ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis Data Science & Marketing Analytics

Tourisms reaction to shock: a study of the effects of COVID-19 measures on tourism in St. Anton am Arlberg

Name student: Esther van der Ent

Student ID number: 424514

Supervisor: A.J. Koning

Second assessor: B. Donkers

Date final version: 11-07-2021

Abstract

The global COVID-19 pandemic has had a big impact on the tourist industries with arrivals dropping 70% in 2020. This has been devastating for destinations that rely on tourism for their income, such as ski towns in Austria. There has been some research on the effect of epidemics and especially COVID-19 on tourism, but the actual effects of the pandemic are yet to be determined as well as a clear strategy on recovery. This thesis aims to determine how the COVID-19 pandemic has affected tourism and how tourism will change after the pandemic. This study used data from 2010 to 2020 on tourism during the winter months in Sankt Anton am Arlberg in Austria. With a Generalized Additive Model and a Vector Autoregression model the data is analyzed and forecasts were made to determine structural changes due to the COVID-19 pandemic. The results showed that both models produced similar forecasts, but the VAR model was best suited for this dataset. The forecasts looked very promising for St. Anton. However, the observations of the 2020 season were generally well below the forecasts for 2020. This showed the big impact of the COVID-19 pandemic. Moreover, the pandemic has changed views on travel and thus the preferences of tourists. In order to recover from this crisis, St. Anton should play into the new preferences of tourist. Moreover, St. Anton should take the lead and use the pandemic as a springboard for structural changes to its tourism industry to create an industry that is sustainable for the future.

Table of Contents

Introduction

At the beginning of 2020, the COVID-19 virus took over the world and governments were quick to respond. One of the biggest impacts of the response to the COVID-19 crisis is the closing of hospitality and leisure venues, which means a halt to travel and tourism. Tourism rates have dropped by 70% due to COVID-19 restrictions and the tourism industry is one of the hardest-hit industries (UNWTO, 2020). For places that rely on tourists and travelers for most of their income, this can be detrimental. As many people living in these areas are dependent on tourism for their livelihood and income. For instance Austrian ski towns, these towns are built around the mountains and the tourists these mountains attract. Especially winter tourists who come to ski and snowboard are important for these towns, but most measures shut out tourism especially in winter. Without tourism, the people living in these towns have no income.

When the virus first broke out in March, the Austrian government was quick to send all tourists home and close all venues cutting the winter season short. These measures were held in place in 2021 as well, leading to basically no tourism for the 2021 winter season. Now that it seems that the government has the virus under control and more and more people are getting vaccinated, the ski towns can think about opening up again for tourists. But it is still unclear as to what they can expect as, despite virus numbers going down and people getting vaccinated, the virus can break out again at any time. Moreover, this crisis led to an opportunity for structural change in the tourism industry, meaning that going back to the situation before COVID is most likely not going to happen. This means that the tourism industry is not only recovering from this crisis but also creating a "new normal" that will sustain well into the future. This new normal should take into account the COVID-19 virus and the measures are taken to stop it, but could also take into account other factors such as the environment and the structural changes due to the COVID-19 crisis. This could be the kick-start needed to rethink tourism and to create a more sustainable industry (UN, 2020).

The COVID-19 crisis has led many people to rethink a lot of aspects of their lives, one of them being travel. Even though young people have a reportedly higher desire to travel (Euronews, 2020), many of the travels in 2020, and probably in the future as well, were close to home sometimes not even leaving the country. Moreover, not only the distance has changed, more people choose private transportation over public transportation such as airplanes, trains and busses (Abdullah et al., 2020). So to create a good game plan to recover from this crisis, travel destinations should not only try to regain capacity but could use this pandemic as a kickstart to re-invent themselves to create a flow of tourism that can be sustained in the long run. A big factor in this will be the way these destinations market themselves toward their target audience. Forecasts will help here as they will give an idea as to what could be expected. Based on these forecasts, St. Anton can identify their target audience and create marketing strategies.

In the years leading up to 2020, a strong positive trend was present in the number of tourists coming to St. Anton. But due to the pandemic, this number fell back in 2020. Not only did the crisis stop the flow of tourists in 2020, but it also set in motion structural changes to the tourist industries.

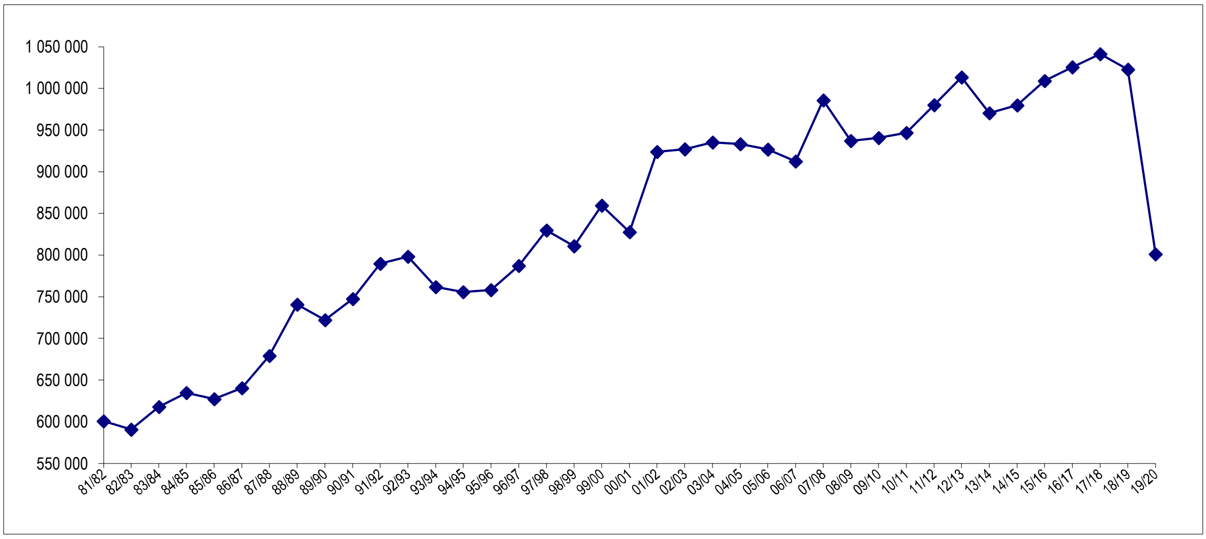

Figure 1: Changes in overnight stays per year in St. Anton am Arlberg 1981 - 2020

As discussed above, the preferred mode of transport has changed, people stay closer to home but young people have more desire to travel. This poses several questions as to what will happen in the future. How will these structural changes play out? Will the number of tourists get back to its pre-COVID level? Will more people stay in hotels or apartments?

As more research is conducted and more becomes clear about the spread of the COVID-19 virus. It is increasingly unlikely that travel will get back to its pre-COVID levels. It is likely that the number of tourists and the distance travelled will not be as high as before the crisis, as international travel is found to be pivotal in the spread of COVID-19 (Pana et al., 2021). This could spur tourist to pick destinations closer to home, avoid public transportation and book less far ahead (Brock, 2020). With travel playing such an important role in the spread of the virus, it is obvious that this crisis will have a lasting effect on the tourism industry. The changes happening now will be structural and will shape the future of the tourism industry. These structural changes might also play out in the type of accommodation chosen and the length of a stay. For instance, people might prefer an apartment over a hotel or the other way around and they might want to take multiple short trips or one longer trip.

As it is still unsure as to how the structural changes will play out, it is hard for businesses to determine how they should market themselves. Knowing this will help businesses develop a clear marketing strategy to recover from the COVID-19 crisis. Therefore, the proposed research question is:

What are the structural changes to tourism due to COVID-19 and how will they affect tourism?

This thesis will attempt to forecast tourism for Sankt Anton am Arlberg in Austria based on a time series analysis of the past 10 years and the structural changes set in motion by the COVID-19 crisis. This research will try to identify and model the structural changes and how they will affect tourism in St. Anton. Moreover, this research will try to generalize the findings to determine how tourism in Sankt Anton am Arlberg will react to a shock.

The structure of the rest of this thesis will be as follows. Firstly, past literature on forecasting and tourism forecasting and the methodology will be discussed. Secondly, a time series analysis will be done based on tourism data of the past 10 years and forecasts will be made for the coming years. Based on these predictions and deviations, recommendations will be given regarding a marketing strategy that could help St. Anton recover from the crisis.

Literature review

Because the corona crisis is a very recent event, there is little research on how it will affect travel and tourism. However, research on tourism crises and the effects of other epidemics is more readily available. This chapter will discuss literature to shape a theoretical framework for the research in this thesis.

In the last few years, a dichotomy among tourism scholars has arisen, between scholars who are pro-growth and scholars who are calling for more responsible and sustainable tourism. The COVID-19 crisis has made the divide even bigger. Some scholars calling for strategies to recover and return to a pre-covid situation, while others see the pandemic as an opportunity to rethink tourism and learn from the past. Higgins-Desbiolles (2020) sets apart these two ideologies. A crisis, however, does not just create a moment for either reflection or change, it can create moments for both as crises reveal dynamics that may not have been noticed before. Certain scholars, such as Butcher (2020), fear that reform of the tourism industry after the COVID-19 crisis will lead to negative effects on the tourism industry such as loss of jobs and decline of the industry as a whole. Others, such as Roy (2020) and Zizek (2020), see this crisis as an opportunity to limit tourism and strive for more sustainability, equality and justice. With his paper, Sigala (2020) emphasized the need for research on the effects of COVID on tourism. The author found that due to the differences in both demand groups and suppliers, the impacts and implications should be studied according to these different target markets. Moreover, he found that negative impacts of the gig economy such as increased pressure and work stress due to lack of job security, insurance and other benefits, have become stronger and more evident due to the COVID-19 pandemic. Workers in the gig economy might not even be entitled to government support during the pandemic, making them even more vulnerable. Additionally, people studying tourism now have been confronted with a sudden standstill in their industry, with cancelled internships and recruitment and an insecure future in the industry. Sigala (2020) looked at how the COVID-19 pandemic could be an opportunity for transformation and the role of research in fostering this change. In the past tourism has proved to be resilient to negative shocks, however, this pandemic is an unprecedented situation. This means that the pandemic could have different and long-lasting effects. Not only will the impact of the COVID-19 pandemic be enormous, moreover the effects of the pandemic are uneven with not all industries experiencing the same impact at the same time and some industries not even feeling an impact. Tourism is one of the hardest-hit industries with tourist arrivals dropping 78% in 2020. This impact is seven times larger than the impact of 9/11 (UNWTO, 2020). This also means that so-called old examples and models do not work in this context. Sigala (2020) focused not so much on how to implement recovery and build resilience for future crises but instead focused on how crisis can foster change and how companies can turn disruption into innovation. Change can be selective and previously crisis-led change has depended on whether and who is affected by the crisis. In order to change the understanding and prediction of the COVID-19 aftermath, research has to understand the drivers of stakeholders and their actions and reactions to the COVID-19 pandemic (Sigala, 2020). To do this, research should look at stakeholders' experiences as well as consciousness and willingness to act on the pandemic. In order to generalize the effects of such research, the knowledge should be formulated in a way that it's suitable for informing and shaping crisis-led transformations.

Skare et al. (2021) was one of the first papers discussing the effects of the COVID-19 crisis on tourism. The authors found that previous crises mostly led to shocks on the individual level and domestic tourism revived as soon as domestic shocks were resolved. However, this crisis is different and it is expected that recovery will take more time. COVID-19 shows that pandemics have a bigger impact and are more destructive than previous research indicates. But even now, many policymakers underestimate the impact COVID-19 will have on the tourism industry. One of the biggest problems due to a decrease in travel and tourism is job risk for the many people working in the tourism industry and GDP loss for countries with a big tourism industry. Skare et al. (2021) conclude that in order to fight a pandemic in the future, the industry needs a crisis-readiness mechanism based on empirical knowledge of the extent of the COVID-19 pandemic. The COVID-19 pandemic is expected to have long-lasting negative effects on tourism and the best way to recover is cooperation between businesses instead of competition (Skare et al., 2021).

Focusing on city tourism in Austria, Jiricka-Pürrer et al. (2020) argue that recovery from the COVID-19 pandemic offers an unprecedented opportunity to change tourism and make it more environmentally friendly. Short city trips are very environmentally unfriendly as most city-trippers take an airplane to get to their destination, moreover increasing demand for short city-trips has led to a situation of over-tourism. The authors argue that flights, especially short-distance, will remain unattractive and this could be enhanced with taxes and policies to discourage air travel. Secondly, the social distancing imperatives to battle COVID-19 have led to a reconsideration of free space in cities. This also became clear during hotter periods when citizens tried to escape the city by going to nearby recreational areas in the suburbs. In a post-COVID situation, the challenge will be to contain crowds at these locations as both citizens and tourists want to escape the heat of the city. This calls for a new way of thinking about creating free, open spaces and limiting crowds (Jiricka-Pürrer et al., 2020). The pandemic put a harsh stop to travel for the time being and this gives a unique opportunity to create a new normal, where not only COVID measures can be taken into account but also the environment. The paper by Jiricka-Pürrer et al. (2020) makes it clear that many measures taken to battle COVID-19 are also in line with a more environmentally friendly way of tourism. Nepal (2020) also emphasizes the fact that especially countries relying heavily on tourism, have to rethink the way they go about tourism and have to find a way to reconcile tourism with COVID restrictions. The COVID crisis has made it clear that just aiming for more tourists every year is not a sustainable strategy. The focus of tourism destinations should be less on numbers and more on the experience, this way crowds can be contained while still providing tourism.

The travel restrictions put in place to fight the COVID-19 pandemic may have set in motion a deglobalization, but people still have a desire to go out and explore. This has introduced a lot of people to local tourism. This has led to an increase in local tourism and a preference for destinations with lower population density and more natural environments. Jeon and Yang (2021) analyzed these changes as seen in Korea and discussed their implications. They found that people not only choose destinations closer to home but also rather have one destination than multiple and prefer destinations with a lot of free space. These changes are a direct effect of the measurements taken to fight COVID-19, with most countries restricting international travel and promoting social distancing. Jeon and Yang (2021) suggest that destinations should promote eco-friendly and socially distant activities such as driving routes, walking tours and bike tours. They also see an opportunity for activities and destinations focused on mental health improvement.

Fotiadis et al. (2020) used a Long Short Term Neural Network and a Generalized Additive Model to predict tourism demand for the next year. Not only is COVID-19 one of the deadliest pandemics, but it has also brought along crises in healthcare and economies as well. This has led to a loss of between 910 billion and 1.2 trillion dollars. As hospitality venues such as restaurants, hotels and event location are seen as hotspots for spreading the virus, these locations have been forced to close down completely and will like to feel the impact of the pandemic for a long time. Forecasting the impact the pandemic has on the industry has benefits for both businesses and governments. It can provide businesses with useful insights on building and implementing effective recovery systems and decision-making frameworks to ensure quick responses to unforeseen situations. For policymakers, epidemics are not just threats to public health but also represent a threat of economic downturn that need fiscal, monetary and supply-side measures to recover (Fotiadis et al., 2020). With tourism accounting for 9.5% of employment among the 18 to 64-year-old population and food and drinks sector for 19.7% and 58.7% of employment in the tourism industry, it is clear that shocks to this industry have a big impact. Fotiadis et al. (2020) used both models and trained these based on data on the SARS epidemic in 2003-2004, the MERS outbreak in 2013-2014 and the financial crisis of 2008-2009. The authors found that for the next year (2021) the tourism industry will have to deal with losses of around 50% and that these losses will persist at least until the next summer. Due to the COVID-19 pandemic growth in the industry will be undone for as much as 15 years. Moreover, the current season (2020) should be counted as a total loss as most countries were still struggling to contain the virus and find a medical solution to stop the virus from spreading.

The industry, however, is not only facing reduced revenue but also increased costs as businesses have to implement government-imposed measures. This means that businesses should incorporate these increased costs and reduced revenues in their revised business models. Additionally, with the effects of the pandemic persisting for at least a year, businesses should also re-evaluate their profitability and might even have to come up with new products, reshaping the industry. This means that, especially in countries that rely heavily on tourism for their GDP, governments should help in providing the implementation of measures. Government involvement will also help ensure that the measures are correctly implemented.

The available literature on specifically tourism and COVID-19 generally describes an opportunity to rethink and reshape tourism. Due to the pandemic, the industry came to a near halt and now it is up to the industry and the tourists to rebuild tourism in a way that is sustainable in a future where COVID-19 will most likely be present. In this future, the focus will most likely be on destinations closer to home where social distancing can easily be carried out. This means that a destination with a lot of free space and not a lot of people could be more popular than cities.

Data

The research question will be answered by analyzing and forecasting data on overnight stays in Sankt Anton am Arlberg, the entire Arlberg region and overnight stays in different types of accommodations. The data is collected by the St. Anton Tourism Office (Tourismus Büro) and contains information on the length of stay, number of tourists per month and per year, type of accommodation. Multiple models will be built and for each model, a different dataset containing the relevant information will be used. The datasets contain monthly and yearly data on the number of tourists and from the 2010-2011 season until the 2019-2020 season and yearly data on the number of guests per accommodation from the 2010-2011 season until the 2019-2020 season.

A quick analysis of this data shows that before the 2019-2020 season, tourism in St. Anton showed a strong upward trend with 1.04 million overnight stays in the town of St. Anton and 1.27 million overnight stays in the whole Arlberg region in 2018.

