University of Stavanger

FACULTY OF SCIENCE AND TECHNOLOGY,
DEPARTMENT OF SAFETY, ECONOMICS AND PLANNING,
UNIVERISTY OF STAVANGER

MASTER'S THESIS

STUDY PROGRAM:	THESIS IS WRITTEN IN THE FOLLOWING SPECIALIZATION/SUBJECT:
Industrial Economics	Economics (not confidential)

TITLE:

Analysis of driving factors for capacity investment in the Norwegian hotel industry

AUTHOR:		ADVISOR:
Student number: 241950	Name: Håkon Johannessen	Prof. Sigbjørn L. Tveteraas

ACKNOWLEDGE RECEIPT OF 2 BOUND COPIES OF THESIS

Stavanger, 15 / 06 / 2019 Signature administration: ……………………………

Acknowledgements

I would like to dedicate these few introductory sentences to extend my gratitude to a few people that have been instrumental to my success.

First, I'd like to sincerely thank my advisor, Prof. Sigbjørn L. Tveteraas. To me, there is really no telling how this assignment would have turned out if not for his invaluable guidance and dedication to motivate, as well as to educate me throughout this study.

Second, I would like to thank my wonderful girlfriend and companion, for her continued support, encouragement and uninterrupted care for my wellbeing.

Lastly, I would not have accomplished anything without my parents, who from an early age inspired me to pursue my interests and challenged me to live up to my potential. Great thanks to them for their love, support and for their continued faith in me.

Abstract

The purpose of this study is to assess the importance of a range of different economic, financial and locational factors, in how they influence the collective will to invest in capacity in the Norwegian hotel industry. Relations are investigated for the Norwegian market as a whole, and while it does not aim to reflect the individual investors willingness to invest, the selection of dependent variables is made from key driving factors as defined by individual investors. These factors were chosen in coherence with previous literature, like Newell and Seabrook's (2006) investigation of factors influencing hotel investment decision making, and Luo and Lam's (2017) research on urbanizations effect on hotel performance. But, uniquely for the Norwegian hotel market, the present study presents empirical evidence for the effects of three main groups of variables, namely "demand", "financial and economical" and "urbanization". My findings add to Luo and Lam's work, that indicators of urbanization, like GNP per capita and people living in urban areas, have a positive relation on capacity as well as performance. Also, I provide reasoning for the significant positive impacts of variables describing demand, such as population, hotel visitors and air-travelers, as well as variables like GNPB, currency rates and interest rates, that describe the financial and economic state.

Key words: hotel investment, hotel capacity, hotel demand, urbanization

Table of Contents

Chapter 1: Introduction

1.1 Research Objective and Motivation

The Norwegian economy has experienced significant fluctuations over the last couple of decades, mostly due to the financial crisis in 2008/2009 and the oil crisis in 2013/2014. The Norwegian economy is sensitive to developments in the price of oil, as the oil industry is a vital part of the economy. Recessions seen today in a range of Norwegian industries can still be traced back as repercussions from the sudden drop in the price of oil in 2013/2014 (Cappelen, Eika & Prestmo, 2014). The Norwegian hotel industry has also been experiencing fluctuating results, appearing to be in line with those of the economy. Coastal regions like Oslofjorden and Vestlandet, known to be heavily reliant on the oil industry, have suffered some of the most dramatic declines following the price recession. For instance, in the years leading up to the oil crisis the city/municipality of Stavanger was one of the reoccurring top performing regions. With solid results in several key performance metrices, like occupancy rate, room price and revenue per available room (RevPAR), the Stavanger region caught the investors' attention and experienced a high rate development of new hotels. After the oil crisis, there followed a sudden decrease in demand and the newly increased capacity was left mainly un-utilized. Even now, five years later financial reports show that Stavanger has an occupancy rate of 48.1%, the lowest of all the big cities in Norway (Berglihn, 2018). However recent reports also show that the industry as a whole has made quite the recovery, and in fact, some regions are reaching new all-time heights, both in terms of occupancy rate and in room price (Bjørshol, 2017). So, what other factors drive the continued investments in the Norwegian hotel industry?

While there certainly appears to be a significant connection between the economy and the hotel industry, there is likely also other factors that influence the balance between supply and demand for the Norwegian hotel industry. Population, urbanization and, for instance, may all be useful indicators of demand. Norway is experiencing a heightened rate of urbanization and growth in population. Over the course of the last decade alone, Norway's population has increased with more than half a million people. From 4.7 million in 2008, to 5.3 million in 2018 (Statistics Norway, 2019), and the portion of the population living in built-up urban areas has increased from 78.6% to 81.5% (Juel, 2017). Numbers from Statistics Norway (2019) show that the tourism consumption in Norway from accommodation services was more than 60% higher in 2017 than in 2007, from 115 to 176 million NOK, and that the contribution from non-domestic visitors increased from 37.5% to 45.5% in the same period. This may indicate that the national demand for capacity in the Norwegian hotel industry is becoming larger.

Part of the motivation for the present study, comes from the desire to relate the situation in Stavanger to the rest of the Norwegian hotel industry, and the purpose of this study is to assess the importance of a range of different economic, financial and locational factors in how they influence the collective will to invest in capacity in the Norwegian hotel industry. Thus, the goal is to describe the development of the Norwegian hotel market in terms of these driving factors of investment. The study will also assess the effects of these factors on both municipality and county levels of detail.

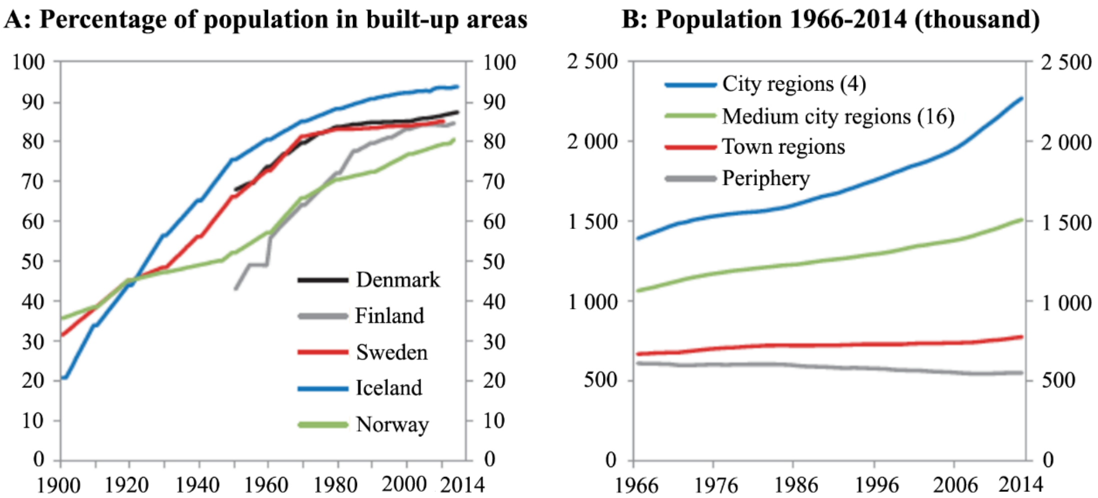
1.2 Importance and Uniqueness of the Study

There is only limited research available concerning the hotel sector. While there are studies assessing the factors driving hotel investment decision making, and studies describing different factors' effect on hotel performance empirically, there appears to be no such studies that tackle the Norwegian hotel industry. Thus, by combining the two, the present study makes a unique assessment of the empirical evidence of factors driving hotel investment for the Norwegian market.

Chapter 2: Theoretical Background

Access to more densely populated areas and more potential manpower is generally a prerequisite, for both the production and the service industry, to maintain higher efficiency and operations on a larger scale. Urbanization, defined as an increase in the percentage of a population living in built-up areas, has historically been slower in Norway than for its other northern neighbors (figure 2.1.A, next page). But after the last world war, and particularly over the course of the last three decades, Norway has experienced urbanization at an increased rate. Figure 2.1.B on the next page shows that the urbanization has been especially relevant for population growth in the four major city regions, Oslo, Stavanger, Bergen and Trondheim. Some growth has also been present in the medium city regions, while the population in the smaller town regions have been more or less stationary, only exhibiting a weak population growth. One explanation for this behavior is the booming economic growth in Norway after the discovery of oil. With the increase in the economy came an increase in demand for labor, which in turn lead to an increase in population through labor migration (Juel, 2017).

Figure 2.1 (Adapted from Juel, 2017 – Figure 2, page 4)



According to Statistics Norway (2019) a collection of houses qualifies as a built-up area if it has at least 200 occupants and the distance between each of the houses are within 50 meters, though exceptions can be made for houses around areas that cannot be populated or are otherwise uninhabitable. This includes parks, sporting arenas, industrial sites or natural obstacles such as rivers or farmlands. Smaller clusters of houses that naturally belongs to a built-up area can be included if not further away than 400m.

Lou and Lam (2017) discovered a relationship between urbanization and hotel performance, particularly regarding hotel occupancy rate (HOR). To express the urbanization in China, they applied the following four dimensions as measures for the level of urbanization; geographical landscape, economic, population and social cultural dimension. They showed that the economic dimensions gross domestic product per capita (GDPpc) and the service industry's share in GDP (SSGDP) were positively related to occupancy rate. As the income of individuals in the region increase and as the service industry becomes relatively more important, the demand of hotel accommodation will increase accordingly. Furthermore, as the number of people moving into the urban areas increase, the demand for hotel accommodation will also increase (Lou & Lam, 2017).

The population dimension was measured by non-agricultural population proportion (NAPP) and was also shown to have a statistically significant positive relationship to the occupancy rate. According to Lou and Lam (2017) non-agricultural population was selected as a measurement of urban population because most of the agricultural workers lived in rural areas, and thus this would serve as a good approximation of the population living in urban areas.

Both the social cultural dimension, number of hospital beds (NHB), and the geographical landscape dimension, area of garden and green (AGG) were shown to have a negative relation to the occupancy rate. Health is a prerequisite for the increase in productivity and can be seen as a vital component of development and growth. As population increase, the number of people who will express a demand for healthcare will also increase, hence increasing the demand for NHB (Lou & Lam, 2017). Lou and Lam argue that an explanation for the negative relation is that an increase in NHB can cause an increase in tourists' concern on the living conditions of the regions, and thereby lead to lowered occupancy rates. Lastly, as a city develops, when the urbanization rate increases, the demand for land for industrialization will increase. AGG reflects this impact of the urbanization.

Lou and Lam's (2017) proposed model of the study:

(equation 2.1)

$$HOR_{i,t} = a + b_1 GDPpc_{i,t} + b_2 SSGDP_{i,t} + b_3 NAPP_{i,t} + b_4 NHB_{i,t} + b_5 AGG_{i,t} + \epsilon_{i,t}$$

where the subscript i, t represents the i-th region at time t.

According to Newell and Seabrook's (2006) study conducted in Australia, financial factors (37.0 per cent) had the highest weight for investors and hotel owners. Hotel investments are primarily prioritized based on underlying financial performance (e.g. forecast ROI, gross operating profit, RevPAR), which in turn is strongly influenced by local market conditions via the location factors (e.g. site attributes, hotel supply and demand); hence the strong link between the financial factor (37.0 per cent) and the location factor (29.9 per cent). The relationships factor (e.g. stakeholder alignment, asset management) (6.6 per cent) was least important (Newell & Seabrook, 2006).

Overall, individual factors and sub-factors that influence hotel investment decision making can be arranged into three levels of importance; The first level include financial and location factors accounting for a total of 66.9 per cent of respondent weightings. The second level include economic and diversification factors accounting for 26.5 per cent of respondent weightings and the final level incorporate relationships, accounting for only 6.6 per cent of respondent weights.

