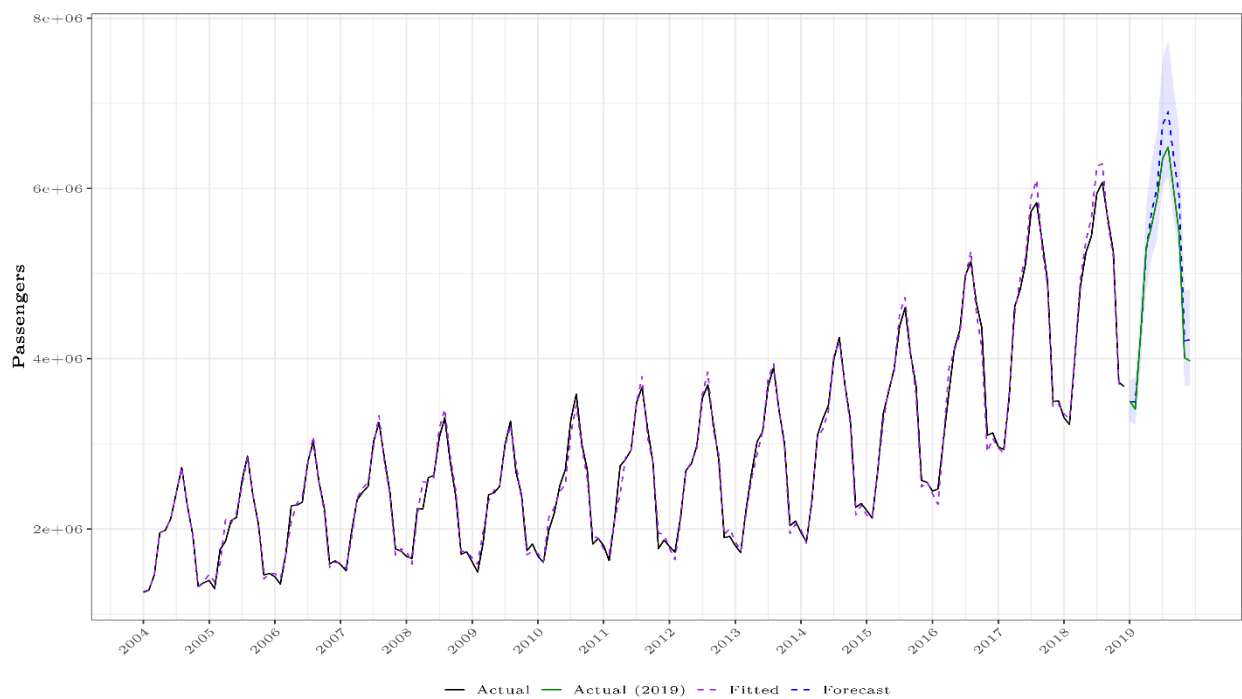


Figure A - 16: Fitted and forecasts from Model Auto ARIMA.