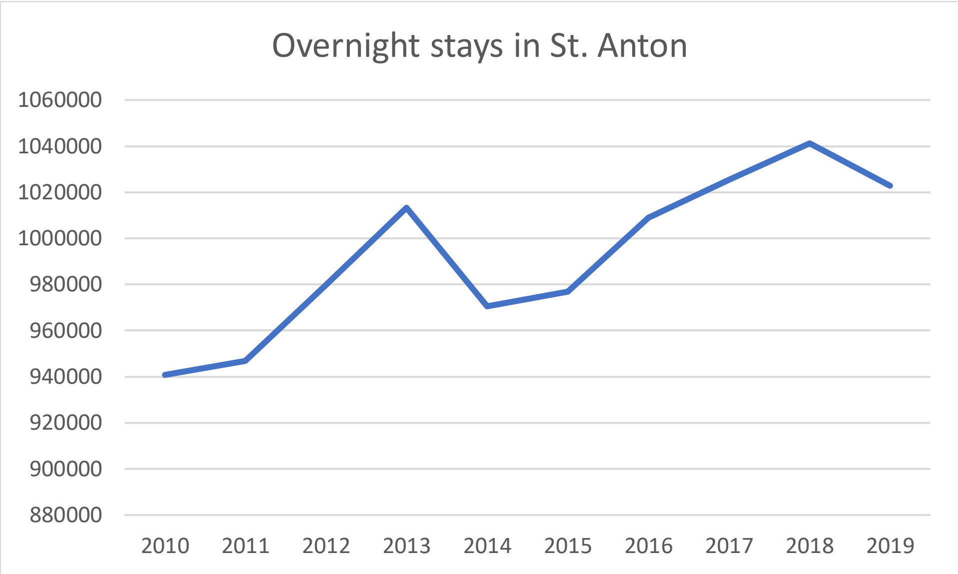

Figure 2: Number of overnight stays in St. Anton per year

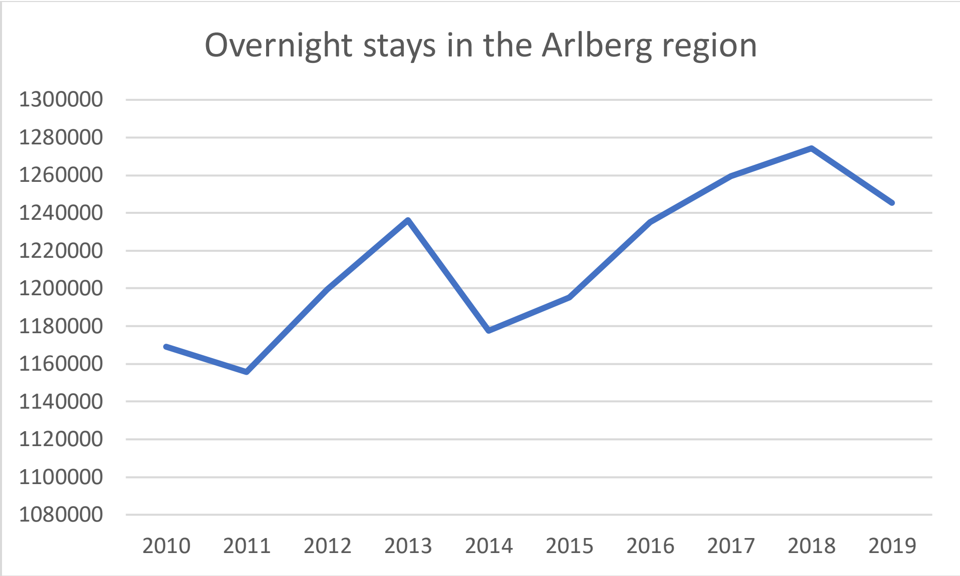

Figure 3: Overnight stays in the Arlberg region per year

When looking at the countries that these tourists come from, it can be determined that most tourists come from within the European Union. With two-thirds of the tourists coming from Germany, Great Britain, Austria or the Netherlands. More than half of these tourists stayed in commercial accommodations such as hotels and pensions as well as chalets owned and operated by travel agencies (figure 5).

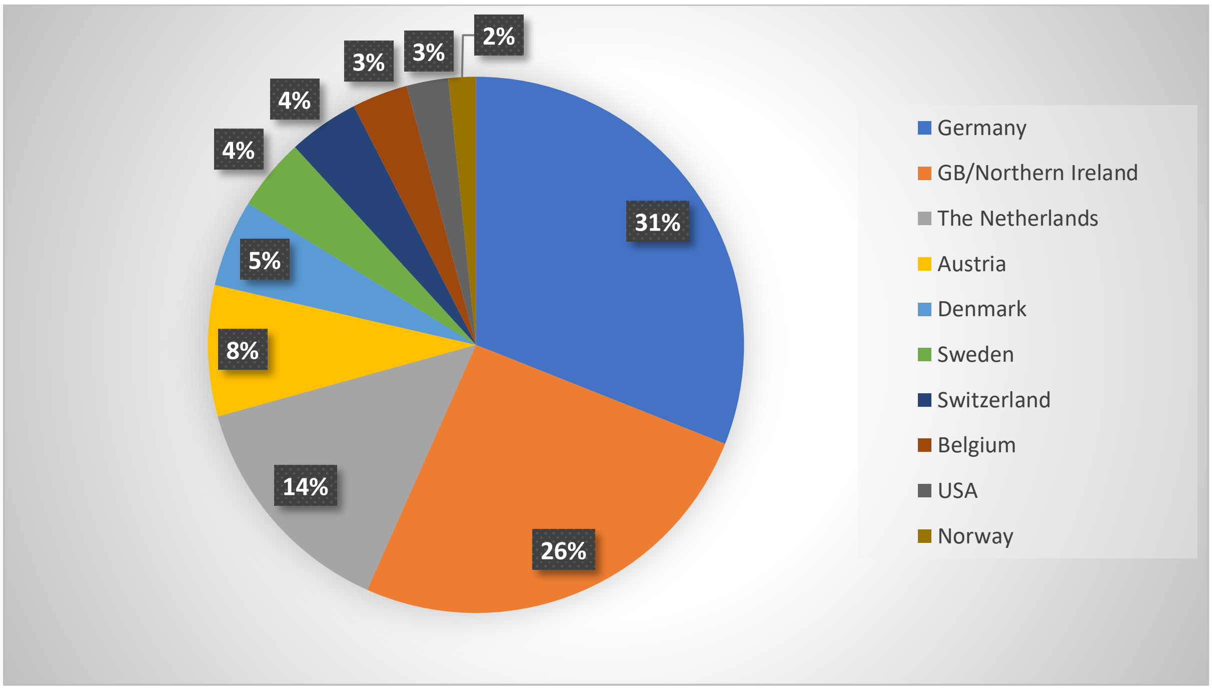

Figure 4: Overnight stays per country of origin for the 2018-2019 (Tourismus Büro St. Anton, 2020)

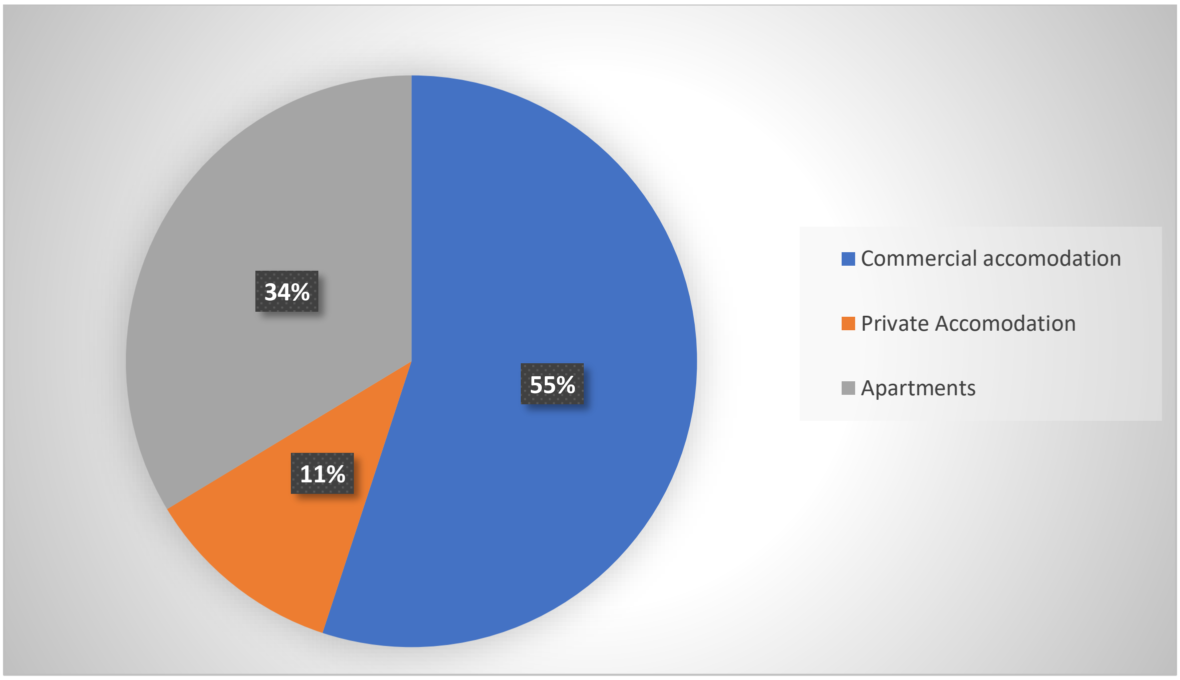

Figure 5: Yearly overnight stays per accommodation category for the 2018-2019 season (Tourismus Büro St. Anton, 2020)

These datasets contain information on multiple variables over time and are multivariate time series. Multivariate times series are time series where each variable is dependent on its past values but is also in some way dependent on the other variables. For instance when you have a dataset on ocean tides that also includes data on rain, wind, moon phases and season. Each variable is dependent and related to its past values but all variables are also dependent on each other. For instance, the tides depend on the moon, wind, season as well as past tides. But wind depends on the season and other weather factors as well as past wind. This way you get a time series where everything is related to everything.

Before any model can be built, the data has to be cleaned and prepared. This means removing missing values and checking for stationarity. Once the data is prepared, the data is split into a training and validation set. 80% of the data is used for training, 20% for validation and then predictions are made based on the whole dataset.

The analysis will be done in two stages: in the first stage analyses and forecasts will be made based on data until the 2018-2019 season, so without the presence of COVID-19, and in the second stage analyses and forecasts will be compared to the observed data from the 2019-2020 season.

Methodology

Introduction

For analyzing and forecasting multivariate time series, there are several statistical models. Some of these have already been used in recent studies researching the effects of the COVID-19 pandemic on tourism. This thesis will use two models that have previously been used in studying the effects of COVID-19 on tourism, namely a Vector Autoregression model and a Generalized Additive Model. This chapter will discuss the methodology behind both models.

Vector Autoregression

One of the most obvious ways to analyze and forecast multivariate time series is a vector autoregression (VAR) model. These models have proven to be the most successful and flexible models to use in forecasting multivariate time series (Stock & Watson, 2001). And since its introduction by Sims (1980), it has become a fundamental model in economic research. VAR models can be used to capture relationships between multiple variables that change over time. Each variable has an equation that models the variable's change over time, this equation contains past values of the variable itself and the other variables as well as an error term. VAR models are linear models with n equations and n variables. This framework is very simple and provides a systematic way to interpret multivariate time-series dynamics. In this thesis, a structural VAR will be used based on Pedroni (2013) and Skare et al. (2021). Structural VARs look for simultaneous links based on economic theory, to interpret the relationships as causal assumptions that need to be identified (Stock & Watson, 2001). Panel VAR applies the VAR framework to panel data.

Before diving deeper into the exact methodology of the vector autoregression understanding of the univariate autoregression, or simply AR is necessary. An AR model is an autoregressive model with one variable.

$$y_t = a_1 y_{t-1} + \varepsilon_t$$

In this model, the current value of the variable depends on its own lagged value. With being the coefficient for lag 1. The model depends on p lags of y or is autoregressive to the order p. The error term is assumed to be normally distributed with a mean of zero and a variance of (Lütkepohl, 2007). Before estimating any autoregressive model, stationarity has to be tested to make sure the right model is chosen. Non-stationary data cannot be properly analyzed and forecasted with autoregressive models.

Creating a model where the value of a variable is dependent solely on its previous values, is often too restricting. Assuming there are more variables that influence a variable's current value is more realistic. This assumption calls for models that not only included lagged values of the dependent variable but also includes exogenous variables. However, in the data used in this thesis, these exogenous variables also depend on lagged values of the endogenous variables. This means, that the AR model has to be expanded to a VAR or vector autoregression model.

The VAR model will treat each variable as a linear function that depends on past values of that variable and past values of all other variables. A VAR model is usually denoted as VAR(p), where p is the number of lags in the model. For a VAR model with n variables each forecasted variable $y_{i,t}$ is:

$$y_{i,t} = a_{i1} y_{i,t-1} + \cdots + a_{in} y_{n,t-1} + e_t$$

Each variable can be written as a separate equation, but it is only a VAR model when all these equations are combined. In a VAR it is assumed that the error terms follow a normal joint distribution.

For a VAR with p lags and k variables, this will give:

$$\begin{bmatrix} y_{1,t} \\ \vdots \\ y_{k,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix} + \begin{bmatrix} a_{1,1}^1 & \cdots & a_{1,k}^1 \\ \vdots & \ddots & \vdots \\ a_{k,1}^1 & \cdots & a_{k,k}^1 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ \vdots \\ y_{k,t-1} \end{bmatrix} + \cdots + \begin{bmatrix} a_{1,1}^p & \cdots & a_{1,k}^p \\ \vdots & \ddots & \vdots \\ a_{k,1}^p & \cdots & a_{k,k}^p \end{bmatrix} \begin{bmatrix} y_{1,t-p} \\ \vdots \\ y_{k,t-p} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ \vdots \\ e_{k,t} \end{bmatrix}$$

This model implies that all variables depend on all variables and each row can be written as a separate equation. Therefore a VAR model could be deconstructed into multiple individual autoregressive distributed lag (ADL) models.

VAR for post-COVID tourism

By using a VAR model this thesis will build on research done by Skare et al. (2020) who used a panel structural auto-regression (PSVAR) to estimate the impact of the COVID-19 crisis on tourism. Skare et al. (2021) used data on previous epidemic outbreaks of viruses to model shocks to tourism. However, previous epidemics such as the SARS virus, H1N1, Ebola, bird flu and MERS were very different from the current COVID-19 pandemic, therefore they used knowledge on the reproduction number of COVID-19 to calibrate the variances. With their PSVAR model, they were able to model the impact of COVID-19 on multiple factors such as sector stability, vulnerability to external shocks, health system stability, competitiveness and region-specific socioeconomic and environmental circumstances. The authors were able to divide the effects into common and individual or country-specific shocks. This thesis, however, does not use panel data and will therefore simplify the model from Skare et al. (2021) to a structural VAR.

One of the main advantages of VAR models is that it only needs a list of the variables that can be hypothesized to have relationships with each other. VAR models are often used to model and forecast economic and financial time series. The forecasts of these models can be made dependent on potential future paths of specified variables, this makes the forecasts very flexible. Forecasting using a VAR model can be done without many a priori expectations and restrictions.

Estimating parameters

To start building the model, first stationarity has to be determined. This is done mathematically with a Phillips Perron test and visually with a plot. This test will allow to determine stationarity without defining the number of lags. The Phillips Perron test builds on the Dickey Fuller test. This test tests the null hypothesis of $\rho = 1$ in

$$\Delta y_t = (\rho - 1)y_{t-1} + u_t$$

In this equation Δ denotes the first difference operator.

The Phillips Perron test takes into account that the process of generating data for y_1 can have a higher order of autocorrelation than is reflected in the equation which makes y_{t-1} endogenous. The Phillips Perron test accounts for this by not defining the number of lags. The Phillips Perron test has the null hypothesis that the variable contains an unit root. The alternative hypothesis states that the variable was generated through a stationary process. Therefore, to determine stationarity the Phillips Perron test p-value should be greater than your alpha as this will allow to reject the null hypothesis. After determining stationarity, the model can be defined and the parameters and covariance can be estimated.

The estimation of the parameters and the covariance matrix of a simple VAR model is not exactly straightforward. The VAR model consists of linear equation, which make it seem as OLS is a good estimator for the parameters. However, OLS estimates are biased because the error term of the model is not unrelated to the current and future values of y. Therefore, maximum likelihood is a better way to estimate the parameters.

To get the maximum likelihood estimator (MLE), let us take θ as a set of potential parameters. And *f* (y_1, y_2, ..., y_n|θ) that describes the probability density of (y_1, y_2, …, y_n). In other words, for a specific observed sample *f* is the likelihood of this sample occurring if the true value of the parameters equals θ. The maximum likelihood estimator is θ^{MLE} that maximizes that likelihood function for the observed data. These estimators can only be obtained with numerical methods.

Suppose a set of observations is obtained from a normal distribution with mean μ and standard deviation σ. The estimators are the values of μ and σ that maximize the joint likelihood of all the observations. For a VAR with no time series this gives:

$$f(y_1, y_2, \dots, y_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(y_i - \mu)^2}{2\sigma^2}\right]$$

The joint unconditional distribution for a time series is a little more complex. But the general rule is that each y_i is dependent on y_{i-1}. This gives the following joint density function, where θ = (ρ, σ):

$$f(y_1, \dots, y_n | \theta, y_1) = f(y_n | \theta, y_{n-1}) f(y_{n-1} | \theta, y_{n-2}) \dots f(y_2 | \theta, y_1)$$

$$f(y_1, \dots, y_n | \theta, y_1) = \prod_{i=2}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(y_i - \rho y_{i-1})^2}{2\sigma^2}\right]$$

Transforming to log:

$$\log(y_1, \dots, y_n | \theta, y_1) = -(n-1)\left(\frac{\log 2\pi}{2} + \log\sigma\right) + \sum_{i=2}^{i}\left[\frac{-(y_i - \rho y_{i-1})^2}{2\sigma^2}\right]$$

The joint likelihood function is then:

$$L(\theta) = f(Y_1, \dots, Y_n | \theta) = \prod_{i=2}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(Y_i - \rho Y_{i-1})^2}{2\sigma^2}\right]$$

For Y_2 given Y_1 this gives:

$$Y_2 | Y_1 = y_1 \sim N\left(\mu_2 + \frac{\sigma_{12}}{\sigma_{22}}(y_1 - \mu_1), \sigma_{22}(1-\rho^2)\right)$$

After calculating the maximum likelihood estimator, the best parameters can be chosen with a likelihood ratio test. The likelihood ratio test compares goodness-of-fit between different models based on the likelihood ratio, which is the ratio of the models' respective likelihoods.. If the likelihoods of the different ratios differ more than one sampling error, the null hypothesis should be rejected. Mostly, the null hypothesis states that θ should be a subset of Θ.

The final parameter to be determined, is the optimal lag length or number of lags. The optimal number of lags can be obtained through model comparison. This is usually done based on information criteria like AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) or HQ(Hannah-Quinn). AIC has small forecasting features, which makes it a very attractive information criterion to use. However, BIC and HQ work better in large samples and are consistent estimators of the true order. This means that BIC and HQ prefer the true order of the model opposed to the order that gives the best forecasts. In this thesis the AIC as well as BIC and HQ will be calculated for each model to determine the optimal amount of lags.

Once a final VAR model is chosen, its estimated parameter values have to be interpreted. Since all variables in a VAR model depend on each other, individual parameter values only provide limited information on the reaction of the system to a shock. To get a better intuition of the model's dynamic behaviour, impulse responses (IR) are used. They give the reaction of a response variable to a one-time shock in an impulse variable. The trajectory of the response variable can be plotted, which results in those wavy curves that can be found in many macro papers.

VAR models apply economic theory to find patterns and links over time between variables. To interpret these correlations causally, several assumptions have to be met. So, before any interpretation can be done these assumptions have to be checked. The assumptions are:

1. **Error terms should be non-correlated with previous periods.**
2. **No heteroskedasticity**
3. **Normal distribution of error terms.**

Each assumption has its tests to see if the assumption is satisfied. The first assumption can be tested with an asymptotic Portmanteau test. This is a statistical test that tests if the autocorrelations of the error terms of the model are different from zero. This test is suitable for the first assumption because only one hypothesis has to be explicitly defined, in this case, the null hypothesis states that error terms are not correlated to previous periods.