Table 2.1 – Hotel investment multi-criteria decision-making model (Newell & Seabrook, 2006)

Level	Factors	Sub-factors	Driver/outcome
First	Financial	Forecast five-year return on investment	Outcome
	Location	Site attributes	Driver
	Financial	Gross operating profit	Driver
	Location	Hotel supply	Driver
	Location	Demand volatility	Driver/outcome
	Financial	Historical rates of return	Driver
	Financial	RevPAR	Driver/outcome
	Diversification	Segment diversification	Driver
	Location	number of domestic visitors	Driver
	Relationships	Alignment with stakeholders	Driver
	Financial	Unsystematic risk	Driver
	Economic	Business spending patterns	Driver
Second	Financial	Economies of scale advantages	Driver
	Diversification	Geographic diversification	Driver
	Location	number of international visitors	Driver
	Economic	Interest rates	Driver
	Location	Age of target hotel	Driver
Third	Economic	Extent market is emerging	Driver
	Economic	Tourist spending patterns	Driver
	Diversification	Link to target property	Driver
	Economic	Extent market is mature	Driver
	Relationships	Independent asset management	Driver
	Diversification	Brand diversification	Driver
	Economic	Employment growth (office)	Driver
	Relationships	Regulatory influence	Driver

Table 2.1 shows the priority order of which factors and sub-factors are evaluated, based on degree of importance identified by hotel investors, owners and operators. Factors and sub-factors are also indicated as either drivers or outcomes, with drivers being characteristics (factors/sub-factors) that contribute to income. In some cases, factors or sub-factors can be both drivers and outcomes; for example, RevPAR is a driver of return on investment and the interaction of supply and demand will influence the performance of RevPAR (Newell & Seabrook, 2006).

Hotel investors were also shown to place greater importance on location attributes that they can specifically identify themselves, these include hotel supply and demand and site attributes, in contrast to macroeconomic impacts including business and tourist spending patterns and growth patterns in employment (Newell & Seabrook, 2006). Whilst financial performance indicators such as forecasted five-year return on investment and RevPAR drives individual hotel analysis, Newell and Seabrook (2006) concludes that hotel investors are cognizant of the importance of geographic diversification to reduce their risk exposure and segment diversification to reduce property-specific occupancy risk.

RevPAR is an essential part of hotel revenue management. RevPAR is short for "revenue per available room" and represents the average revenue generated by each available guest room during a specific period of time (Hayes & Miller, 2011). It is commonly used as a performance metric to make an assessment regarding a hotel's operations, and its ability to fill its available rooms at an average rate. RevPAR can be calculated as the average daily rate (ADR) multiplied with the occupancy rate (OR), or equivalently simply by dividing accommodation revenue by rooms available:

(equation 2.2)

$$RevPAR = ADR \times OR = \frac{Accomodation\ Revenue}{Rooms\ Sold} \times \frac{Rooms\ Sold}{Rooms\ Available} = \frac{Accommodation\ Revenue}{Rooms\ Available}$$

Volatility and uncertainty in demand is also important to consider when making investment decisions in the hotel industry (Newell & Seabrook, 2006). It is not unusual for hotels to have customers fail to show up for their booking reservations. The purpose of yield management in hotels is to reduce the high frequency and fluctuation of uncertain demand by selling rooms and services to the right people at the right time and the right price (Chen & Lin, 2013). The relationship between uncertain demand and firm capacity has been discussed in several previous economic literatures, however empirical evidence supporting the relationship is lacking in the hotel industry. According to Chen and Lin (2013) their main empirical findings shows a significant positive association between demand uncertainty and hotel capacity decisions. This type of relationship implies that more uncertainty leads to higher investments in hotel capacity.

Chapter 3: Research Methodology

This chapter introduces the approach and theory applied to achieve the goal of the present study, namely, to express and give empirical estimates of the underlying relations and factors that drive investments in the Norwegian hotel industry. These investments are defined as an increase in hotel capacity, measured in number of hotel rooms available, instead of the numeric magnitude of resources invested. This includes all methods and models used and their limitations, as well as a presentation of relevant theory. The empirical study model consisted of first determining which variables to focus on and gather historical quantitative data for these variables. Collecting accurate data has been a major part of the present study. The detail level of the data varied from municipality to national level, and monthly to yearly. Once collected, the data was arranged as panel data, and screened for deviations. Descriptive statistics such as *mean, min, max, st.dev, correlations* and *autocorrelations*, along with a graphical analysis were used to better represent the full implications of the data. The variables were then sorted into different regression models, to determine their impact on hotel capacity and performance. In the present study, pooled OLS regression were the main model used, as it is a simple yet powerful tool for doing regression on panel data. Most of the proposed models were estimated through regression for both county and municipality level, and only using yearly aggregated data, as the correlation matrices showed that relations between variables stayed approximately the same for monthly and yearly observations. Lastly, common theory for hypothesis testing were applied to assess the validity of the results, and to determine the level of statistical significance of the findings.

Figures in this chapter are mostly to provide some basic visual context, thus readability may be somewhat compromised. Important figures are presented again under the results chapter of the present study.

3.1 Population and Sample

An important distinction to make when processing statistical data, is that between a population and its sample. The population is a large group of cases from which a sample is picked out and which is stated in theoretical terms. Sample is a smaller set of cases, results from which are generalized to the population it was drawn from (Neuman, 2007). Or in other words, the results of analyzing the sample data are used to estimate properties of the entire population. For the present study, sampled factors only really convey accurate information about their respective regions, within their respective timeframe, but this information is generalized to serve as an approximation of the relation for Norway in its entirety and for all points in time.

Conforming with previous research, the present study samples some financial, economic, and locational sub-factors deemed by Newell & Seabrook (2006) to be relevant in hotel investment decision making. The sampled populations also include variables that according to Lou and Lam (2017) indicate level of urbanization in a region.

In the present study the population parameters and their samples are defined as follows:

- **Hotel industry key figures** – Sample consists of *monthly* quantitative data from 2008 to 2018, covering regions from municipality[*] to national level. Sample variables include *RevPAR (in 1000 NOK), total rooms available, total number of hotels, total beds available, domestic visitors* and *international visitors* (Statistikknett, 2019; Statistics Norway, 2019).
- **Population and geographic** – Sample consists of *yearly* quantitative data from 2008 to 2018, covering regions from municipality[*] to national level. Sample variables include *population at end and beginning of year, average population, region land area (in* km^2*) and region average population density (in people/*km^2*)*(Statistics Norway, 2019).
- **Gross National Product** – Sample consist of *yearly* quantitative data from 2008 to 2017, only covering county and national levels. Sample variables include *GNP Basis value (in million NOK) and GNPB volume change (in %-change)* (Statistics Norway, 2019).
- **Exchange rate and Interest rate**– Sample consist of *monthly* quantitative data from 2008 to 2018, only available on national level. Sample variables include *GBP, SEK, DKK, EUR, USD, weighted currency,* and *key policy rate* (Norges Bank, 2019). The weights for *weighted currency* were computed using county level data on origin country of visitors, presented in appendix table A2 (Statistics Norway, 2019).
- **Air traffic** – Sample consist of *monthly* quantitative data from 2014 to 2018, covering 44 airports owned by Avinor and 6 private airports. Sample variables include *scheduled domestic flights, scheduled international flights,* and *total scheduled flights* (Avinor, 2019). All data is presented in number of passengers.

[*] Municipality regions defined in Appendix table A1

- **Brent oil price** – Sample consist of *monthly* quantitative data from 2008 to 2018, only available on national level. Sample variable included is *brent oil price* (IndexMundi, 2019). Data represented in NOK per barrel of crude oil and was converted from USD to NOK using the exchange rate data collected earlier.

In the present study, the data collected was further sorted into groups, that would later be used in regression models as the explanatory variables of three main effects on hotel capacity:

- **Demand –** Variables chosen to represent demand consist of domestic and international hotel visitors, average population, as well as scheduled domestic and international flight passengers. Domestic and international visitors are driving factors for hotel investment, according to Newell and Seabrook (2006), and the remaining are assumed to also be significant indicators of demand.
- **Financial and Economic –** Variables chosen to represent financial and economic factors are GNPB, weighted average currency, crude oil brent price and key policy rate. GNPB and key policy rate are, according to Newell and Seabrook (2006), driving factors for hotel investment. Weighted average currency rate, or the strength of the Norwegian Krone compared to the currency of the most common origin countries of visitors (appendix A2), is closely related to GNPB and key policy rate, and is therefore assumed to be a significant factor. The price of crude oil brent was also included as a factor, as it is a vital part of the Norwegian economy.
- **Urbanization –** Variables chosen to represent level of urbanization are GNP per capita, and number of people living in urban/built up areas. These are proposed by Lou and Lam (2017) and proven to be driving factors for hotel performance in China. Lou and Lam also purposed other variables that have little implications in the Norwegian market and society, and thus has been excluded.

3.2 Data Collection

The data collected in the present study can be characterized as *panel data*, as it exhibits components of both *time series data* and *cross-sectional data,* each with their own benefits and limitations. A *time series data* set consist of observations on a single or several variables that changes over time. Because past events can influence future events and lags in behavior are prevalent in the social sciences, time is an important dimension in a time series data set (Wooldridge, 2014). Thus, when analyzing *time series data* individual observations can rarely, if ever, be assumed to be independent across time. This *autocorrelation* within each of the variables can be a potential pitfall when trying to establish relationships between them. *Autocorrelation* is also generally related to the discussion of *stationarity* in a time dependent data series, this will be addressed later in this chapter. In contrast, *cross-sectional data* generally focus on values from individual units. These units might refer to people, companies or countries, or as in the case of the present study, regions, counties or municipalities. *Cross-sectional data* has no time dimension, even if the date of data collection varies somewhat, this is ignored.

A *panel data* set, also known as a *longitudinal data* set, consist of a *time series* for each *cross-sectional* member of that data set. One advantage of using *panel data* is a larger number of data points, which in turn increases the degrees of freedom and contribute to reduce *collinearity* among the explanatory variables – hence improving the efficiency of econometric estimates (Wooldridge, 2001). *Panel data*, by design, also allows for increased control for omitted (unobserved or mis measured) variables.

Most macroeconomic data is collected through a system of national accounts, made available in printed and, increasingly, digital form in university and government libraries (Koop, 2000). Luckily, the availability of comprehensive and detailed digital historical data archives has massively improved over the last decades. In the present study most of the data has been retrieved directly, or through different reproductions of the digital data archives from Statistics Norway (SSB). According to their own official website, SSB is the national statistical institute of Norway and the country's main producer of official statistics. They are responsible for collecting, producing and communicating statistics related to the Norwegian economy, population and society at national, regional and local levels. Their statistics are mainly prepared using raw data from two sources: administrative registers and survey questionnaires. In addition, an increasing amount of information is collected directly from businesses and local authorities own computer systems (Statistics Norway, 2014).

For the collection of hotel data, the digital archives of *Statistikknett* (SN) was used. While they do not produce their own foundational statistics, this was a natural choice as all data presented by SN builds on SSB's official statistics. SSB themselves only publish hotel related data on regional or county levels but allows other actors like SN to purchase more detailed data to publish on their own. By doing so SN is able to make available standardized and comparable statistics for smaller regions and municipalities (Statistikknett, 2015). One of the strengths of the SSB statistics is that the aggregated values for the smaller regions lines up with the published statistics for county and regional levels. While statistical oddities and deviations are inevitable in data sets of this magnitude, it does not really affect the validity of the aggregated data. However, this weakness of the data becomes more important when looking at smaller regions and municipalities. In order to prevent unnecessary skewness in the analysis, outliers and empty data cells have been omitted from the data that forms the foundation for this study.

