Time Series Analysis of Tourism in the Netherlands 1990-2022

National and Foreign Tourism in Campsites, Caravans, and similar facilities

Ágatha del Olmo Tirado

20/02/2024

Introduction

This study focuses on analyzing tourist accommodation data in the Netherlands from 1990 to 2022 obtained from Eurostat, specifically focusing on the total number of tourists, both domestic and foreign, who have stayed in caravan parks, campsites, and similar facilities. This study is particularly interesting considering that the Netherlands is characterized by its famous canals and picturesque landscapes, making it an ideal destination for exploration by caravan.

To study the behavior of the series, we will analyze each of its components separately: trend, seasonality, cycle, intervention, and remainder. After loading the necessary libraries and dating the database, we have truncated the series until December 2019 to avoid the effect of Covid on the overall study.

```
completeData <- read.csv2("./Netherlands_TO_OT.csv",
                          header = TRUE)

completeData <- ts(completeData[, 1],
             start = 1990,
             frequency  = 12)

data <- window(completeData, end = c(2019, 12))
```

After preparing the data, we can observe the pattern of the time series, which can be either additive or multiplicative. Additive series imply that each component adds its effect to the others, $y_t = T_t + S_t + C_t + I_t + R_t$, whereas multiplicative series assume a percentage increase over the other components, $y_t = T_t \cdot S_t \cdot C_t \cdot I_t \cdot R_t$.

Figure 1. Tourism in the Netherlands 1990-2019

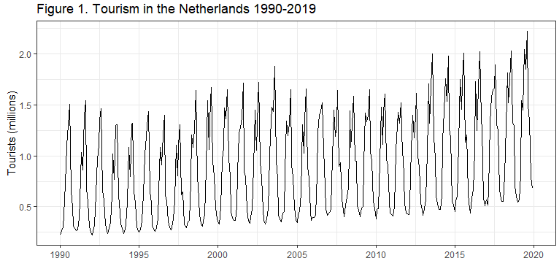
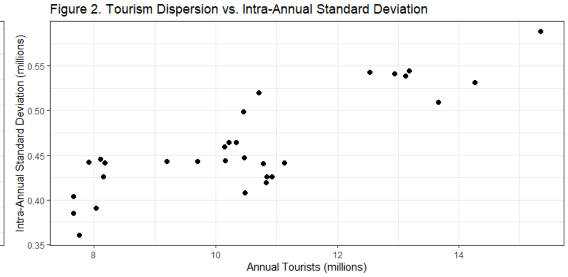


Figure 2. Tourism Dispersion vs. Intra-Annual Standard Deviation



As the series scheme is not entirely clear in the first figure, we analyze a second one, where it is evident that the series has a multiplicative scheme, meaning that with more tourists, there is a higher standard deviation or dispersion. Understanding the scheme is essential to know which methods we can use and how to interpret them.

Trend

The trend is the series' behavior over time (always longer than a year). We aggregate the monthly series to annual, so the graph has less noise, and its trend is clearer.

Figure 3. Trend of the Tourism Series in the Netherlands 1990-2019

We observe a general increasing trend. Specifically, the first decade remained stagnant around 8 million tourists, but in 1997, the numbers surged and remained at 11 million during the following decade until 2012, when they increased significantly until 2019.

To explain this trend, there are two possible simple causes regarding domestic tourism. Firstly, the country's population has increased by approximately 2.5 million people from 1990 to 2019, making it easy for domestic tourism to grow in absolute numbers. Additionally, per capita GDP has increased by about €30,000 during the same period, allowing domestic tourism to grow due to the economic capacity and leisure time of the modern Dutch. However, more information would be needed to determine if domestic tourism represents a significant enough proportion of the total for these events to have an effect.

Seasonality

Seasonality is the repetitive or cyclical change in the series' behavior (less than a year). To study seasonality rigorously, we must decompose the series into its different components, so we can extract it.

There are several methods for obtaining decomposition in RStudio: tapply, decompose (moving averages), and stl (weighted local regressions). The first option will always remain in the shadow compared to the others as it is the poorest or least cared for, the second is used with series without intervention or outliers and can be used with both additive and multiplicative schemes, and the last one can take into account intervention but is only valid for additive schemes (without performing "tricks").

Previously, we had anticipated the problem of intervention by eliminating the last years that contained the effect of Covid, so we can use the decompose option without any problems, specifying that it is of multiplicative type. We check that we have done it correctly by adding the seasonality of each month, which should be equal to the order, in this case, 12 (months of the year).