The second assumption, no heteroskedasticity, can be tested with the Auto-Regressive Conditional Heteroskedasticity (ARCH) effect test. This is a test to test the null hypothesis that error terms portray no conditional heteroskedasticity. In other words, the null hypothesis of this test states that there is no heteroskedasticity so that when the null hypothesis cannot be rejected, the assumption of no heteroskedasticity is met. When variables are highly correlated, performing an ARCH test can be challenging. In that case, absence of heteroskedasticity can be determined with a residual plot. Figure 1 shows an example of a residual plot on random data showing heteroskedasticity. In case of no heteroskedasticity, the residual plot should show no clear fan or cone shape in the error terms.

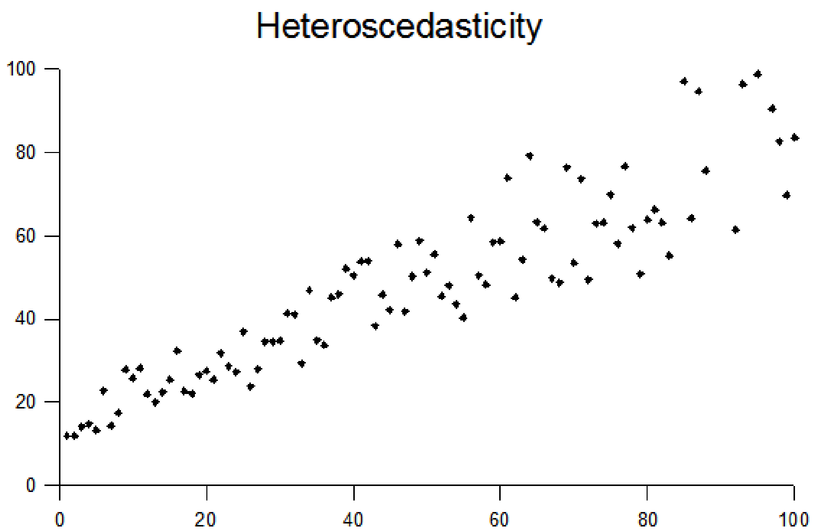

Figure 6: Heteroskedasticity in random data

The third and final assumption calls for a normal distribution of error terms, this assumption can be tested in multiple ways such as the Jarque-Bera test, Kurtosis, skewness and can be visualized with a QQ-plot. The easiest way is to use a QQ plot. A QQ plot shows the error terms plotted against a normal distribution. It plots the quantiles of the error terms against the quantiles of the distributions. If the error terms are normally distributed, the QQ plot should show a straight diagonal line, curvature in a QQ plot reflects skewness of the distribution of the error terms.
The final step before interpreting the model is to determine that there are no structural breaks in the data. Structural breaks are points where the data deviates from its trend, this happens when an exogenous variable changes unexpectedly for instance if there is an economic crisis. If the data contains structural breaks the forecasts made by the model are no longer accurate. Stability is tested with a cumulative sum test, this test uses the cumulative sum of OLS error terms. The null hypothesis of this test states that the cumulative sum of the error terms has a mean of zero (Ploberger & Kramer, 1992). Once the parameters are estimated, all assumptions are satisfied and stability is established, the results of the VAR model can be causally interpreted.

First, Granger causality and instantaneous causality is tested between the dependent variables. Next, impulse response functions are drawn. Impulse response function plots show the effect of an impulse in one variable on another variable over time. For instance, a plot could show the effect of a shock in chocolate sales on marshmallow sales for the next 5 years. This allows for the assessment of dynamic effects and To substantiate the conclusion from these plots, the functions are bootstrapped. To do this the VAR model is estimated with OLS and the residuals are calculated. From these calculated residuals, samples are drawn. For each sample, a time series with an estimator is calculated for each sample. Based on the covariance between y and its lags, a covariance matrix is calculated. Based on these statistics, a confidence interval is calculated.

Generalized Additive Model

Another great way of analyzing and forecasting time series is by using a Generalized Additive Model (GAM). GAM is a regression method that can model a dependent variable based on independent variables that are in the form of smooth functions called splines. In other words, our response variable depends on unknown smooth functions of the predictor variable. This model is a mix of a generalized linear model and an additive model, first created by Hastie & Tibshirani (1990). The main idea is that the relationships between the predictive variables and the dependent variable follow smooth patterns, these can either be linear or non-linear. These relationships can be estimated simultaneously and added up to predict the dependent variable. This way a GAM model finds the right balance between an easily interpretable but biased linear model and very flexible and complex black-box machine learning algorithms. In an additive linear model, interpretation of a partial derivative does not depend on the values of other variables in the model. This makes it possible to formulate statements based on the output of the model that is understandable to people with little knowledge about the mathematical background of a model. As the interpretation of the model is the same as for a simple or multiple linear regression. Moreover, GAM models can avoid irregular predictor functions by adjusting the smoothness level tackling the variance/bias trade-off upfront. This imposes a belief that predictor functions are smooth by nature, which makes the results more easily interpretable. Another advantage is that, because predictor functions are automatically derived during model estimation, common nonlinear patterns that would be missed by regular linear models are captured.

The Kolmogorov-Arnold representation theorem (Kolmogorov, 1957 & Schmidt-Hieber, 2021) proved that a multivariate function can be represented as a sum of univariate functions. In the form of:

$$f\left(\underset{x}{\rightarrow}\right) = \sum_{q=0}^{2n} \Phi_q \left(\sum_{p-1}^{n} \phi_{q,p}(x_p) \right)$$

However, this theorem assumes that there is an already existing multivariate function and does not give the tools needed to construct this multivariate function. To make the function more suitable for modeling, the generalized additive model simplifies this function by not including the outer sum. Here, x_g is the same as x_p in the previous equation. Φ is a smooth function representing the variables.

$$\mathrm{f}\left(\underset{\mathrm{x}}{\rightarrow}\right) = \Phi\left(\sum_{p=1}^{n} \phi_p(x_g)\right)$$

With g for the inverse of Φ :

$$\mathrm{g}\left(f\left(\underset{x}{\rightarrow}\right)\right) = \sum_{\mathrm{i}} \mathrm{f_i}(\mathrm{x_i})$$

Here g is a function representing the inverse of Φ and $\mathrm{f_i}(\mathrm{x_i})$ represents the function i of variable $\mathrm{x_i}$.

Then the standardized formulation of the generalized additive model is then derived from the formula for approximating an expected value of an observed unit:

$$\mathrm{g}(\mathrm{E}(\mathrm{Y})) = \beta_0 + \mathrm{f_1}(\mathrm{x_1}) + \cdots + \mathrm{f_m}(\mathrm{x_m})$$

The smooth function f consists of the sum of the known basis functions $b_i(x_i)$ of the smoothing splines, and their regression coefficients β:

$$f(x) = \sum_{i=1}^{q} b_i(x)\beta_i$$

In this equation q is the basic dimension.

Smoothing splines are functions that are piecewise defined by polynomial functions, in other words each part of the smoothing spline is defined by a polynomial function. These polynomial functions connect at the knots. For GAM a penalized regression is used to regularize the smoothness of a spline. This allows for a linear formulation of the model:

$$g(E(y)) = \beta X + \varepsilon$$

With X being the matrix representing the model and β being the vector of coefficients. This means the goal is to minimize:

$$\|y - \beta X\|^2 + \lambda \int_0^1 [f"(x)]^2 dx$$

With:

$\int_0^1 [f"(x)]^2 dx = \beta^T S\beta$

Where S is a matrix of known coefficients. This means that the regression coefficients can be obtained by:

$$\hat{\beta} = (X^T X + \lambda S)^{-1} X^T y$$

The estimate of β, $\hat{\beta}$, is obtained through penalized iterative re-weighted least squares (P-IRLS).

Once the smoothing parameters are chosen, the degrees of freedom need to be established. In a normal Generalized Linear Model (GLM) the degrees of freedom is the same as the number of parameters to be estimated. For GAM it is different due to the use of smoothing splines. The effective degrees of freedom will then be:

$$model\ edf = tr(B(B'WB + P)^{-1}B'W)$$

Where P is the penalty matrix in the form of a diagonal block matrix, B is the model matrix of the basis functions and W is the weighting matrix for the coefficients. This is similar to calculating degrees of freedom for a GLM, where the trace of the hat matrix is taken.

A GAM model does not automatically filter out the variables with the biggest causal relation to the dependent variable, like a regression. For a GAM model it is necessary to chooses which variable to include as predictor variables. This has several reasons, one of the biggest being that GAM is susceptible to multicollinearity. Moreover, GAMs are parameter heavy, meaning that including a lot of predictor variables can lead to strange and untrustworthy results. Therefore, dependent variables need to be screened beforehand in order to filter out variables that will never have a significant contribution to the model. This is done by measuring the univariate strength of the variables.

Measuring univariate strength can be done with the Information Value (IV) and Weight of Evidence (WOE), these are some of the most powerful measures. Both of these measures show statistical strength of findings, WOE shows the predictive power of data and IV ranks the variables based on their importance. With this framework it is possible to detect any relationship between a predictor and the dependent variable, both linear and nonlinear, as well as visualize the relationship. Moreover, this framework can assess the predictive value of missing values and seamlessly compare strength of variables. The WOE/IV framework is based on the relationship that says that a conditional logit, given x_j, can be written as an overall log-odds plus log-density ratio, which is the weight of evidence. This means that estimating the weight of evidence and fitting a logit model is the same. This way, WOE and x_j and WOE can be plotted against each other to see how x affects Y.

$$\log\frac{P(Y=1|x_j)}{P(Y=0|x_j)} = \log\frac{P(Y=1)}{P(Y=0)} + \log\frac{f(x_j|Y=1)}{f(x_j|Y=0)}$$

The predictive strength of x can be measured by leveraging WOE, through the information value:

$$IV = \int \log\frac{f(x_j|Y=1)}{f(x_j|Y=0)} \left(f\left(x_j\middle|Y=1\right) - f\left(x_j\middle|Y=0\right)\right) dx$$

This way the information value is basically the weighted sum of all WOE values. The weights take into account the differences between the numerators and denominators. This means that to find the univariate strength, WOE has to be calculated.

To calculate WOE, the conditional densities have to be known or calculated. The most common approach is to bin x_j in a k x 2 table where k is the number of bins. The cells in the matrix count the number of times that Y equals 1 and 0 respectively. The column percentages from this table give the conditional densities. When B_1, …, B_k denote the bins for x_j, the WOE for bin i is:

$$WOE\left(x_j\right)_i = \log\frac{P(X_j \in B_i|Y=1)}{P(X_j \in B_i|Y=0)}$$

So, then if the IV is the weighted sum of all k WOE values. This gives:

$$IV\left(x_j\right) = \sum_{i=1} k\left(P(X_j \in B_i|Y=1\right) - (X_j \in B_i|Y=0)) \times WOE\left(x_j\right)_i$$

Through pre-screening, the pool of possible variables in the model is reduced, but this does not define the final model. Once pre-screening is done, multivariate selections of variables inside a GAM model have to be applied for the final model. In R there are two options for variable selection: stepwise selection and shrinkage. Stepwise selection is simply eliminating or adding variables until the most optimal model is reached. Shrinkage adds penalties to the model by adding constants to the diagonal of the penalty matrix, this way smoothers the are very wiggly are shrunk to 0. Thus in shrinkage variables are not actually removed of added, but weak smoothers are set to 0.

Once all parameters are determined and the variables are screened the model is fitted through a backfitting algorithm. There are different parameters that can be used here, namely Generalized Cross Validation Criteria (GCV) and Restricted Maximum Likelihood, which is a mixed model approach (REML). REML converges faster and GCV is likely to under smooth the predictor variables.

The goal in fitting a GAM, with both GCV and REML, is to maximize the penalized likelihood function. The penalized likelihood function is:

$$2l(\alpha, s_1(x_1), \dots, s_p(x_p)) - penalty$$

Where $l(\alpha, s_1(x_1), \dots, s_p(x_p)$ is the standard log likelihood function and the penalty can be based on derivatives:

$$penalty = \sum_{j=1}^{p} \lambda_j \int s'_j(x_j)^2 dx$$

The parameters $\lambda_1, \dots, \lambda_n$ are the smoothing functions mentioned above. These control the penalty in the model. In other words, the smoothing functions define the level of smoothness that is imposed.

Next, different VAR and GAM models will be created and interpreted. First, forecasts will be made based on the period from 2011 until 2019, so without the COVID affected season, this will be the first and second stage. And then after that, in the third stage, forecasts will be made based on the 2019-2020 season, so based on the COVID affected season.

Results

Vector Autoregression

In this thesis, multiple time series were analyzed and for each time series analysis and the prediction was done with a VAR and a GAM model. In each model, the 2019/2020 season was left out as it was highly affected by covid and therefore could give a distorted image for the future. For both St. Anton and the entire Arlberg region the total number of overnight stays was forecasted based on the total number of overnight stays for both St. Anton and the Arlberg region for 2010 until 2019. Second, the overnight stays in different categories of accommodation for St. Anton based on yearly data on these variables from 2010 till 2019 were forecasted. Third, the amount of guests and the amount of overnight stays per month for St. Anton based on monthly data of 2010 till 2019 was forecasted. Finally, the observations for the 2019-2020 season were compared to the forecasts for each model.

St. Anton versus Arlberg Region

For the first model a typical, unrestricted VAR was run. The number of lags in the model was 1, as this was found to be the optimal number of lags for our model. A model with 1 lag has the lowest AIC. Since VAR models deal with multiple intercorrelated variables, the coefficients do not give a lot of information. Therefore these are not interpreted separately, instead, the results of the applications are interpreted.

However, before applying and interpreting our model, diagnostics should be run to make sure all assumptions are satisfied. The first assumption states that the error terms should be non-correlated with previous periods. This is tested with an asymptotic Portmanteau test, this test had a p-value of 0.002 (table 1) which is smaller than 0.05 which means the null hypothesis that the error terms are correlated is rejected. Thus, the first assumption is satisfied.

The second assumption is the absence of heteroskedasticity. This is tested with an ARCH test that can identify clustered volatility areas. The p-value for this test is 0.4363 (table 1), which is not lower than 0.05 which means the null hypothesis that there is heteroskedasticity cannot be rejected. Moreover, figure 1 shows the residual plot of this first model and shows slight heteroskedasticity. Therefore the second assumption, of no heteroskedasticity, is not satisfied. However, this problem can be fixed by using Heteroskedasticity Robust Standard Errors. This way the second assumption is satisfied after all.

Test	Chi-squared	p-value
Portmanteau	5.3033	$<2.2e^{-16}$
ARCH	0.4363	0.4363

Table 1: Model Diagnostics

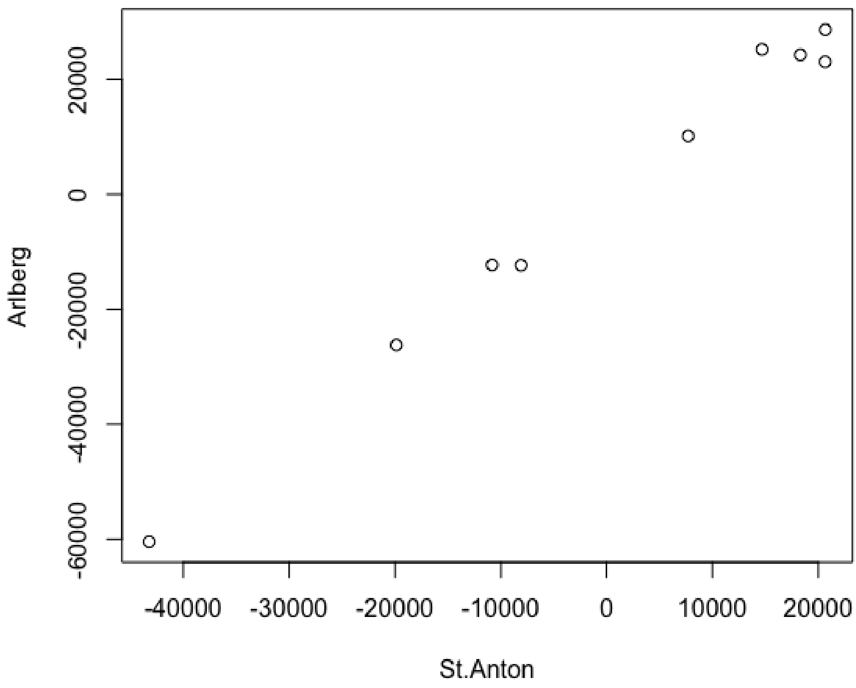

Figure 7: Residual plot for overnight stays

The final and third assumption to be tested is a normal distribution of the error terms. This is determined with a QQ plot, which is shown in figure 2. The QQ plot does not show a big curvature. The straight diagonal line fits the error terms pretty well. Therefore, the assumption of normally distributed error terms is satisfied.

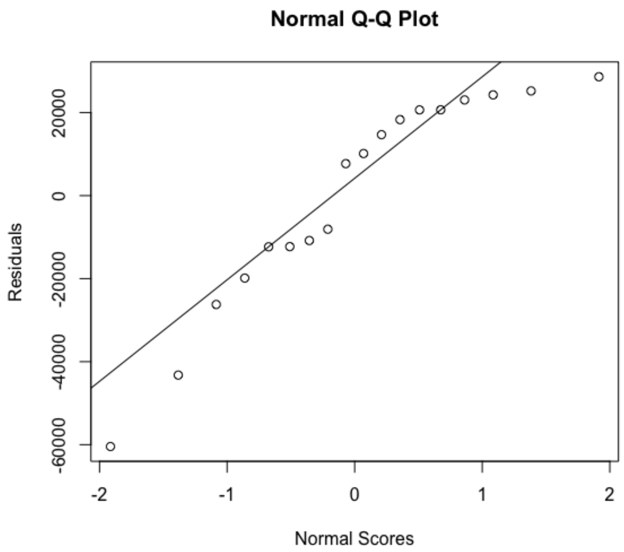

Figure 8: QQ plot for overnight stays error terms

Next, the presence of structural breaks can be tested with a stability test. If structural breaks go unnoticed, the estimations made by the model are no longer accurate, therefore this is visually tested with a plot of the sum of recursive residuals. If the line representing this sum goes outside of the red bounds, a structural break was observed. These plots are shown in figure 3. The plots show that for the time series used in this model, there were no structural breaks.