Air traffic data was collected directly from Avinor. According to their own official website Avinor is a wholly owned state limited company under the Norwegian Ministry of Transport and Communications and is responsible for 44 state-owned airports. Avinor's role in society is to own, operate and develop a national network of airports for the civilian sector and joint air navigation services for the civilian and military sectors (Avinor, 2017). They collect their own data and makes available the monthly aggregates for each Avinor owned airport in Norway. Data on exchange rate and interest rate was collected from *Norges Bank* (NB). NB is Norway's central bank and is tasked with promoting economic stability in Norway. NB also manages the Government Pension Fund Global and the bank’s own foreign exchange reserves (Norges Bank, 2016). They also collect their own data and makes available daily or monthly averages of all currencies traded at their exchange.

Brent oil price was collected from IndexMundi. According to their website, IndexMundi’s mission is to turn raw data from all over the world into useful information for a global audience. They capture statistics that are scattered or otherwise hidden and present them via user-friendly maps, charts, and tables which allow visitors to understand complex information at a glance.

3.3 Data Structure

Table 3.3.1 – Data structure

County	Municipality	Year	Month	Accommodation Revenue (1000 NOK)	Revenue per Room (NOK)	...
Akershus	Bærum	2008	1	9788	940	...
Akershus	Bærum	2008	2	9105	989	...
Akershus	Bærum	2008	3	7049	1003	...
Akershus	Bærum	2008	4	10206	995	...
Akershus	Bærum	2008	5	11291	943	...
Akershus	Bærum	2008	6	12541	917	...
Akershus	Bærum	2008	7	6391	637	...
Akershus	Bærum	2008	8	11678	873	...
Akershus	Bærum	2008	9	12883	966	...
Akershus	Bærum	2008	10	11038	1008	...
Akershus	Bærum	2008	11	10794	1007	...
Akershus	Bærum	2008	12	5993	971	...
Akershus	Bærum	2009	1	8344	997	...
Akershus	Bærum	2009	2	8894	1101	...
Akershus	Bærum	2009	3	8930	942	...
Akershus	Bærum	2009	4	6106	1016	...
Akershus	Bærum	2009	5	9171	940	...
Akershus	Bærum	2009	6	11971	1007	...
...	...	...	...	...	...	...

Having a data set were time periods differ in length (e.g. monthly or yearly) between variables, the data set is inherently *unbalanced*. As the hotel industry and driving factors for investment is the main focus of the present study, the remaining sample data was fitted to the structure of the hotel data. This entails that all data on higher levels than monthly municipality are simply duplicated and repeated down to fit the structure (table 3.3.1), to regain *balance*. As an example, *yearly* data would be repeated twelve times to fit with the *monthly* structure of the hotel data, the same goes for data adopted from aggregated to individual regions. Further, data outside of the timeframe 2008-2018 is cut, and when setting up regression models, all data is limited to the smallest timeframe of the included variables. Again as an example, when doing regression with air traffic data, only datapoints within the timeframe 2014 to 2018 is used. Lastly, due to confidentiality SSB require that a minimum of three different hotel businesses must be operational in any region where statistics are published. This has forced SN to aggregate some smaller regions where minimum number of operational businesses were not met, this structure is shown in appendix table A1.

3.4 Data Analysis

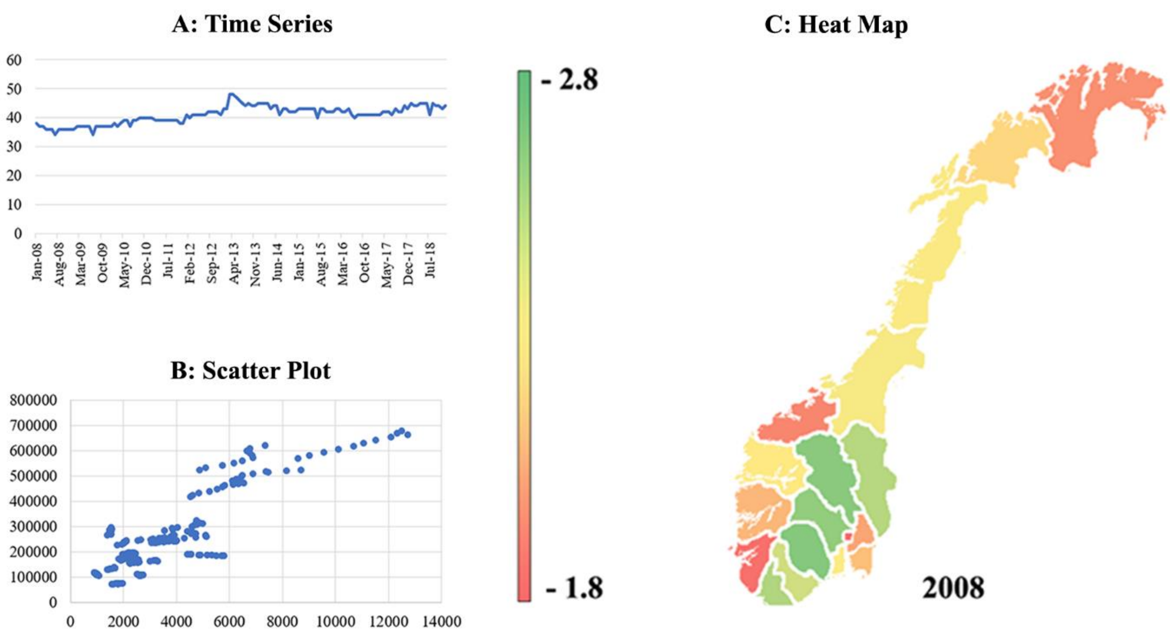
Once the data was collected, the next important step was to have it summarized. As Koop (2000) describes, one can think of the whole field of econometrics as one devoted to the development and dissemination of methods whereby information in data sets is summarized in informative ways. So, in order to more efficiently convey the information contained in the data sets, a combination of both *graphical* and *descriptive analysis* was used.

3.4.1 Graphical Analysis

As is often the case when working with historical data, most of the raw data sets that builds the foundation for the present study are very large. In the present study some variables consist of more than 20000 observations – far too many to be presented as raw numbers for a reader to comprehend. Charts and tables are very useful ways of presenting such large datasets, as well as give a visual overview of their main features. There are many different types of charts, but some of the perhaps most commonly used are *time series graphs, scatter plots*, and *heat maps,* all of which are used in the present study.

- **Time series graph** (figure 3.4.1.A) is a traditional plot that shows how some variable, y-axis, evolves over time, x-axis.

Figure 3.4.1



- **Scatter plots** (figure 3.4.1.B) are a way of modeling the nature of the relationship between two or more variables. Each dot on the chart represents a point using corresponding pairs of information from the X and Y variables. If Y tends to grow as X grows, that indicates that there may be a positive relationship between the two variables. Similarly, if one tends to decrease as the other increases this indicates a potential negative relationship (Koop, 2000).

- **Heat maps** (figure 3.4.1.C) are a way of presenting data in the form of a map or a diagram in which data values are represented as colors. The scale indicates of a represented value is high or low compared to the others.

3.4.2 Descriptive Analysis

While graphs and plots provide an immediate visual representation of the data, descriptive analysis methods serve as an important, as well as more numerically precise addition to the graphical analysis. A very useful first observation is to find numeric values for where the "mass" of a data distribution lies. Such values are commonly referred to as a distribution's *measures of location*, and the word "location" is meant to convey the idea of the center of a distribution. There are, according to Trochim (2001) three main ways to estimate central tendencies of a distribution:

1. The *mean* is the statistical term for the average of the numeric data values, and it is the simplest measure of location of a distribution. It is given by the following mathematical formula:

$$\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} \qquad \textbf{(equation 3.4.2.1)}$$

   where n is the number of data points in the sample (*sample size*) and Y is the sampled variable with mean $\bar{Y}$.
2. The *median* is quite simply the middle value of the data set. That is, it is the value that splits the distribution into two equal halves (Koop, 2000). For distributions with an odd number of sampled values, the *median* is calculated to be the average of the two middle values.
3. The *mode*, like the *mean* and *median* is another common measure of location of a distribution. It represents the most common value, the value that appear most frequently in the data set.

Of course, these measures of location fail to provide any account for the spread of the distribution, and therefore hide a great deal of variability. One of the simplest measures of variability of a distribution is its *dispersion*. By looking at the distance between the minimum (*min)* and maximum (*max)* value contained in a sample we can begin to form an idea of how dispersed the distribution is. However, using these values alone as guidelines for dispersion can be unreliable. A simple example would be how statistical outliers, values that are substantially lower or higher than the other values in the data set, can cause unwanted skewness when trying to measure dispersion in this manner. Therefore, the present study will utilize *min* and *max* along with a more common measure of dispersion, that is the *standard deviation.*

A data set's *standard deviation* can be derived directly or through the *variance* of that data set. Informally, *variance* measures how far the observations in a distribution are spread out from the *mean*. A more rigid definition is that *variance* is the expected value of the squared deviation from the *mean*, and is given by the following mathematical formula:

$$var(Y) = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n}$$ **(equation 3.4.2.2)**

where n is the *sample size* and Y is the sampled variable with mean $\bar{Y}$. A distribution's *std.dev.* is a standardized measure of dispersion, and thus can be interpreted in a comparative sense. That is, if one were to compare the standard deviations of two different distributions, the one with the smaller standard deviation will always exhibit less dispersion (Koop, 2000). The interpretation of *std.dev* is that it will be a low value when the data is close to the sample mean, and when the data is more spread, the *std.dev* is high. Mathematically, it is defined to be the square root of the *variance*:

$$\sigma_Y = \sqrt{var(Y)} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n}}$$ **(equation 3.4.2.3)**

where *std.dev.* of Y is denoted σ_Y, n is the *sample size* and Y is the sampled variable with mean $\bar{Y}$. As variance can be derived directly from *std.dev*, which is also standardized for better comparisons, the present study will only provide numeric values for the *std.dev*.

3.4.3 Correlations

While graphical presentations such as *scatter plots* may give an approximate indication of the nature of the relationship between variables, it is often more useful to also find a quantitative way of describing this relationship. Correlation is an important way of numerically quantifying the relationship between two variables. For the present study, a table displaying the correlation between all pairs of variables was computed, using monthly and yearly data for both regional levels in the data (municipality and county). As correlations were consistent moving from monthly to yearly data, only the yearly data was used in further analyses. However, municipality and county level results showed some distinctions to each other, so both municipality and county level were used in the further analyses. Between two variables, X and Y the correlation is calculated by using the following mathematical formula:

$$r_{X,Y} = corr(X, Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n\sigma_X\sigma_Y}$$ **(equation 3.4.3.1)**

Where the correlation between X and Y is denoted $r_{X,Y}$, $\bar{X}$ and $\bar{Y}$ are the *mean*, σ_X and σ_Y are the *standard deviations* and n is the *sample size*. In the present study, the table of correlations between the variables were created using the data analysis tool in Microsoft Excel.

Whenever correlation is calculated, the resulting value of r always lies between -1 and 1, which may be written as $-1 \leq r \leq 1$. Positive values of r are interpreted as indications of a positive relationship between the variables. Similarly, negative values of r indicate a negative relationship. Larger positive values of r indicate stronger positive correlation, and larger negative values a stronger negative correlation. If $r = 1$ or $r = -1$ this indicate perfect positive or perfect negative correlation respectively. Lastly, when $r = 0$ or is very close to 0, it means that the correlation is absent, or very weak.

