Master

Data Analytics for Business

Master's Final Work

Dissertation

Measuring the Impact of Data Anomalies on Tourism Demand Forecasts

Rosanna Mueller

March - 2023

Master

Data Analytics for Business

Master's Final Work

Dissertation

Measuring the Impact of Data Anomalies on Tourism Demand Forecasts

Rosanna Mueller

Supervision:
Prof. Nuno Ricardo Martins Sobreira

DOCUMENT SPECIFICALLY CREATED FOR OBTAINING A MASTER'S DEGREE

March - 2023

GLOSSARY

AICc – Corrected Akaike's Information Criterion.

AR – Autoregressive.

ARIMA – Autoregressive Integrated Moving Average.

ARIMA_auto – ARIMA model using the automatic ARIMA modelling function.

ARIMA_d0D1 – ARIMA $(p, 0, q)(P, 1, Q)_{12}$.

ARIMA_d1D1 – ARIMA $(p, 1, q)(P, 1, Q)_{12}$.

ETS – Exponential Smoothing.

ETS_auto – Exponential Smoothing Model using the automatic modelling function.

HWA – Holt Winters' Additive.

Lisbon M.A. – Lisbon Metropolitan Area.

MA – Moving average.

MFW – Master's Final Work.

Recursive CV – Recursive window cross-validation.

RMSE – Root Mean Squared Error.

Rolling CV – Rolling window cross-validation.

SARIMA – Seasonal ARIMA.

SNaïve – Seasonal Naïve Model.

STL decomposition – Seasonal-Trend decomposition using LOESS.

TSLM – Time Series Linear Model.

ABSTRACT

Time series models have proven to be powerful tools for forecasting tourism demand. However, the recent Covid-19 outbreak has severely impacted the tourism industry, and models which were previously able to provide accurate forecasts may no longer be viable. This work aims to further analyse this situation by measuring the impact of data anomalies caused by the Covid-19 pandemic on the forecasting performance of different time series models. For this purpose, the monthly number of tourist Overnight Stays per region in Portugal from 2000 to 2022 is used and forecasting competitions are performed on three selected time series.

These forecasting competitions contain various approaches, from simple methods to different variants of Autoregressive Integrated Moving Average and Exponential Smoothing models. The forecasting performance of the models is assessed firstly by excluding the Covid-19 pandemic from the time series and secondly by including this period. In addition, a logarithmic transformation of the forecast variable is performed as well as different types of cross-validation approaches are used.

The results reveal that Autoregressive Integrated Moving Average and Exponential Smoothing models showed superior performance before the Covid-19 outbreak, but a significant loss of performance in the months thereafter. In contrast, the Naïve method produced comparatively good forecasts during these months due to its simplicity. Moreover, the Drift model applied to seasonally adjusted data was able to compete with the best models and displayed a lower deterioration in prediction accuracy following the Covid-19 outbreak. Besides, this work provides evidence that regions with a higher percentage of Portuguese tourists displayed lower declines in tourism demand during the Covid-19 period.

KEYWORDS: Time Series Models; Forecasting Tourism Demand; Change in Model Performance; Data Anomalies; Covid-19.

JEL CODES: C12; C22; C52; C53; C88; Z32.

RESUMO

Os modelos de séries temporais provaram ser instrumentos eficazes a prever a procura turística. Contudo, o recente surto de Covid-19 afectou gravemente a indústria do turismo, e os modelos que anteriormente eram capazes de produzir previsões com uma reduzida margem de erro podem já não ser viáveis. Este trabalho visa analisar esta situação ao medir o impacto das anomalias de dados causadas pela pandemia de Covid-19 no desempenho preditivo de alguns modelos de séries temporais. Para este efeito, é utilizado o número mensal de estadias nas várias regiões de Portugal entre 2000 e 2022 e são realizadas previsões com vários modelos para três séries temporais distintas.

Estes modelos de previsão incluem vários modelos, desde métodos simples a diferentes variantes de modelos Autoregressivos Integrados de Média Móvel e modelos de Suavização Exponencial. Em primeiro lugar, o desempenho da previsão destes modelos é avaliado excluindo à série temporal o período da pandemia Covid-19 e, em segundo lugar, incluindo este mesmo período. Além disso, é realizada uma transformação logarítmica na variável prevista e, de seguida, são implementados diferentes tipos de abordagens de validação cruzada.

Os resultados revelam que os Modelos Autoregressivos Integrados de Médias Móveis e Suavização Exponencial apresentam um melhor desempenho antes do pandemia Covid-19, mas uma perda significativa de desempenho nos meses seguintes. Em contraste, o método Naive, devido à sua simplicidade, consegue produzir previsões relativamente boas durante estes meses. Além disso, o modelo Drift aplicado a dados ajustados sazonalmente foi capaz de competir com os melhores modelos, revelando uma menor deterioração na precisão das suas previsões durante a pandemia da Covid-19. Além disso, este trabalho fornece provas de que as regiões com uma maior percentagem de turistas portugueses apresentaram menores declínios na procura turística durante o período de Covid-19.

PALAVRAS-CHAVE: Modelos de séries temporais; Previsão da procura turística; Alteração no desempenho do modelo; Anomalias de dados; Covid-19.

JEL: C12; C22; C52; C53; C88; Z32.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express my gratitude to Professor Sobreira for his contributions and support throughout this journey.

I would like to thank Rodrigo Conceição, Sofia Pêga and João Sivestre for their friendship and constant willingness to help.

Finally, I would like to extend my gratitude to my family for their unwavering support and belief in me.

1. INTRODUCTION

Planning is always crucial for the tourism industry, and an essential part of the planning process is forecasting tourism demand. Due to their ease of use, non-causal time series models are frequently applied in forecasting time series data. These models are based on historical patterns in the time series, such as trend and seasonality, and make forecasts based on the assumption that these patterns will persist into the future.

However, in a rapidly changing world, sudden external changes can occur, leading to changes in patterns and data anomalies in the time series. In these situations, models that perform well during periods with regular patterns in the time series may perform worse, while other models may perform better. Only recently, with the outbreak of the Coronavirus disease, the tourism industry experienced such a sudden external change that led to a sharp decline in tourism demand and a slow recovery.

This Master's Final Work (MFW) aims to further analyse this situation by measuring the impact of data anomalies caused by the Covid-19 pandemic on the forecasting performance of various time series models. For this purpose, the monthly number of Overnight Stays in touristic accommodations from 2000 to 2022 per region of Portugal is gathered from the Statistical Institute of Portugal ("Instituto Nacional de Estatística") and forms the data basis of this work.

In recent decades, Portugal has experienced an extraordinary growth in tourism, making it one of the country's most important economic sectors (Turismo de Portugal, 2021). Portugal is renowned for its diversity. City tours, historical landmarks, wine regions, culinary specialities, religious sites and long sandy beaches are just some of the reasons why tourists travel to Portugal. However, the Covid-19 pandemic had a severe impact on Portugal's tourism. As a result, Portugal's tourism data offers interesting characteristics that may vary from region to region and will be analysed in more detail in this thesis.

The forecasting performance of the different models is evaluated firstly by excluding the time period affected by the Covid-19 pandemic from the time series and secondly by including this period. In addition, the evolution of the performance over the years will be analysed. The purpose of this work is not to predict the future development of tourism, but to find out how the performance of the models changed during the Covid-19 pandemic. In particular, this MFW addresses the following research question: "How does the forecasting performance of time series models for tourism demand change with the presence of data anomalies induced by the Covid-19 pandemic?"

In the past decades, several studies have been conducted on forecasting tourism demand. The main developments in this field may be found in literature reviews (Song et al., 2019; Jiao & Chen, 2019). The diversity of forecasting models and the increased use of modelling combination techniques are described as two key developments.

However, most studies evaluated the forecasting performance of models during periods with regular patterns in the time series. The performance in situations of sudden external changes has not been given much attention, even though tourism data is characterised by a high sensitivity to external changes. Therefore, the addressed research question is relevant, and the present work can contribute to the literature by complementing these studies. To the author's knowledge, this MFW is the only academic work that uses recent tourism demand data from different regions in Portugal and analyses the impact of the Covid-19 pandemic on the predictive performance of different time series models. A detailed analysis of the literature on forecasting tourism demand can be found in chapter 2 of this MFW.

In the present work, forecasting competitions are performed, which are often described as the application of different forecasting models on the same time series to evaluate which method produces the most accurate point forecasts. Forecasting competitions were also frequently used by the statistician Rob J. Hyndman, who is responsible for important developments in the field of forecasting time series (e.g. Athanasopoulos et al., 2011; Hyndman & Athanasopoulos, 2018). This confirms that the chosen approach of this MFW is appropriate for the problem to be solved.

Since the software R provides useful functions and packages for time series analysis, the implementation for this thesis is made in R. The RMD file and the knitted HTML file are available on Github[1]. In addition, a programming walkthrough, which provides a description of the organisation of the code, can be found in Appendix B.

[1] https://github.com/RM-Mueller/MFW_l56280.

This MFW concludes that sophisticated time series models showed superior performance during periods with regular patterns in the time series, but recorded a significant loss of performance in the months following the Covid-19 outbreak. In contrast, the Naïve method is able to produce relatively good forecasts during this period due to its simplicity.

Moreover, the Drift model applied to seasonally adjusted data was often able to compete with the best models. A smaller performance decrease following the Covid-19 outbreak, allowed it to be among the top performers for predictions from 2010 to 2022 for all three time series, whereas the linear regression and Prophet models have been among the inferior forecasting models during the whole time period. In addition, the exploratory analysis of the time series provided evidence that regions with a higher share of Portuguese tourists were less affected by strong declines in tourism demand during the Covid-19 period.

The remainder of this work is structured as follows. The next section reviews recent developments in the tourism forecasting literature. This is followed by an exploratory data analysis of the time series and a description of the models used in the forecasting competitions. Thereafter, the forecasting procedure and evaluation will be explained. Finally, the results are presented before a conclusion is drawn in the last section.

2. LITERATURE REVIEW

Due to the sustained growth of the tourism market, tourism forecasting has received increasing attention. Over the last few decades, more than 600 studies have been published on modelling and forecasting tourism demand (Song et al., 2019). Many studies focus on developing forecasting techniques to improve the accuracy of predictions, which can help decision makers to enhance the efficiency of their strategic planning and minimize the risk of wrong decisions (Chen et al., 2019).

There are numerous forecast variables available to measure tourism demand. It is common to use tourist arrivals from an origin country to a destination country as a measure (Jiao & Chen, 2019). However, other variables such as the number of tourist Overnight Stays (Athanasopoulos et al., 2009; Bigović, 2012), average hotel occupancy (Pan & Yang, 2017) and tourist expenditures (Song et al., 2013) have been applied as well.

The main developments in tourism forecasting in recent decades involve the increased diversity of forecasting models, the combination of models and the improvement in forecasting accuracy (Song et al., 2019). Methodological approaches to forecast tourism demand can be divided into quantitative and qualitative methods. Time series models belong to the first category (Hyndman & Athanasopoulos, 2018).

Despite numerous studies in the field of tourism demand forecasting, there is still not a consensus on which forecasting methods are most accurate under which circumstances (Peng et al., 2014). In this dissertation the focus will be on non-causal time series models, which forecast tourism demand based on its historical patterns. Due to their ease of implementation and ability to adequately capture historical patterns, these models remain popular in the literature on tourism demand forecasting (Song et al., 2019; Jiao & Chen, 2019).

There are two types of time series approaches: Basic time series models and advanced time series models (Peng et al., 2014). The former includes the Naïve, the single exponential smoothing, and the historical mean models. The Naïve models are most often applied in tourism forecasting literature (Song et al., 2019). Despite its simplicity, the Naïve (or no change) technique is accurate in predicting annual tourism data one year ahead (Athanasopoulos et al., 2011; Witt & Witt, 1995).

The advanced time series models can be distinguished from the basic models by the fact that they integrate additional time series features, like trend and seasonality. Various types of exponential smoothing models and autoregressive integrated moving average (ARIMA) methods belong to this category (Song et al., 2019). A variety of ARIMA models have been widely used in time series analyses of tourism demand, and especially the seasonal ARIMA (SARIMA) models often prove to be the most accurate due to the seasonality of tourism data (Song & Li, 2008; Saayman & Botha, 2017).

Time series models have been used extensively in recent decades, with some general trends becoming established. Traditional time series models are often considered as benchmark models to evaluate the forecasting performance of new models. However, several new models based on ARIMA and SARIMA models have been developed with the aim of improving forecasting accuracy (Song & Li, 2008). Models, such as AR fractionally integrated moving average (ARFIMA) (Chu, 2008a), autoregressions of AR

(ARAR) (Chu, 2008b) and ARMAX models (Pan & Yang, 2017), are examples of extensions to the traditional ARIMA models. Moreover, a range of “partial linear” time series models and nonlinear time series models have been applied to forecast tourism demand (Jiao & Chen, 2019).

A major challenge in forecasting tourism demand is the high sensitivity of demand to external shocks such as terrorism, earthquakes, and diseases (Song et al., 2013). Most studies evaluated forecasting performance in time periods without external shocks (Kourentzes et al., 2021). However, a few studies have analysed the impact of a pandemic on tourism in a particular region and highlighted the difficulties in predicting tourism during a pandemic (Choe et al., 2020; Kuo et al., 2008; Mao et al., 2010).

This thesis aims to contribute to the literature by complementing these studies through analysing the performance of different time series models on forecasting tourism data during a period with data anomalies. In particular, it examines the impact of the Covid-19 pandemic on tourism demand forecasts generated by various time series models for several regions in Portugal. According to the author's knowledge, this is the first academic work addressing this issue.

3. METHODOLOGY

The above disclosure of existing research in tourism forecasting is followed by an exploratory data analysis of the time series. Based on the results, appropriate time series models are chosen, and their theoretical framework will be explained. Finally, the forecasting procedure and evaluation will be described.

3.1. Exploratory Data Analysis

The variable used for the analysis is Overnight Stays in touristic accommodations. The data set contains the monthly number of Overnight Stays from January 2000 to June 2022 per region of Portugal. According to NUTS 2 (“Nomenclatura das Unidades Territoriais para Fins Estatísticos”), Portugal is divided into seven regions: Azores, Alentejo, Algarve, Central, Lisbon Metropolitan Area (Lisbon M.A.), Madeira and North.