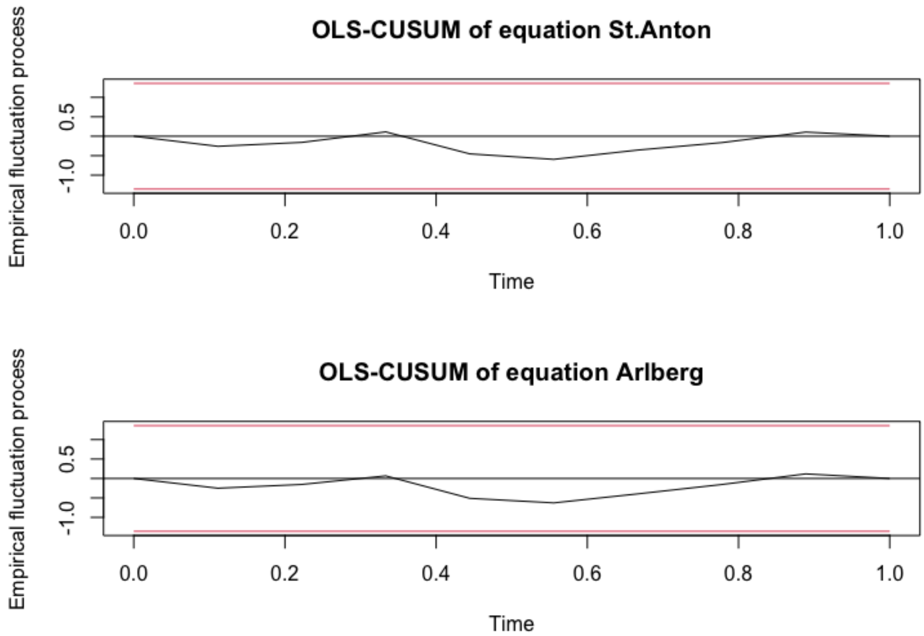

Figure 9: OLS-CUSUM for St. Anton and Arlberg region

Now that all assumptions and stability are tested, changes can be simulated. This is done by testing the Granger causality between the variables, the results are shown in table 2. The results show that St. Anton yearly overnight stays do not Granger cause the yearly overnight stays in the Arlberg Region, neither do the overnight Stays in the Arlberg Region Granger cause the overnight stays in St. Anton. This is shown in table 2 where it can be seen that the p-values for Granger causality are larger than 0.05, which means that we cannot reject the null hypothesis that one variable does not Granger cause the other variable. For instantaneous causality the p-value is smaller than 0.05, this means that the null hypothesis of no instantaneous causality is rejected. This means that there is instantaneous causality between St. Anton and Arlberg, meaning that knowing what St. Anton will do in the future will help in predicting what Arlberg will do. In other words, in the case of instantaneous causality a model using past, current and future values of St. Anton to predict Arlberg has a smaller prediction error than a model only using past and current values.

H0	*F-test*	*Chi-squared*	*p-value*
St.Anton do not Granger-cause Arlberg	0.16824		0.6889
No instantaneous causality between: St.Anton and Arlberg		4.4875	0.03414
Arlberg do not Granger-cause St.Anton	0.011868		0.9151
No instantaneous causality between: Arlberg and St.Anton		4.4875	0.03414

Table 2: Granger Causilty and instantaneous causality

Next, impulse response functions are implemented and impulse response plots are drawn. These plots show the effect of an impulse in one variable on another variable. In this case it shows the effect of an impulse in overnight stays on the guests and an impulse in guests on the overnight stays. The impulse response function was calculated for 5 years ahead, the plots are shown in figures 4 and 5. The x-axis shows the years ahead, with 1 being the first year after the impulse, and the y-axis shows the difference with St. Anton. So the black line shows how the overnight stays in the Arlberg region will react after an impulse in overnight stays in St. Anton relative to the overnight stays in St. Anton. Both graphs show that a shock to the other variable, does not have a big effect to the other variables. Both graphs show that when one variable shows an impulse, the other variable will slightly increase over time.

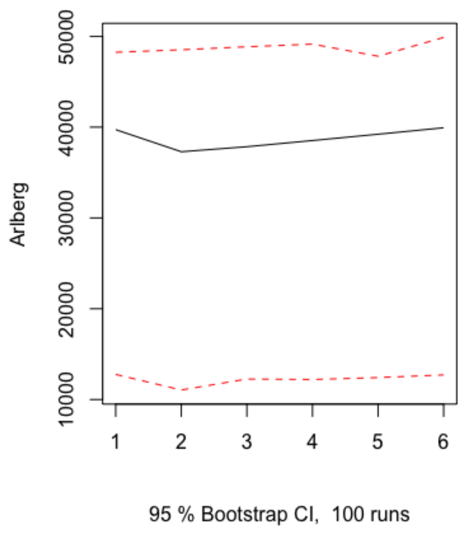

Figure 10: Impulse Response Function of St. Anton to Arlberg

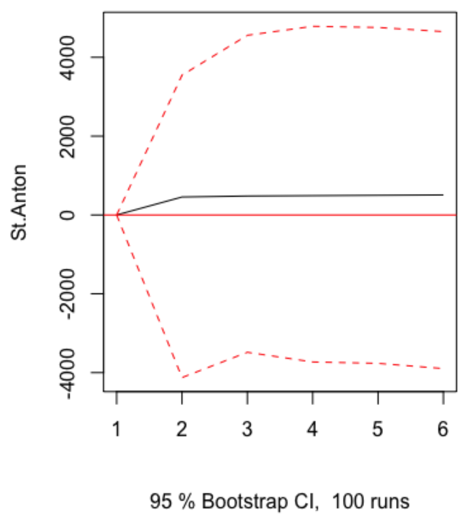

Figure 11: Impulse Response Function of Arlberg to St. Anton

Finally, a forecast can be created. In this research a forecast was estimated for 5 years ahead with a 95% confidence interval. The fan charts in figures 6 and 7 display the forecasts for the 95% confidence interval from year 1, which is 2010, to year 15, which is 2025. Both graphs show that the amount over yearly overnight stays will flatten out in the future with a slight downward trend.

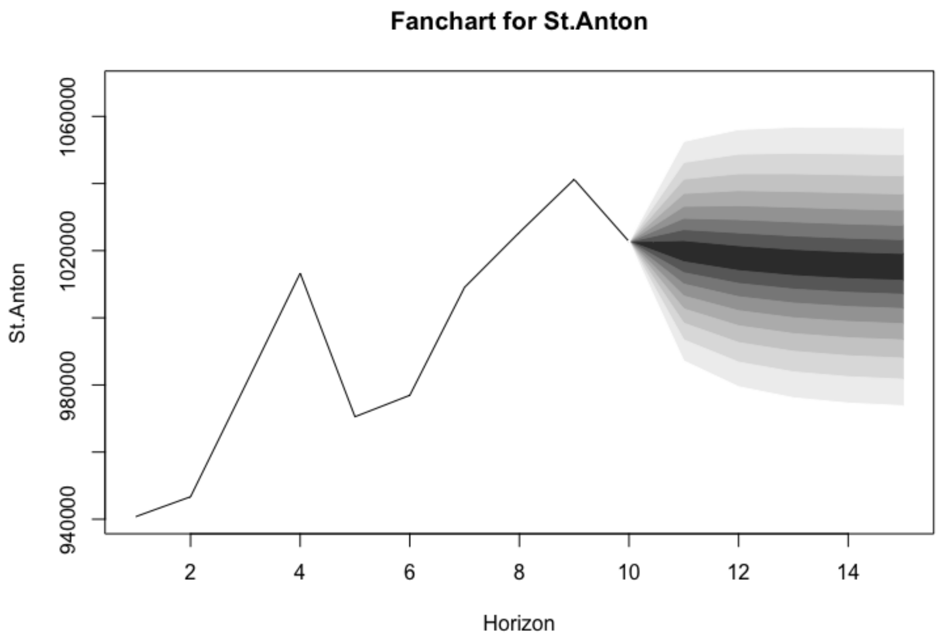

Figure 12: 1st stage forecast for St. Anton

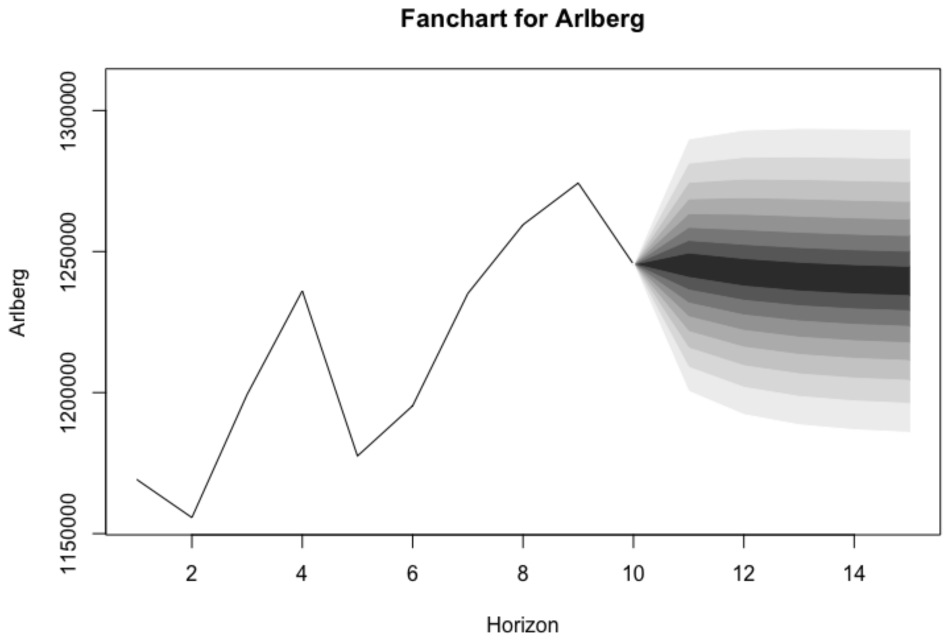

Figure 13: 1st stage forecast for Arlberg

Accommodation category

The second model, looked at different categories of accommodation. The categories are commercial, private, apartment and other. Commercial accommodations are hotels and pensions as well as chalets owned and operated by travel agencies; private accommodations are accommodations that are privately owned for instance apartments that the guests own or locals renting out rooms in their house; and apartments are apartments rented by the guests via travel agencies and websites such as Airbnb.

However, before applying and interpreting our model, again diagnostics should be run to make sure all assumptions are satisfied. The first assumption states that the error terms should be non-correlated with previous periods. This is tested with an asymptotic Portmanteau test, this test had a p-value of less than 0.002, as shown in table 3, which is smaller than 0.05. This means the null hypothesis that the error terms are correlated is rejected. Thus, the first assumption is satisfied.

Test	*Chi-squared*	*p-value*
Portmanteau	6.436	$< 2.2e^{-16}$

Table 3: Portmanteau test

The second assumption is the absence of heteroskedasticity, this can easily be tested with a residual plot. Figure 8 shows the residual plot of this third model, the plot shows no cone or fan shape and thus it can be said that there is no heteroskedasticity.

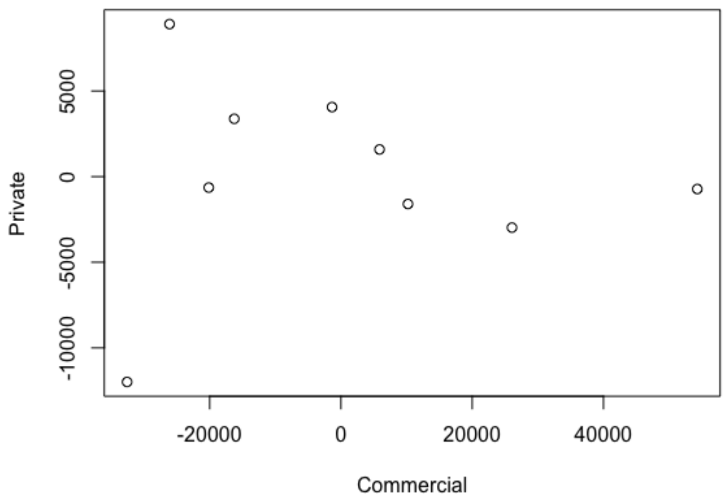

Figure 14: Residual plot for accommodation category model

The final and third assumption to be tested is a normal distribution of the error terms. This is tested with a QQ plot, which is shown in figure 9. The QQ plot does not show a big curvature. The straight diagonal line fits the error terms pretty well. Therefore, the assumption of normally distributed error terms is satisfied.

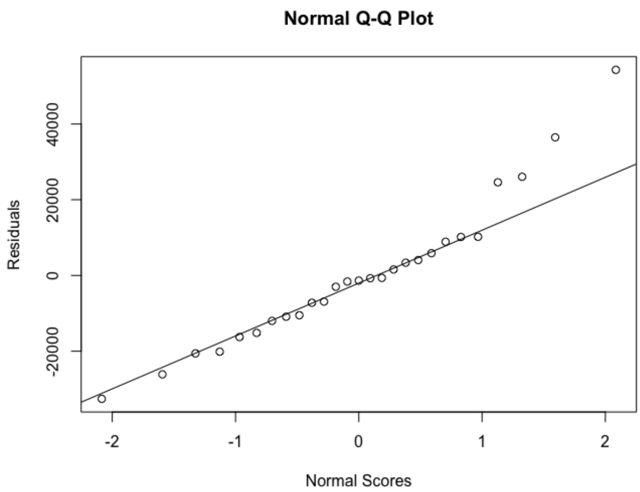

Figure 15: QQ plot for category model error terms

Next, the presence of structural breaks can be tested with a stability test. If structural breaks go unnoticed, the estimations made by the model are no longer accurate, therefore this is visually tested with a plot of the sum of recursive error terms. If the line representing this sum goes outside of the red bounds, a structural break was observed. The plots in figure 10 show that for the time series used in this model, there were no structural breaks.

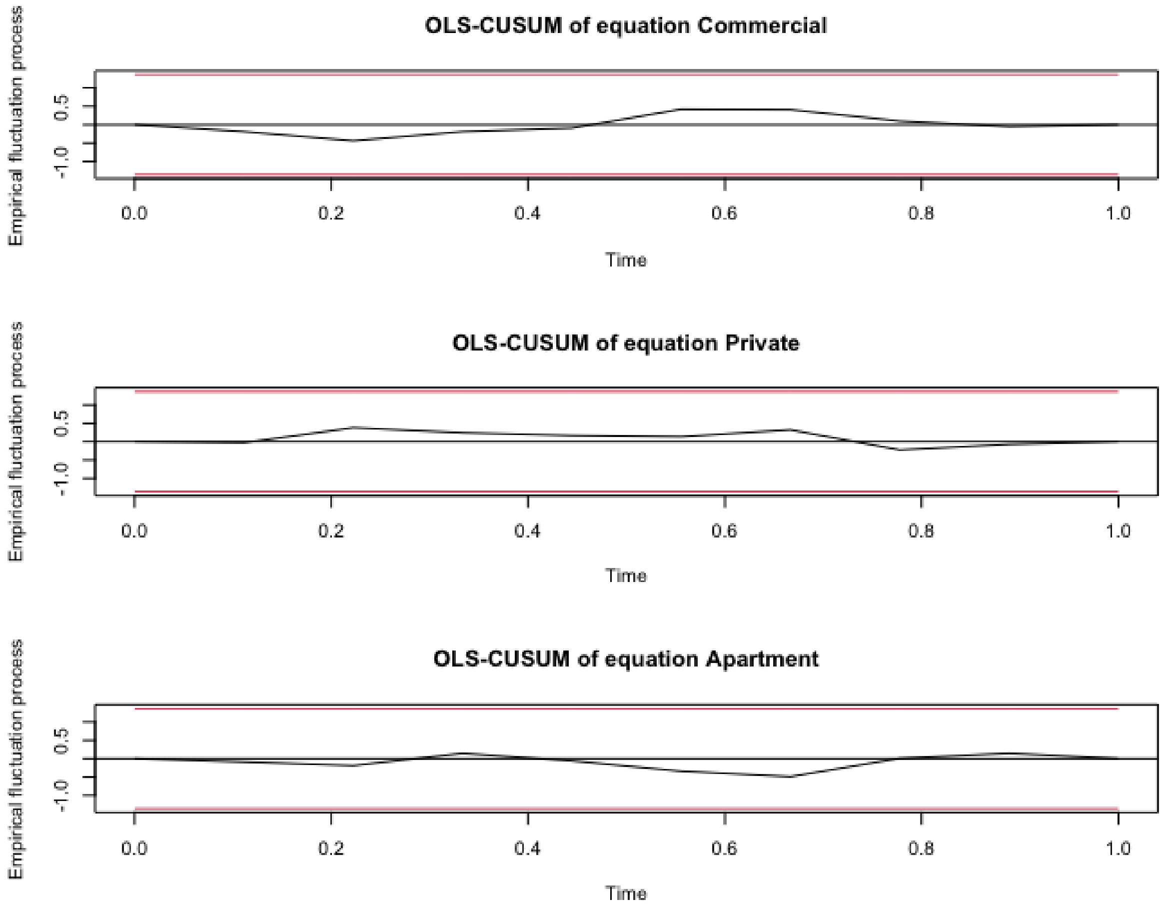

Figure 16: OLS-CUSUM for different categories

Now that all assumptions and stability are tested, changes can be simulated. This is done by testing the Granger causality between the variables. The results in table 4 show that overnight stays in commercial accommodation do not Granger cause the amount of overnight stays in apartments, private accommodation or other accommodation. However, overnight stays in private accommodation Granger cause the overnight stays in commercial accommodation, apartments and other accommodation; overnight stays in private accommodation Granger cause overnight stays in the three other types of accommodation; and overnight stays in other accommodation Granger cause overnight stays in the other categories. For this model there is no instantaneous causality.

H0	*F-test*	*Chi-squared*	*p-value*
Commercial do not Granger-cause Private, Apartment	0.24686		0.7844
No instantaneous causality between: Commercial and Private, Apartment, Other		1.8715	0.3923

Private do not Granger-cause Commercial, Apartment, Other	2.6808		0.1011
No instantaneous causality between: Private and Commercial, Apartment		2.8568	0.2397
Apartment do not Granger-cause Commercial, Private	2.2528		0.1395
No instantaneous causality between: Apartment and Commercial, Private		3.1869	0.2032

Table 4: Granger causality and instantaneous causality

Next, impulse response functions are implemented and impulse response plots are drawn. These plots show the effect of an impulse in one variable on another variable. In this case it shows the effect of an impulse in overnight stays in one category on the overnight stays in other categories over the 6 following years. The impulse response function was calculated for 5 years ahead, the plots are shown in figure 12. The x-axis shows the years ahead, with 1 being the first year after the impulse, and the y-axis shows the difference with St. Anton. So the black line shows how the overnight stays in the Arlberg region will react after an impulse in overnight stays in St. Anton relative to the overnight stays in St. Anton. The first graph shows that a shock to overnight stays in commercial accommodations will result in a slightly higher number of overnight stays in private accommodations. The second graph shows that a shock to commercial accommodations will slowly increase the overnight stays in apartments in the following three years. Next, it shows that an impulse in overnight stays in private apartments does not really have an effect on overnight stays in apartments. Thirdly, the graphs shows that an impulse in overnight stays in apartments leads to a decrease in overnight stays in private accommodations. The last graph shows that an impulse in apartments will initially decrease the overnight stays in commercial accommodations, but after about two years the overnight stays in commercial accommodations will increase again after which it will fluctuate around its initial level. This shows that in general when the overnight stays in one category increase, they will decrease in another category. This could mean that guests that stayed in one type of accommodation the one year, stay in a different type of accommodation the next year.