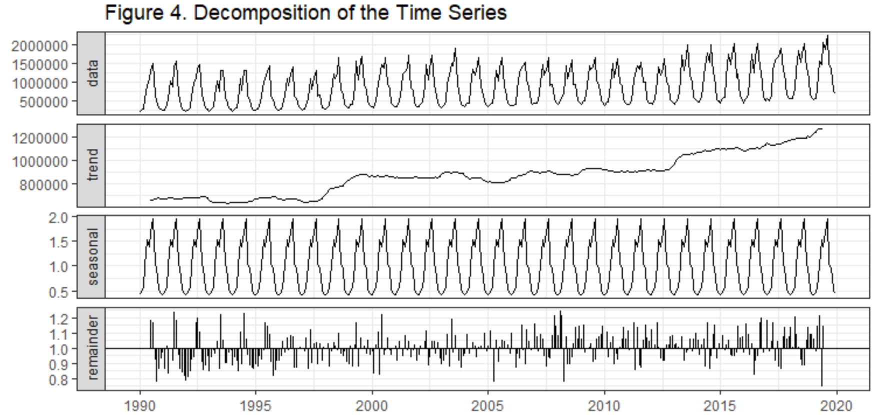
```
dataDecomp <- decompose(data,
                        type = "mult")
round(dataDecomp$figure,2)
```

```
 [1] 0.42 0.47 0.56 1.00 1.53 1.39 1.66 1.94 1.06 0.95 0.54 0.46
```

```
sum(dataDecomp$figure)
```

```
[1] 12
```

Figure 4. Decomposition of the Time Series



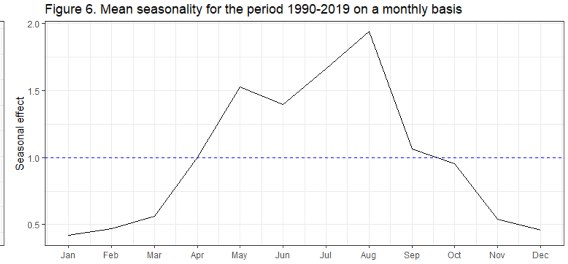
This plot shows each of the decomposed components of the series graphically. Right now, we are focusing on the third one, seasonality.

The seasonal components represent how much each month deviates from the annual average and are expressed around 1. Winter months stand out - especially January, where it exceeds half of the usual - with the lowest number of tourists, and summer -

particularly August, where it doubles - with the highest tourism. We can see it more clearly in the following graphs:

Figure 5. Seasonality in the Tourism Series in the Netherlands 1990-2019

Figure 6. Mean seasonality for the period 1990-2019 on a monthly basis

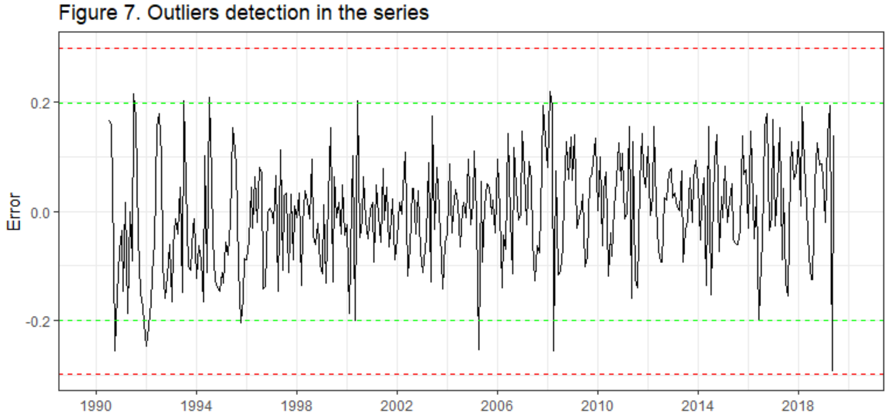


Intervention and Residual

Intervention and residual (the stochastic component of the series) are studied together, and we can measure if the cause of a seemingly noticeable rise or fall in the series is due to the randomness of the error or an intervention. The method consists of considering a value as an outlier when an error exceeds at least 3 times its standard deviation.

```
error <- log(remainder(dataDecomp))
sderror <- sd(error, na.rm = TRUE)
```

Figure 7. Outliers detection in the series




```
time(error)[!is.na(error) & abs(error) > 2.9 * sderror]

[1] 2019.333
```

As we can see, there are no errors that exceed 3 times their standard deviation (it should be noted that if we had used the complete series, there would have been a clear intervention caused by Covid), but in the graph, we see that at the end of the series, there is a month where visits to the country decrease notably despite not reaching the red line. To know exactly when this happens, instead of surpassing 3 times the standard deviation, we ask that it surpasses 2.9 times, and R detects April 2019.