The data from 2000 to 2015 was gathered from the Statistical Institute of Portugal (Instituto Nacional de Estatística, 2022). As the website of the Statical Institute only provides the monthly number of Overnight Stays per region until 2015, the data from

2016 onwards is obtained from Portugal's National Tourism Authority (Turismo de Portugal, 2022). However, Portugal's National Tourism Authority cites the Statistical Institute of Portugal as the data provider. Therefore, the data source is identical.

3.1.1. All Time Series

Initially, all seven regions were examined. Figure 7 in the appendix shows the time series plots, which reveal that the time series are characterised by a strong seasonality and an upward trend until the end of 2019. There is a major change in 2020, characterised by a sharp decline in Overnight Stays, followed by a slow recovery in the subsequent years. The event, which caused these data anomalies, was the spread of Covid-19, which was declared as a pandemic by the World Health Organization on 11 March 2020 (World Health Organization, 2023). Since travel restrictions have been imposed in many countries in March 2020, this work refers to March 2020 as the month of the Covid-19 outbreak.

Although the main characteristics of the seven time series are similar, a closer analysis reveals differences. The decline in Overnight Stays after March 2020 is particularly high for the regions Azores, Lisbon M.A., and Madeira. Therefore, the recovery of tourism demand in these regions is slow in the following months. The average of the year-over-year percentage changes from April 2020 to December 2020 confirms this observation, as it is the highest for these regions at more than -78%.

Similarly, the regions Algarve and North showed a big decline in Overnight Stays. However, the time series plots indicate a faster recovery compared to the previously mentioned regions. This observation is reinforced by an average of the year-over-year percentage changes from April 2020 to December 2020 of about -71%. Tourism in Alentejo and Central recovered fastest from the Covid-19 outbreak. The average of the year-over-year percentage changes from April 2020 to December 2020 is the lowest for these two regions at below -64%.

The pandemic led to reduced air travel and international travel restrictions (Abdullah et al., 2020), which may partly explain the different recovery rates among the seven Portuguese regions. In particular, for Azores and Madeira, the reduced air traffic had a major impact on tourism demand. Moreover, most people travelled within their own country due to international travel restrictions. Therefore, there seems to be a correlation between the share of Portuguese tourists in a given region and the decline in tourism due to the Covid-19 pandemic. The following hypothesis can be made: The lower the proportion of Portuguese tourists in a given region, the higher the decline in tourism in that region due to the Covid-19 pandemic.

As the Statistical Institute of Portugal also provides information about the residence countries of the tourists (Instituto Nacional de Estatística, 2022), the validity of the hypothesis can be further assessed. In order to quantify the weight of domestic tourism and the decline of tourism during the Covid-19 period in each region, three different measurements are employed. An overview of the different measurements can be found in Table I.

All nine possible combinations between the measures of the two variables are analysed. First, a scatter plot is generated for each pair to confirm the linear relationship between the variables and the absence of outliers. In addition, for each of the measures, it is checked whether the data is drawn from a normal distribution. Pearson’s correlation tests are then performed on the nine pairs. The p-value of all correlation tests is below the 5% significance level. This means that the null hypothesis of no linear relationship between the two variables is rejected. The correlation coefficient ranges between 0.80 and 0.96, indicating a strong positive relationship between the two variables. These results prove that there is indeed significant evidence of a correlation between the proportion of Portuguese tourists in a given region and the decline in tourism due to the Covid-19 pandemic among the Portuguese regions.

Finally, this chapter has shown that there are three groups of regions according to different rates of recovery from the Covid-19 pandemic. In the following, the focus of the MFW will be on the analysis of one region per group: Madeira, Algarve and Alentejo.

3.1.2. Madeira

The time series plot for Madeira in Figure 1 reveals a significant seasonality of the time series. The seasonal subseries plot, which can be found as Figure 8 in the appendix, shows that the number of Overnight Stays is on average highest in August and lowest in December. In 2019, for example, Madeira recorded 828,426 Overnight Stays in August, but only 445,470 Overnight Stays in December. This corresponds to a reduction of 46.23%.

Moreover, the graphs indicate an upward trend until February 2020, which is particularly strong from 2010 until 2019. From 2000 to 2009, Madeira recorded an average annual increase of 1.29% in the number of Overnight Stays, while this figure rose to 3.42% from 2010 to 2019.

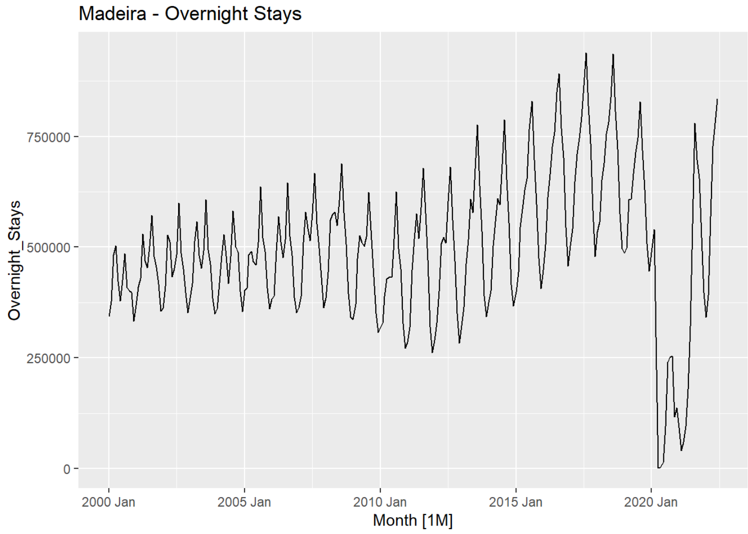

Figure 1 – Monthly Overnight Stays in Madeira.

Furthermore, the time series plot for Madeira reveals a seasonal variation that increases proportional with the level of the time series from 2000 to 2019. The difference between the month with the highest number of Overnight Stays, August, and the month with the lowest number of Overnight Stays, December, was 151.675 Overnight Stays in 2000. By contrast, the difference between these two months in 2019 was 382,956 Overnight Stays, which corresponds to more than a doubling of the variation.

However, after the Covid-19 outbreak, these patterns changed. In March 2020, Madeira recorded 297,649 Overnight Stays, which was 50.93% lower than in March 2019. In April 2020, the number of Overnight Stays decreased even further to 2,191 Overnight Stays, which corresponds to a year-over-year percentage change of -99.64%.

Therefore, the regular seasonal pattern, which was observed before, did not continue in the months after March 2020. The month October corresponds in 2020 to the peak month with 253,895 Overnight Stays. The following low point of the data can be found with 40,883 Overnight Stays in February 2021. In the following months, the seasonality seems to converge back to the previous pattern: August 2021 corresponds to the peak with 780,123 Overnight Stays and January 2022 to the low with 341,744 Overnight Stays.

In addition, the time series indicates an upward trend after the Covid-19 outbreak. This is illustrated by Figure 9 in the appendix, which shows the time series with the superimposed trend-cycle component by applying Seasonal-Trend decomposition using LOESS (STL decomposition). In 2020, Madeira reported an annual number of Overnight Stays of 2,441,536, but in 2021 the figure had risen to 4,395,765 Overnight Stays. This is a relative increase of approximately 80%.

Despite the fact that not all data is available for 2022, it can be stated that the number of Overnight Stays from January to June 2022 for each month is higher than in the same month of the previous year. As an example, in June 2022, the number of Overnight Stays was 833,955, approximately 159% higher than in the previous year and even 16.79% higher than in 2019.

To conclude, the Covid-19 outbreak had a significant impact on the time series of Madeira. The sharp drop in the number of Overnight Stays in March/April 2020 also affected the subsequent months. Therefore, no seasonal pattern could be observed in 2020. However, tourism recovered fast and in the following years the seasonality seemed to converge back to the previous patterns. Similarly, an upward trend pattern is visible. For some months in 2022, the number of Overnight Stays is even higher than in 2019.

3.1.3. Algarve

The strong seasonal pattern of tourism in the Algarve is highlighted by the time series plot in Figure 2. There is a big difference between the summer months and the winter months. The seasonal subseries plot in Figure 10 in the appendix shows that on average December is the month with the lowest and August the month with the highest number of Overnight Stays. For example, in August 2019 Algarve reported 3,439,271 Overnight Stays, whereas in December 2019 the number of Overnight Stays was 602,860. This corresponds to a reduction of 82.47%.

Moreover, the time series plot reveals an upward trend before the Covid-19 outbreak. This is particularly evident between 2010 and 2020, when the Algarve recorded an average annual increase of 4.96% in the number of Overnight Stays. Before 2010, there was even a slight downward trend, with an average annual decrease of 1.22% between 2000 and 2009.

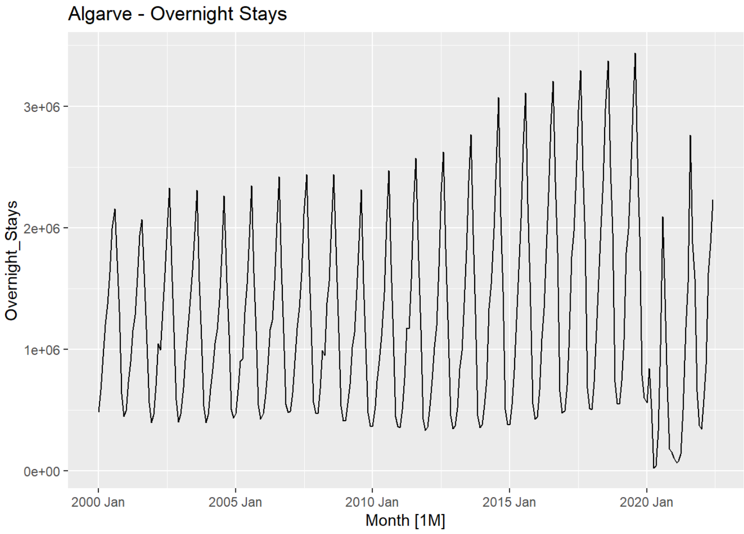

Figure 2 – Monthly Overnight Stays in Algarve.

In addition, an increasing seasonal variation with the level of the time series before March 2020 can be detected by looking at the time series plot. In 2000, the number of Overnight Stays decreased by -1,707,565 from August to December, while in 2019 it decreased by -2,836,411.

The Covid-19 outbreak resulted in a big decline in Overnight Stays. In March 2020, the number of Overnight Stays was 565,558 lower than the year before, which corresponds to a reduction of 52.42%. In April the number decreased even further, such that there were only 25,614 Overnight Stays, which is a reduction by 98.57% compared to April 2019.

However, the time series plot shows that there was still a seasonal pattern, which is consequently at a lower level of the time series. In 2020 and 2021, August remains the peak month, with 2,090,569 and 2,762,537 Overnight Stays, respectively. The lowest number of Overnight Stays remains in winter. Although there is a slight shift backwards. In 2021, February was the month with the lowest number of Overnight Stays and in 2022 it was January.

Furthermore, an upward trend after the Covid-19 outbreak can be detected. An illustration of the trend can be found in Figure 11 in the appendix, which shows the time series with the superimposed trend-cycle component by applying STL decomposition. In 2021, the annual number of Overnight Stays increased by 37.81% to 10,874,036.

In 2022, the data is only available until June. However, by looking at the seasonal differences, an increase in Overnight Stays can be detected every month. For example, in June 2022, the number of Overnight Stays was 2,230,207, which is approximately 89% higher than in the previous year and only 8.06% lower than in 2019.

In summary, the Covid-19 outbreak led to a sharp decline in the number of Overnight Stays in Algarve. However, the same patterns can be observed in the data thereafter, namely an upward trend and a strong seasonality. Moreover, the number of Overnight Stays in 2022 appears to be almost at the same level as before the outbreak.

3.1.4. Alentejo

The time series plot for Alentejo in Figure 3 shows a significant seasonality. This observation is confirmed by the seasonal subseries (see Figure 12 in the appendix), which reveals, that on average the number of Overnight Stays is lowest in January and highest in August. For example, in August 2019 Alentejo recorded 484,403 Overnight Stays, whereas in January 2020 this figure decreased to 126,349 Overnight Stays. This represents a reduction of approximately 74%.

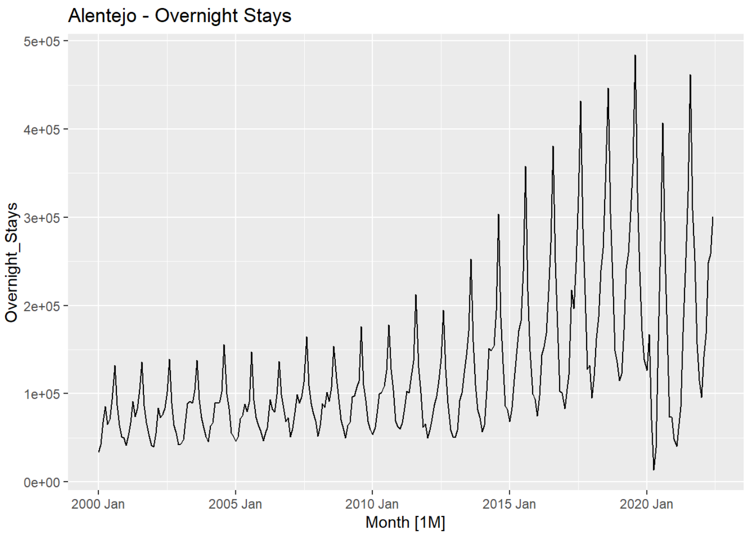

Figure 3 – Monthly Overnight Stays in Alentejo.

Furthermore, the time series plot shows an upward trend before the Covid-19 outbreak. It reveals that, particularly from 2010 to 2019, the annual number of Overnight Stays increased. From 2000 to 2009, the annual number of Overnight Stays increased on average by 3.12%, whereas from 2010 to 2019 an average increase of 10.61% was reported.

In addition, the time series plot for Alentejo shows before the Covid-19 outbreak a seasonal variation that increased strongly with the level of the time series. Alentejo reported a difference of 90,732 Overnight Stays between August 2000 and January 2001. In contrast, the difference between August 2019 and January 2020 was 358,054 Overnight Stays, which corresponds nearly to a fourfold increase in seasonal variation.

However, the Covid-19 outbreak led to a sharp decline in Overnight Stays. The number of Overnight Stays was 70,530 in March 2020, approximately 59% lower than in March of the previous year. In April 2020, the figure fell further to 13,892, which corresponds to a year-over-year percentage change of 94.25%.

Despite the pandemic, the seasonal pattern remained relatively stable. In 2020, August was the month with the highest number of Overnight Stays with 407,020, followed by a low point in February 2021 with 40,443 Overnight Stays. Thereafter, August 2021 and January 2022 clearly correspond to the highest and lowest points.