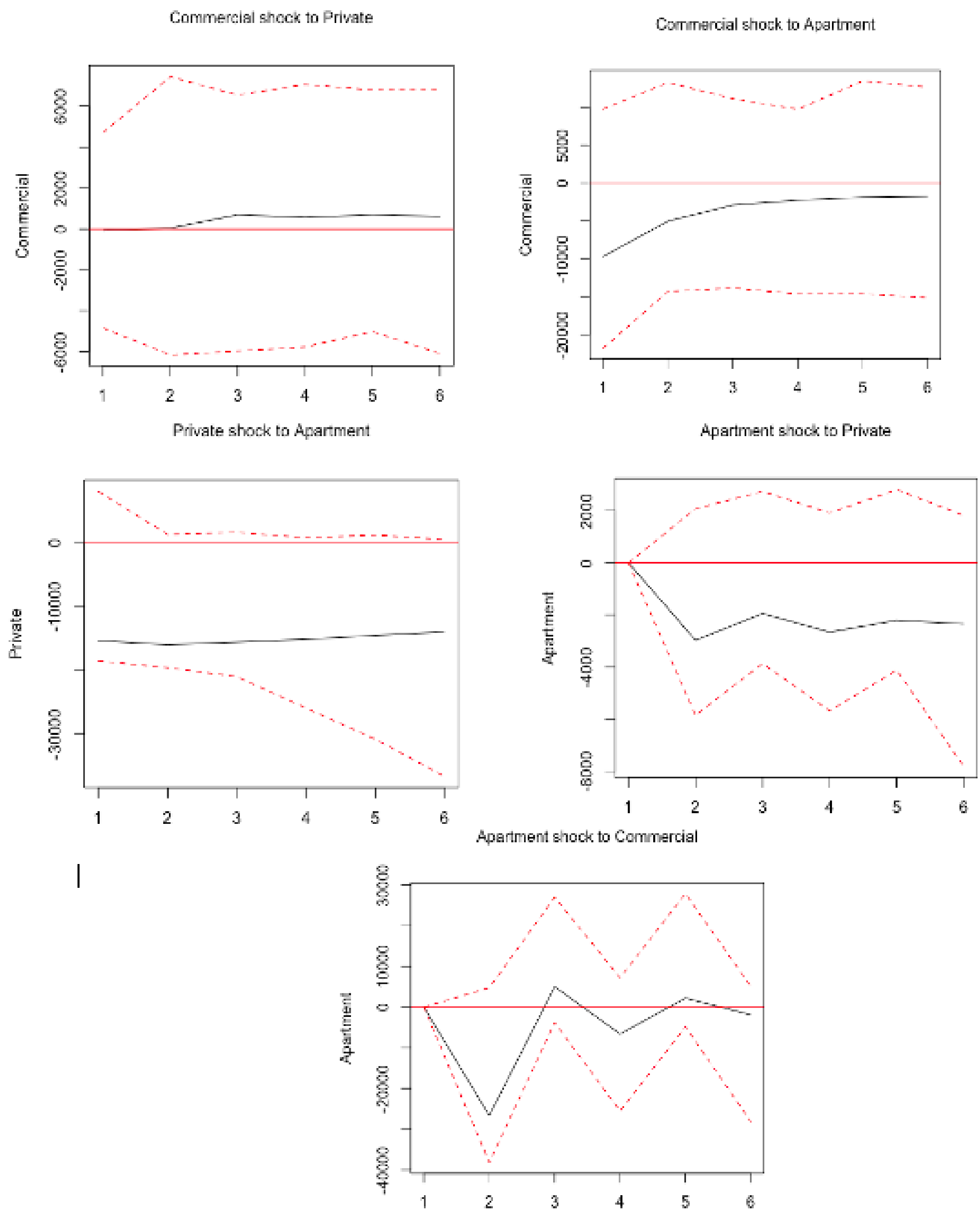

Figure 17: Impulse response function for Category model

Finally, a forecast can be created. In this research a forecast was estimated for 5 years ahead with a 95% confidence interval. The fan charts display the forecasts for the 95% confidence interval, from year 1, which is 2010, to year 15, which is 2025. These fan charts are shown in figures 13, 14 and 15. For apartments, the forecast shows a steep increase continuing the upward trend of the previous years. For private accommodations, the forecast shows a decrease in the upcoming 5 years which continues the downward trend of the previous years. For commercial accommodations, the forecast shows fluctuation between the years but an overall constant trend after a slight increase in the first couple of years. This is similar to the trend in the previous years.

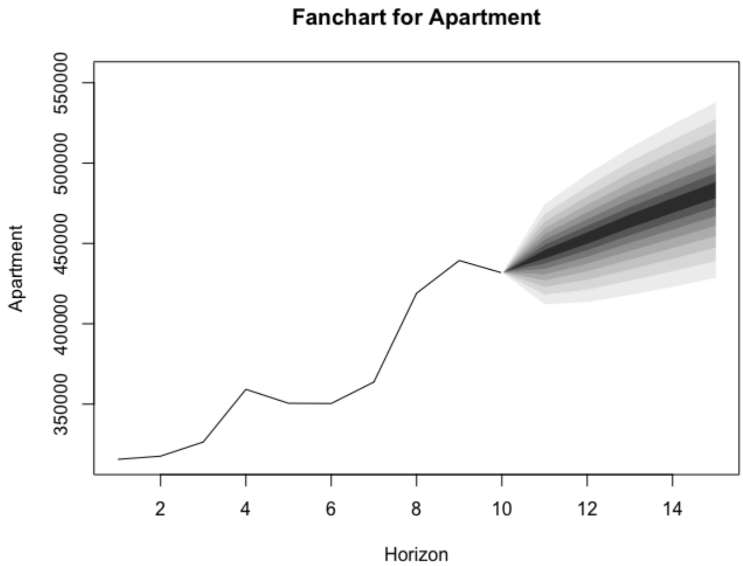

Figure 18: 1st stage forecast for category Apartment

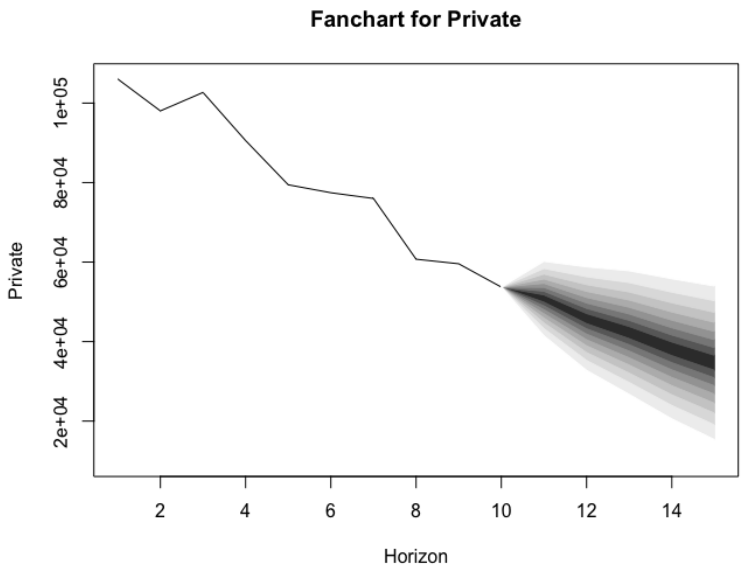

Figure 19: 1st stage forecast for category Private

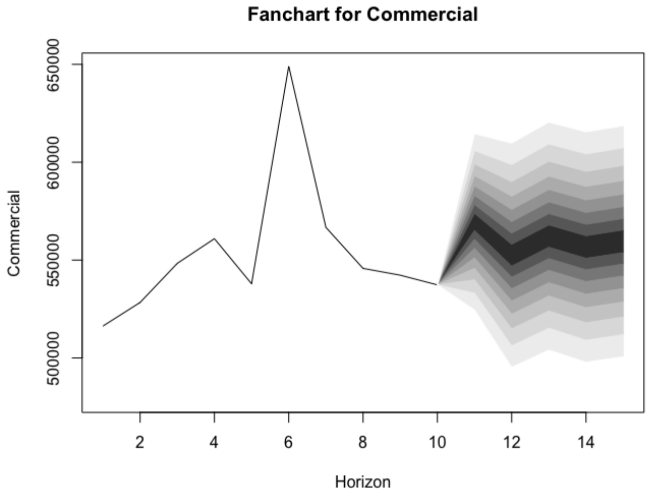

Figure 20: 1st stage forecast for category Commercial

Overnight stays versus amount of guests

The third model ran a typical unrestricted VAR on a time series containing information on the number of guests and the number of overnight stays per month per season in St. Anton. Based on the AIC test this model had 1 lag. Before the model could be applied to make a forecast, the assumptions had to be tested.

The first assumption states that the error terms should be non-correlated with previous periods. This is tested with an asymptotic Portmanteau test, this test had a p-value smaller than 0.002 (table 3) which is smaller than 0.05 which means the null hypothesis that the error terms are correlated is rejected. Thus, the first assumption is satisfied.

Test	*Chi-squared*	*p-value*
Portmanteau	1.1417	$<2.2e^{-16}$

Table 3: Model Diagnostics

The second assumption is the absence of heteroskedasticity, this can easily be tested with a residual plot. Figure 15 shows the residual plot of this first model, this plot also shows no heteroskedasticity.

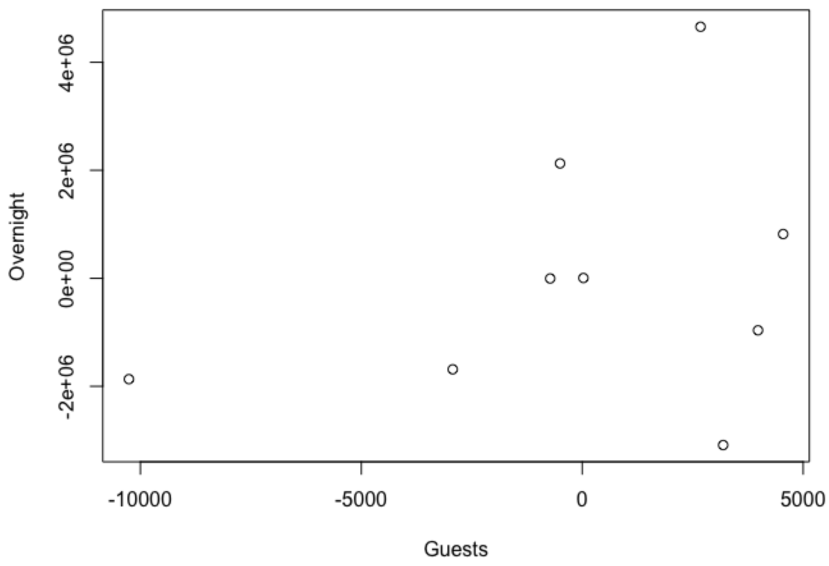

Figure 21: Residual plot for overnight stays

The final and third assumption to be tested is a normal distribution of the error terms. This is tested with a QQ plot, which is shown in figure 17. The QQ plot does not show a big curvature, but it does show some outliers. After removing these outliers, the straight diagonal line fits the error terms pretty well. Therefore, the assumption of normally distributed error terms is satisfied.

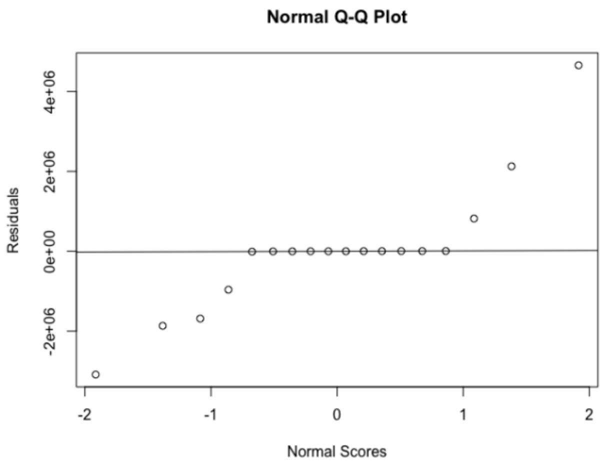

Figure 22: QQ plot for overnight stays error terms

Next, the presence of structural breaks can be tested with a stability test. If structural breaks go unnoticed, the estimations made by the model are no longer accurate, therefore this is visually tested with a plot of the sum of recursive error terms. If the line representing this sum goes outside of the red bounds, a structural break was observed. The plots are shown in figure 18 and show that for the time series used in this model, there were no structural breaks.

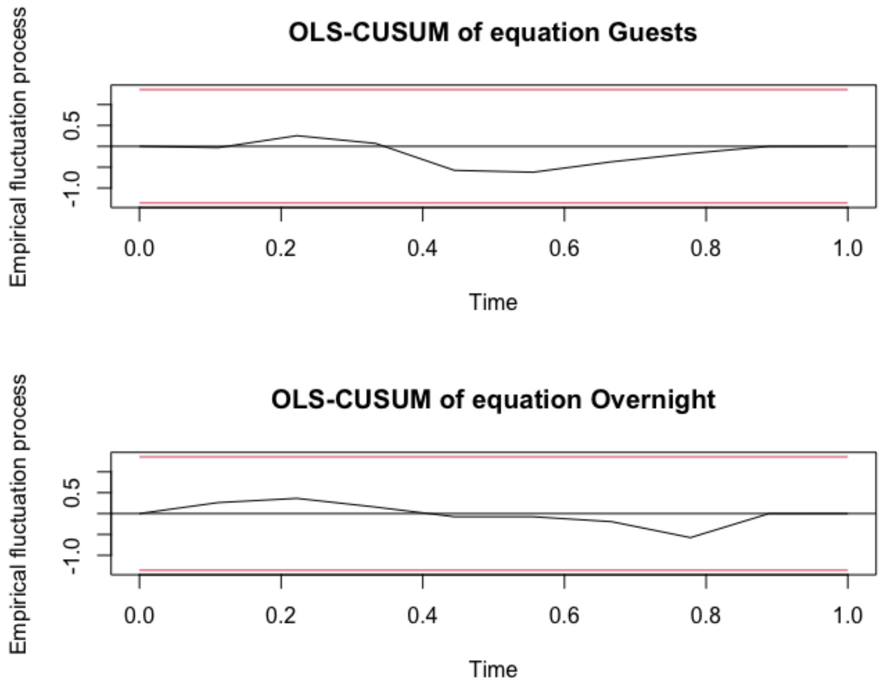

Figure 23: OLS-CUSUM for overnight stays and number of guests

Now that all assumptions and stability are tested, changes can be simulated. This is done by testing the Granger causality between the variables. The results in table 4 show that overnight stays Granger cause the amount of guests in St. Anton, since the p-value is equal to 0.05 which means we reject the null hypothesis the there is no Granger causality. But the guests in St. Anton do not Granger cause the overnight stays in St. Anton, this p-value is larger than 0.05 which means we cannot reject the null hypothesis. For instantaneous causality the p-value is larger than 0.05, this means that the null hypothesis of nu instantaneous causality cannot be rejected. This means that there is no instantaneous causality between Overnight and Guests,. Instantaneous causality means that knowing what Overnight will do in the future will help in predicting what Guests will do. In other words, in the case of instantaneous causality a model using past, current and future values of Overnight to predict Guests has a smaller prediction error than a model only using past and current values.

H0	*F-test*	*Chi-squared*	*p-value*
Overnight do not Granger-cause Guests	0.46997		0.05099
No instantaneous causality between: Overnight and Guests		0.73315	0.3919
Guests do not Granger-cause Overnight	0.061164		0.8088

No instantaneous causality between: Overnight and Guests	0.73315	0.3919

Table 5: Granger Causilty and instantaneous causality

Next, impulse response functions are implemented and impulse response plots are drawn. These plots show the effect of an impulse in one variable on another variable. In this case it shows the effect of an impulse in overnight stays on the guests and an impulse in guests on the overnight stays. The impulse response function was calculated for 5 years ahead, the plots are shown in figures 19 and 20. The x-axis shows the years ahead, with 1 being the first year after the impulse, and the y-axis shows the difference with St. Anton. So the black line shows how the overnight stays in the Arlberg region will react after an impulse in overnight stays in St. Anton relative to the overnight stays in St. Anton. The first graph shows that a shock to the overnight stays in St. Anton will result in a lower number of guests initially, but it also shows that almost immediately the amount of guests will return to the level before the shock. Translating this to tourism in St. Anton gives that this could mean that at the moment of the shock the increase in overnight stays does not come from more guests, but from fewer guests staying for a longer period. However, quite quickly the amount of guests will recover suggesting the despite some people staying in St, Anton for longer, there are also more guest in total coming to St. Anton. In the second graph it can be seen that a shock to the guests stays in the Arlberg region will lead to a decrease in the overnight stays, this decrease will last for a long time with the amount of guests eventually being the same as the amount of overnight stays. This could mean that instead of more people spending the same amount of overnight stays in St. Anton, there are more people coming to St. Anton but fewer people staying overnight in St. Anton. This could be that these guests have an accommodation outside of St. Anton or visit St. Anton for a shorter period, for instance 4 days instead of a week.

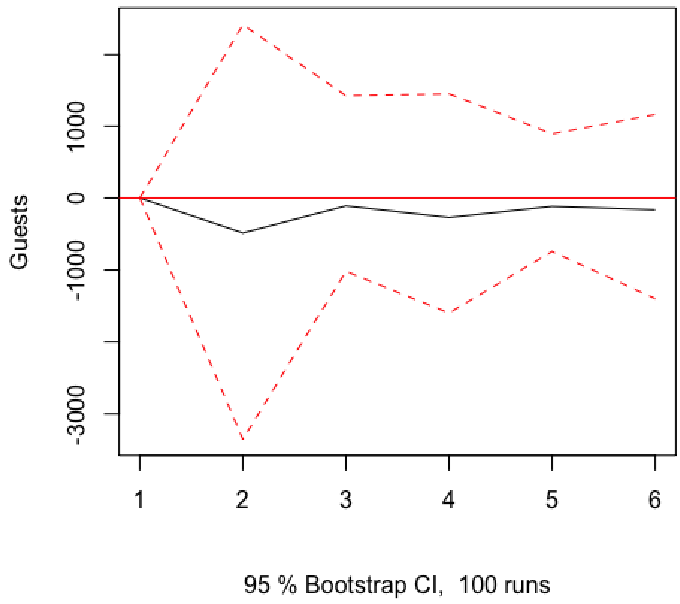

Figure 24: Impulse Response Function for Overnight to Guests

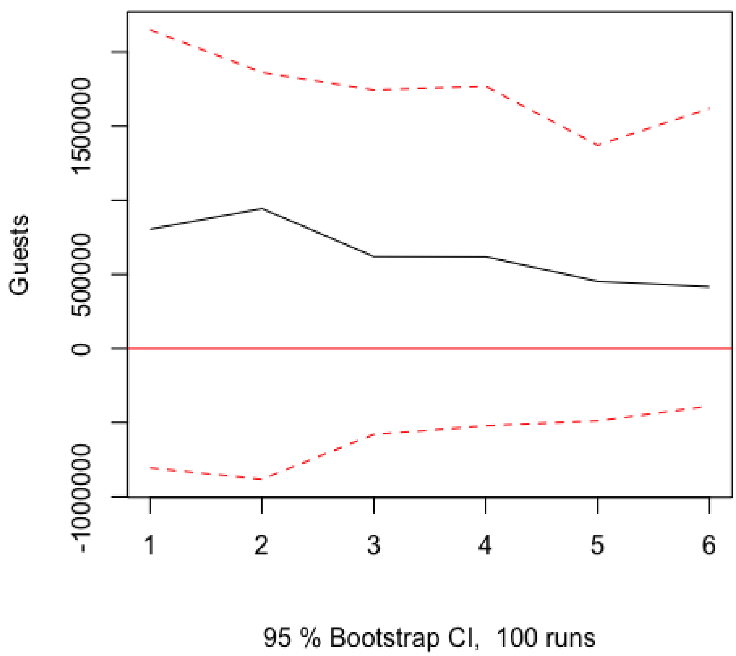

Figure 25: Impulse Response Function for Guests to Overnight

Finally, a forecast can be created. In this research a forecast was estimated for 5 years ahead with a 95% confidence interval. The fan charts display the forecasts for the 95% confidence interval, from year 1, which is 2010, to year 15, which is 2025. The fan charts are displayed in figures 21 and 22. The first chart, is the forecast for the amount of guests, this charts shows that the amount of guest will stay the same for the next 5 years, which continues the constant trend seen from year 9 onwards. The second chart shows the forecast for the number of overnight stays. This chart shows that after the overnight stays decreased in year 10 (2019), it will increase again after which the amount of overnight stays will fluctuate between the years. Overall this forecast shows a constant trend.