It is important to emphasize that correlation only provides an indication that there is a relationship between the two variables, it does not however indicate that one variable *causes* the other. If two variables X and Y are dependent on each other (*direct causality*), such that X causes Y or vice versa, correlation may be falsely large. Also, the possibility of high correlation due to a third variable (*indirect causality*), Z, should always be considered (Pallant, 2010). Thus, it is important to explore the context of the variables and seek to ensure their independency. This is the difference between correlation and causality

3.5 Regression Models

A related concept to correlation, covered in the next sections of this chapter, is regression, which is essentially an extension of correlation to cases of three or more variables that introduces an aspect of causality (Koop, 2000). Regression is arguably the most important tool economists can use to better understand the relationship among two or more variables, and so it is also a critical part of the present study. Due to its nature, it is particularly useful when there are many variables and the interactions between them are complex, which is often the case. In fact, much of econometric analysis begins with the following premise: X and Y are two variables, representing some population, and the analysts are interested in “explaining Y in terms of X”, or in “studying how Y varies with changes in X” (Wooldridge, 2014). Thus, the purpose of simple multiple regression is to look for informative (non-trivial) linear combinations of multiple explanatory variables X, that approximate a dependent variable Y. This section will cover the models used and their limitations, as well as some key theoretical definitions.

3.5.1 Ordinary Least Squares Regression (OLS)

As a way of introducing regression, it is beneficial to begin with a simpler case, using only two variables, and then to follow up by expanding the model to be capable of handling multiple variables. All regression done in the present study is based on producing the best fitting linear relationship which minimizes the sum of the squared residuals. Estimates found in this way are called *least squares* estimates, or *ordinary least squares* (OLS) (Koop, 2000).

For the simple case, using only two variables, the true relationship can be described using the following simple regression model:

$$Y = a + b_1 X + \varepsilon \qquad \textbf{(equation 3.5.1.1)}$$

here Y is the *dependent variable* and X is the *independent* or *explanatory variable,* a is the constant term, b_1 is the partial slope with respect to X and ε is the total error term. However, it is impractical or impossible to calculate these coefficients exact, and so the OLS regression provides an approximation of Y:

$$\hat{Y} = \alpha + \beta_1 X \qquad \textbf{(equation 3.5.1.2)}$$

in this case $\hat{Y}$ will be the approximation of Y, α and β_1 approximate the constant term (a) and the partial slope (b), and the total error term ε has been omitted. Throughout the present study the Greek letters α and β are used consistently to represent the approximations obtained through OLS regression.

Of course, it is often useful to compute a number to summarize how well the OLS regression line fits the data. By looking at actual (Y) versus fitted ($\hat{Y}$) values, a rough estimate for the regression model's "goodness of fit" can be obtained. *R* and *R-square* are the most important values describing this property of the model. *R-square* is simply the squared value of *R* and represent the ratio of the explained variation compared to the total variation; thus, it is interpreted as the *fraction of the sample variation in Y that is explained by X* (Woolridge, 2014). *R* itself is the correlation between the at actual (Y) and fitted ($\hat{Y}$) values and can be calculated using the following formula:

$$R = corr\left(Y, \hat{Y}\right) = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(\hat{Y}_i - \bar{\hat{Y}})}{n\sigma_Y\sigma_{\hat{Y}}} \qquad \textbf{(equation 3.5.1.3)}$$

where $\bar{X}$ and $\bar{Y}$ are the *mean*, σ_X and σ_Y are the *standard deviations* and n is the *sample size*.

Figure 3.5.1.1

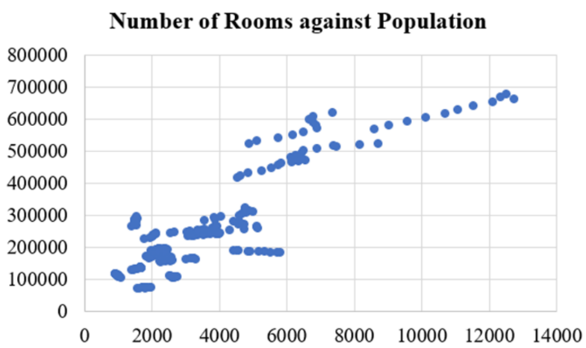


Figure 3.5.1.2

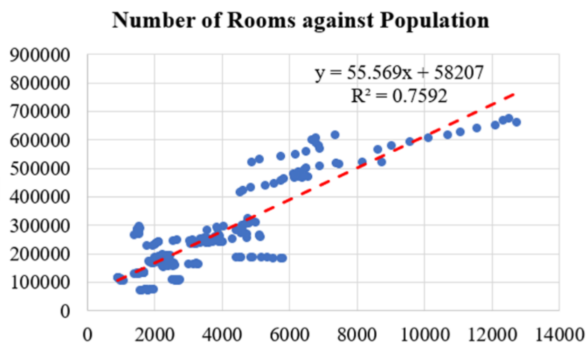


Figure 3.5.1.1 shows a scatter plot that indicates some positive relation between X and Y, in this case *number of hotel rooms* against *region average population*. Figure 3.5.1.2 shows the best fitted line (dashed red) using OLS regression, it also shows the linear relationship of X (x) and $\hat{Y}$ (y), and the calculated *R-square* value for the regression.

A very common transformation of the simple regression estimate, applied to both the *dependent* and the *explanatory* variables, is the logarithmic transformation. This can be put mathematically:

$$ln(\hat{Y}) = \alpha + \beta_1 ln(X) \quad \textbf{(equation 3.5.1.4)}$$

In such regressions the β_1 describe *elasticity* instead of the partial slope. And so, Y would tend to change β_1-percent for a one percent change in X. Because of this property, the present study will mainly use regressions transformed in this manner.

3.5.2 Pooled OLS Regression

Most of the features of simple regression can easily be expanded upon to be able to handle multiple variables. In the case of the present study, because the panel data set consists of a relatively short period of observation (11 years) and a larger cross section (18 counties, 166 municipalities), it is common to employ a pooled OLS method. Hence this will be used for the main models. When pooling, or combining multiple *cross-sectional* variables for the regression model, the results describe the relation for the composition of units, as a whole, and not necessarily the relation for each individual unit. Since multiple regression implies the existence of more than two variables, trying to plot the relationships may quickly require high-dimensional graphs, thus it is usually not possible to display these relationships visually. However, the strategy and intuition for finding approximations for the coefficient is exactly the same as for the simple regression model. The multiple regression model can be formulated mathematically:

$$Y = a + b_1 X_1 + b_2 X_2 + \ldots + b_k X_k + \varepsilon \quad \textbf{(equation 3.5.2.1)}$$

then it can be log-transformed:

$$ln(Y) = a + b_1\, ln(X_1) + b_2\, ln(X_2) + \ldots + b_k\, ln(X_k) + \varepsilon \quad \textbf{(equation 3.5.2.2)}$$

and estimated with the pooled OLS method:

$$ln(\hat{Y}) = \alpha + \beta_1 ln(X_1) + \beta_2 ln(X_2) + \ldots + \beta_k ln(X_k) \quad \textbf{(equation 3.5.2.3)}$$

here k represents the total number of *explanatory variables* $(X_1, X_2, \cdots, X_k)$ used to produce the estimate $\hat{Y}$, and $\beta_1, \beta_2, \cdots, \beta_k$ represent all the elasticities:

$$\frac{\partial ln(\hat{Y})}{\partial ln(X_i)} = \beta_i \quad \textbf{(equation 3.5.2.4)}$$

3.5.3 Differenced Variables and Time Lag

In the case of time series data, it is not uncommon for the effect of the independent variables to take some time to manifest itself. This implies that the value of the dependent variable at a given point in time (Y_t) should depend not only on the value of the independent variable at the same point in time (X_t), but also on the past values of the independent variable ($X_{t-1}, \ldots, X_{t-m}$). Using these kinds of *lagged* variables is not only a simple way of beginning to capture this dynamic, but it is also a fundamental concept to more advanced analyses of time series data (Koop, 2000). The simplest model used to put this concept in the language of regression, is the *distributed lag model*:

$$Y_t = a + b_0 X_t + b_1 X_{t-1} + \ldots + b_m X_{t-m} + \varepsilon_t \qquad \textbf{(equation 3.5.3.1)}$$

in this model, the right-hand side variables are the *lagged* variables, and m is the *lag order* or *lag length*.

Another property of time series data is the existence of correlation across observations for the same variable. While the *distributed lag model* accounts for the effect of the past values of the independent variables (X), the dependent variable (Y) may also depend on its own past values (Y_{t-m}). This is referred to as *autocorrelation*. A common tool for researchers to better understand the properties of a time series is the *autocorrelation*-function:

$$r_m = Corr(Y_t, Y_{t-m}) \qquad \textbf{(equation 3.5.3.2)}$$

where r_m represents the autocorrelation between Y_t and Y_{t-m} at *lag length* m.

Time series that exhibit high *autocorrelation* and *trend* behavior is also likely to exhibit *non-stationary* behavior. Generally, we do not want to include such variables in regression models as they may cause misleading estimation results. The next section on limitations cover this in more detail. For the present study, a simple time step transformation was used to help combat this issue of *non-stationarity.* By instead calculating the percent change of a variable from $t - 1$ to t, the resulting time series will be *stationary*. Because of this trait, such variables are often referred to as *difference stationary*. So, if the variable Y is assumed to be *non-stationary*, we will want to difference it and use ΔY instead. For this calculation, the present study has used the following approximation:

$$ln(Y_t) - ln(Y_{t-1}) \approx \%\Delta Y_{t|t-1} \qquad \textbf{(equation 3.5.3.3)}$$

here the notation " $\%\Delta Y_{t|t-1}$" is used to emphasize that the value represents the percent change in Y from $t-1$ to t, throughout the present study the shorthand notation ΔY_t (or dY) will be used to represent the same change. This approximation holds up well for smaller percentage changes, and can be proven using the first order Taylor expansion of $ln(x) \approx x - 1$:

(equation 3.5.3.4)

$$ln(Y_t) - ln(Y_{t-1}) = ln(\frac{Y_t}{Y_{t-1}}) \approx \frac{Y_t}{Y_{t-1}} - 1 = \frac{Y_t - Y_{t-1}}{Y_{t-1}} = \%\Delta Y_{t|t-1}$$

Combining equations 3.5.3.3 and 3.5.2.2 forms the log-differenced multiple regression model:

(equation 3.5.3.5)

$$\Delta Y_{\mathrm{t}} = a + b_1 \Delta X_{1,\mathrm{t}} + b_2 \, \Delta X_{2,\mathrm{t}} + \ldots + b_k \, \Delta X_{k,\mathrm{t}} + \varepsilon$$

Where the notation $\Delta Y_{\mathrm{t}} = ln(Y_t) - ln(Y_{t-1})$, and similarly, $\Delta X_{1,\mathrm{t}} = ln(\mathrm{X}_{1,t}) - ln(\mathrm{X}_{1,t-1})$.

3.5.4 Limitations

While the issue of *non-stationarity* (or the existence of a *unit root)* in the data set, is a fundamental limitation to *time series* data analysis, it is also, inherently, a fundamental limitation to panel data analysis. Data sets that exhibit high *autocorrelation* and *trend* behavior, will often also display high correlation between residuals, rendering the OLS regression method imprecise. In these cases, variants of generalized least squares (GLS) regression is often used, as GLS does not require residuals to be uncorrelated. Software such as "XLSTAT" or "Stata", that is more oriented towards advanced statistical analyses than Microsoft Excel, can perform many useful variants of GLS, as well as other regressions. Some that could have been appropriate to incorporate in the present study include Feasible GLS, fixed effect, random effect and quantile regression. However, only OLS regression were used in the present study, as it is still the most commonly applied. Most of the regression models in the present study were also proposed both on level form (presumed stationary), and on a differenced form (stationary).