Furthermore, the time series shows an upward trend after the Covid-19 outbreak. This is visualized in Figure 13 in the appendix, which shows the time series with the superimposed trend-cycle component by applying STL decomposition. In 2021, the annual number of Overnight Stays was 2,280,089, which is approximately 25% higher than in the previous year.

The observed months in 2022 confirm this observation, because all months record a higher number of Overnight Stays than the same month of the previous year. For example, Alentejo reported 300,520 Overnight Stays in June 2022, which is 18.20% higher than in June 2021 and only 1.81% lower than in 2019.

In summary, the Covid-19 outbreak led to a sharp decline in the number of Overnight Stays in Alentejo. However, tourism in Alentejo recovered relatively fast and the same core patterns can be observed in the data thereafter, namely an upward trend and a strong seasonality. In addition, the number of Overnight Stays seems to reach nearly the same level again in 2022.

3.2. Time Series Models

In this section, the details of the models applied to forecast the number of Overnight Stays will be described. This MFW focuses on univariate time series models that use only information on the variable being forecast. Therefore, all models are based on the same data to allow a direct comparison between the performance of the various models.

The exploratory analysis for Madeira, Algarve and Alentejo revealed patterns in the time series, which are considered to choose appropriate models for the forecasting competition. In order to enable comparability, the same forecasting models are selected for each region. Moreover, the author has chosen simple as well as more sophisticated forecasting models to increase diversity in the forecasting competition. The following paragraphs describe which models are included in the forecasting competition for which reasons. A detailed explanation of the respective method is provided in the following subchapters.

The time series plots of Madeira, Algarve and Alentejo show regular patterns before the Covid-19 outbreak: A strong seasonality and an upward trend. For this reason, the Seasonal Naïve method, which captures the seasonal pattern of the time series, and the Random Walk with Drift method applied to seasonally adjusted data obtained from a STL decomposition, which considers both trend and seasonality, are chosen. Since the patterns changed after the Covid-19 outbreak and in particular for Madeira no trend, no seasonality and no stable level in 2020 could be identified, the Naïve method is included in the forecasting competition as another simple forecasting model.

In addition, Exponential Smoothing (ETS) and ARIMA models are chosen since they capture both seasonality and trend. The Holt-Winters' Additive Method is selected as an ETS model, which assumes seasonal variations that are roughly constant throughout the series. The reason for this is that all models are also included in the forecasting competition with the log-transformed variable. Therefore, the logarithmic transformation already considers an increasing seasonal variation. Also, the automatic ETS modelling function (Hyndman et al., 2002; Hyndman et al., 2008) is used in order to evaluate if it provides more accurate predictions.

Furthermore, the automatic ARIMA modelling function provided by the $fable$ package is applied. This function uses a combination of Hyndman-Khandakar algorithm

(Hyndman & Khandakar, 2008) to obtain the ARIMA model with the lowest Corrected Akaike's Information Criterion (Hyndman & Athanasopoulos, 2018). A fully automated ARIMA model, where all parameters are defined automatically, is added to the forecasting competition. In addition, the following two ARIMA models are implemented with predefined parameters as well as automatically selected parameters using the automatic ARIMA modelling function: ARIMA $(p, 1, q)(P, 1, Q)_{12}$ and ARIMA $(p, 0, q)(P, 1, Q)_{12}$.

Lastly, a linear regression model including trend and seasonality components and a Prophet model is included in the forecasting competition, which is able to capture trend, seasonality and holiday effects. The model characteristics of the Prophet model are automatically selected.

3.2.1. Naïve and Seasonal Naïve Model

The Naïve model, which is also called Naïve I, is a simple forecasting method which uses the most recent observation as a forecast. It is frequently used as a benchmark against sophisticated models, but can also be useful for data following a random walk (Hyndman & Athanasopoulos, 2018). Despite its simplicity, the Naïve model was able to outperform more complex models in several studies (Goh & Law, 2011).

The Seasonal Naïve model (SNaïve) sets each prediction equal to the last observed value from the same season. This method is useful for time series with strong seasonality (Hyndman & Athanasopoulos, 2018).

3.2.2. Drift Model applied to seasonally adjusted Data

The Drift model, which is also called the constant change model, is appropriate for time series with a significant trend component. The forecast is the last observation plus a drift term. The mean change in historical data is used as the value of the drift term (Hyndman & Athanasopoulos, 2018). Therefore, the forecast for time $T + h$ is given by:

(1) $$\hat{y}_{T+h|T} = y_T + h \left(\frac{y_T - y_1}{T-1}\right),$$

where historical data is denoted by $y_1, \ldots, y_T$.

If the Drift model is applied to seasonally adjusted data, the model is able to capture both, trend and seasonality. The STL decomposition was developed by Cleveland et al.

(1990) and uses the LOESS method to divide a time series additively into trend, seasonal and remainder components. Therefore, it is able to provide seasonally adjusted data. In this MFW the STL decomposition is used with its default parameters. By using the $decomposition\_model()$ function in R the forecasts will be "reseasonalised" automatically by adding in the seasonal naïve forecasts of the seasonal component.

3.2.3. Prophet Model

The Prophet model was implemented by Facebook's Core Data Science team (Taylor & Letham, 2018) and was originally intended to predict daily data with weekly and annual seasonality, as well as holiday effects. However, it has since been expanded to predict several types of seasonality (Satrio et al., 2021; Hyndman & Athanasopoulos, 2018).

It works best if the times series shows a strong seasonality and there are several seasons of historical data available. The Prophet model can be considered as a nonlinear regression model, which decomposes the time series into three main model components: Trend, seasonality, and holidays (Hyndman & Athanasopoulos, 2018). They are combined in the following equation:

$$y_t = g(t) + s(t) + h(t) + \epsilon_t \ , \tag{2}$$

where $g(t)$ is a piecewise-linear trend, $s(t)$ represents the seasonal pattern, and $h(t)$ captures the holiday effects, which may occur at irregular intervals on one or more days. The last part ϵ_t represents the error term, which accounts for any unusual changes not accommodated by the model.

The formulation is similar to a generalized additive model, which is a class of regression models in which typically non-linear smoothing factors are applied to the regressors. It enables easy decomposition and new components can be added, if necessary. For example, if a new source of seasonality is identified. In the case of the Prophet model, the only regressor is time and several linear and nonlinear functions of time can be used as components.

The key advantages of the Prophet model are that outliers are handled relatively well and it is robust to missing data as well as shifts in the trend. It is available via the $fable.prophet$ package in R. In this MFW the automatic selection of the parameters is used. Only the period of the seasonal component is set to 12, which stands for monthly data (Hyndman & Athanasopoulos, 2018).

3.2.4. Linear Regression Model

A Time Series Linear Model (TSLM) is a regression model which assumes a linear relationship between the forecast variable y and the predictor variables x. The simplest case involves only one predictor variable (Hyndman & Athanasopoulos, 2018) and is represented by the following equation:

(3) $$y_t = \beta_0 + \beta_1 x_t + \epsilon_t \quad ,$$

where β_0 represents the intercept and β_1 denotes the slope of the line. The last part ϵ_t is the error term, which captures everything that may affect the forecast variable y_t other than x_t.

Trend and seasonality components can be added to the simple linear regression model as predictor variables (Tiwari et al., 2017). They are created from the characteristics of the time series data. The variable "trend" represents a linear trend and "season" is a factor indicating the season, which depends on the frequency of the data. For example, there will be eleven dummy variables for monthly data. Therefore, the TSLM for monthly data with trend and seasonality components is given by:

(4) $$y_t = \beta_0 + \beta_1 t + \beta_2 d_{2,t} + \beta_3 d_{3,t} + \beta_4 d_{4,t} + \beta_5 d_{5,t} + \beta_6 d_{6,t} + \beta_7 d_{7,t} + \beta_8 d_{8,t} + \beta_9 d_{9,t} + \beta_{10} d_{10,t} + \beta_{11} d_{11,t} + \beta_{12} d_{12,t} + \epsilon_t \quad ,$$

where β_0 represents the intercept. β_1 denotes the trend and $t = 1, \dots, T$. The variables $d_{i,t}$ are dummy variables which are equal to 1 if t is in month i and 0 otherwise. The last part ϵ_t is the random error term.

In this MFW, the $trend()$ and $season()$ components within the TSLM function are used allowing R to automatically estimate a model from the data that minimizes the sum of squared errors (Hyndman & Athanasopoulos, 2018).

3.2.5. Holt-Winters' Additive Model

The Holt-Winters' Additive model belongs to the ETS family of models. ETS is a technique that assigns a weight to each observed value in order to predict upcoming values. More recent observations have a higher weight and there is a decay to the older observations, which will have a lower impact on the predictions. This is based on the idea that the most recent observations should explain the future value of the variable better than older observations (Hyndman & Athanasopoulos, 2018). ETS was introduced in the 1950s (Brown, 1959; Holt, 1957; Winters, 1960) and has been widely used in both academia and practice ever since. It is popular mainly because of its transparency and relatively good performance. Furthermore, it has been shown to perform well in forecasting tourism arrivals (Kourentzes et al., 2021).

The Holt-Winters' Method (Holt, 1957; Winters, 1960) is a specific ETS model, which is able to capture seasonality and trend. The framework consists of one forecast equation and three smoothing equations, which are for level, trend and seasonality components. There are two types of Holt-Winters' models, which differ in the nature of the seasonal variation: The additive and the multiplicative method. While the additive method is appropriate, when the data shows a seasonal variation that is roughly constant with the level of the time series, the multiplicative method should be used, if the seasonal variation is increasing with the level of the time series (Hyndman & Athanasopoulos, 2018).

The equations of the Holt-Winters' Additive Method are as follows (Hyndman & Athanasopoulos, 2018):

(5) $\hat{y}_{t+h|t} = l_t + hb_t + s_{t+h-m(k+1)}$,

(6) $l_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1})$,

(7) $b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1}$,

(8) $s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$,

where $y_1, \dots, y_t$ represents historical data, l_t the level of the time series, b_t the trend, s_t the seasonal component and $\hat{y}_{t+h|t}$ the forecast for h periods ahead. m denotes the period of the seasonality, i.e. the number of seasons in a year. The parameter k is the integer part of $(h-1)/m$, ensuring that the estimates of the seasonal indices used for the forecasts are from the last year of the sample. The smoothing parameters α, β, γ are estimated from 0 to 1 and used to minimize mean squared errors. The parameter α controls the weight given to each observation based on the available information at a specific point of time. The level equation is a weighted average between the seasonally adjusted observation $(y_t - s_{t-m})$ and the non-seasonal forecast $(l_{t-1} + b_{t-1})$ for time t. The trend equation reflects a linear trend. Lastly, the seasonal equation is a weighted average of the current seasonal index $(y_t - l_{t-1} - b_{t-1})$ and the seasonal index of the same season of the previous year.

The forecasting competitions of this MFW include the Holt-Winters' Additive model. In addition, the automatic ETS modelling function (Hyndman et al., 2002; Hyndman et al., 2008) is used, which selects the model by minimising the Corrected Akaike's Information Criterion (AICc) and optimises the parameter values by maximum likelihood estimation.

3.2.6. Seasonal ARIMA Model

Autoregressive integrated moving average (ARIMA) is a model used for predictions, which aims to describe the autocorrelations in the data. It considers the lagged values of the predicted variable as well as a moving average component. Box & Jenkins (1970) popularised ARIMA modelling in the early 1970s. For time series forecasting, ARIMA models are one of the most popular approaches (Hyndman & Athanasopoulos, 2018).

Since the time series for Madeira, Algarve and Alentejo show a strong seasonal pattern, this MFW focuses on the SARIMA model, which accounts for the seasonality of the data. SARIMA is composed of non-seasonal and seasonal autoregressive (AR) and moving average (MA) models. It can be written as a composite model of the form:

(9) $$ARIMA(p,d,q)(P,D,Q)_m \ ,$$

where p, d and q represent the non-seasonal AR order, non-seasonal differencing and non-seasonal MA order, respectively, whereas P, D and Q denote the seasonal AR order, seasonal differencing and seasonal MA order, respectively. The model parameter m is the seasonal period, which corresponds to the number of observations per year.

Mathematically, the general form of the SARIMA is given by:

(10) $$\begin{aligned}&\left(1-\phi_1 B-\cdots-\phi_p B^p\right)(1-\Phi_1 B^m-\cdots-\Phi_P B^{Pm})(1-B)^d(1-B^m)^D y_t \\ &= c+\left(1+\theta_1 B+\cdots+\theta_q B^q\right)(1+\Theta_1 B^m+\cdots+\Theta_Q B^{Qm})\varepsilon_t \ ,\end{aligned}$$

where $\left(1-\phi_1 B-\cdots-\phi_p B^p\right)$ is the non-seasonal AR(p). $(1-\Phi_1 B^m-\cdots-\Phi_P B^{Pm})$ represents the seasonal AR(P). $(1-B)^d$ is the non-seasonal difference and $(1-B^m)^D$ the seasonal difference. $\left(1+\theta_1 B+\cdots+\theta_q B^q\right)$ represents the non-seasonal MA(q) and

$(1 + \Theta_1 B^m + \cdots + \Theta_Q B^{Qm})$ the seasonal MA(Q). Moreover, y_t is the time series observation at time period t and B is the backward shift operator, which reflects the time series lags. The parameter ε_t is a sequence of the error term with mean zero and constant variance.

A common obstacle in using ARIMA models for forecasting is the initialisation of the parameters, which is subjective and can be difficult to apply (Ponnam et al., 2016). Motivated by this observation, there have been several attempts to automate ARIMA modelling. In this work the automatic ARIMA modelling function provided by the $fable$ package in R (Hyndman & Khandakar, 2008) is used. The function applies a variant of the Hyndman-Khandakar algorithm, which combines unit root tests, minimisation of the AICc and maximum likelihood estimation to generate an ARIMA model. It is possible to have all parameters determined by the automatic ARIMA modelling function or to define certain parameters and use the automatic function for the remaining parameters (Hyndman & Athanasopoulos, 2018).

The automatic ARIMA modelling function is applied in various ways in this MFW. In addition to a fully automatically determined ARIMA model, several "semi-automatic" ARIMA models are implemented, where only certain parameters are defined.

In order to determine the parameter d, the automatic ARIMA modelling function applies the KPSS test, which tests the null hypothesis that the data is stationary against the alternative that it has a unit root (Kwiatkowski et al., 1992). However, there is empirical evidence that the KPSS test does not perform well in selecting the appropriate order differencing in certain circumstances (Müller, 2005; Shin & Schmidt, 1992). Therefore, this work aims to be independent of the results of a statistical test and considers the two most common situations: $d = 1$ and $d = 0$.