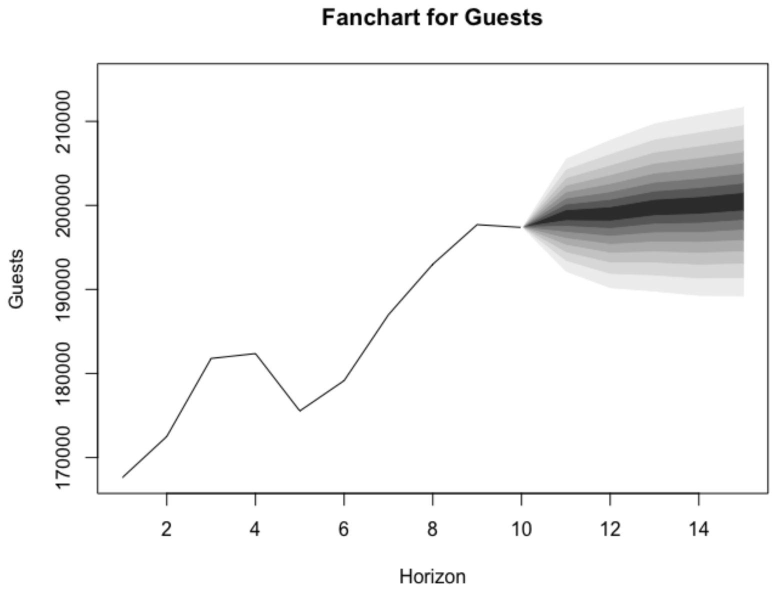

Figure 26: 1st stage forecast for Guests

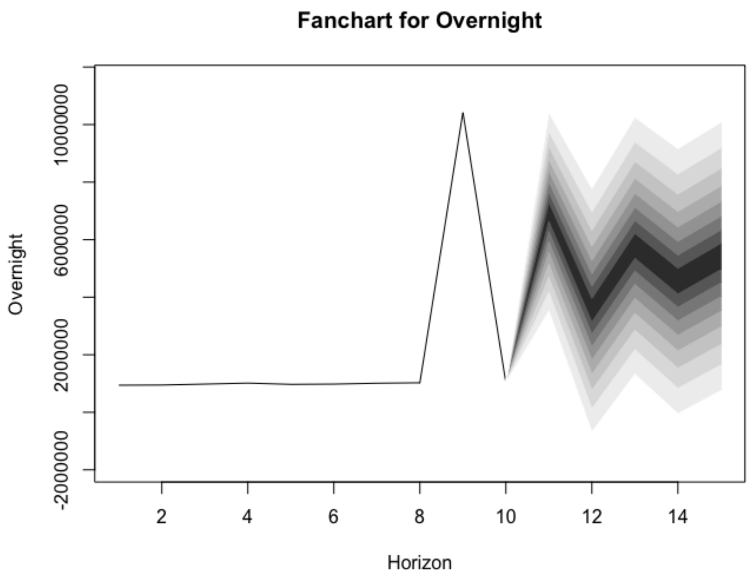

Figure 27: 1st stage forecast for Overnight stays

For completeness these models were also calculated with more lags than 1. However, adding lags lead to a higher AIC, BIC and HQ. Moreover, with more lags the models did not satisfy the assumptions needed for a statistically interpretable model.

Generalized Additive Model

Next, the results from the different GAM models that were implemented will be discussed. For each subset multiple models were created, each with different independent variables. Then for each subset the model with the highest statistical inference was chosen, this decision was based on adjusted R-squared and explained deviance. Then forecasts were made with each model for the following 10 years.

St. Anton and the Arlberg region

Firstly, the overnight stays in St. Anton and the Arlberg region were analyzed and forecasted with the GAM model. For both St. Anton and the Arlberg region the independent variable was the first lag of the dependent variable, the overnight stays in respectively St. Anton and the Arlberg region. The coefficient of first lag of St. Anton is significant at 95% confidence level, with the p-value of the lag smaller than 0.05, and the coefficient of the first lag for the Arlberg region is significant for the 90% confidence level, with the p-value for the lag being smaller than 0.1. This means that both models are statistically significant and can be used for forecasting.

	Parametric Coefficient	*Std. Error*	*t-value*	*Pr(>|t|)*
Constant	998418	7882	126.7	5.04e-13
Approximate significance of smooth terms	**edf**	**Ref. df**	**F**	**P-value**
Lag(Commercial)	1	1	6.853	0.0345
Adjusted R-squared	0.423	**Deviance Explained**	49.5%	

Table 6: Summary of GAM for St. Anton

	Parametric Coefficient	*Std. Error*	*t-value*	*Pr(>|t|)*
Constant	1219819	10966	111.2	1.25e-12
Approximate significance of smooth terms	**edf**	**Ref. df**	**F**	**P-value**
First lag	1	1	4.668	0.0676

Adjusted R-squared	0.314	**Deviance Explained**	40%

Table 7: Summary of GAM for Arlberg

Next, both models were used to make a forecast for the next 10 years based on the data without the COVID affected 2019-2020 season. Both forecasts show a slight decrease in year 5, after which the overnight stays will increase. Overall the trend is increasing. With the forecast predicting up to 1.03 million overnight stays in St. Anton in year 10, which is 2030. And a forecast of over 1.25 million overnight stays in the entire Arlberg region 2030.

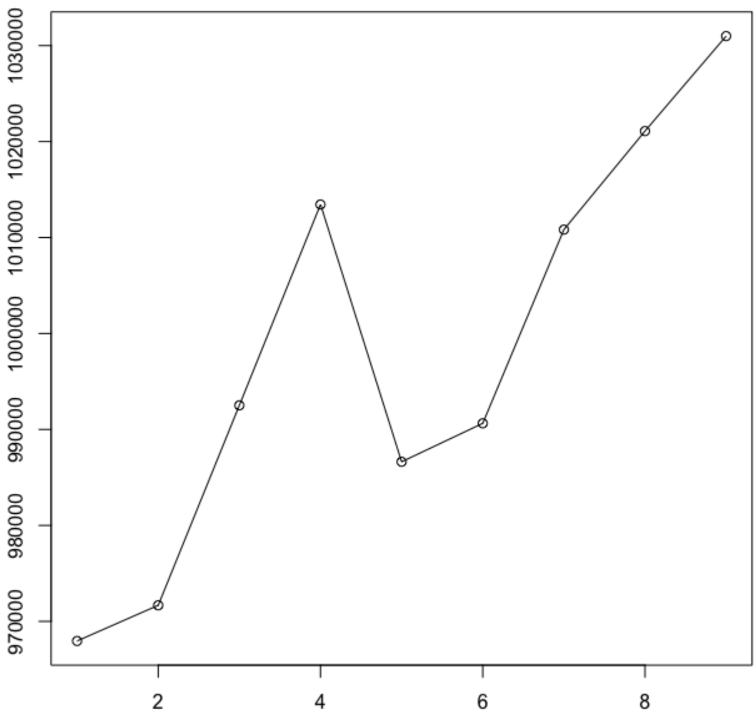

Figure 28: Forecast for St. Anton

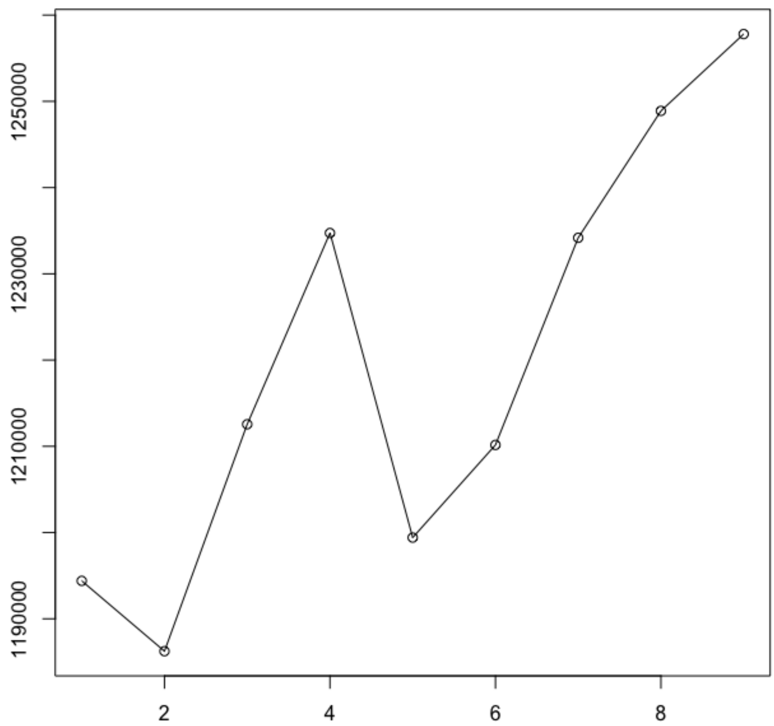

Figure 29: Forecast for Arlberg region

Accommodation categories

Next, the GAM models for the different accommodation categories were created. For each category, Commercial, Private and Apartments, a model was created. Each model contained the lag of that concerned category as well as the current values of the other two categories as the independent variables. These models were also created using lags of the other categories, but these had a lower adjusted R-squared and explained deviance. For Commercial accommodation and Private accommodation, the models including the other two categories had the highest statistical significance, for apartments the model with only the lag of the overnight stays in apartments had the highest significance, as measured by adjusted R-squared.

The GAM model for commercial accommodations, showed that none of the included independent variables are statistically significant, meaning that they cannot be interpreted separately. The adjusted R-squared is below zero, meaning that this model has no explanatory power. So, with this model we cannot say anything about the effects of the lag of the overnight stays in commercial accommodation on its future values.

	Parametric coefficient	*Std. Error*	*t-value*	*Pr(>|t|)*

Constant	557395	12798	43.55	3.17e-07
Approximate Significance of smooth terms	**edf**	**Ref. df**	**F**	**p-value**
Lag(Commercial)	1	1	0.220	0.659
Private	1	1	0.994	0.365
Apartment	1.374	1.608	0.632	0.418
Adjusted R-Squared	-0.117	**Deviance Explained**	35.4%	

Table 8: GAM model for Commercial Accommodations

The second model is for overnight stays in private accommodations. Here not only the lag of the overnight stays in private accommodations is included as an independent variable, but also the current values for overnight stays in commercial accommodations and apartments. All the independent variables in this model are statistically significant, moreover this model has an adjusted R-squared of 0.969. This means that predictions from this model should be very accurate.

	Parametric coefficient	*Std. Error*	*t-value*	*Pr(>|t|)*
Constant	77574	1015	76.41	3.54e-06
Approximate Significance of smooth terms	**edf**	**Ref. df**	**F**	**p-value**
Lag(Private)	1	1	7.530	0.0711
Commercial	1.944	1.996	5.866	0.0970
Apartment	1.959	1.997	14.410	0.0304
Adjusted R-Squared	0.969	**Deviance Explained**	98.8%	

Table 9: GAM model for Private Accommodations

The third model created was for the overnight stays in apartments. For this model only the lag of the overnight stays in apartments was used as an independent variable as this gave the highest adjusted R-squared. This model has an adjusted R-squared of 0.766, meaning that 76.6% of the variance is explained by this model.

	Parametric coefficient	*Std. Error*	*t-value*	*Pr(>|t|)*
Constant	373093	7328	50.91	2.95e-10
Approximate Significance of smooth terms	**edf**	**Ref. df**	**F**	**p-value**
Lag(Apartment)	1	1	27.12	0.00124
Adjusted R-Squared	0.766	Deviance Explained	79.5%	

Table 10: GAM model for Apartments

Next, these models were used to predict the overnight stays per accommodation category for the next 10 years. Because of its low adjusted R-squared, the model for Commercial accommodations was not used for predictions. As the model described above was the model with the highest R-squared, it has to be concluded that this dataset contains too little information to predict the overnight stays in Commercial accommodations.

However, for Private accommodations and Apartments forecasts were made. For private accommodations the model predicts a strong decrease, with a small plateau in years 4 to 6. In 2020 the forecasted overnight stays are at 100 000, but in 2030 this will only be 60 000. So over the next 10 years this will decrease with 40%.

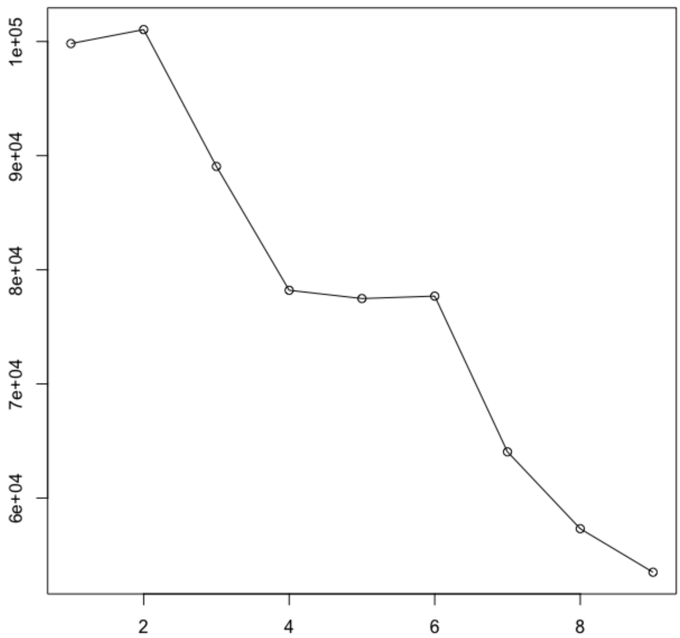

Figure 30: Forecast Private Accommodations

For Apartments, the forecast shows a strong increase. With the forecast line almost going straight up from year 6 on. The overnight stays in apartments start at only 340 000 but end at 440 000, an increase of almost 30%. Here, too, a small plateau with a minimal decrease can be seen from year 4 to 6. This leads to guessing that the loss in overnight stays in Private accommodation might be made up by the increase in overnight stays in Apartments.

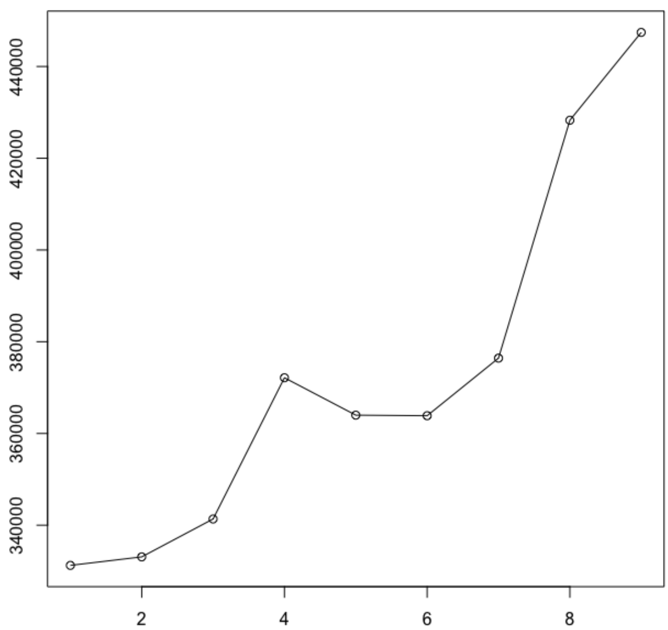

Figure 31: Forecast Apartments

Overnight stays vs. guests

Finally, a model was created for the overnight stays and the amount of guests in St. Anton am Arlberg. The first model is for the number of guests in St. Anton. This model includes the current value for Overnight stays as well. The model is statistically significant and has an adjusted R-squared of 0.918, which is very high.

	Parametric coefficient	*Std. Error*	*t-value*	*Pr(>|t|)*
Constant	185159.4	879.7	210.5	1.48e-09
Approximate Significance of Smooth terms	***edf***	***Ref. df***	***F***	***p-value***
Lag(Guests)	1.855	1.978	3.753	0.1144
Overnight	1.975	1.999	9.140	0.0321
Adjusted R-Squared	0.918	**Explained Deviance**	95.7%	

Table 11: GAM model for number of guests

The second model is for the overnight stays in St. Anton. This model includes both a lag of the overnight stays and the current value of the amount of guests. In this model, both independent variables are statistically significant, however the adjusted R-squared is only 0.546 meaning that only 54.6% of the variance in the overnight stays is explained by the model.

	Parametric coefficient	*Std. Error*	*t-value*	*Pr(>|t|)*
Constant	2040485	705316	2.893	0.0324
Approximate Significance of Smooth terms	**edf**	**Ref. df**	**F**	**p-value**
Lag(Guests)	1.782	1.952	6.146	0.0500
Overnight	1	1	5.406	0.0676
Adjusted R-Squared	0.546	**Explained Deviance**	70.4%	

Table 12: GAM model for number of overnight stays

Next, based on these two GAM models the number of guests and overnight stays were forecasted for the next 10 years. Figure … show that the number of guests is expected to increase, with a small dip in year 3. The overall trend is increasing and the number of expected guests in 2030 is 195 000, this is 11% higher than the number of guests in 2020.

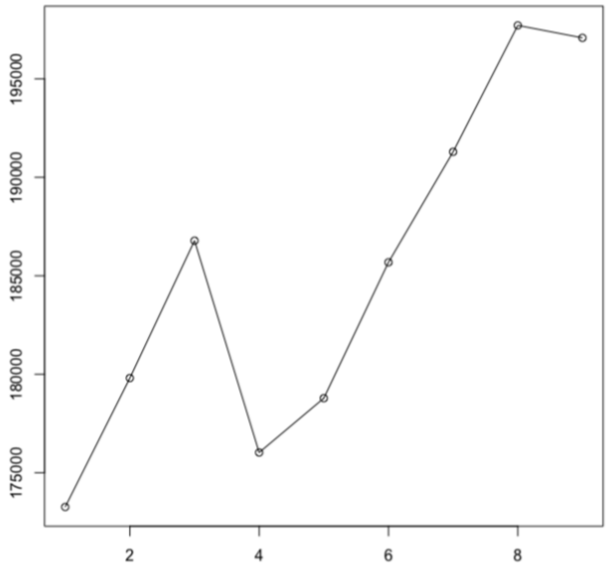

Figure 32: Forecast for number of guests

For the overnight guests, the expected number is expected to roughly stay the same. Between years a fluctuation is visible, with a big peak at year 8, but overall the number of overnight stays will stay the same at around 1 million overnight stays.