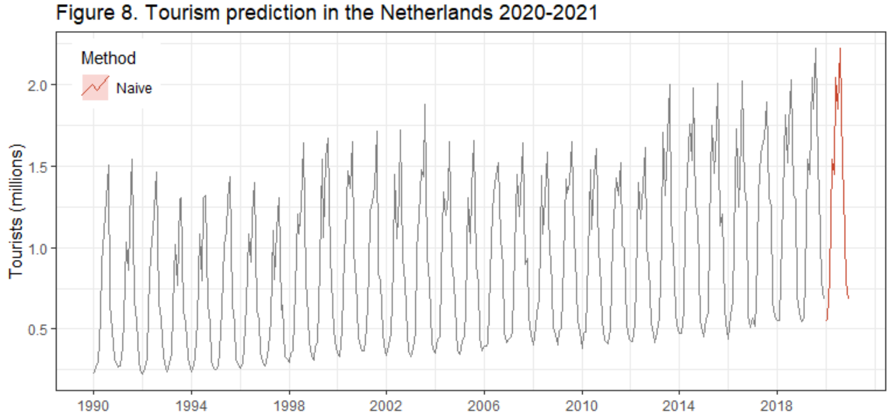
This decrease in tourism in the Netherlands can be attributed to a tragic event that occurred on the morning of March 18, 2019: the Utrecht attack, where three people died and seven were seriously injured. This attack was considered the worst Islamic terrorist attack the country had experienced up to that point, and the threat level of the city rose to level 5. Considering Utrecht is the fourth most important city in the country, it is understandable that the effect the attack had on visits to the country in the following weeks was noticeable.

Simple Prediction Methods After studying the characteristics of the series, we can predict with the most appropriate simple method, and if something is very clear about tourism in the Netherlands, it is that there is a strong seasonal component. This implies that we should use the naive method with seasonality.

This method consists of using the last observation of the same month available to predict the next one, $\hat{y}T + h = yT - m(k + 1)$. For example, to predict June 2022, we would use June 2019 since it is the last available data point in our series.

```
snaivedata <- snaive(data, h = 12)
```

Figure 8. Tourism prediction in the Netherlands 2020-2021



After prediction, it is interesting to measure if the model we have used provides us with good predictions or not.

Prediction Evaluation A good model should not have bias, should have good fit quality, trustworthy confidence intervals, and should improve over the simplest method.

Bias: ME and MPE -> less than 1% has no bias Fit Quality: RMSE, MAPE, and MAE -> should approach 0 *Confidence Intervals: ACF1 (correlation between* $\hat{e}t$ *and* $\hat{e}t - 1$*) -> less than 0.1* Improvement: MASE -> if greater than 1, a simpler method would be preferable

```
accuracy(snaivedata)

                    ME     RMSE      MAE      MPE     MAPE MASE       ACF1
Training set 21377.64 127584.6 84734.29 1.897998 9.120059    1 -0.1737397
```

Regarding bias, we see that there is a 1.9%, so there is some bias, and being positive, we know it is bias from below, meaning that the predictions made tend to be systematically lower than the actual values.

Regarding the fit quality, it is wrong by 127,584 visitors according to RMSE, and according to MAE, 84,734 (MAE is always smaller and not comparable to RMSE), that is, 9.12%, which represents a considerable part of the total.

About the confidence intervals, we cannot rely on them, as ACF1 is above 0.1 in absolute value, so we should use bootstrapping.

Finally, MASE is exactly 1 because it is compared to itself, since the simplest method in this case is precisely the naive one with seasonality.