A common method for testing this behavior in time series more manually, would be the *autoregressive model (AR).* For panel data, the more basic approach of testing with the AR-model would have had to be performed on the time series of each *cross-sectional* member of the data set. In the context of the present study, while the autoregressive model otherwise was used to model certain relations, testing for stationarity in this manner was not deemed feasible, and so the stationarity of the variables was not determined. But, to enable some intuition, the method for conducting these tests are described in this section, nonetheless. The general *autoregressive model* AR(m) of m-th order can be expressed mathematically, with the formula:

$$Y_t = a + \Phi_1 Y_{t-1} + \ldots + \Phi_m Y_{t-m} + \varepsilon_t \quad \textbf{(equation 3.5.4.1)}$$

where the coefficients $\Phi_1, \ldots, \Phi_m$ represent the influence of each lagged subset of Y, and m is the *lag length.* For different values of Φ these models can allow for the random fluctuating behavior typical of growth rates of many macroeconomic time series; for the trend behavior typical of the macroeconomic series themselves; or for intermediate cases between these extremes (Koop, 2000). Subtracting Y_{t-1} from both sides of the equation (3.5.4.1) makes the determination of *unit root* behavior more convenient, and with some rearranging* we obtain:

$$\Delta Y_t = a + \rho Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \ldots + \gamma_{m-1} \Delta Y_{t-m+1} + \varepsilon_t \quad \textbf{(equation 3.5.4.2)}$$

Where the coefficients in this regression, ρ, γ_1, …, γ_{m-1} are simple functions of Φ_1, …, Φ_m. Rephrasing the AR(m) model in this way, the equation is still in the form of a regression model and the value of ρ can be more easily computed. If $\rho = 0$ this implies that the AR(m) time series Y contains a *unit root* and is *non-stationary*; however, if $-2 < \rho < 0$ this implies that the time series does not contain a *unit root* and is *stationary* (Koop, 2000).

Further, the data may exhibit a trend behavior as an exact function of time, referred to as *deterministic trend*. To account for this the AR(m) model can be expanded with the term δt to represent this time dependent trend, and is commonly referred to as the *AR(m) with deterministic trend model*:

$$\Delta Y_t = a + \rho Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \ldots + \gamma_{m-1} \Delta Y_{t-m+1} + \delta t + \varepsilon_t \quad \textbf{(equation 3.5.4.3)}$$

Lastly, by looking at the regression estimated value of ρ, we can determine whether or not the time series Y includes a unit root, and therefore has to be omitted or differenced in the regression. Unfortunately, Microsoft Excel does not correctly provide a t-stat for the OLS estimate of ρ so the Dickey-Fuller test, or rather their rule of thumb, as described by Koop (2000) is presented as a way of determining unit root:

---

* Each step in the derivation of this equation only involves simple algebra, however there are many steps and the method can quickly become quite messy, as such the derivation was not included in the present study.

- If the time series regression **includes** a *statistically significant deterministic trend*, the Dickey-Fuller critical value is approximately -3.45. This entails that the unit root hypothesis should be rejected if the *t-stat* on ρ is more negative than -3.45, otherwise conclude that the series has a unit root.
- If the time series regression **does not include** a *statistically significant deterministic trend*, the Dicker-Fuller critical value is approximately -2.89. This entails that the unit root hypothesis should be rejected if the *t-stat* on ρ is more negative than -2.89, otherwise conclude that the series has a unit root.

Other Issues that may arise when designing regression models are *multicollinearity* and *endogeneity*. *Multicollinearity* occurs when one explanatory variable in the multiple regression model is highly correlated with the others and tends to inflate the variable of the slope coefficient estimated through the regression. This will reduce the significance of each individual variable, but the independent variables, as a whole, may still be significantly explaining the dependent variable (Kennedy, 2003). *Endogeneity* most often occur when there is a reverse causality between the dependent variable and at least one of the dependent variables (Wooldridge, 2014). That is, if the dependent variable is causing the independent variable, and oppositely, the independent variable is causing the dependent variable. In the present study, an example of *endogeneity* would be that off hotel revenue and hotel rooms available. While it is assumed to be true that higher hotel revenue causes an increase in hotel rooms available, it is also likely that increasing the number of available hotel rooms will also increase the hotel revenue. Both *multicollinearity* and *endogeneity* affect the accuracy of the regression results and needs to be taken into consideration.

3.6 Hypothesis Testing with ANOVA

In the present study, regressions were performed using the "regression" data analysis tool in Microsoft Excel. Along with the coefficients and R-values, Excel also compute t-stat and p-values for the *explanatory variables* and a "analysis of variance"-table. These are used as reasoning to either *reject* or *fail to reject* the *null hypothesis* (H_0) for the model. The overall two tailed hypothesis test can be put mathematically:

$$\begin{cases} H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0 \\ H_1 : at\ least\ one\ \beta_i \neq 0;\ for\ i = 1, 2, \cdots, k \end{cases}$$

if H_0 is rejected, we will proceed to test each β_i individually:

$$\begin{cases} H_0 : \beta_i = 0 \\ H_1 : \beta_i \neq 0 \end{cases}$$

Nonetheless the interpretation is the same: H_0 assumes no useful (non-trivial) linear relationship between Y and $X_1, X_2, \cdots, X_k$, while H_1 as an alternative hypothesis purpose that Y can be described through a linear combination of at least one X_i (for $i = 1,2, \cdots, k$). The t-stat (*t*) measures how many estimates standard deviations β_i is from the hypothesized value of β_i, while the p-value (*p*) can be interpreted as the probability that the result was coincidental, or due to randomness. These convey much of the same information, as p-value is derived from a t-distribution, and so for hypothesis testing there is no added benefit of choosing one over the other. In the present study, both the *t* and *p*-value of a regression model will be presented, but only the *p*-value will be used directly for the hypothesis testing. If the *p*-value is less than the chosen critical value (*c)*, H_0 will be rejected and H_1 will be preferred. The default significance level is chosen to be 5%, *c* = 0.05, implying a default confidence level of 95%.

Pallant (2010) describes that analysis of variance (ANOVA) is so called because it compares the variance of the dependent variable Y (believed to be due to the independent variables), with the variability within each of the independent variables (believed to be due to chance). Perhaps the most useful result of conducting this analysis is the obtained *F ratio* for the regression model, which represents the variance of Y divided by the total variability within X. Table 3.6.1 on the next page illustrates how an ANOVA-table is typically structured for a multiple regression model.

Table 3.6.1 – ANOVA structure

	Degrees of freedom (*df*)	Sum of squares (*SS*)	Mean sum of squares (*MS*)	F
Regression	k	$SS_{Tot} - SS_{res}$	$\frac{SS_{reg}}{k}$	$\frac{MS_{reg}}{MS_{res}}$
Residual	n – k – 1	$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$	$\frac{SS_{res}}{n-k-1}$	
Total	n – 1	$\sum_{i=1}^{n}(Y_i - \bar{Y})^2$		

The *degrees of freedom* are simply defined as presented in table 3.6.1, where n is the total number of observations and k is the number of independent variables. For the sum of squares calculation, Y_i and $\hat{Y}_i$ are the value of the *i*-th actual and estimated observations, while $\bar{Y}$ is the mean of the actual observations. The variances are formulated as the *mean sum of squares* (MS), derived as shown in table 3.6.1, and *F* is computed as the ratio between the two. Along with the *degrees of freedom* for both the regression and the *residuals*, the value for *F* in the ANOVA corresponds to a *p-value* from the F-distribution. This *p-value* is often called the *statistical significance* of *F* and is used in the present study to determine whether or not to reject the general *null hypothesis* for any OLS regression estimate.

Chapter 4: Results

This chapter is dedicated to present the descriptive characteristics of the data, the graphical representations of the data and the proposed regression models, as well as to give a brief contextualization of them. The implications and interpretations of the findings will be discussed in the next chapter.

4.1 Descriptive Data

Abbreviations used for the variables, as well as a description and their detail level, is described in table 4.1.1 below:

Table 4.1.1 - Abbreviations

Data set	Abbreviation	Description	Detail level*
Hotel	RevPAR	Hotel revenue divided by number of hotel rooms available, in NOK	M, M
	AH	Number of available hotels	M, M
	AB	Number of available hotel beds	M, M
	AR	Number of available hotel rooms	M, M
	DV	Domestic visitors	M, M
	IV	International visitors	M, M
Population	AvP	Average population, in number of people, calculated as average of	M, Y
	LA	Land area of region, in km2	M, C
	PUA	People living in urban/built-up areas	C, Y
GNP	GNPB	Gross national product basis value, in million NOK	C, Y
	GNPpc	GNP per capita, GNP divided by population	C, Y
Currency	CW	Currency weighted; GBP, SEK, DKK, EUR and USD weighted by origin country of visitors	C, M
Flight	SD	Schedueled domestic flight passengers	C, M
	SI	Schedueled international passengers	C, M
Oil Price	BP	The price per barrel of brent oil	N, M
Interest Rate	KPR	Key Policy Rate, as set by the Norwegian bank, in percent	N, M

* Detail level is presented with regionality (M=municipality, C=county, N=national) first, and then regularity (M=monthly, Y=yearly, C=constant).

Table 4.1.2 – Descriptive Statistics

Level	Data set	Mean	Min	Max	Std.Dev	n
Municipality	RevPAR	395.01	6.00	1,553.00	174.18	18,391
	AH	9.03	1.00	60.00	7.65	18,391
	AB	1,401.21	50.00	14,105.00	1,701.28	18,391
	AR	647.25	24.00	6,721.00	818.51	18,391
	DV	7,108.08	20.81	159,332.90	10,649.63	17,173
	IV	2,546.03	0.92	123,249.63	6,352.83	17,173
	AvP	37,354.66	918.50	366,509.00	53,756.98	1,826
	LA	2,473.69	58.00	18,818.00	2,937.36	166
County	RevPAR	429.35	171.00	1,116.00	138.72	2,508
	AH	50.89	18.00	145.00	22.31	2,508
	AB	7,994.12	1,718.00	25,993.00	4,858.94	2,508
	AR	3,747.65	686.00	13,356.00	2,462.06	2,508
	DV	40,968.70	4,807.12	186,867.61	31,802.88	2,415
	IV	14,938.17	248.78	180,351.79	22,947.22	2,415
	AvP	266,461.11	72,445.50	677,268.00	156,013.41	209
	LA	16,006.63	426.00	45,755.00	11,464.84	19
	PUA	213,810.05	53,194.00	668,752.00	156,147.72	190
	GNPB	109,941.25	19,429.00	549,572.00	100,594.78	190
	CW	5.95	1.63	10.03	1.43	2,508
	SD	190,079.26	11.00	764,450.00	210,134.79	875
	SI	109,850.12	1.00	1,572,111.00	237,544.90	905
National	BP	522.07	271.99	712.78	114.70	132
	KPR	1.67	0.50	5.75	1.30	132

Table 4.1.2 shows descriptive statistics for the variables used in the regression models. The variables highlighted in grey, indicates that their data set has been adapted to fit the corresponding level. In this case, the descriptive statistics for the variables from RevPAR to LA, is presented on both municipality level and aggregated to county level. The number of observations, n, only accounts for number of unique data points (duplicates removed). For variables with detail level "M, M" (166 municipalities, 12 months), over the course of 11 years, the resulting maximum observations is defined as $n = 166 * 12 * 11 = 21\,912$. Similarly, for data with detail level "C, M" (19 counties, 12 months), the number of maximum observations is $n = 19 * 12 * 11 = 2508$. However, as many of the variables include 0-valued and other mis-measured data points that will be omitted, the actual number of observations for the variables are usually smaller than the maximum.