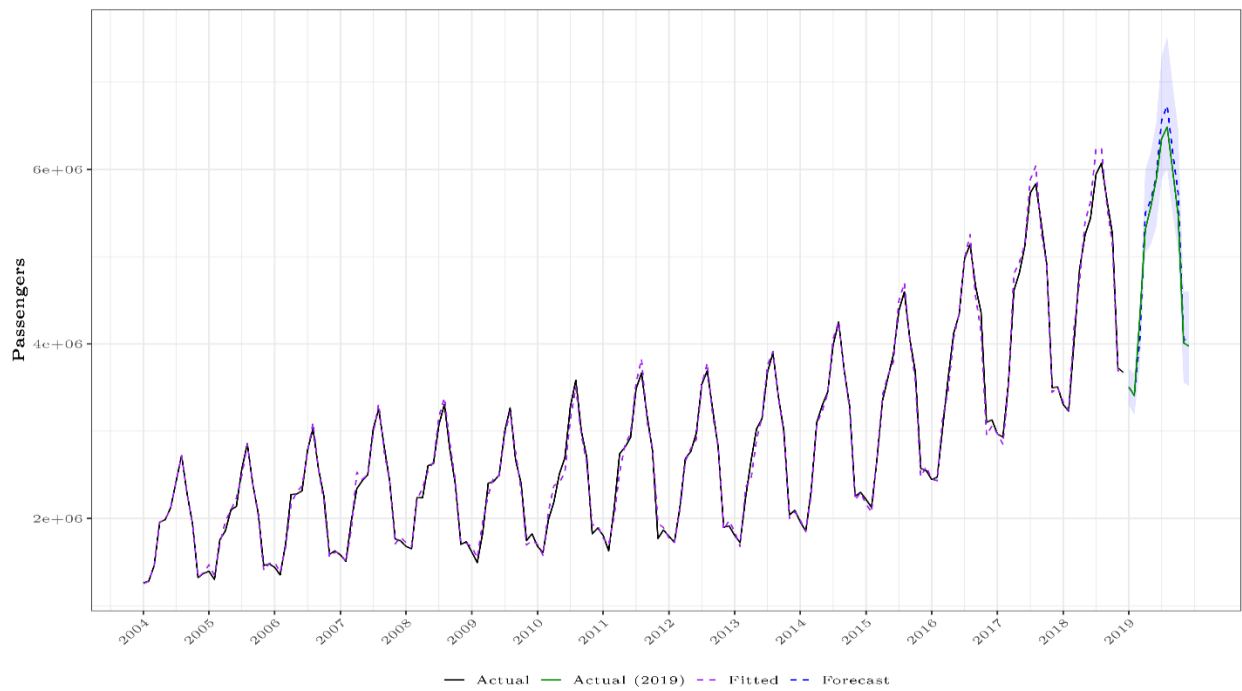


Figure A - 17: Fitted and forecasts from Model 10.

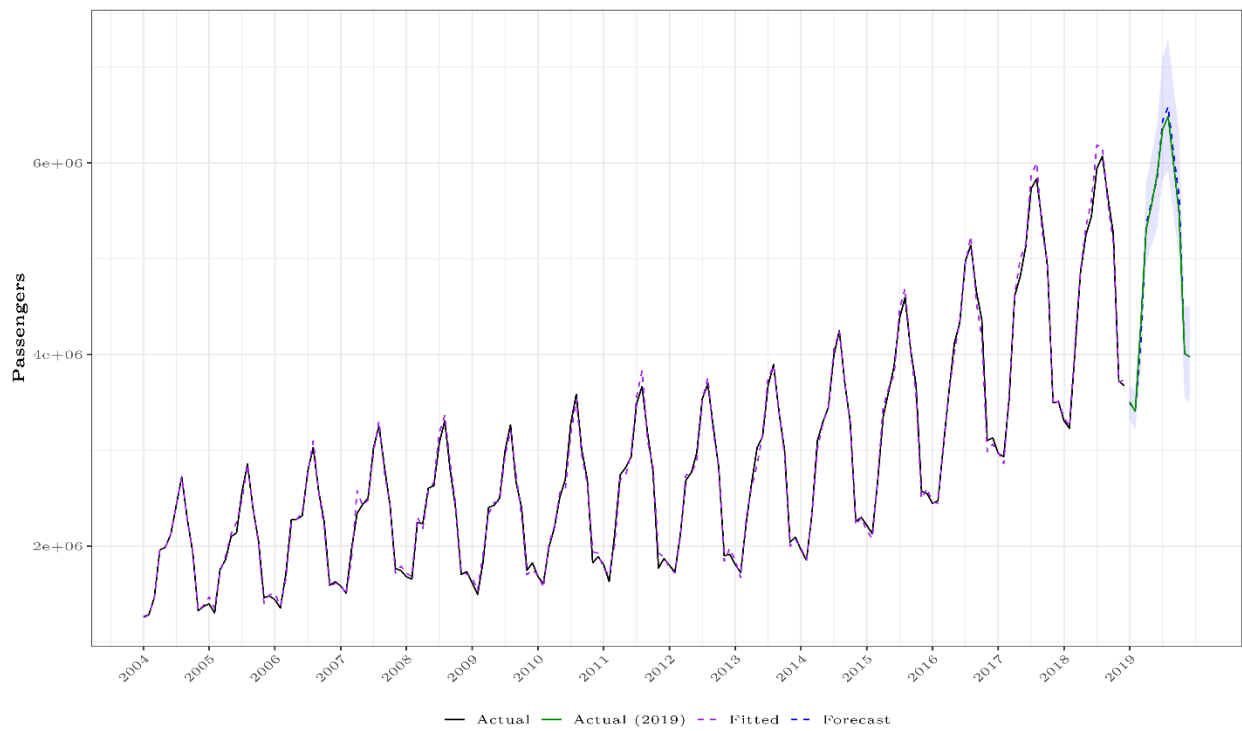


Figure A - 18: Fitted and forecasts from Model 12.