Following the recommendations of Hyndman & Athanasopoulos (2018) the $nsdiffs()$ function of the $fable$ package is applied, which is based on the seasonal strength of a STL decomposition, to determine whether seasonal differencing is necessary. Since $F_s \geq 0.64$ for all time series, one seasonal difference is suggested. Therefore, the parameter D is set to 1 for the “semi-automatic” ARIMA models.

3.3. Forecasting Procedure and Evaluation

The exploratory data analysis revealed that the seasonal variation of the time series increased until March 2020. However, no increase in seasonal variation could be observed during the pandemic, as the time span is too short. Since the logarithmic transformation of the predicted variable is a suitable means to stabilise the variance (Box & Jenkins, 1970; Lütkepohl & Xu, 2012), each model appears twice in the forecasting competition: Once with forecasts based on the time series with the original variable, and once with forecasts based on the log-transformed variable.

In order to separate the available data dynamically into training and test data, two types of time series cross-validation are used: Recursive window and rolling window cross-validation. The function $stretch\_tsibble$ enables recursive window cross-validation (recursive CV) in R. In this work the initial window size is 120, which means that the initial training set contains observations from January 2000 to December 2009. The number of Overnight Stays for January 2010 is the first test observation and will be compared to the forecasted value calculated by the various models. The window size increases incrementally by 1 month. This procedure will be continued until the test data set contains observations from January 2000 to May 2022 and the number of Overnight Stays in June 2022 will be the test observation.

The function $slide\_tsibble$ is used for rolling window cross-validation (rolling CV) and shifts a fixed-length window through the data. The window size is set to 120 months and the step size to 1, which means that for each validation fold the training and test windows are shifted by one month. The first training and test sample produced by the rolling CV is the same as produced by a recursive CV. However, the window slides through the data, and therefore the remaining training data sets differ. As an example, the last training data set contains observations from June 2012 to May 2022 and uses the number of Overnight Stays of June 2022 as test observation. It should be noted that in the case of the automatic ARIMA and ETS modelling functions, cross-validation may result in the selection of different models for the different training data sets.

As this MFW is interested in the performance of different models in predicting the next observation, the forecast horizon is set to $h = 1$, which represents forecasts of one month ahead. To assess the forecast accuracy, point forecasts are compared to the actual number of Overnight Stays in a given month and region. The evaluation of the forecast accuracy is based on two metrics: The Root Mean Squared Error (RMSE) and the Root Mean Squared Scaled Error (RMSSE). The forecast errors of the RMSE are on the same scale as the data and therefore scale dependent. The equation is given by:

(11) $$RMSE = \sqrt{mean(\epsilon_t^2)} ,$$

where ϵ_t is the absolute error at time t.

Since the three time series showed a significantly lower level of Overnight Stays in 2020, the situation can be compared to a different scale. Therefore, the RMSSE, which was introduced in 2006 by Hyndman & Koehler, is used as a scale-independent measure. For seasonal time series the errors are scaled based on the in-sample MAE from the Seasonal Naïve forecast method. Thus, a scaled error is defined as:

(12) $$q_j = \frac{\epsilon_j}{\frac{1}{T-m}\sum_{t=m+1}^{T}|y_t - y_{t-m}|} ,$$

where ϵ_j is the absolute error at time j and m is the seasonal period.

Consequently, the RMSSE is given by:

(13) $$RMSSE = \sqrt{mean(q_j^2)} .$$

RMSSE is smaller than one if the forecast is better than the average one-step seasonal Naïve forecast calculated from the training data. Conversely, it is bigger than one if the forecast is worse than the average one-step seasonal naïve forecast calculated from the training data. Due to the squaring portion of RMSE and RMSSE, larger errors are penalised more than smaller errors.

4. RESULTS

In this chapter the results of the several forecasting competitions will be described and analysed in different dimensions. The first three sections summarise the forecasting results for Madeira, Algarve, and Alentejo and describe how the models perform according to the RMSE. In order to assess the forecasting performance of the methods, three aspects of the results are considered: (i) Development of absolute forecasting errors; (ii) development of overall RMSE by including more years in the time series and according ranking of the models; and (iii) development of yearly RMSE in particular during the Covid-19 pandemic. The main focus will be on the second aspect, with consideration of the other aspects since they may lead to a different view of the results. As the same figures and tables appear in all three subchapters, an explanation of their structure is given once for Madeira. Thus, the sections for Algarve and Alentejo focus only on the results.

In the following two chapters, additional dimensions of the results are analysed. Firstly, the robustness of the results to a scale-independent evaluation metric, the RMSSE, is examined. Secondly, the performance of the models after varying the estimation window is analysed. Lastly, the final chapter provides an insight into another area of research, analysing how to improve forecasting performance following data anomalies.

In order to increase the readability of this work and to use consistent wording, the same short names for the models are used as in the R code. The two ARIMA models, ARIMA $(p, 1, q)(P, 1, Q)_{12}$ and ARIMA $(p, 0, q)(P, 1, Q)_{12}$, are called "ARIMA_d1D1" and "ARIMA_d0D1", respectively. The fully automatic models using the automatic ARIMA modelling function and the automatic ETS modelling function are called "ARIMA_auto" and "ETS_auto". The Drift model applied to seasonally adjusted data using STL decomposition is shortened as "Drift_STL". Moreover all models, which are based on a log-transformed variable end with "_log". Consequently, all models are based on the original variable, if they do not end with "_log".

4.1. Forecasting Results: Madeira

This section describes the forecasting results of the various models for Madeira. As mentioned above, the forecasting competition is performed twice for Madeira. The difference is that the division of the time series into training and test data is based on two different types of time series CV: Recursive and rolling CV. Firstly, this chapter focuses on the results obtained based on a recursive CV and thereafter, a comparison to the results based on a rolling CV will be made.

Figure 14 in the appendix shows the point forecasts of the models versus the actual number of Overnight Stays, which is displayed in black. As the number of models is high, it is difficult to differentiate each model. However, first insights can be gained. The performance of all models before 2020 seemed to be relatively stable, whereas during the Covid-19 pandemic the forecasts of most models were inaccurate. Particularly, the

SNaïve, TSLM and Prophet methods generated forecasts, which are far away from the actual number of Overnight Stays.

Now, the performance of the different models is evaluated by using the RMSE to assess the point forecast accuracy. Figure 4 shows how the RMSE developed by including more years in the time series and how the corresponding ranking of the models changed. In the table on the left, the two columns headed “2010-2019” show the RMSE per model based on the forecasts from January 2010 to December 2019 and the corresponding ranking. The two columns headed “2010-2022” display the RMSE per model based on the January 2010 to June 2022 forecasts and the corresponding ranking.

For clarity, it is important to note that each model appears only once in the following analysis. That is, if the model based on the log-transformed variable achieved a lower RMSE for the forecasts from 2010 to 2022, only that model appears. Conversely, the model based on the original variable would only appear if it achieved a lower RMSE. However, for the interested reader, all figures and tables of this MFW are also available with the full range of models in the files provided on Github.

The line plot on the right side of Figure 4 shows the development of the RMSE by including more years in the time series. The RMSE is always calculated from 2010 and the x-axis shows up to which year the RMSE is calculated. Therefore, this plot reveals which years’ inclusion changed the ranking of the models.

	2010-2019		2010-2022	
Model	Rank	RMSE	Rank	RMSE
ARIMA_d0D1	1	24,386.04	1	59,729.41
Drift_STL	4	26,899.35	2	60,088.61
ARIMA_d1D1	2	24,724.45	3	60,178.35
ARIMA_auto	3	24,913.02	4	60,451.53
HWA_log	5	30,719.70	5	65,805.38
ETS_auto_log	6	34,099.55	6	66,826.92
Naive	9	82,367.68	7	91,659.07
Prophet_log	8	67,677.75	8	171,804.12
TSLM	10	90,460.21	9	180,367.90
SNaive	7	53,807.75	10	204,508.35

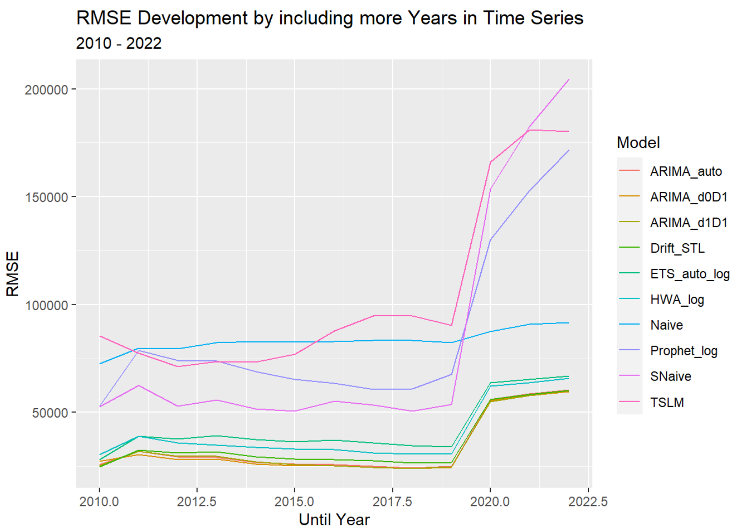

Figure 4 – Development of RMSE of Madeira’s Forecasts based on Recursive CV.

By looking at the line plot, it can be depicted that the RMSE stayed at a relatively stable level for all models before the Covid-19 outbreak. The performance of all ARIMA and ETS models as well as the Drift_STL model was close to each other and their RMSE stayed lowest during this period. Whereas TSLM, SNaïve, Naïve and Prophet_log produced more inaccurate forecasts as their RMSE is consistently at a higher level.

The table shows that the RMSE of the forecasts from 2010 to 2019 for the ARIMA, ETS and Drift_STL models is between 24,000 and 35,000, while the RMSE varies between 53,000 and 91,000 for the remaining models. By the end of 2019, the RMSE is with 24,386.04 lowest for ARIMA_d0D1 followed by ARIMA_d1D1 and ARIMA_auto.

But the Covid-19 outbreak and in particular the inclusion of 2020 in the time series led to a big increase in the RMSE for all models. An exception is the Naïve Model and therefore the forecasting performance of Naïve was able to get closer to the models which are more competitive.

Table II in the appendix presents the absolute and relative differences between the calculated RMSE based on the forecasting errors from 2010 to 2019 and the calculated RMSE based on the forecasting errors from 2010 to 2020. Accordingly, it provides us with information on the extent of the increase in the RMSE by including 2020 in the time series. It reveals that the RMSE only increased by 6.29% for the Naïve Model, but by 185.67% for the SNaïve model. However, also the three ARIMA models, which showed superior performance before the Covid-19 outbreak, recorded with more than 122% a big increase in the RMSE.

Moreover, the line plot in Figure 4 shows that the inclusion of 2021 and 2022 did not lead to another big increase in the RMSE for the ARIMA, ETS and Drift_STL models. By contrast, Prophet_log, TSLM and SNaïve seem to be inferior to the other methods during the Covid-19 period, since their RMSE increased further.

Finally, the ranking of the models did not significantly change by including the Covid-19 period in the time series. One reason for this is that the RMSE in Figure 4 always includes the forecasts from 2010 onwards for the calculation, and the Covid-19 period is only a relatively short period of 2.5 years. The three ARIMA models together with Drift_STL still achieve the best overall performance. They are followed by the two ETS models, which remain on rank 5 and 6. Much closer to them is now the Naïve model which is ranked 7th. The Prophet_log, TSLM and SNaïve are still ranked at the bottom.

But how precise were the forecasts of the models for each particular year? Figure 15 in the appendix gives an answer to this question as it shows the RMSE per year. The line plot emphasizes the main conclusions drawn above. However, the performance of the models in 2020, 2021 and 2022 is particularly interesting as it provides new findings. The graph reveals that with an RMSE of about 128,000 the Naïve method was able to produce the most accurate forecasts in 2020. It is followed by the ARIMA, ETS and Drift_STL models whose performance is close to each other, with an RMSE between 165,000 and 182,000. This number seems acceptable, but it corresponds to a relative increase in the RMSE for 2019 to 2020 of above 430%. Prophet_log, SNaïve and TSLM performed worst in 2020.

In 2021, all models were able to produce better forecasts than in 2020, as their RMSE of 2021 is lower compared to their RMSE of 2020. The performance of the ARIMA, ETS and Drift_STL models increased strongly, resulting in the best performance in 2021 and consequently a better performance than the Naïve model. However, in 2022 the performance of the Naïve model was again able to get closer to the more competitive models. During the whole Covid-19 period Prophet_log, SNaïve and TSLM performed worst.

By comparing these results with those obtained using a rolling CV (see Figure 16 in the appendix), mainly the same conclusions can be drawn. Therefore, this paragraph only describes the major differences. TSLM performed better before the Covid-19 outbreak and was consequently able to produce more accurate forecasts than Naïve during this period. As a result, Naïve was ranked lowest, when considering the RMSE from 2000 to 2019. Once again, the ARIMA, ETS and Drift_STL models competed for the best ranks, with their performances consistently close to each other from 2010 to 2022. The only exception is the ARIMA_d0D1 model, which showed a higher increase in the RMSE by including 2020 in the time series.

4.2. Forecasting Results: Algarve

The forecasting results for Algarve will be described in this chapter. Firstly, the focus will be on the forecasting results based on a recursive CV. Thereafter, the results will be compared to the results obtained based on a rolling CV.

The line plot in Figure 17 in the appendix shows the point forecasts of all 20 models and the actual number of Overnight Stays from 2010 to 2022 in black. The number of models is large. Therefore, only first fundamental insights can be gained. The performance of all models seemed to be relatively stable from 2010 to 2019, but the forecasts became less accurate, due to the outbreak of Covid-19. Particularly in 2020, the performance of all models appeared to be poor, but improved in the subsequent years.

In the following, the performance of the different models is evaluated by using the RMSE to assess the point forecast accuracy. Figure 5 shows how the RMSE developed by including more years in the time series and how the corresponding ranking of the models changed. The table and the line plot follow the same structure as in Figure 4.