Table 4.1.3 – Correlation Matrix for Municipality (monthly data points)

	RevPAR	AH	AB	AR	DV	IV	AvP	LA	PUA	GNPB	CW	SD	SI	BP	KPR
RevPAR	1.000														
AH	0.193	1.000													
AB	0.270	0.842	1.000												
AR	0.311	0.800	0.976	1.000											
DV	0.410	0.703	0.905	0.935	1.000										
IV	0.393	0.561	0.670	0.680	0.687	1.000									
AvP	0.296	0.555	0.725	0.805	0.752	0.488	1.000								
LA	-0.085	0.362	0.068	0.052	0.021	0.042	-0.029	1.000							
PUA	-	-	-	-	-	-	-	-	-						
GNPB	0.273	0.143	0.291	0.350	0.330	0.238	0.416	-0.224	-	1.000					
CW	0.263	0.119	0.079	0.131	0.141	0.230	0.099	0.129	-	0.288	1.000				
SD	0.246	0.409	0.671	0.678	0.704	0.498	0.289	-0.266	-	0.592	-0.110	1.000			
SI	0.120	0.138	0.447	0.447	0.514	0.445	0.044	-0.251	-	0.510	-0.200	0.893	1.000		
BP	-0.029	0.004	-0.007	-0.014	-0.005	-0.013	-0.018	-0.021	-	-0.042	-0.254	-0.004	-0.015	1.000	
KPR	-0.058	-0.012	-0.046	-0.054	-0.048	-0.059	-0.038	-0.070	-	-0.127	-0.193	-0.012	-0.022	0.244	1.000

Table 4.1.3 shows all the pairs of correlations between the different variables, using data at “M, M” (municipality, monthly) detail level. The variables highlighted in grey have been adapted to the appropriate detail level. In this case, county level data and data with lower regularity, has been duplicated down to fit the “M, M”-structure, and is therefore less accurate.

Table 4.1.4 – Correlation Matrix for County (monthly data points)

	RevPAR	AH	AB	AR	DV	IV	AvP	LA	PUA	GNPB	CW	SD	SI	BP	KPR
RevPAR	1.000														
AH	0.202	1.000													
AB	0.439	0.708	1.000												
AR	0.516	0.619	0.974	1.000											
DV	0.649	0.510	0.897	0.937	1.000										
IV	0.620	0.438	0.686	0.726	0.741	1.000									
AvP	0.486	0.338	0.791	0.861	0.835	0.578	1.000								
LA	-0.194	0.185	-0.268	-0.301	-0.322	-0.186	-0.526	1.000							
PUA	0.470	0.205	0.707	0.796	0.774	0.573	0.932	-0.548	1.000						
GNPB	0.529	0.317	0.828	0.917	0.878	0.662	0.923	-0.471	0.903	1.000					
CW	0.373	0.184	0.150	0.230	0.242	0.352	0.126	0.129	0.131	0.215	1.000				
SD	0.496	0.288	0.821	0.842	0.866	0.593	0.863	-0.383	0.831	0.844	0.024	1.000			
SI	0.408	0.084	0.757	0.766	0.808	0.638	0.819	-0.529	0.846	0.809	-0.081	0.902	1.000		
BP	0.015	0.035	0.010	0.000	0.017	0.003	-0.006	-0.001	-0.010	-0.025	-0.277	0.019	0.008	1.000	
KPR	-0.032	0.006	-0.083	-0.080	-0.069	-0.077	-0.040	-0.004	-0.043	-0.101	-0.169	-0.027	-0.035	0.233	1.000

Table 4.1.4 shows all the pairs of correlation between the different variables, using data at “C, M” (county, monthly) detail level. The variables highlighted in grey have been calculated from a smaller data set, as flight data only were collected from 2014 to 2018 and may therefore be less accurate.

Table 4.1.5 – Correlation Matrix for Municipality (yearly data points)

	RevPAR	AH	AB	AR	DV	IV	AvP	LA	PUA	GNPB	CW	SD	SI	BP	KPR
RevPAR	1.000														
AH	0.209	1.000													
AB	0.347	0.843	1.000												
AR	0.410	0.800	0.978	1.000											
DV	0.462	0.719	0.944	0.976	1.000										
IV	0.386	0.710	0.863	0.873	0.847	1.000									
AvP	0.442	0.567	0.733	0.809	0.791	0.661	1.000								
LA	-0.129	0.368	0.060	0.045	0.012	0.054	-0.033	1.000							
PUA	-	-	-	-	-	-	-	-	-						
GNPB	0.400	0.155	0.309	0.366	0.364	0.331	0.423	-0.225	-	1.000					
CW	0.369	0.148	0.108	0.172	0.155	0.225	0.144	0.181	-	0.376	1.000				
SD	0.384	0.454	0.723	0.729	0.789	0.695	0.374	-0.237	-	0.571	-0.060	1.000			
SI	0.213	0.169	0.476	0.478	0.552	0.519	0.100	-0.203	-	0.493	-0.250	0.910	1.000		
BP	-0.082	-0.008	-0.015	-0.024	-0.021	-0.059	-0.025	-0.027	-	-0.050	-0.402	-0.006	-0.008	1.000	
KPR	-0.101	-0.023	-0.059	-0.067	-0.062	-0.086	-0.046	-0.071	-	-0.139	-0.244	-0.004	-0.003	0.250	1.000

Table 4.1.5 shows all the pairs of correlations between the different variables, using data at “M, Y” (municipality, yearly) detail level. The variables highlighted in grey have been adapted to the appropriate detail level. In this case, county level data has been duplicated down to fit the “M, Y”-structure, and is therefore less accurate. Data sets that have a higher regularity than yearly, has been aggregated/averaged to yearly detail level.

Table 4.1.6 – Correlation Matrix for County (yearly data points)

	RevPAR	AH	AB	AR	DV	IV	AvP	LA	PUA	GNPB	CW	SD	SI	BP	KPR
RevPAR	1.000														
AH	0.184	1.000													
AB	0.555	0.709	1.000												
AR	0.675	0.620	0.974	1.000											
DV	0.742	0.512	0.936	0.979	1.000										
IV	0.727	0.494	0.858	0.913	0.897	1.000									
AvP	0.715	0.351	0.803	0.870	0.899	0.773	1.000								
LA	-0.291	0.184	-0.274	-0.304	-0.349	-0.250	-0.525	1.000							
PUA	0.689	0.212	0.715	0.801	0.832	0.764	0.930	-0.545	1.000						
GNPB	0.775	0.330	0.839	0.925	0.945	0.887	0.923	-0.469	0.901	1.000					
CW	0.549	0.194	0.175	0.273	0.260	0.357	0.153	0.167	0.162	0.267	1.000				
SD	0.748	0.267	0.805	0.840	0.904	0.775	0.868	-0.350	0.836	0.857	0.164	1.000			
SI	0.637	0.066	0.745	0.766	0.843	0.731	0.841	-0.521	0.871	0.833	-0.074	0.922	1.000		
BP	-0.056	0.024	-0.003	-0.013	-0.013	-0.056	-0.011	0.000	-0.015	-0.033	-0.404	-0.050	-0.041	1.000	
KPR	-0.061	0.013	-0.087	-0.084	-0.081	-0.088	-0.047	0.000	-0.051	-0.107	-0.197	-0.064	-0.053	0.254	1.000

Table 4.1.6 shows all the pairs of correlation between the different variables, using data at “C, Y” (county, yearly) detail level. The variables highlighted in grey have been calculated from a smaller data set, as flight data only were collected from 2014 to 2018 and may therefore be less accurate.

Table 4.1.7 – Autocorrelation Matrix for Municipality (yearly data points)

	RevPAR	AH	AB	AR	DV	IV	AvP	LA
r1	0.912	0.994	0.991	0.995	0.995	0.978	1.000	1.000
r2	0.830	0.980	0.976	0.989	0.987	0.961	1.000	1.000
r3	0.784	0.966	0.964	0.985	0.986	0.950	1.000	1.000
r4	0.730	0.951	0.946	0.978	0.982	0.938	1.000	1.000
r5	0.721	0.933	0.926	0.969	0.981	0.928	1.000	1.000
r6	0.687	0.916	0.909	0.963	0.982	0.948	1.000	1.000

Table 4.1.8 – Autocorrelation Matrix for County (yearly data points)

	RevPAR	AH	AB	AR	DV	IV	AvP	LA	PUA	GNPB	CW	SD	ST	BP	KPR
r1	0.941	0.997	0.995	0.997	0.991	0.978	1.000	1.000	0.999	0.999	0.938	1.000	1.000	0.388	0.990
r2	0.867	0.984	0.985	0.993	0.992	0.980	1.000	1.000	0.999	0.998	0.876	1.000	0.999	-0.217	0.969
r3	0.808	0.966	0.971	0.990	0.991	0.974	1.000	1.000	0.998	0.998	0.843	1.000	0.998	-0.547	0.938
r4	0.779	0.941	0.951	0.983	0.990	0.969	1.000	1.000	0.998	0.997	0.851	0.999	0.998	-0.778	0.902
r5	0.780	0.912	0.929	0.974	0.987	0.972	0.999	1.000	0.998	0.997	0.766	-	-	-0.423	0.864
r6	0.749	0.884	0.906	0.965	0.986	0.970	0.999	1.000	0.998	0.998	0.778	-	-	0.511	0.826

Tables 4.1.7 and 4.1.8 shows the values from the autocorrelation function, to a maximum lag length of 6 years, using yearly data for both county and municipality levels. The variables highlighted in grey have been calculated from a smaller data set, as flight data only were collected from 2014 to 2018 and could therefore only be accurately calculated to a maximum lag length of 4 years.

4.2 Graphical Data

This section is dedicated to highlight the time series for some of the variables, to better show the relation between the volatile monthly and the more stable yearly data. This section will also, present some useful visual representations of correlations and national tendencies.