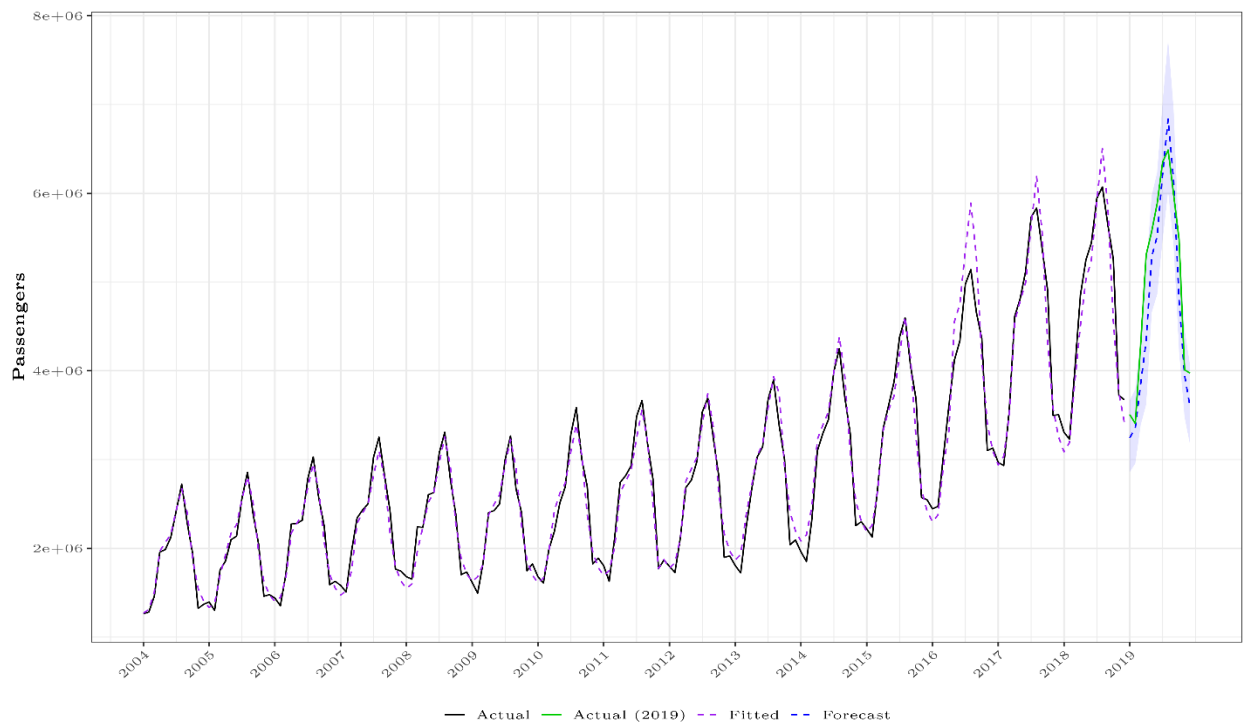


Figure A - 19: Fitted and forecasts from Model 11.

ANNEX A: TRAFFIC FORM



FORMULÁRIO ANAC 1 - FORMULÁRIO DE TRÁFEGO (ANAC FORM 1 - TRAFFIC FORM)

AEROPORTO / AIRPORT

1. MOVIMENTO / MOVEMENT				2. OPERADOR / OPERATOR	
	1.1 NÚMERO / Number	1.2 HORA UTC / Hour UTC	1.3 DATA / Date		
CHEGADA / Arrival					
PARTIDA / Departure					

3. VOO / FLIGHT					
	3.1 NÚMERO / Number	3.2 NATUREZA / Type of Flight	3.3 CLASSIFICAÇÃO / Classification	3.4 TIPO DE OPERAÇÃO / Type of Operation	3.5 HORÁRIO UTC / Scheduled UTC
CHEGADA / Arrival					
PARTIDA / Departure					

4. AERONAVE / AIRCRAFT						
	4.1 MATRÍCULA / Registration	4.2 TIPO / Type Code		4.3 P.M.D. (Ton. Métrica) / MTOW (Metric Ton)	4.4 LUGARES OFERECIDOS / Available Seats	4.5 CERTIFICADO DE RUÍDO / Noise Certificate
		ICAO	ATA			
CHEGADA / Arrival						
PARTIDA / Departure						

5. DISTRIBUIÇÃO DE TRÁFEGO / TRAFFIC DISTRIBUTION									
5.1 ESCALAS / Route Stops			5.2 PESSOAS A BORDO / People on Board				5.4 CARGA / Freight (kg)		5.4 CORREIO / Mail (kg)
			LOCALS / Local		DIRETO / Direct	TRANSFERÊNCIA / Transfer			
			SEM TAXA / Without Charge	COM TAXA / With Charge			SEM TAXA / Without Charge	COM TAXA / With Charge	SEM TAXA / Without Charge
CHEGADA / Arrival	ORIGEM / Origin								
	PARA-ESCALAS INTERMEDIÁRIAS / Intermediate Stops								
TOTAL									
PARTIDA / Departure	PARA-ESCALAS INTERMEDIÁRIAS / Intermediate Stops								
	TOTAL								
DESTINO / Destination									

OPERADOR / Operator	AEROPORTO / Airport	ALFÂNDEGA / Customs	CONTROLO DE PASSAP. / Passport Control

6. HOUVE MUDANÇA DE EQUIPAMENTO / Equipment change ☐ SIM / YES ☐ NÃO / NO

Motivos Técnicos / Technical reasons ☐

Outros / Others ☐

OBSERVAÇÕES / Remarks

Figure A - 20: Traffic Form