	2010-2019		2010-2022	
Model	Rank	RMSE	Rank	RMSE
Drift_STL	7	116,436.33	1	192,094.8
ARIMA_auto	1	91,874.23	2	201,552.5
ARIMA_d0D1	2	92,494.37	3	202,684.9
ARIMA_d1D1	3	96,707.66	4	208,197.1
HWA_log	5	113,161.96	5	231,042.4
ETS_auto_log	4	110,633.52	6	234,012.4
Prophet_log	8	191,870.60	7	415,313.8
TSLM	9	294,055.70	8	439,795.6
SNaive	6	114,889.37	9	464,522.7
Naive	10	517,678.97	10	519,314.1

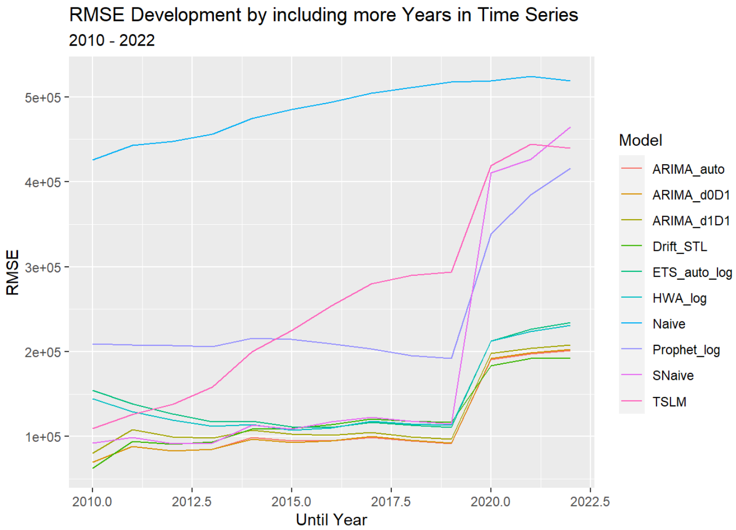

Figure 5 – Development of RMSE of Algarve’s Forecasts based on Recursive CV.

The line plot reveals that the performance of all models was at a relatively stable level before the Covid-19 outbreak. There are only two exceptions: Naïve and TSLM recorded a steady increase in the RMSE, which was particularly high for TSLM.

According to the two illustrations in Figure 5, the performance of the models by the end of 2019 can be divided into three groups: The best performing models, the second best performing models and the worst performing models. The three ARIMA models are the best performing models as their RMSE is the lowest between 91,000 and 97,000. The second group form the two ETS, the SNaïve and the Drift_STL models with an RMSE between 110,000 and 117,000. Lastly, the worst performing models from 2010 to 2019 are Prophet_log, TSLM and Naïve. In particular the forecasts of the Naïve model resulted in a high RMSE with above 517,000.

However, the inclusion of the year 2020 led to a big increase in the RMSE, which was highest for SNaïve and lowest for Naïve. Table III in the appendix displays the absolute and relative increase in the RMSE due to the inclusion of 2020. While the RMSE of the Naïve model increased by only 0.21%, the RMSE of the SNaïve rose by 257.74%. With an increase of above 104% also the performance of the three ARIMA models has gotten significantly worse. They are followed by the two ETS models, which recorded an increase in the RMSE of around 90%. The relative difference of Drift_STL is the third smallest with 57.44%.

Nevertheless, the line plot in Figure 5 shows that the inclusion of 2021 and 2022 did not lead to such a further strong increase in the RMSE. For the ARIMA, ETS, Naïve and Drift_STL models, the overall RMSE remained at a relatively stable level. For TSLM, SNaïve and Prophet_log, the RMSE increased further, with the largest increase for Prophet_log.

According to the forecasting performance based on the RMSE from 2010 to 2022, the table in Figure 5 reveals that Drift_STL was able to slide up to rank 1. The main reason for this is that the Drift_STL method experienced a relatively small increase in the RMSE compared to the ARIMA and ETS models by including the Covid-19 period in the time series. In contrast, the RMSE of the SNaïve method increased significantly, dropping it to rank 9. The ranking of the other models remained similar.

Figure 18 in the appendix presents the RMSE per year, thus it shows the performance of the various models for each year separately. In particular, the performance in 2020 is interesting as it shows that, contrary to Madeira’s forecasting results, the Naïve method was not able to produce the best forecasts. The performance of Naïve is only slightly better than the ARIMA and ETS models and not able to beat the Drift_STL method, which produced the most accurate forecasts. Prophet_log, TSLM and SNaïve performed worst in 2020.

A comparison is now made with the results of the forecasting competition based on a rolling CV. Figure 19 in the appendix shows how the RMSE based on a rolling CV developed by including more years in the time series. Since the results are very similar, only two main differences will be described. TSLM was able to forecast more accurately until 2019 as the RMSE of approximately 212,000 was 82,000 lower than with a recursive CV. In addition, the HWA model performed better during the Covid-19 pandemic and was able to slide up to rank 1 by the end of 2022, meaning that it outperformed Drift_STL.

4.3. Forecasting Results: Alentejo

This chapter describes the forecasting results for Alentejo and follows the same structure as sections 4.1 and 4.2. Hence, it focuses first on the forecasting results based on a recursive CV and thereafter a comparison to the forecasting results based on a rolling CV is made.

Figure 20 in the appendix shows the point forecasts of all models for the period 2010-2022 as well as the actual number of Overnight Stays. Therefore, a first impression of the forecasting accuracy can be gained. At the beginning of this forecasting period, almost all models seemed to produce forecasts which were close to the actual number of Overnight Stays. But as the years went by, more models produced inaccurate forecasts, and the Covid-19 outbreak exacerbated this. Particularly in 2021 and 2022, many models produced forecasts that were very distant from the actual numbers. In 2022, however, several models were able to improve their forecasting performance.

More details are provided by Figure 6, which shows how the RMSE developed by including more years in the time series and how the corresponding ranking of the models changed. It follows the same structure as the previous two figures above. The line graph shows that even before the Covid-19 outbreak, the RMSE of all models was not at a stable level. The RMSE increased steadily for all models, whereby the increase was particularly strong for Naïve and TSLM_log. The ranking at the end of 2019 confirms that these two models were inferior to the other models, as their RMSE of over 47,000 is more than twice as high as the RMSE of the other methods. This time, the Prophet_log model is competitive with SNaïve and Drift_STL as their performances at the end of 2019 are close, with an RMSE between 22,000 and 24,000. The three ARIMA models and ETS_auto_log produced the most accurate forecasts, followed by the HWA model.

	2010-2019		2010-2022	
Model	Rank	RMSE	Rank	RMSE
ARIMA_d0D1_log	1	14,564.93	1	26,522.64
ARIMA_auto_log	2	14,847.27	2	26,808.36
Drift_STL	6	22,338.22	3	28,040.79
ARIMA_d1D1	3	15,731.55	4	28,349.73
ETS_auto_log	4	15,873.19	5	29,933.59
HWA	5	17,819.80	6	34,305.99
Prophet_log	7	22,594.65	7	46,907.16
SNaive	8	23,520.64	8	50,152.41
TSLM_log	9	47,252.29	9	52,720.38
Naive	10	56,208.84	10	61,485.90

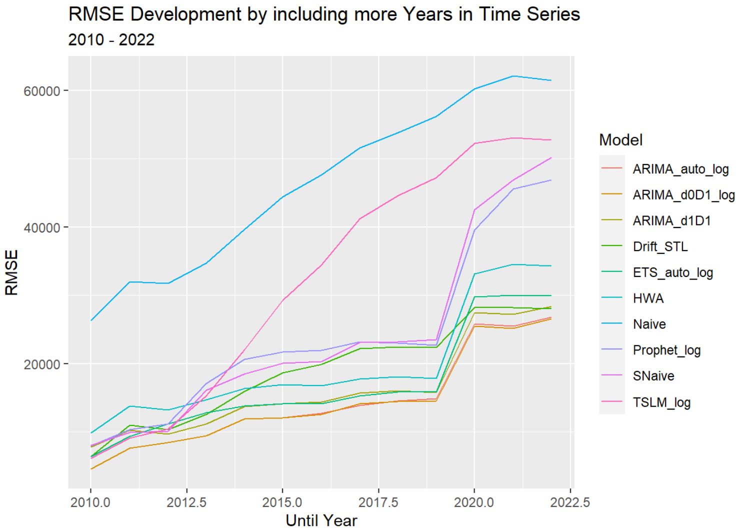

Figure 6 – Development of RMSE of Alentejo’s Forecasts based on Recursive CV.

However, the inclusion of 2020 in the time series led to a big increase in the RMSE, which is lowest for TSLM_log and Naïve. The absolute and relative differences between the RMSE of 2019 and 2020 for each model is provided in Table IV. Remarkably, the relative increase in the RMSE for Alentejo is on average lower than for Algarve and Madeira. The ETS and SNaïve models recorded the highest increase in RMSE, with values ranging from 80% to 88%. The increase was slightly lower for the ARIMA and Prophet_log models, with values between 74% and 75%. Naïve recorded the smallest increase at around 7%, followed by TSLM_log and Drift_STL at around 11% and 26%, respectively.

The line plot in Figure 6 reveals that thereafter, the RMSE stayed at a relatively stable level for all models. There are only two models whose RMSE increased further: SNaïve and Prophet_log. However, the ranking of the models according to their RMSE based on forecasts from 2010 to 2022 (see table in Figure 6) remained mostly unchanged. The only difference is that the forecasting performance of Drift_STL got closer to the ARIMA models and therefore, this methodology was able to slide up from rank 6 to rank 3. Consequently, ARIMA_d1D1, ETS_auto_log and HWA moved down a rank.

But does this change when looking at the performance of the models for each specific year? An answer to this question can be found by analysing the annual RMSE, which is presented in Figure 21 in the appendix. The line plot shows that Drift_STL was able to produce the most accurate forecasts for 2020, which explains the change in the ranking. Contrary to the results obtained for Madeira and Algarve, the Naïve model was not able to compete with the best models in 2020. With an RMSE of 91,140.01 in 2020, it is more than 30% higher than the RMSE of Drift_STL. In particular, SNaïve and Prophet_log were inferior to the other models in 2020 as their RMSE is with above 110,000 highest. Nevertheless, all methods produced more accurate forecasts in 2021, according to a much lower RMSE in 2021. For most models, the RMSE per year continued to decrease in 2022. Only the ARIMA and SNaïve models performed worse than in 2021.

Since the results of the forecasts based on a rolling CV are similar, only the main differences are presented. Figure 22 in the appendix shows how the RMSE based on a rolling CV developed by including more years in the time series. Throughout the entire forecast period, the performance of the ARIMA and ETS models was consistently very close to each other. Before the Covid-19 outbreak, Drift_STL was not able to compete with the ARIMA and ETS models. However, the increase in RMSE due to the Covid-19 outbreak was significantly lower, allowing it to move up to rank 1. Moreover, TSLM_log produced consistently more accurate forecasts than based on a recursive CV and was therefore able to outperform SNaïve by the end of 2022.

4.4. Robustness of Results to scale-independent Forecasting Measure

Since the number of Overnight Stays in the months after the Covid-19 outbreak is very low, it can be compared to a different scale. As the previous analysis focused only on the RMSE, which is a scale-dependent measure, a comparison to a scale-independent measure should be made. The scale-independent counterpart to the RMSE is the RMSSE.

However, considering the RMSSE, the ranking of the models is exactly the same as the ranking according to the RMSE at the end of 2019 and at the end of 2022. Therefore, the core statements remain unchanged, and a more detailed analysis will be omitted in this MFW. For interested readers, an analysis of the development of the RMSSE is available in the files on Github.

4.5. Model Performance after varying the Estimation Window

The previous analysis of the forecasting results focused mainly on two estimation windows: January 2010 to December 2019 and January 2010 to June 2022. In the second case, the proportion of observations during the Covid-19 pandemic was relatively small. This may lead to smaller differences between the rankings based on the two estimation windows. But how do the models perform when the estimation window is changed?

To answer this question, the ranking is analysed based on different estimation windows for both recursive and rolling CV for the three regions. The first two estimation windows begin in January 2015. The ranking of the models according to the RMSE based on the forecasts from January 2015 to December 2019 is compared with the ranking according to the RMSE based on the forecasts from January 2015 to June 2022. The next two estimation windows start in January 2017 and the last two estimation windows start in January 2019. In each case, the proportion of observations during the Covid-19 pandemic gradually increases. Therefore, the impact of these observations on the change in ranking also increases. This is confirmed by the results, as the later the estimation window starts, the more the ranking of the models changes by including the Covid-19 period in the time series.

The comparatively good performance of Drift_STL during the Covid-19 period is reinforced by this analysis. After including the Covid-19 period in the time series, it always moved up in the rankings. For example, for Alentejo with recursive CV, it ranks from 6th to 3rd when the estimation window starts in 2010, but from 7th to 1st when the estimation window starts in 2017. The same behaviour is observed for the Naïve model. In fact, the difference between the two rankings for Naïve increases the later the estimation window starts. However, despite the inclusion of the observations during the Covid-19 pandemic in the time series, it remains one of the inferior models. During this period, Naïve can compete with the best models only for Madeira.

When the estimation window starts in 2017, in most cases the three ARIMA models move down in the ranking after the inclusion of the Covid-19 period in the time series. In addition, the two ETS models do not show a consistent behaviour for the estimation window starting in 2017: In some cases the ranking improves, in others it deteriorates. An exception are the results of Algarve with rolling CV. In this case, both ETS models performed well during the Covid-19 period as their ranking improved.

The biggest change in the rankings shows SNaïve for Algarve with rolling CV when the estimation window starts in 2019. By including the Covid-19 period in the time series, SNaïve drops down from 1st to 10th place. Finally, the Prophet and TSLM models are still outperformed by the majority of the models, regardless of when the estimation window starts.

4.6. Improving Forecasting Accuracy following Data Anomalies

The previous analysis revealed that the data anomalies induced by the Covid-19 outbreak led to higher forecasting errors, and therefore a higher RMSE. In 2020, the annual RMSE was particularly high, and in 2021 and 2022, it was still higher than before the Covid-19 outbreak. But how can forecasting accuracy be improved after the data anomalies?

Many studies have been conducted in the field of data anomaly detection in time series, also known as outlier detection (Blázquez-García et al., 2021; Wu, 2016). However, these studies do not focus on how to deal with outliers once they have been identified. In the present work, the removal of outliers is not possible, as some modelling functions, e.g., STL decomposition or ETS, do not allow for missing values without any problems (Hyndman & Athanasopoulos, 2018). Therefore, the procedure chosen is to adjust the outliers.