Figure 4.2.1

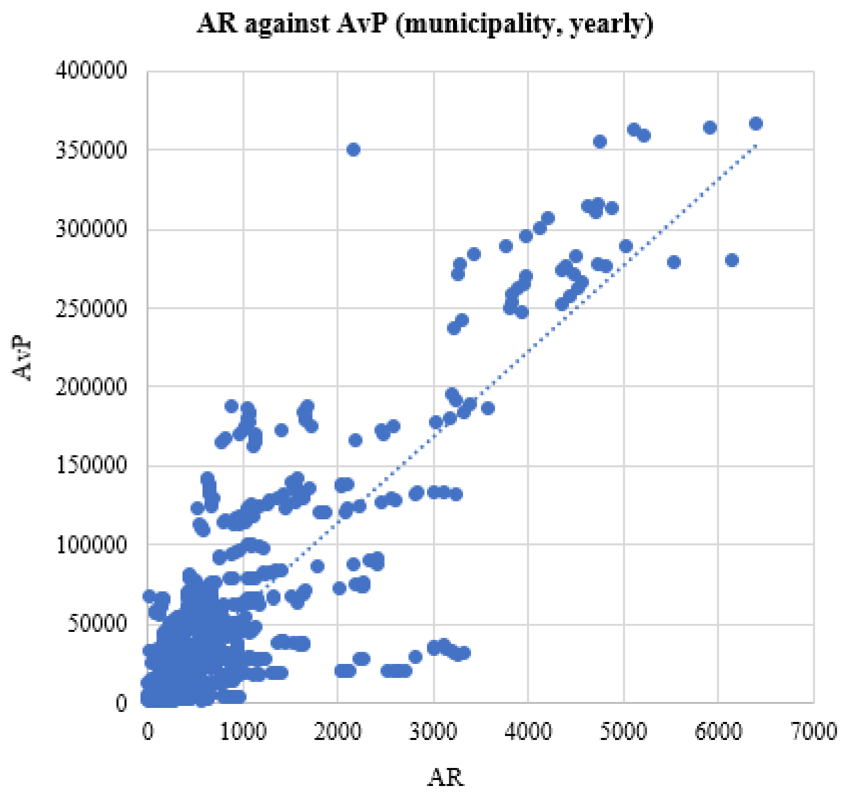


Figure 4.2.2

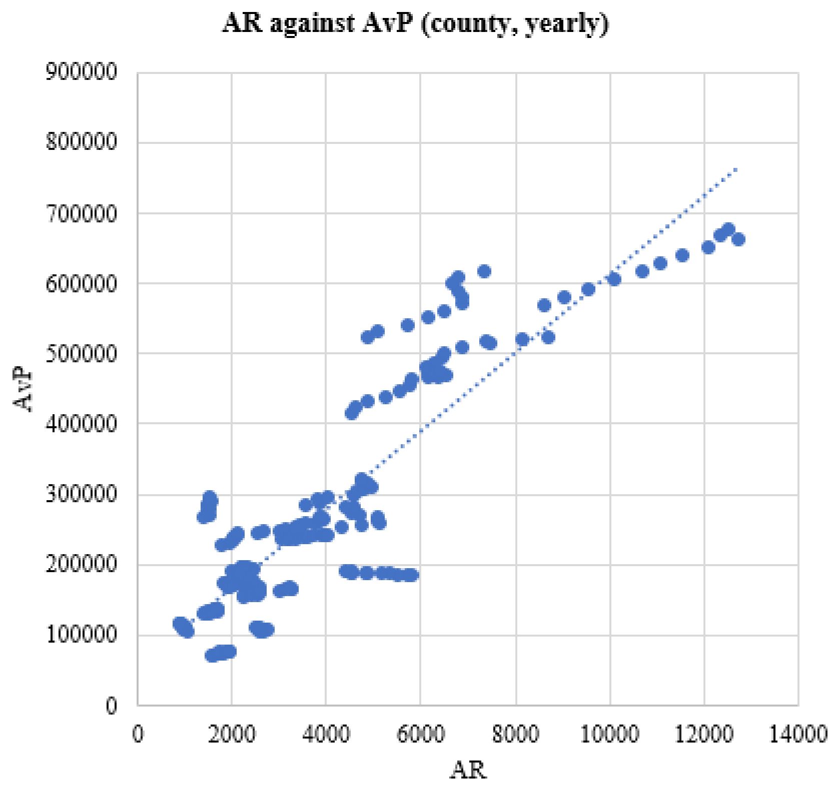


Figure 4.2.3

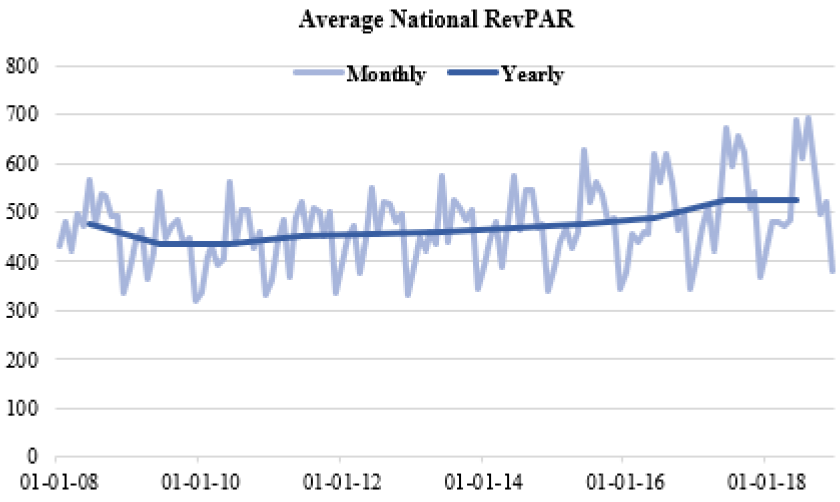


Figure 4.2.4

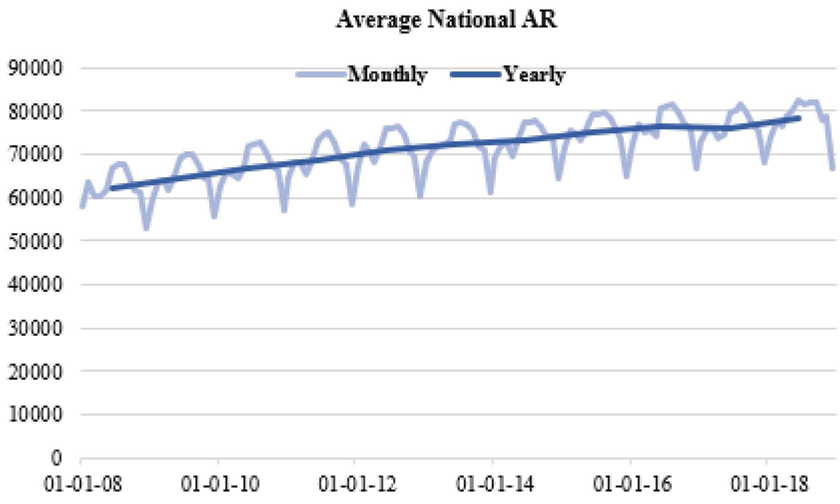


Figure 4.2.5

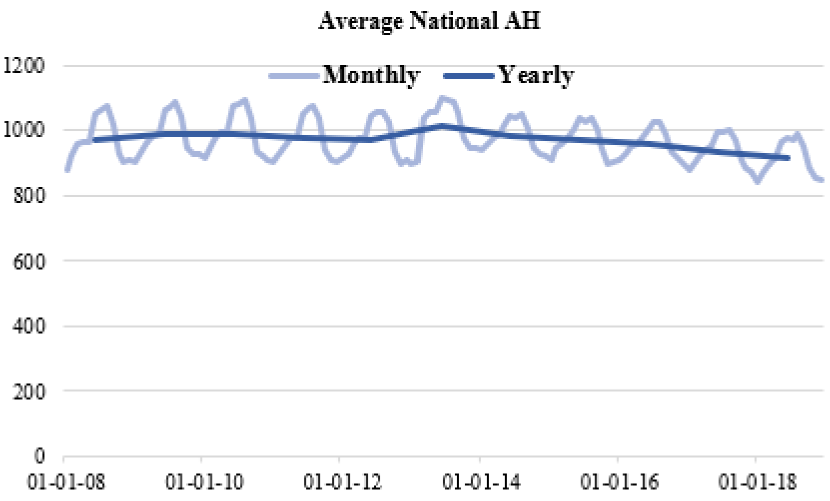


Figure 4.2.6

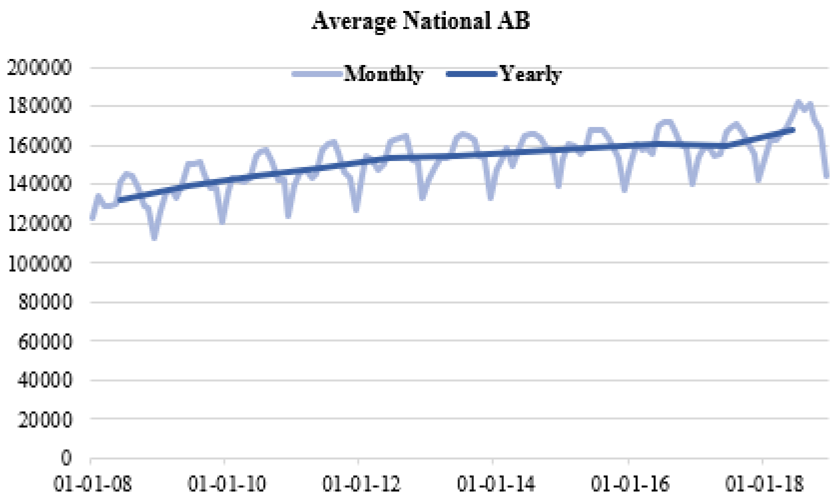


Figure 4.2.5

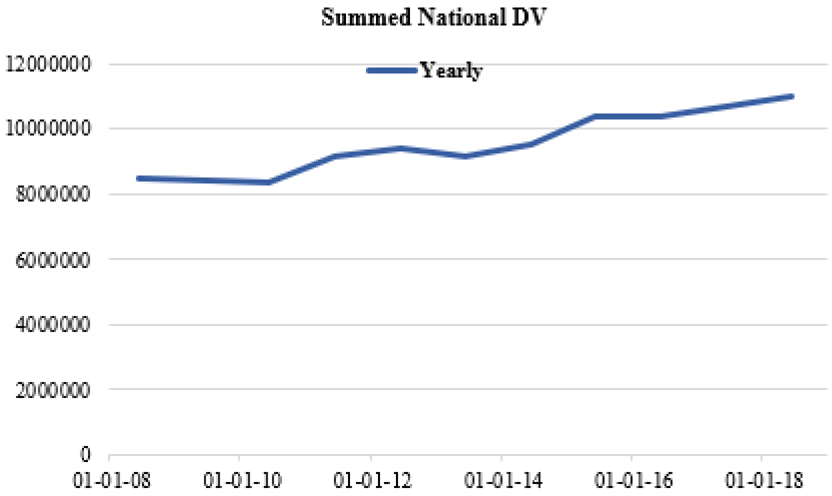


Figure 4.2.6

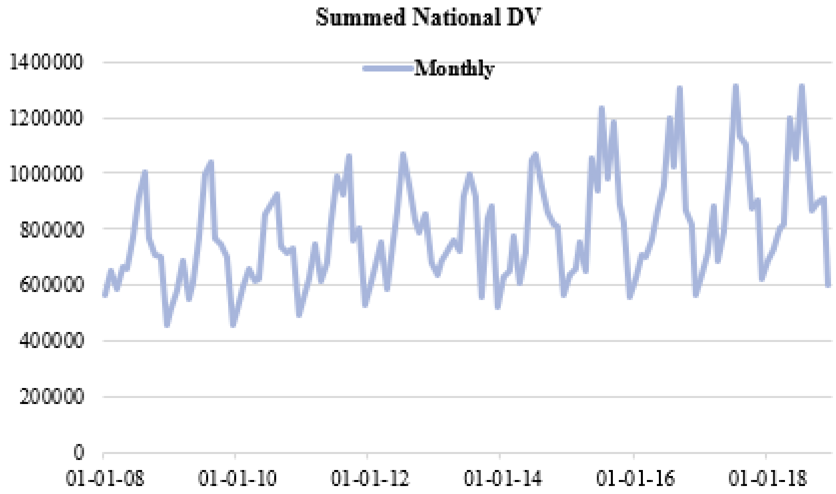


Figure 4.2.7

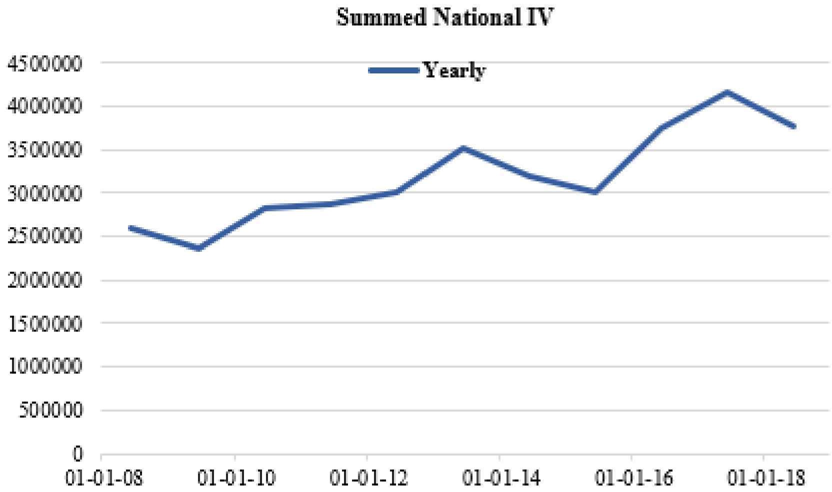