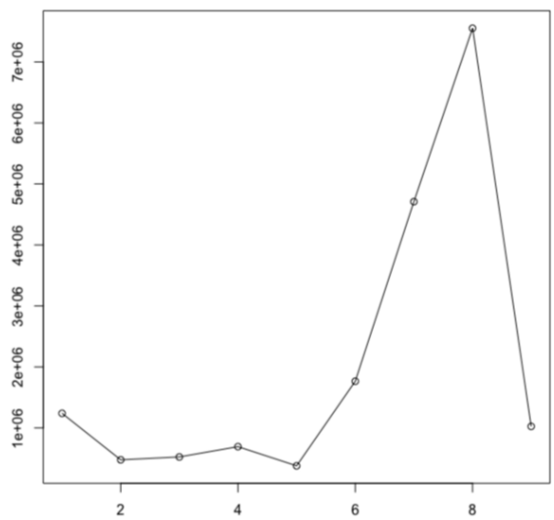

Figure 33: Forecast for overnight stays

The COVID-19 season

In the second stage of this research, the time series including the 2019-2020 season will be compared to the forecasts made in the first stage. The 2020-2021 season was left out, since the ski resort was not open for tourists during the season. In order to compare the current situation with COVID to the forecast without COVID, the time series with COVID were plotted. These plots will then be compared to the fan charts displaying the forecasts made in the first stage.

St. Anton and the Arlberg Region

The first model compared the overnight stays in St. Anton with the overnight stays in the entire Arlberg region. Figures 34 to 36 give the fan charts and forecasts for St. Anton and the Arlberg region, red dots in the fan charts reflect the observed overnight stays in the different categories during the 2019-2020 season. Figure … shows that in the 2019-2020 season the total number overnight stays in St. Anton dropped to 800,000, while the season before the total number of overnight stays was almost 1,050,000. The overnight stays in the Arlberg region went from 1,260,000 to 1,000,000. Both forecasts made with VAR in the first stage show a slight downward trend in the number of overnight stays, for St. Anton the forecast is between 1,060,000 and 980,000 and for the Arlberg region the forecast is between 1,300,000 and 1,200,000. This shows that even in the worst case scenario but without COVID, the forecasted numbers are higher than in the situation with COVID. The GAM model, shows a different picture with both forecasts going up, albeit with a slight plateau for the year 5. This plateau is where the fan chart forecasts stop and both fan chart show a plateau at the end. For the GAM forecasts, the actual numbers of the COVID season are below what was forecasted, just like with the VAR forecasts.

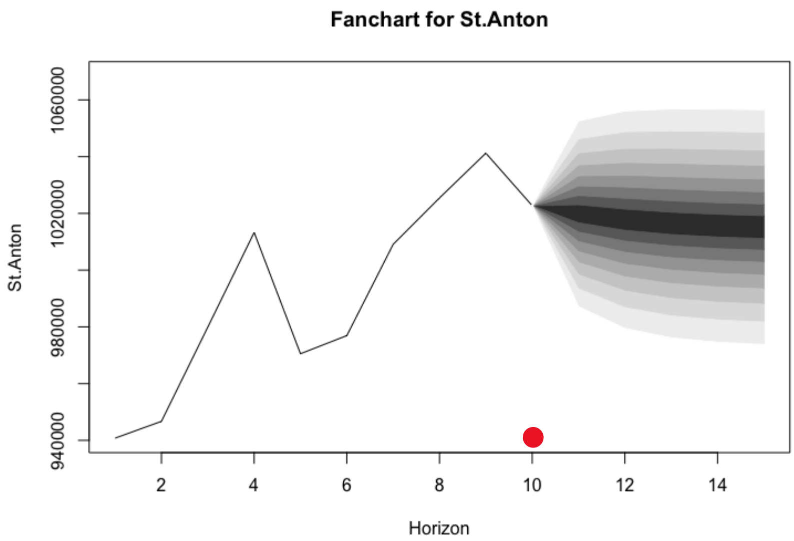

Figure 34: VAR Forecast for overnight stays in St. Anton am Arlberg

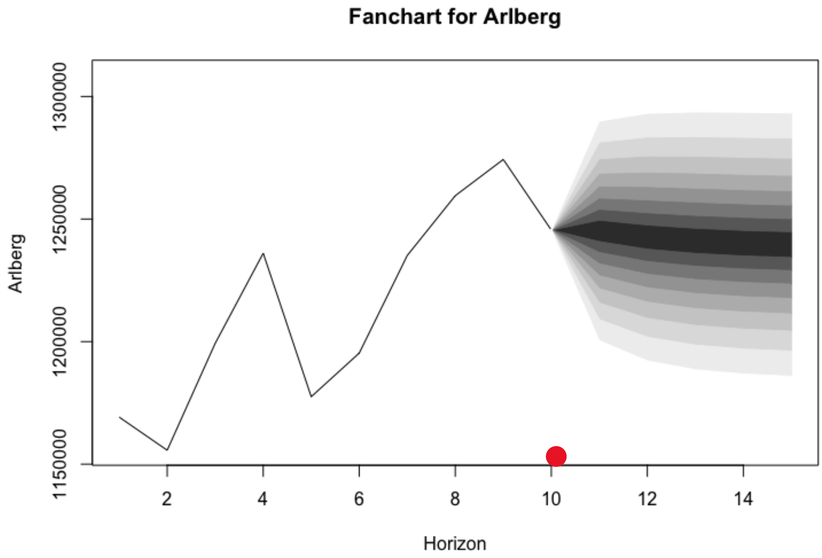

Figure 35: VAR Forecast for overnight stays in the Arlberg region

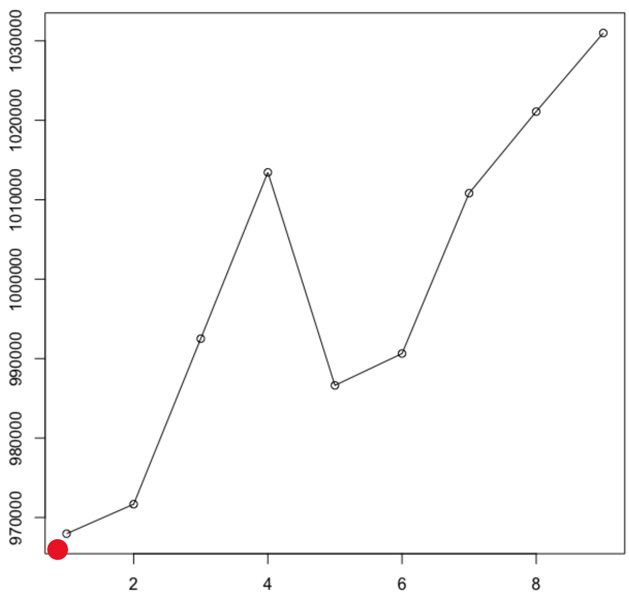

Figure 36: GAM Forecast for St. Anton

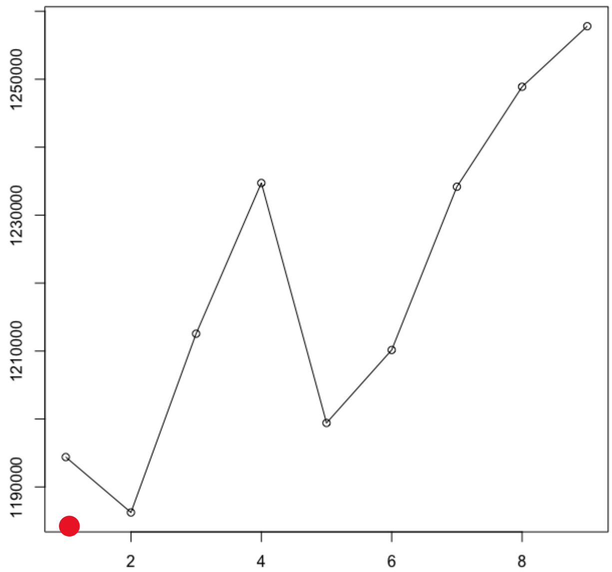

Figure 37: GAM Forecast for Arlberg region

Accommodation categories

The second model compared the overnight stays in different accommodation categories in St. Anton. Here, too, a big drop in the amount of overnight stays can be seen in the 2019-2020 season. This is the result of the lockdown restrictions put into place in February and early March due to the pandemic. For commercial accommodations, the forecast without COVID shows a range between 500,000 and 600,000 for the 5 seasons after 2018-2019. The red dots in the fan charts, reflect the observed overnight stays in the different categories during the 2019-2020 season. The graph for commercial shows that the overnight stays dropped below 450,000. For private accommodations, there is a downward trend even without COVID, as can be seen in figure 39, with the forecast expecting somewhere between 20,000 and 60,000 overnight stays in private accommodations. Due to COVID the overnight stays dropped to 50,000 which is within the forecasted range. For apartments, the forecasted range is between 400,000 and more than 450,000 overnight stays in 2020-2021, with a strong upward trend to 450,000 to 550,000 forecasted overnight stays in the 2024-2025 season. The observed amount of overnight stays in 2019-2020 was only 320,000. So, for apartments the 2019-2020 season was well below the forecasted amount.

For commercial accommodations, the GAM model was not used to make a forecast as the model was not significant. However, for the other two categories a model was created and used to create a forecast. The trends in the GAM forecasts match the trends in the VAR forecasts with a downward trend for Private accommodations and an upward trend in apartments. The same goes for the observed values for the 2020 season. For private accommodations, the observed value is very much below the forecasted value, but for apartments the observation and forecast are quite similar.

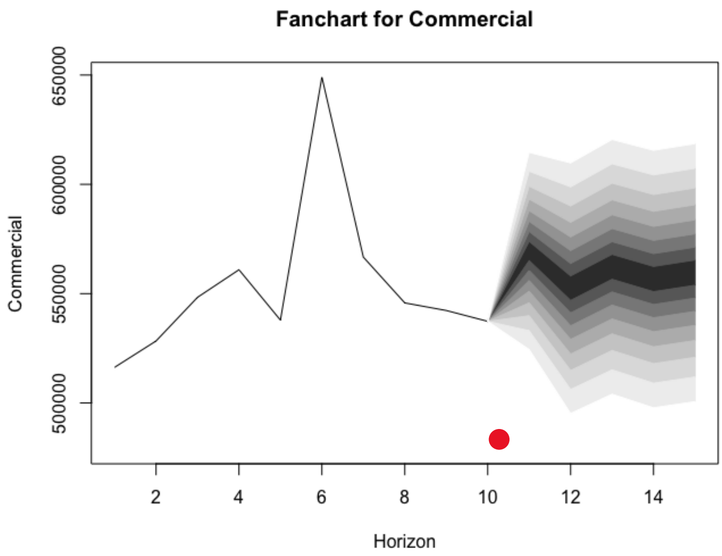

Figure 38: VAR Forecast for overnight stays in commercial accommodations

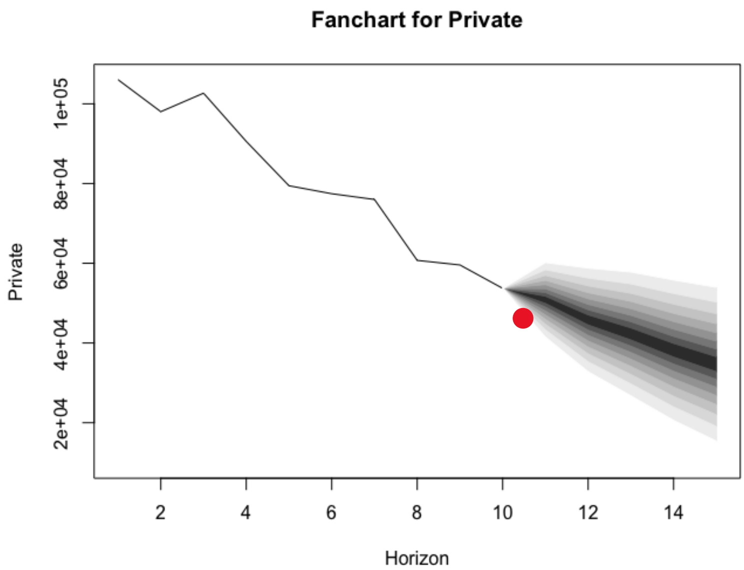

Figure 39: VAR Forecast for overnights stays in private accommodations

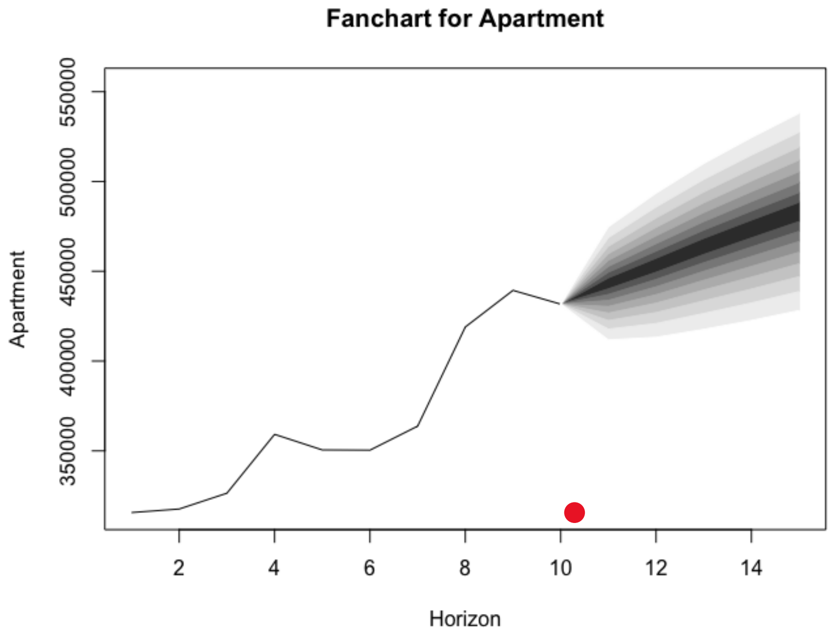

Figure 40: VAR Forecast for overnights stays in apartments

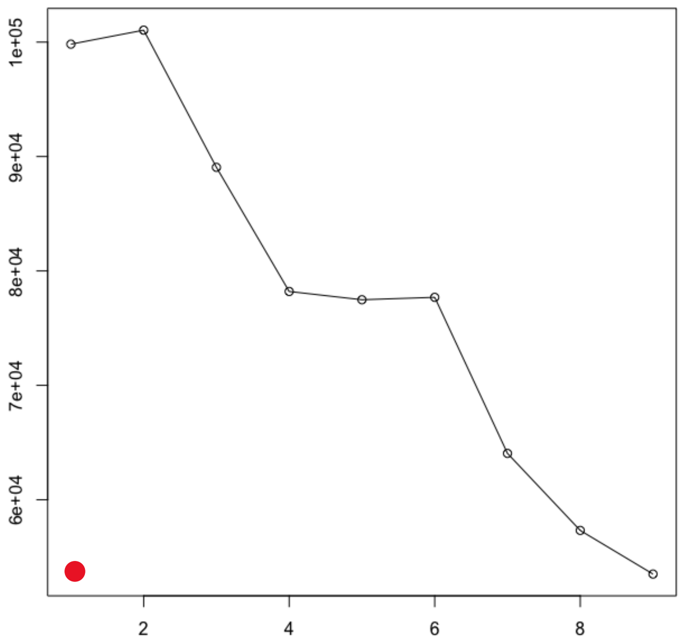

Figure 41: GAM Forecast Private Accommodations

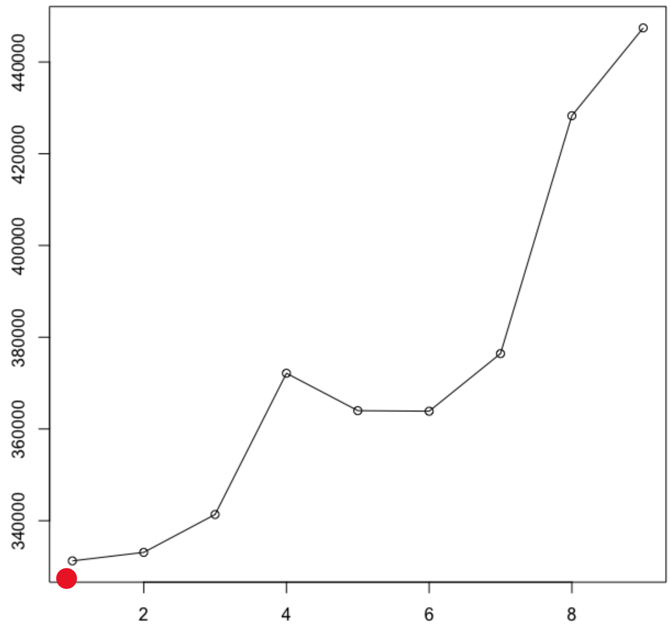

Figure 42: GAM Forecast Apartments

Guests and Overnight stays

The third model compared the number of guests with the number of overnight stays in St. Anton am Arlberg. Figures 43 to 46 show the fan charts and line charts for the forecasted overnight stays and the number of guests based on 2010 until 2019, the red dots in the graphs reflect the values for the 2019-2020 season. The forecast for the number of guests ranges between 190,000 and 210,00 for each year until the 2024-2025 season. During the 2019-2020 season the number of guests dropped to 150,000, which is at the bottom of the graph. It can be seen that this is well below the time series from 2010 till 2019 and the forecast from 2020 till 2025. This means that due to the lockdown imposed to counter the COVID-19 pandemic, the amount of guests in St. Anton diverged from the previous years and the forecast based on those years. For the amount of overnight stays the forecast ranges from 2 million to 10 million, with large fluctuations between the seasons. In 2019-2020 a total of 2 million overnights stays were observed in St. Anton. So, this is just within the forecasted range.

The GAM forecasts for the amount of guests differs from the VAR forecast. With GAM there is a strong upward trend, a dip at year 5 followed by a strong upward trend again. Whereas the VAR forecast shows just a slight upward trend. Moreover, the predicted value for 2020 according to GAM is quite close to the observed value, where is with VAR the observed value is well below the predicted value. For overnight stays the two predictions are quite similar, with both predictions showing some fluctuation between years until it ends with an upward trend. Also, the observed value versus predicted value is similar for both of them, with the observed value being almost the same as the predicted value.