Six different approaches are evaluated for the recursive CV datasets of the three regions: Adjustment of outliers by 100% to the mean per month of the previous 20, 10 and 5 years; and adjustment of the outliers by 50% to the mean per month of the previous 20, 10 and 5 years. Moreover, these six approaches are applied to two cases: In the first case, the observations from March 2020 to June 2021 are classified as outliers and the forecasts are performed for July 2021 to June 2022; and in the second case, the observations from March 2020 to December 2021 are classified as outliers and the forecasts are performed for January to June 2022. The model selection is made based on which model produced the best forecasts for each region in the period 2010 to 2022.

Table V in the appendix summarizes the results. In some cases the adjustments improved the RMSE only slightly or even led to an increase in the RMSE, whereas in other cases the RMSE decreased significantly. The highest reduction of the RMSE was achieved by classifying the observations from March 2020 to December 2021 as outliers. In this case, for example, the RMSE for Alentejo was 20,489.61, which is almost 50% lower than the RMSE based on the original outliers.

This chapter is intended to provide an insight into another area of research. However, as further research in this area does not contribute to the research question of this thesis, no further analysis will be conducted.

5. CONCLUSION

In this work, forecasting competitions were designed for three regions of Portugal using data from the tourism sector. The time series contain the monthly number of tourist Overnight Stays from 2000 to 2022 and therefore, a period with regular patterns in the time series and a period with data anomalies induced by the Covid-19 outbreak can be found. Madeira, Algarve and Alentejo were selected as the time series of interest because the tourism demand of these three regions showed different patterns, particularly after the Covid-19 outbreak. The forecasting methods considered in this MFW are pure time series approaches, ranging from simple methods, such as Naïve models, to more sophisticated models, such as ARIMA and ETS models.

It can be concluded that methods which perform well during a time period with regular patterns in the time series, will not necessarily show superior performance during a time period with data anomalies. This work showed that the ARIMA and ETS models have always been among the models with the best forecasting performance from 2010 to 2019, but their performance deteriorated sharply in 2020. However, these methods were able to achieve fast improvements in 2021 and 2022. A possible reason for this could be that the time series in 2021 and 2022 displayed some patterns again, and these more sophisticated methods were able to capture these new patterns quickly. For example, the ETS model weights more recent observations more heavily than older observations, which is a useful feature in this situation.

Moreover, the SNaïve method was able to compete with more sophisticated models before the Covid-19 outbreak. This was observed in particular for the Algarve, whose time series showed a strong seasonality. By contrast, the Naïve method showed poor performance during this period. However, the Covid-19 breakout reversed this situation: While the SNaïve model recorded a sharp deterioration in prediction accuracy, the performance of the Naïve model reduced only slightly. Particularly in 2020, the Naïve model was able to compete with more complex models. For Madeira the Naïve model recorded in 2020 even the most accurate forecasts. The main reason for this is that the SNaïve method relies only on the seasonal pattern in the data, and the seasonality behaved differently in 2020 than before. The Naïve model, on the other hand, is not based on any patterns in the time series and performs particularly well when the time series follows a random walk.

Another conclusion is that TSLM and Prophet have always been among the inferior forecasting models during the whole time period. Therefore, these methodologies do not seem to be suitable for predicting the number of Overnight Stays in the selected regions. In contrast, Drift_STL was consistently able to compete with the best models and, due to the smaller drop in performance following the Covid-19 outbreak, it was able to rank among the top models for predictions from 2010 to 2022 for all three time series. Changing the estimation window to increase the proportion of observations during the Covid-19 pandemic further enhanced this effect, and Drift_STL was ranked first in most cases.

These conclusions were mainly drawn from the development of the RMSE and are based on a recursive CV. However, a comparison was made with the results based on a rolling CV and the results obtained using the RMSSE, a scale-independent accuracy measure. Moreover, the model performance was evaluated after varying the estimation window. The robustness of the results was confirmed as the main conclusions remained unchanged. In addition to the results of the forecasting competitions, the exploratory data analysis revealed another finding: There is evidence that regions with a higher percentage of Portuguese tourists were less affected by strong declines in tourism demand during the Covid-19 period.

For further investigation, it would be of interest to enlarge the forecasting competition with models that incorporate explanatory variables. After performing a broad forecasting competition, Athanasopoulos et al. (2011) confirmed the conclusion already drawn from previous studies: Pure time series approaches forecast tourism demand data more accurately than methods that use explanatory variables. Nevertheless, these studies were based on time series with regular patterns. The performance of the models could change during time periods with data anomalies. Furthermore, in this MFW two accuracy measures are used that penalise large errors more than small errors. Since this is not favourable in all situations, further accuracy measures could be used to evaluate forecast results.

All in all, this work was motivated by the observation that there has been limited research on forecasting tourism demand in the presence of data anomalies. The recent global pandemic of Covid-19, with its significant impact on the tourism industry, provided an opportunity to further explore this research area and extend the existing literature. In a rapidly changing world, models cannot always be applied to time series with regular patterns without data anomalies. Hence, it is crucial to increase the focus on this field of study to ensure that research is prepared for the future.

REFERENCES

Abdullah, M., Dias, C., Muley, D. & Shahin, M. (2020). *Exploring the impacts of COVID-19 on travel behavior and mode preferences*. Transportation Research Interdisciplinary Perspectives 8, 100255.

Athanasopoulos, G., Ahmed, R.A. & Hyndman, R.J. (2009). *Hierarchical forecasts for Australian domestic tourism*. International Journal of Forecasting 25 (1), 146-166.

Athanasopoulos, G., Hyndman, R.J., Song, H. & Wu, D.C. (2011). *The tourism forecasting competition*. International Journal of Forecasting 27 (3), 822-844.

Bigović, M. (2012). *Demand forecasting within Montenegro tourism using Box-Jenkins methodology for seasonal ARIMA models*. Tourism and Hospitality Management 18 (1), 1-18.

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), 1-33.

Box, G.E.P. & Jenkins, G.M. (1970). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.

Brown, R.G. (1959). *Statistical forecasting for inventory control*. New York: McGraw-Hill Book Company.

Chen, J.L., Li, G., Wu, D.C. & Shen, S. (2019). *Forecasting Seasonal Tourism Demand Using a Multiseries Structural Time Series Method*. Journal of Travel Research 58 (1), 92-103.

Choe, Y., Wang, J. & Song, H. (2020). *The impact of the Middle East respiratory syndrome coronavirus on inbound tourism in South Korea toward sustainable tourism*. Journal of Sustainable Tourism 29 (7), 1117-1133.

Chu, F.-L. (2008a). *A fractionally integrated autoregressive moving average approach to forecasting tourism demand*. Tourism Management 29 (1), 79-88.

Chu, F.-L. (2008b). *Analyzing and forecasting tourism demand with ARAR algorithm*. Tourism Management 29 (6), 1185–1196.

Cleveland, R.B., Cleveland, W.S., McRae, J.E. & Terpenning, I. (1990). *STL: A seasonal-trend decomposition*. Journal of Official Statistics 6 (1), 3-73.

Goh, C. & Law, R. (2011). *The Methodological Progress of Tourism Demand Forecasting: A Review of Related Literature*. Journal of Travel & Tourism Marketing 28 (3), 296-317.

Holt, C.C. (1957). *Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages*. ONR Memorandum No. 52, Carnegie Institute of Technology, Pittsburgh.

Hyndman, R.J. & Athanasopoulos G. (2018) *Forecasting: Principles and practice.* 3rd Ed. Melbourne: OTexts.

Hyndman, R.J., & Khandakar, Y. (2008). *Automatic Time Series Forecasting: The forecast Package for R*. Journal of Statistical Software 27 (3), 1-22.

Hyndman, R.J. & Koehler, A.B. (2006). *Another look at measures of forecast accuracy.* International Journal of Forecasting 22 (4), 679-688.

Hyndman, R.J, Koehler, A.B., Ord, J.K. & Snyder, R.D. (2008). *Forecasting with exponential smoothing: the state space approach*. Heidelberg: Springer Science & Business Media.

Hyndman, R.J., Koehler, A.B., Snyder, R.D. & Grose, S. (2002). *A state space framework for automatic forecasting using exponential smoothing methods*. International Journal of forecasting 18 (3), 439-454.

Instituto Nacional de Estatística (2022). Turismo, Portal do INE [Online]. Available from: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_publicacoes&PUBLICACOESrevista=00&PUBLICACOEStema=55581 [Accessed: 21/09/2022].

Jiao, E.X. & Chen, J.L. (2019). *Tourism forecasting: A review of methodological developments over the last decade*. Tourism Economics 25 (3), 469-492.

Kourentzes, N., Saayman, A., Jean-Pierre, P., Provenzano, D., Sahli, M., Seetaram, N. & Volo, S. (2021). *Visitor arrivals forecasts amid COVID-19: A perspective from the Africa team*. Annals of Tourism Research 88, 103197.

Kuo, H.-I., Chen, C.-C., Tseng, W.-C., Ju, L.-F. & Huang, B.-W. (2008). *Assessing impacts of SARS and avian flu on international tourism demand to Asia*. Tourism Management 29 (5), 917-928.

Kwiatkowski, D., Phillips, P.C., Schmidt, P. & Shin, Y. (1992). *Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?*. Journal of econometrics 54 (1-3), 159-178.

Lütkepohl, H. & Xu, F. (2012). *The role of the log transformation in forecasting economic variables.* Empirical Economics 42, 619-638.

Mao, C.-K., Ding, C.G., Lee, H.-Y. (2010). *Post-SARS tourist arrival recovery patterns: An analysis based on a catastrophe theory*. Tourism Management 31 (6), 855-861.

Müller, U. (2005). *Size and power of tests for stationarity in highly autocorrelated time series.* Journal of Econometrics 128, 195-213.

Pan, B. & Yang, Y. (2017). *Forecasting Destination Weekly Hotel Occupancy with Big Data.* Journal of Travel Research. 56 (7), 957-970.

Peng, B., Song, H. & Crouch, G.I. (2014). *A meta-analysis of international tourism demand forecasting and implications for practice*. Tourism Management 45, 181-193.

Ponnam, L.T., Rao, V.S., Srinivas, K. & Raavi, V. (2016). A comparative study on techniques used for prediction of stock market. In: *2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT)*. Pune, India, 09-10 September 2016. IEEE, 1-6.

Saayman, A. & Botha, I. (2017). *Non-linear models for tourism demand forecasting*. Tourism Economics 23 (3), 594-613.

Satrio, C.B.A., Darmawan, W., Nadia, B.U. & Hanafiah, N. (2021). *Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET*. Procedia Computer Science 179, 524-532.

Shin, Y. & Schmidt, P. (1992). *The KPSS stationarity test as a unit root test*. Economics Letters 38 (4), 387-392.

Song, H., Gao, B.Z. & Lin V.S. (2013). *Combining statistical and judgmental forecasts via a web-based tourism demand forecasting system*. International Journal of Forecasting 29 (2), 295-310.

Song, H. & Li, G. (2008). *Tourism demand modelling and forecasting-A review of recent research*. Tourism Management 29 (2), 203-220.

Song, H., Qiu, R.T.R. & Park, J. (2019). *A review of research on tourism demand forecasting: Launching the Annals of Tourism Research Curated Collection on tourism demand forecasting*. Annals of Tourism Research 75, 338-362.

Taylor, S.J. & Letham B. (2018). *Forecasting at Scale*, The American Statistician 72 (1), 37-45.

Tiwari, S., Bharadwaj A. & Gupta, S. (2017). Stock price prediction using data analytics. In: *2017 International Conference on Advances in Computing, Communication and Control (ICAC3).* 01-02 December 2017, Mumbai, India. IEEE, 1-5.

Turismo de Portugal (2021). *Overview Tourism in Portugal* [Online]. Available from: http://www.turismodeportugal.pt/en/Turismo_Portugal/visao_geral/Pages/default.aspx [Accessed: 29/01/2023].

Turismo de Portugal (2022). Dormidas, TravelBI [Online]. Available from: https://travelbi.turismodeportugal.pt/alojamento/dormidas/ [Accessed: 13/09/2022].

Winters, P.R. (1960). *Forecasting sales by exponentially weighted moving averages*. Management science 6 (3), 324-342.

Witt, S.F. & Witt, C.A. (1995). *Forecasting tourism demand: A review of empirical research*. International Journal of Forecasting 11 (3), 447-475.

World Health Organization (2023). *Coronavirus disease (COVID-19) pandemic* [Online]. Available from: https://www.who.int/europe/emergencies/situations/covid-19 [Accessed: 21/01/2023].

Wu, H.-S. (2016). A survey of research on anomaly detection for time series. In: *2016 13th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP).* 16-18 December 2016, Chengdu, China. IEEE, 426-431.

APPENDICES

Appendix A – Data and Results

Table I – Different Measures for Ratio of Portuguese Tourists and Decline in Tourism due to Covid-19

	Ratio of Portuguese Tourists			Mean year-over-year rel. Change in Overnight Stays		
Region	August 19	Summer 18 & 19	Jan 18 - Dec 19	Apr 20 - Dec 20	Mar 20 - Feb 21	Summer 20 & 21
Alentejo	0.6886	0.6572	0.6493	-0.4696	-0.5160	-0.1820
Algarve	0.3679	0.2953	0.2366	-0.7197	-0.7279	-0.4778
Azores	0.3304	0.3424	0.4300	-0.7857	-0.7628	-0.5209
Central	0.5693	0.5330	0.5602	-0.6334	-0.6580	-0.3761
Lisbon M.A.	0.1931	0.1987	0.2116	-0.8344	-0.8228	-0.6699
Madeira	0.1743	0.1448	0.1189	-0.8069	-0.7921	-0.5137
North	0.4062	0.3875	0.4056	-0.7039	-0.7147	-0.4616

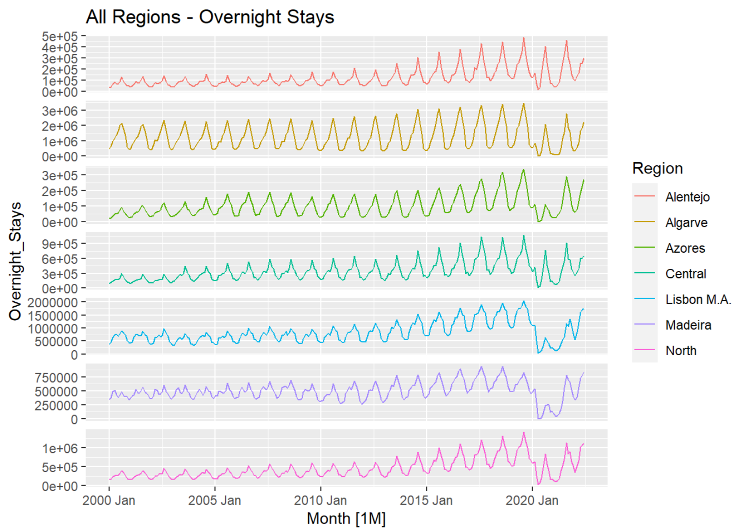

Figure 7 – Monthly Overnight Stays in all Regions.