Moving Averages Prediction

To improve the estimation of quality, we can obtain the MAPE (Mean Absolute Percentage Error) using the moving prediction origin method, also known as cross-validation. This method consists of fitting and predicting for a part of the sample (fitting with the training set and evaluating with the test set) that moves forward and repeats the process k times.

```
k <- 144
h <- 12
TT <- length(data)
s <- TT - k - h

mapeSnaive <- matrix(NA, s + 1, h)

for (i in 0:s) {
 train.set <- subset(data, start = i + 1, end = i + k)
 test.set <- subset(data, start = i + k + 1, end = i + k +h)
 fit <- snaive(train.set, h = h)
 mapeSnaive[i + 1,] <- 100*abs(test.set - fit$mean)/test.set
}

mapeSnaive <- colMeans(mapeSnaive)
round(mapeSnaive,2)

 [1] 8.86 8.84 8.80 8.85 8.94 8.99 8.92 8.94 8.99 8.96 9.00 9.05
```

Figura 9. Evolución del error de predicción

Despite later calculating the actual error committed, this way we have an idea of what our error would be in out-of-sample predictions unaffected by Covid, which is very useful since we do not expect a similar event to occur soon.

We see that the errors in the short term are very similar to those in the long term. At a 1-month horizon, the error is 8.86%, while at a 12-month horizon, the error is 9.05%. But it seems to trend upwards, so the further the time horizon, the worse the prediction seems to be.

Prediction vs. Covid

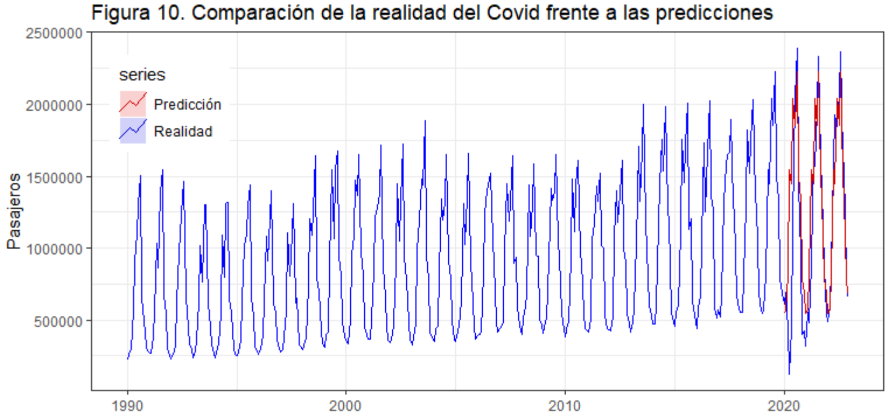
To obtain the effect of Covid, we subtract the prediction we made ignoring Covid from what should have happened (the actual series).

```
snaivedata_nocovid <- snaive(data, h = 36)

aggregate(completeData - snaivedata_nocovid$mean, FUN = sum)/1000000

Time Series:
Start = 2020
End = 2022
Frequency = 1
[1] -4.209511 -1.086880  1.102863
```

The results are expressed in millions: in 2020, there were 4 million fewer tourists than expected, in 2021, one million fewer, and in 2022, there is a recovery with one million more tourists than expected. This improvement in 2022 is probably due to the fact that the naive method with seasonality does not take into account that the series had a growing trend, so it was expected that after recovering from Covid, the tourism figures would be higher than forecasted.

To see if the effect of Covid is significant compared to the rest of the years, we can visualize it more clearly:

Figura 10. Comparación de la realidad del Covid frente a las predicciones



As we can see, the decline due to Covid is not very pronounced; it doesn't seem that its effect was very noticeable in the Netherlands. In fact, the year 2020 closely resembles the figures from the 1990s, while in some European countries, the pandemic caused record-breaking havoc well below historical levels. It should be noted that if we only considered foreign tourists, the difference would probably be even greater.

Conclusion

A steady growth in tourism in the Netherlands has been highlighted from 2019 to the present, with a marked seasonality showing more tourism in the summer, and a notable intervention in April 2019 due to the Utrecht attack. This work emphasizes the need to understand and address tourism fluctuations for making informed decisions, such as improving infrastructure for tourism in caravans and campsites, and implementing marketing policies to attract a greater number of visitors to the country.

Webgraphy

Wikipedia contributors. (2023, September 27). 2019 in the Netherlands. Wikipedia.https://en.wikipedia.org/wiki/2019_in_the_Netherlands Simcox, R. (2019, March 25). The Netherlands' luck is running out. Foreign Policy. https://foreignpolicy.com/2019/03/25/the-netherlands-luck-is-running-out/ Guía completa markdown y su integración con R. (n.d.). https://rstudio-pubs-static.s3.amazonaws.com/330387_5a40ca72c3b14824acedceb7d34618d1.html ¿Qué tan rápido se recupera el turismo del COVID-19? (2022, April 6). Foro Económico Mundial. https://es.weforum.org/agenda/2022/04/que-tan-rapido-se-recupera-el-turismo-de-covid-19/