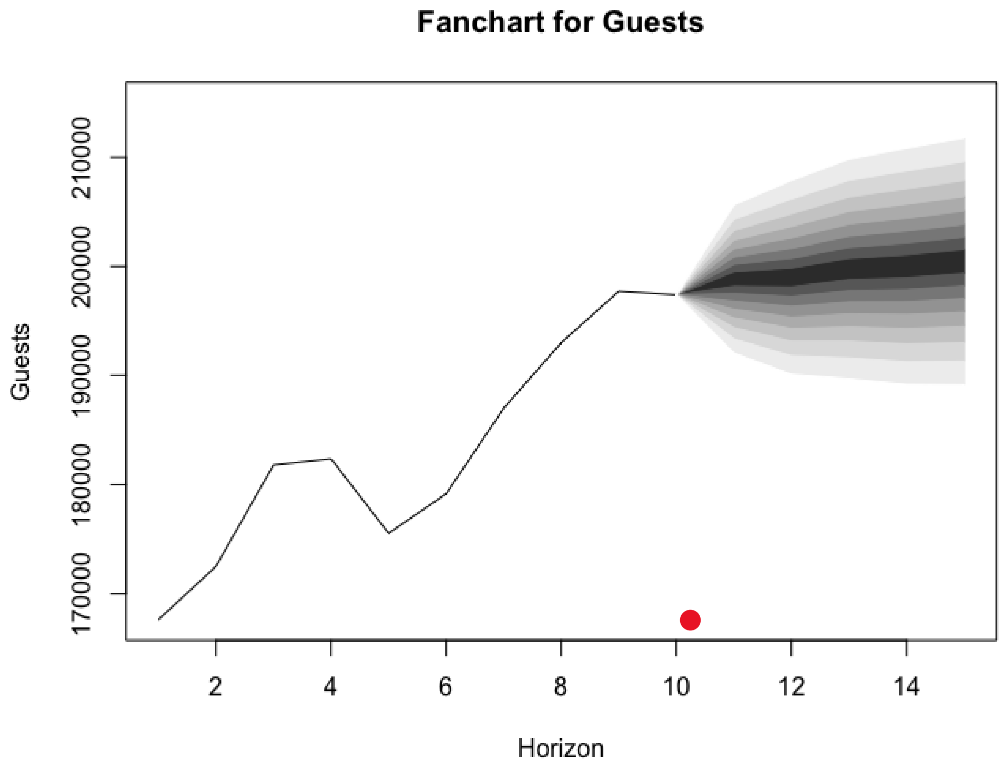

Figure 43: VAR Forecasts for amount of guests in St. Anton am Arlberg

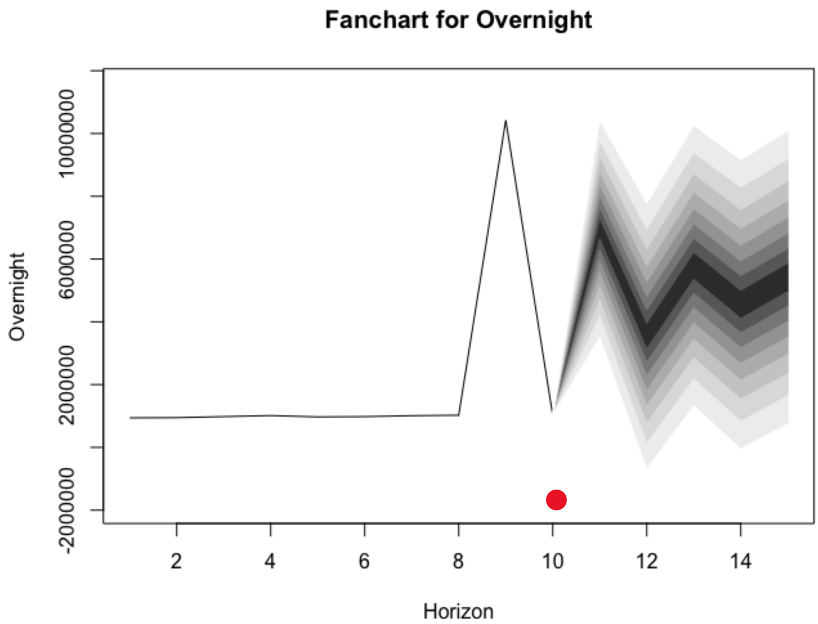

Figure 44: VAR Forecast for overnight stays in St. Anton am Arlberg

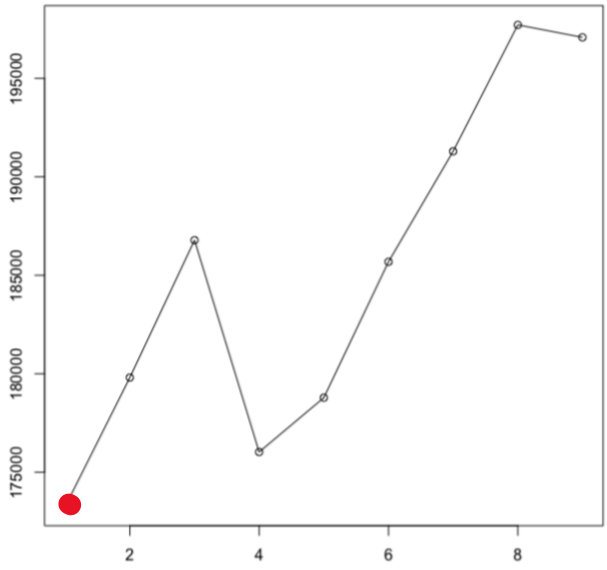

Figure 45: GAM Forecast for number of guests

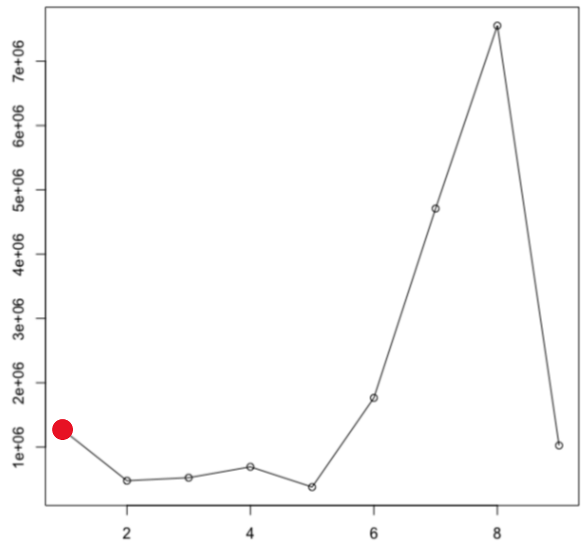

Figure 46: GAM Forecast for overnight stays

The second stage has shown that in the 2019-2020 season almost all numbers were below the forecasts made in the first stage. This proves that COVID had a big impact on tourism in St. Anton. During the most recent season, 2020-2021, the tourism numbers dropped even more as the ski resort was only open for locals. It is clear that due to the COVID-19 pandemic, the forecasts from stage 1 need to be adjusted. To do this correctly, not only the current tourism numbers should be used, but one should also look at the measures taken to prevent spreading of the virus as well as the sentiment among tourists following this pandemic. Moreover, we saw that in general the forecasts made by the VAR model and the GAM model are similar. With forecasts from both models showing overall similar trends and forecasted values. However, there were some differences between the forecasts. Such as the forecasts for overnight stays and the forecasts for St. Anton and the Arlberg region

In the next chapter, the conclusion, these results will be combined with related literature to draw conclusions and determine how tourism will develop in St. Anton and the Arlberg region as the world opens up again after the pandemic.

Conclusion & Discussion

This thesis has made it clear that the tourism industry is at a turning point. The COVID-19 pandemic could be a great kickstart to change the industry. However, everybody relying on tourism for their income and livelihood have been hit hard by the measures taken to stop the spread of the virus. These people and businesses will most likely be focusing on how to regain what was lost due to the pandemic. To help in formulating strategies that suit and help both the tourists and the businesses, this thesis attempted to forecast tourism in Sankt Anton am Arlberg for the next 5 seasons.

Forecasts based on the previous nine years, with both VAR and GAM models but without COVID, showed an overall upward trend in tourism numbers for St. Anton. With downward trends showing in the overnight stays in private accommodations. This would be a good forecast for any tourism destination. As it means that more people will visit and these people will stay in commercial accommodation and rented apartments more than in their private accommodations. However, due to COVID-19 St. Anton was barely able to welcome any tourists in the 2020-2021 season. This means that St. Anton had to deal not only with fewer tourists in the 2019-2020 season but also in the 2020-2021 season. This is a harsh reality for a town that is built around tourism.

The second stage of the analysis showed that for all tested variables, the numbers from the 2019-2020 season were well below what was forecasted. Based on the literature on tourism after the pandemic, it is highly unlikely that tourism numbers will return to their pre-COVID levels within the next 5 years. It is expected that the COVID-19 pandemic has led to structural changes in the tourism industry. This poses a challenge for St. Anton as they will have to market and position themselves to get the most amount of tourists possible. Based on the forecasts made in this thesis and the reviewed literature, it can be concluded that tourists are likely to want to stay closer to home and avoid public transport like planes, trains and busses. This means that St. Anton should market to people in Europe, especially Germany, Switzerland, France and the Netherlands. These are countries that people can easily travel from with their car and these are the countries that most of the tourists came from before the pandemic. These analyses and forecasts were made to try and answer the question:

What are the structural changes to tourism due to COVID-19 and how will they affect tourism?

However, the analysis of the data and the literature showed that this is not an easy question to answer. Perhaps the most plausible answer this thesis can give, is that the structural changes can be defined by St. Anton itself. However, based on these findings and the literature, it could be said that tourism has to be built up from scratch again. With more and more people getting vaccinated, the possibilities of travel are increasing again. Meaning more people can come and visit St. Anton. While it is hard to say what the structural changes will be from the tourists perspective, it has become clear that the tourism industry itself can use this pandemic the start structural changes itself. Even without COVID, the forecasts show changes and trends that were about to happen, especially in the accommodation categories. St. Anton can use these forecasts to create marketing campaigns that play into what St. Anton wants for the future. This could mean promoting more apartments, to play into the expected changes in accommodation categories before COVID. Or promoting nature and preserving nature, playing into the need to deal with the climate crisis. Moreover, St. Anton could focus their marketing efforts on welcoming back tourists, while also sticking to government-imposed measures against COVID. All in all, St. Anton is currently at a tipping point. And while there it cannot force tourists into certain patterns, it can promote tourism in a way that is sustainable for the future.

The use of both a VAR model and a GAM model allowed for the comparison of the two models. In this case, it seems that the VAR model is the best model for this dataset. As this dataset only contains the dependent variable and its lag, it is not necessary to know the relationships between different variables as there are very few variables. If there would be more explanatory variables, the GAM model could be better as this model also gives an idea of the individual effects of independent variables on the dependent variable. Moreover, the interpretation of the GAM model is similar to that of a linear regression, which makes it easily understandable for people with no background in statistics. The main advantage of the VAR model, in this case, is that it requires less work, as the model will help choose the right parameters. Whereas in a GAM some parameters have to be determined with statistical formulas, therefore it is more work to build a GAM model.

Before ending this thesis, the limitations of this research will be discussed in the final chapter of this thesis. This chapter will help improve further research.

Limitations

While the research conducted in this thesis can be relevant to the field of tourism research, there are many possible improvements. Especially on the data side, big improvements could be made. The first limitation of this research is that the data only concerns one town in Austria and this town is not necessarily an average town as it is geared towards high-income tourists or very experienced skiers. This means that the tourists in this town are most likely people that come for the unique descends in the ski area or for the allure that the town has. This means that findings from this research might not apply to other towns that might be more geared towards beginners or middle-income tourists. Therefore, using data from multiple towns might help in generalizing the findings of this research. The second data related limitation is that this research used a time series with only 10 points which is a relatively small dataset. Additionally, the dataset only has the dependent variable and its lag as variables. Adding more time points and more explanatory variables such as demographics, income, whether or not someone has been to that town before could increase the significance and explanatory power of the models used in this research. Moreover, with more time and money any research could be improved, also this research. Furthermore, this pandemic is a very extreme situation. This makes it very hard to draw clear conclusions on how industries will react, as a situation like this has never occurred before. However, this research can be used as a way of signaling. The findings of this research could be used as signals for a crisis in the future.

To conclude this thesis, some suggestions for future research will be discussed. Future research could include more variables and data from more destinations to the dataset. This will most likely make a big improvement to the models used in this research. Moreover, future research could try to simulate exogenous shocks, to give a more general idea of how tourism responds to government-imposed shocks. Finally, future research could focus on particular subgroups of tourists, such as tourist with different skill levels in skiing and snowboarding or tourists from different socio-economic groups. This could give an idea as to how certain subgroups react to exogenous shocks.

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. https://doi.org/10.1016/j.trip.2020.100255

Arundhati Roy (2020, April 3). The pandemic is a portal. *Financial Times.* https://www.ft.com/content/10d8f5e8-74eb-11ea-95fe-fcd274e920ca

Box, G. E. P., & Pierce, D. A. (1970). Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. *Journal of the American Statistical Association*, *65*(332), 1509–1526. https://doi.org/10.1080/01621459.1970.10481180

Brock, S. (2020, October 5). Here are 8 ways travel will change after the pandemic. *Travel.* https://www.nationalgeographic.com/travel/2020/10/heres-how-covid-is-changing-travel-according-to-the-experts/

Butcher, J. (2020, May 4). The war on tourism. *Spiked*. https://www.spiked-online.com/2020/05/04/the-war-on-tourism/

Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., Pastore y Piontti, A., Mu, K., Rossi, L., Sun, K., Viboud, C., Xiong, X., Yu, H., Halloran, M. E., Longini, I. M., & Vespignani, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, *368*(6489), 395–400. https://doi.org/10.1126/science.aba9757

De Vos, J. (2020). The effect of COVID-19 and subsequent social distancing on travel behavior. *Transportation Research Interdisciplinary Perspectives*, *5*, 100121. https://doi.org/10.1016/j.trip.2020.100121

Eloriaga, J. (2020, June 22). *A Deep Dive on Vector Autoregression in R - Towards Data Science*. Medium. https://towardsdatascience.com/a-deep-dive-on-vector-autoregression-in-r-58767ebb3f06

Farzanegan, M. R., Gholipour, H. F., Feizi, M., Nunkoo, R., & Andargoli, A. E. (2020). International Tourism and Outbreak of Coronavirus (COVID-19): A Cross-Country Analysis. *Journal of Travel Research*, 004728752093159. https://doi.org/10.1177/0047287520931593

Fotiadis, A., Polyzos, S., & Huan, T. C. T. (2021). The good, the bad and the ugly on COVID-19 tourism recovery. *Annals of Tourism Research*, *87*, 103117. https://doi.org/10.1016/j.annals.2020.103117

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized Additive Models (Chapman & Hall/CRC Monographs on Statistics and Applied Probability)* (1st ed.). Chapman and Hall/CRC.

Hastie, T., & Tibshirani, R. (1986). Generalized Additive Models. *Statistical Science*, *1*(3). https://doi.org/10.1214/ss/1177013604

Hastie, T., Tibshirani, R., & Friedman, J. H. (2001). *The Elements of Statistical Learning*. Springer Publishing.

Jeon, C.-Y., & Yang, H.-W. (2021). The structural changes of a local tourism network: comparison of before and after COVID-19. *Current Issues in Tourism*, 1–15. https://doi.org/10.1080/13683500.2021.1874890

Jiricka-Pürrer, A., Brandenburg, C., & Pröbstl-Haider, U. (2020). City tourism pre- and post-covid-19 pandemic – Messages to take home for climate change adaptation and mitigation? *Journal of Outdoor Recreation and Tourism*, *31*, 100329. https://doi.org/10.1016/j.jort.2020.100329

Kolmogorov, A. N. (1957). On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. In *Doklady Akademii Nauk* (Vol. 114, No. 5, pp. 953-956). Russian Academy of Sciences.

Lew, A. A., Cheer, J. M., Haywood, M., Brouder, P., & Salazar, N. B. (2020). Visions of travel and tourism after the global COVID-19 transformation of 2020. *Tourism Geographies*, *22*(3), 455–466. https://doi.org/10.1080/14616688.2020.1770326

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*(2), 297–303. https://doi.org/10.1093/biomet/65.2.297

Lütkepohl, H. (2007). *New Introduction to Multiple Time Series Analysis*. Springer Publishing.

Nepal, S. K. (2020). Adventure travel and tourism after COVID-19 – business as usual or opportunity to reset? *Tourism Geographies*, *22*(3), 646–650. https://doi.org/10.1080/14616688.2020.1760926

Palmer, S. (2020, December 29). *Here's what experts are forecasting for the travel industry in 2021*. Euronews. https://www.euronews.com/travel/2020/12/29/here-s-what-experts-are-forecasting-for-the-travel-industry-in-2021

Pana, T. A., Bhattacharya, S., Gamble, D. T., Pasdar, Z., Szlachetka, W. A., Perdomo-Lampignano, J. A., Ewers, K. D., McLernon, D. J., & Myint, P. K. (2021). Country-level determinants of the severity of the first global wave of the COVID-19 pandemic: an ecological study. *BMJ Open*, *11*(2), 1. https://doi.org/10.1136/bmjopen-2020-042034

Pedroni, P. (2013). Structural Panel VARs. *Econometrics*, *1*(2), 180–206. https://doi.org/10.3390/econometrics1020180

Ploberger, W., & Kramer, W. (1992). The Cusum Test with Ols Residuals. *Econometrica*, *60*(2), 271. https://doi.org/10.2307/2951597

Schmidt-Hieber, J. (2021). The Kolmogorov–Arnold representation theorem revisited. *Neural Networks*, *137*, 119-126.

Shapiro, S. S., & Francia, R. S. (1972). An Approximate Analysis of Variance Test for Normality. *Journal of the American Statistical Association*, *67*(337), 215–216. https://doi.org/10.1080/01621459.1972.10481232

Sigala, M. (2020). Tourism and COVID-19: Impacts and implications for advancing and resetting industry and research. *Journal of Business Research*, *117*, 312–321. https://doi.org/10.1016/j.jbusres.2020.06.015

Škare, M., Soriano, D. R., & Porada-Rochoń, M. ł. (2021). Impact of COVID-19 on the travel and tourism industry. *Technological Forecasting and Social Change*, *163*, 120469. https://doi.org/10.1016/j.techfore.2020.120469

Stock, J. H., & Watson, M. W. (2001). *Vector Autoregressions*. Washington University. https://faculty.washington.edu/ezivot/econ584/notes/varModels.pdf

United Nations. (2020). *Policy Brief: COVID-19 and Transforming Tourism*. https://www.un.org/sites/un2.un.org/files/sg_policy_brief_covid-19_tourism_august_2020.pdf

UNWTO. (2021). *International Tourism and COVID 19*. Www.Unwto.Com. https://www.unwto.org/international-tourism-and-covid-19

Yang, Y., Zhang, H., & Chen, X. (2020). Coronavirus pandemic and tourism: Dynamic stochastic general equilibrium modeling of infectious disease outbreak. *Annals of Tourism Research*, *83*, 102913. https://doi.org/10.1016/j.annals.2020.102913

Zizek, S., & Žižek, S. (2020). *Pandemic!* Wiley.