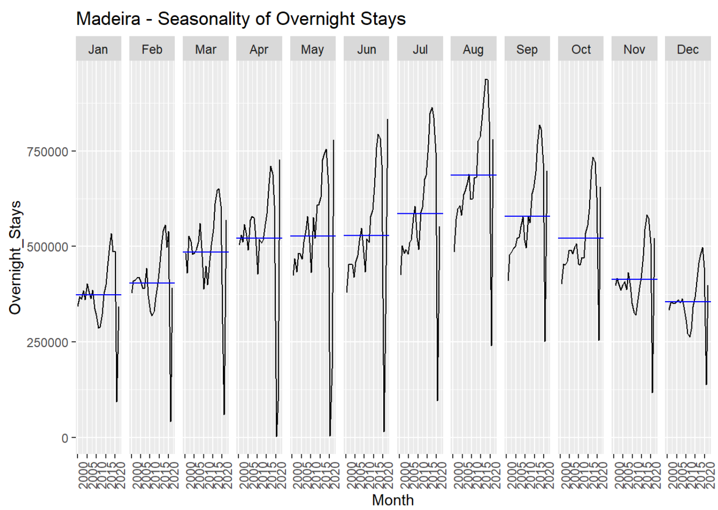

Figure 8 – Seasonality of monthly Overnight Stays in Madeira.

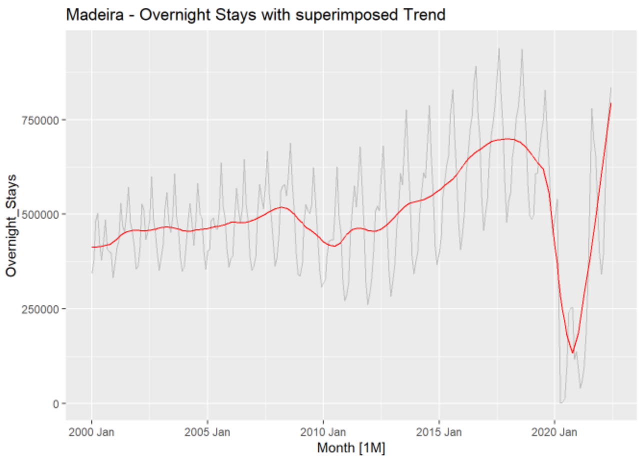

Figure 9 – Madeira’s Overnight Stays with Trend-Cycle Component (red).

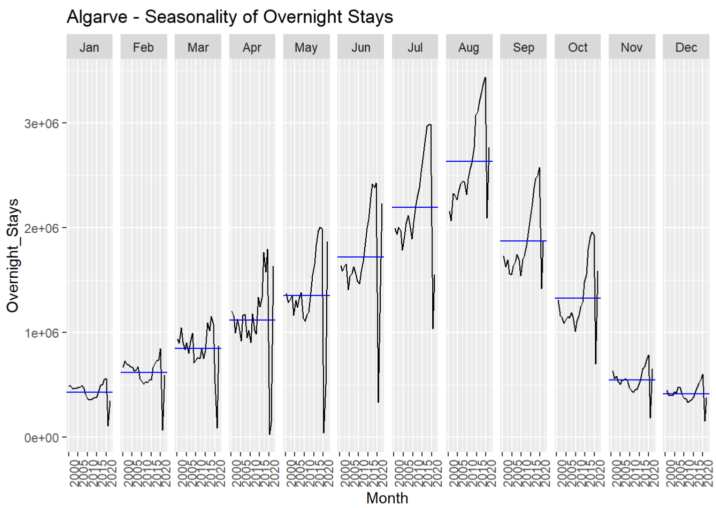

Figure 10 – Seasonality of monthly Overnight Stays in Algarve.

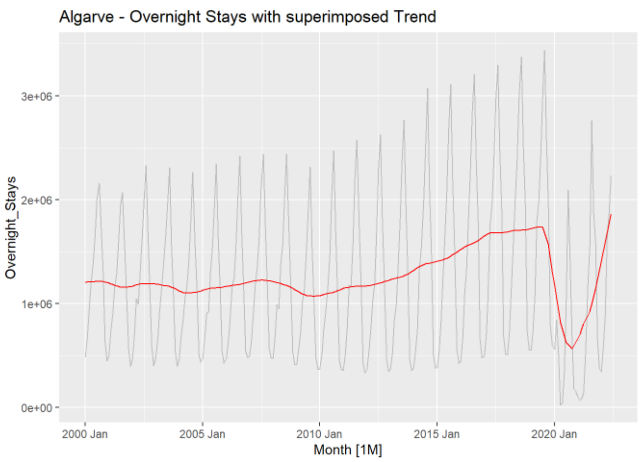

Figure 11 – Algarve’s Overnight Stays with Trend-Cycle Component (red).

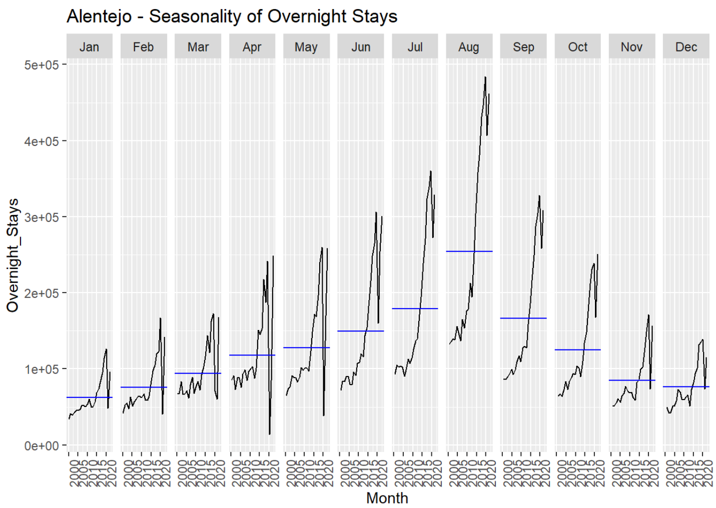

Figure 12 – Seasonality of monthly Overnight Stays in Alentejo.

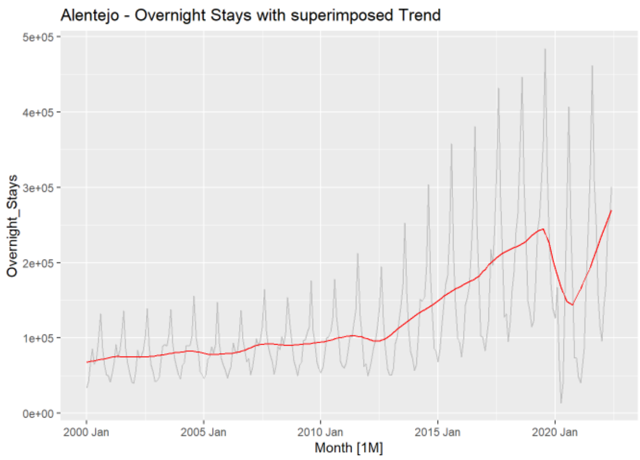

Figure 13 – Alentejo's Overnight Stays with Trend-Cycle Component (red).

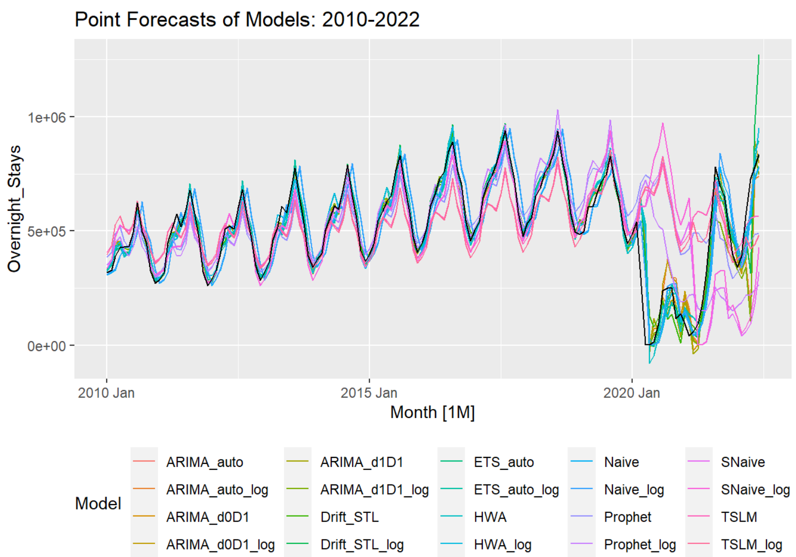

Figure 14 – Madeira's Point Forecasts based on Recursive CV.

Table II – Change of RMSE by including 2020 in Time Series: Madeira's Forecasts based on Recursive CV

Model	RMSE_2019	RMSE_2020	Diff_abs	Diff_rel
Naive	82,367.68	87,551.46	5,183.78	0.0629
TSLM	90,460.21	166,078.42	75,618.22	0.8359
ETS_auto_log	34,099.55	63,815.00	29,715.45	0.8714
Prophet_log	67,667.94	130,256.74	62,588.80	0.9249
HWA_log	30,719.70	62,139.90	31,420.20	1.0228
Drift_STL	26,899.35	56,113.79	29,214.44	1.0861
ARIMA_d1D1	24,724.45	55,087.27	30,362.82	1.2280
ARIMA_auto	24,913.02	56,055.39	31,142.36	1.2500
ARIMA_d0D1	24,386.04	55,580.87	31,194.83	1.2792
SNaive	53,807.75	153,714.16	99,906.41	1.8567

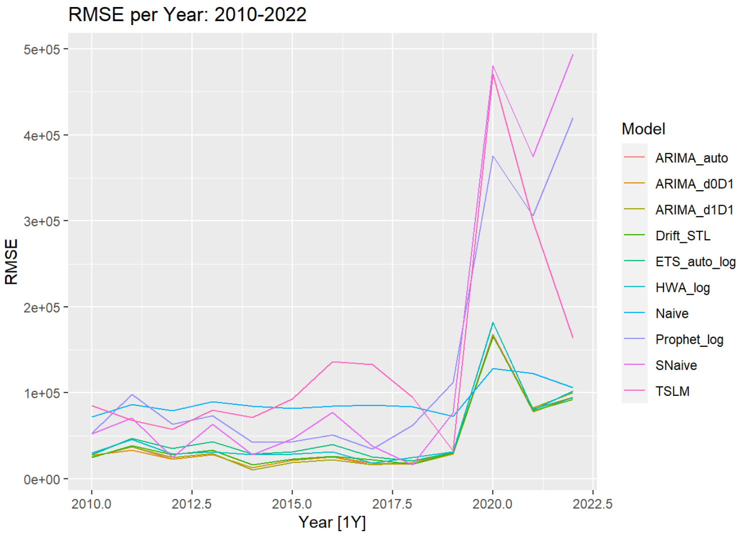

Figure 15 – RMSE per Year of Madeira's Forecasts based on Recursive CV.

	2010-2019		2010-2022	
Model	Rank	RMSE	Rank	RMSE
Drift_STL	4	26,363.61	1	60,452.48
ETS_auto	6	30,306.75	2	62,458.92
ARIMA_auto	2	24,567.87	3	62,917.47
ARIMA_d1D1	3	24,793.46	4	63,294.54
HWA	5	28,906.75	5	67,655.80
ARIMA_d0D1	1	24,247.36	6	79,891.52
Naive	10	82,367.68	7	91,659.07
Prophet	8	57,337.07	8	154,968.90
TSLM	9	79,276.66	9	188,328.39
SNaive	7	53,807.75	10	204,508.35

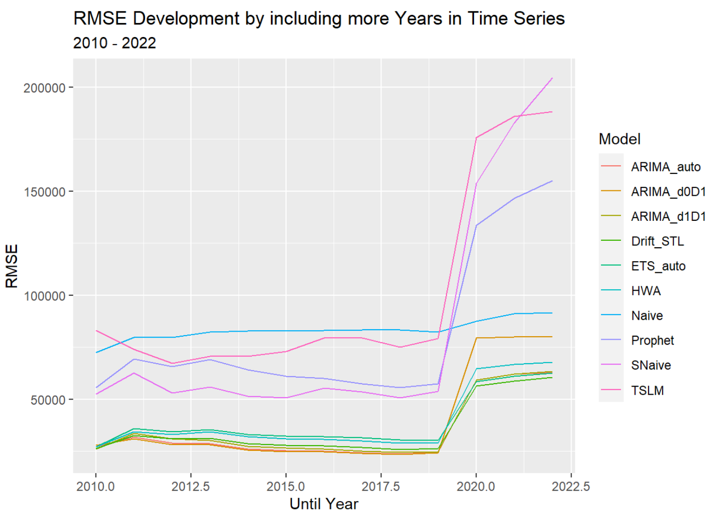

Figure 16 – Development of RMSE of Madeira's Forecasts based on Rolling CV.

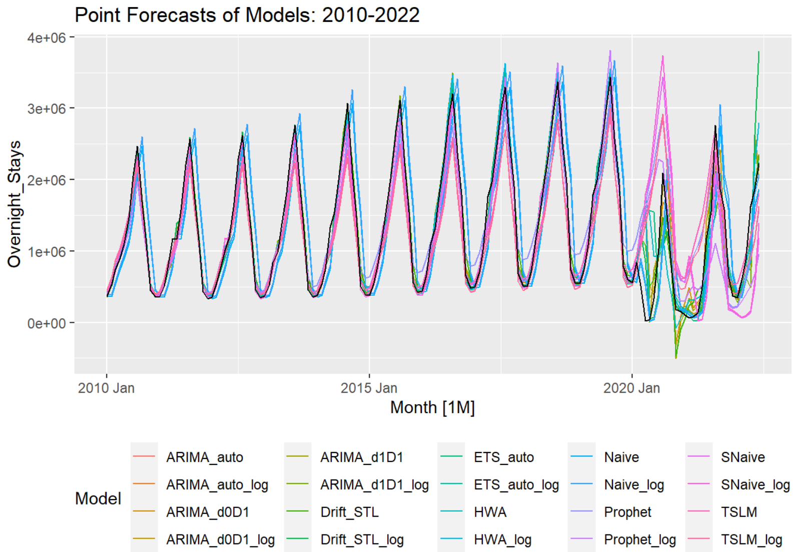

Figure 17 – Algarve's Point Forecasts based on Recursive CV.

Table III – Change of RMSE by including 2020 in Time Series: Algarve's Forecasts based on Recursive CV

Model	RMSE_2019	RMSE_2020	Diff_abs	Diff_rel
Naive	517,678.97	518,752.3	1,073.314	0.0021
TSLM	294,055.70	418,948.9	124,893.248	0.4247
Drift_STL	116,436.33	183,322.9	66,886.579	0.5744
Prophet_log	191,938.74	338,933.9	146,995.191	0.7658
HWA_log	113,161.96	212,664.3	99,502.352	0.8793
ETS_auto_log	110,633.52	212,743.6	102,110.094	0.9230
ARIMA_d1D1	96,707.66	198,239.1	101,531.481	1.0499
ARIMA_auto	91,874.23	190,802.6	98,928.405	1.0768
ARIMA_d0D1	92,494.37	192,160.9	99,666.511	1.0775
SNaive	114,889.37	411,007.5	296,118.087	2.5774

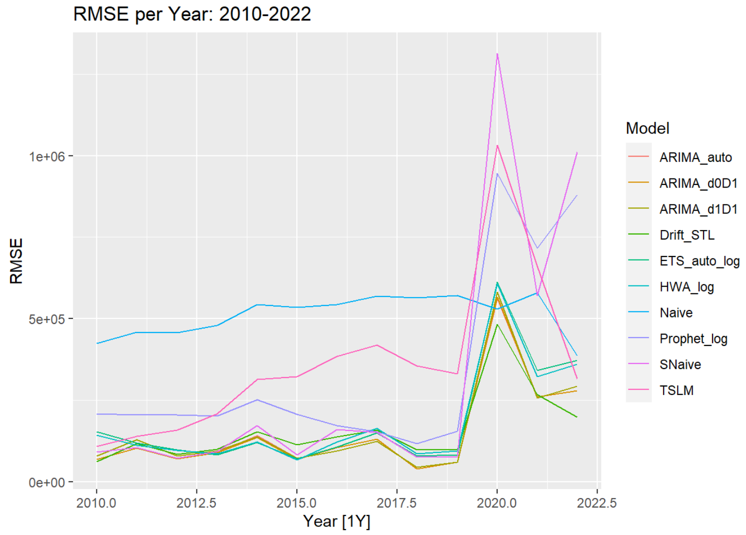

Figure 18 – RMSE per Year of Algarve's Forecasts based on Recursive CV.

	2010-2019		2010-2022	
Model	Rank	RMSE	Rank	RMSE
HWA	7	122,389.91	1	191,442.2
Drift_STL	4	113,320.35	2	193,122.8
ARIMA_d1D1	3	93,811.73	3	215,791.6
ARIMA_auto	1	91,912.66	4	217,384.8
ARIMA_d0D1	2	93,792.07	5	219,343.2
ETS_auto	6	116,099.46	6	220,659.7
Prophet	8	198,633.13	7	406,842.2
TSLM	9	212,112.29	8	444,227.3
SNaive	5	114,889.37	9	464,522.7
Naive	10	517,678.97	10	519,314.1

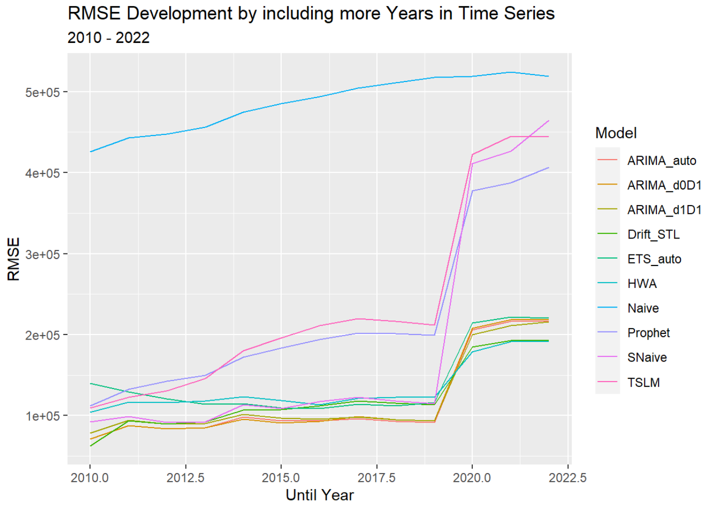

Figure 19 – Development of RMSE of Algarve's Forecasts based on Rolling CV.

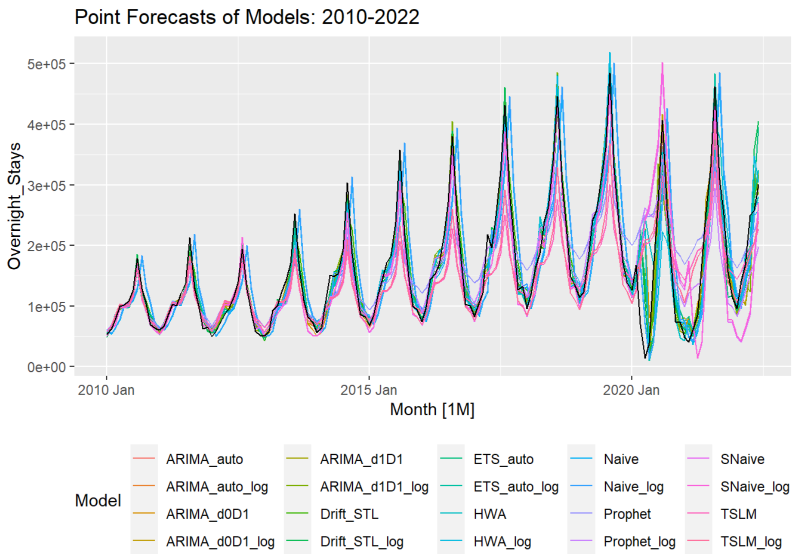

Figure 20 – Alentejo's Point Forecasts based on Recursive CV.

Table IV – Change of RMSE by including 2020 in Time Series: Alentejo's Forecasts based on Recursive CV

Model	RMSE_2019	RMSE_2020	Diff_abs	Diff_rel
Naive	56,208.84	60,227.48	4,018.637	0.0715
TSLM_log	47,252.29	52,246.69	4,994.395	0.1057
Drift_STL	22,338.22	28,160.85	5,822.627	0.2607
ARIMA_auto_log	14,847.27	25,834.83	10,987.558	0.7400
Prophet_log	22,690.97	39,602.17	16,911.203	0.7453
ARIMA_d1D1	15,731.55	27,486.41	11,754.859	0.7472
ARIMA_d0D1_log	14,564.93	25,497.51	10,932.578	0.7506
SNaive	23,520.64	42,516.00	18,995.359	0.8076
HWA	17,819.80	33,174.71	15,354.915	0.8617
ETS_auto_log	15,873.19	29,804.38	13,931.189	0.8777

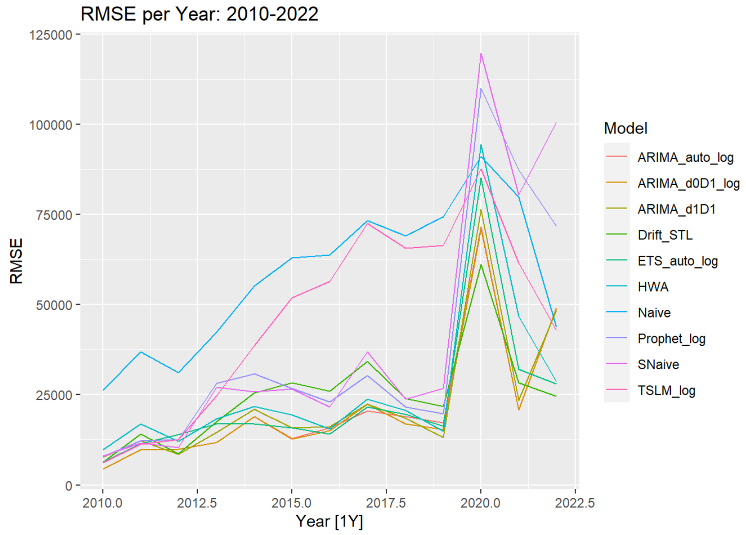

Figure 21 – RMSE per Year of Alentejo's Forecasts based on Recursive CV.

	2010-2019		2010-2022	
Model	Rank	RMSE	Rank	RMSE
Drift_STL	7	21,577.49	1	27,520.32
ARIMA_d1D1	4	17,058.72	2	29,645.78
ARIMA_d0D1	3	16,883.04	3	30,580.20
ETS_auto_log	2	15,918.56	4	30,808.59
ARIMA_auto	5	17,283.48	5	30,821.72
HWA_log	1	15,354.53	6	31,631.44
Prophet_log	6	20,213.57	7	47,514.68
TSLM_log	9	30,199.82	8	48,999.68
SNaive	8	23,520.64	9	50,152.41
Naive	10	56,208.84	10	61,485.90

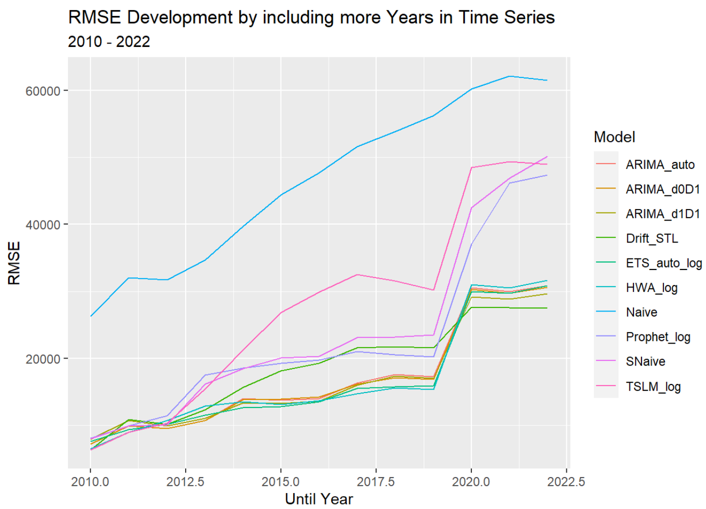

Figure 22 – Development of RMSE of Algarve's Forecasts based on Rolling CV.

Table V – Impact of Outlier Adjustments on RMSE

	RMSE Madeira		RMSE Algarve		RMSE Alentejo	
	Outliers Mar 20 - Jun 21	Outliers Mar 20 - Dec 21	Outliers Mar 20 - Jun 21	Outliers Mar 20 - Dec 21	Outliers Mar 20 - Jun 21	Outliers Mar 20 - Dec 21
Original_Outliers	81,986.40	94,935.39	287,110.3	280,124.0	36,532.77	49,126.65
Adj._Outliers_Mean_20yrs	73,981.28	76,633.71	331,018.1	183,863.7	55,061.13	32,692.21
Adj._Outliers_Mean_10_yrs	83,522.77	78,499.72	360,154.7	156,644.1	38,327.68	28,368.60
Adj._Outliers_Mean_5yrs	100,134.13	86,592.24	407,568.8	118,427.7	19,643.98	20,489.61
Adj._Outliers_50%_Mean_20yrs	74,280.56	87,414.02	295,402.5	238,534.1	30,140.17	29,407.60
Adj._Outliers_50%_Mean_10yrs	75,096.66	87,919.24	306,838.6	226,337.9	28,091.84	29,168.61
Adj_Outliers_50%_Mean_5yrs	79,291.27	92,754.84	322,928.8	238,534.1	29,268.17	31,204.13

Appendix B – Programming Walkthrough

This appendix provides a programming walkthrough that describes the organisation of the R code. The same code is available in two different files on Github. The RMD file is more technical and contains the source code. All chunks must be run first to see the results, e.g. figures and tables. In addition, there is an HTML file containing the figures and tables, and by clicking on the "Code" button the source code will appear. The work is divided into three main sections: "Exploratory Data Analysis", "Fitting and Evaluation Models" and "Improving Forecasting Accuracy following Data Anomalies".

The exploratory data analysis starts with the creation of the tsibble with the Overnight Stays per region from 2000 to 2022, which forms the database of this work. Thereafter, time series plots for all seven regions are shown and seasonal differencing is performed for several months in 2020 to determine the impact of Covid-19 on the number of Overnight Stays per region. Subsequently, a correlation analysis is carried out to determine if there is a relationship between the share of Portuguese tourists in a given region and the decline in tourism in that region due to the Covid-19 pandemic. The last subchapter provides an in-depth analysis of seasonality and trend for the three selected time series, namely Madeira, Algarve and Alentejo.

The second main section, "Fitting and Evaluation Models", is separated into Madeira, Algarve and Alentejo. Each of these three chapters is further divided into a recursive CV and a rolling CV, which always follow the same structure. First, the respective CV is conducted, and a seasonal strength test is performed on the training sets to define the parameter D of the ARIMA models. Thereafter, the models are fitted and one-step ahead forecasts are produced. In the subchapter "Forecast Errors" the point forecasts of the various models are shown in plots for several time periods. The next four chapters focus on the RMSE. First, the ranking of the models based on the RMSE from 2010 to 2019 and from 2010 to 2022 is presented in a table. Second, the development of the RMSE by including more years in the time series and third, the RMSE per year is displayed in several line plots. There are also tables showing the difference between the RMSE of certain years of interest. The fourth chapter explores how the ranking of the models changes when the estimation window is varied. Moreover, the first and second analyses mentioned above are also performed for the RMSSE. In all these subchapters focusing on the RMSE and RMSSE, each analysis is performed both for all models (with original and log-transformed variables) and for ten selected models. For the ten models, each model type appears only once in the analysis with either the original or the log-transformed variable, depending on the lower RMSE for the forecasts from 2010 to 2022.

In the final section, "Improving Forecasting Accuracy following Data Anomalies", the outliers are adjusted in various ways and, using only one forecasting model, the RMSE of the forecasts based on the adjusted time series is compared with that based on the original time series. For each region there are two cases: In the first case, the observations from March 2020 to June 2021 are classified as outliers and the forecasts are performed for July 2021 to June 2022; and in the second case, the observations from March 2020 to December 2021 are classified as outliers and the forecasts are performed for January 2022 to June 2022. Within these two cases, six different types of outlier adjustment were applied to the recursive CV datasets of the three regions. The last subchapter "Overall Summary" summarises the results of this section.