Figure 4.2.7

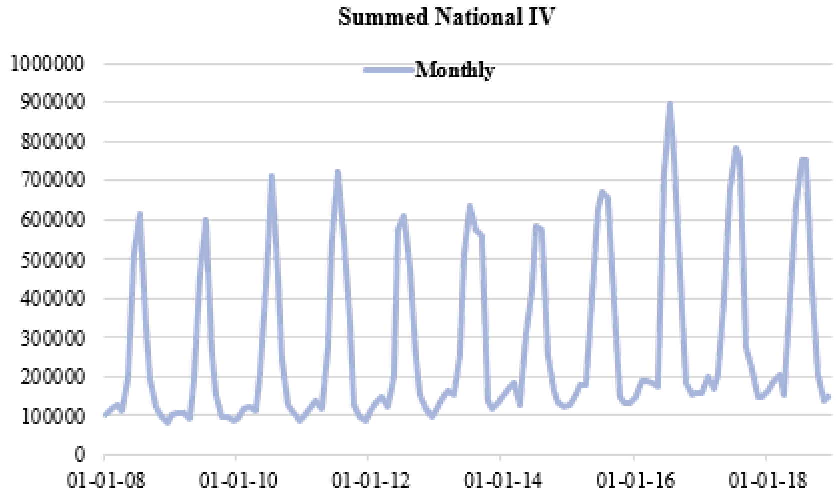


Figure 4.2.8

Average Number of Hotels (by County)

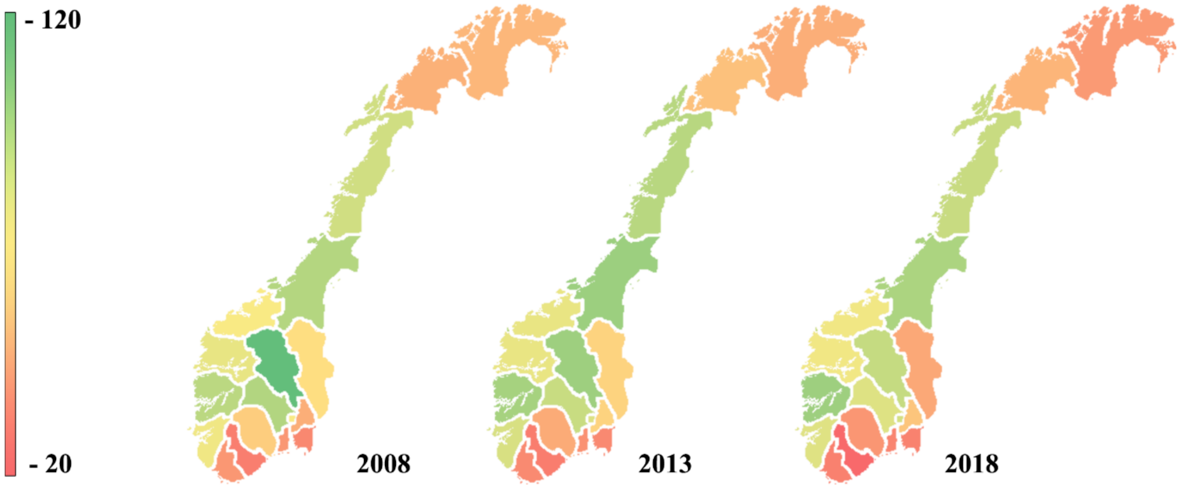


Figure 4.2.9

Average Number of Hotel Rooms (by County)

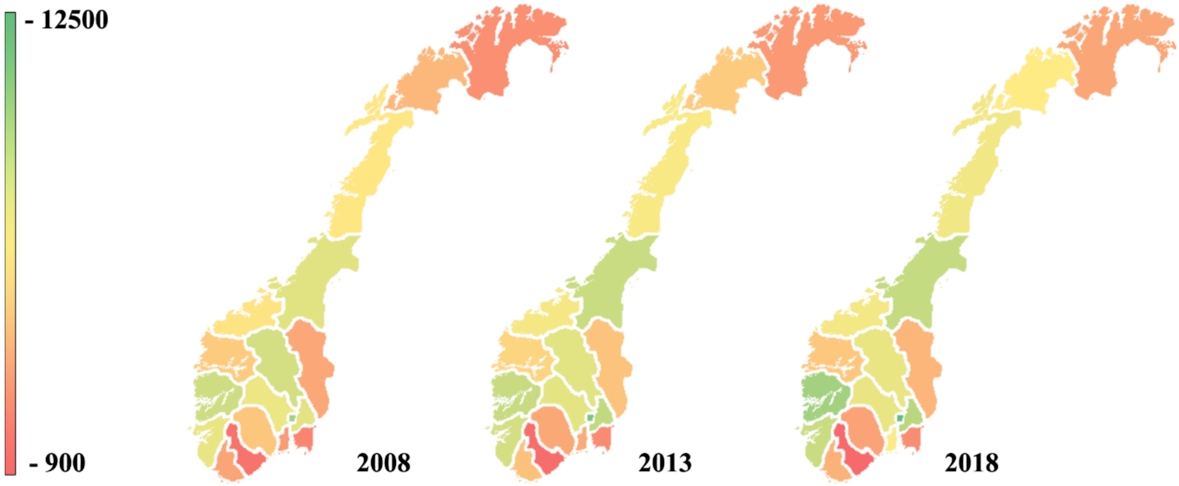


Figure 4.2.10

Average Size of Hotel Rooms (in Beds per Room, by County)

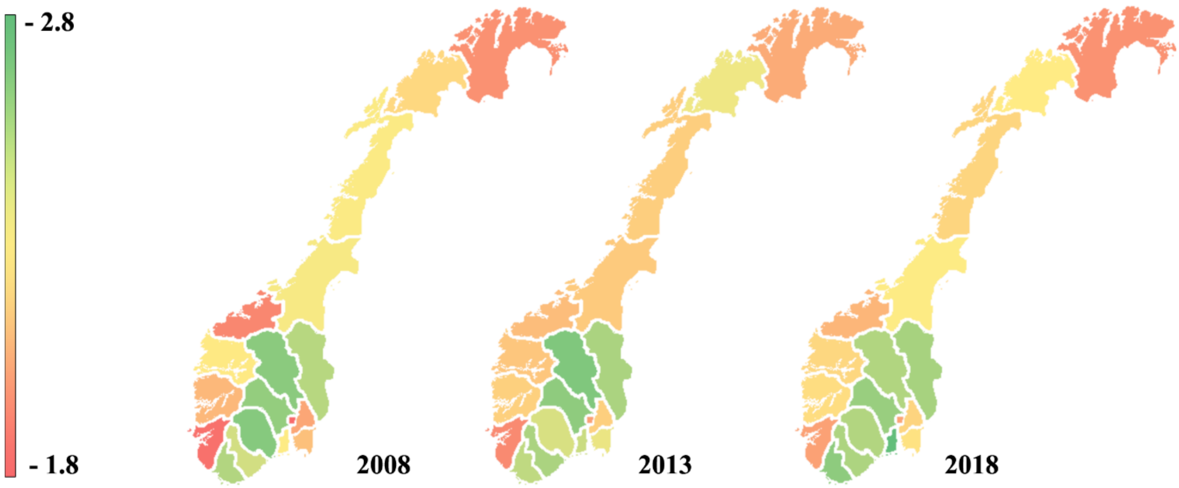


Figure 4.2.11

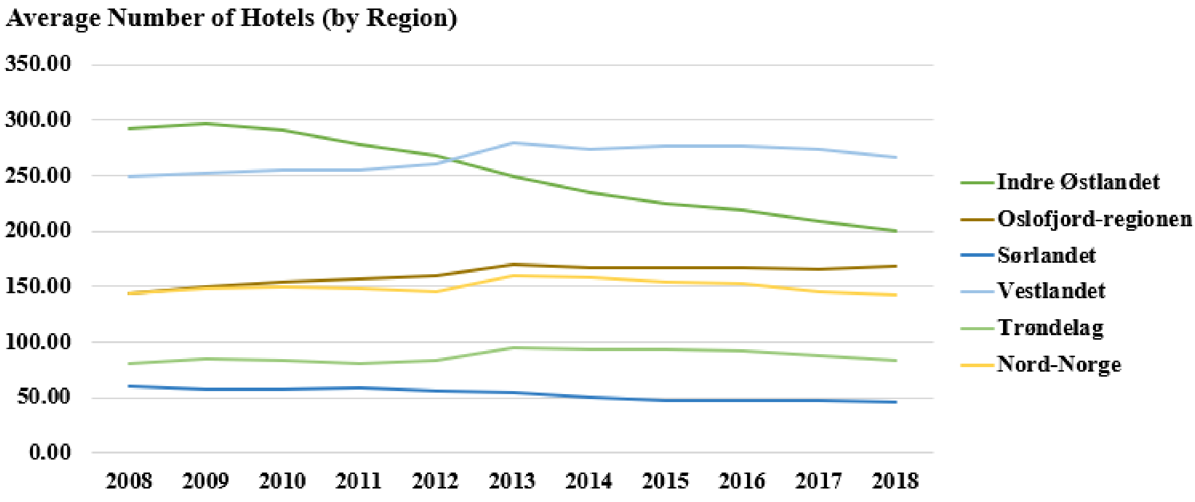


Figure 4.2.12

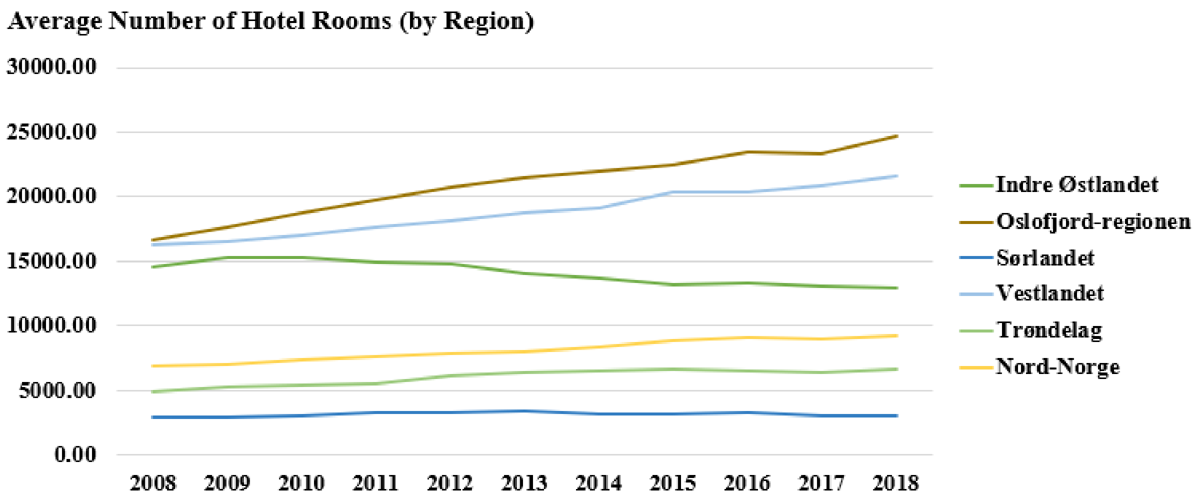
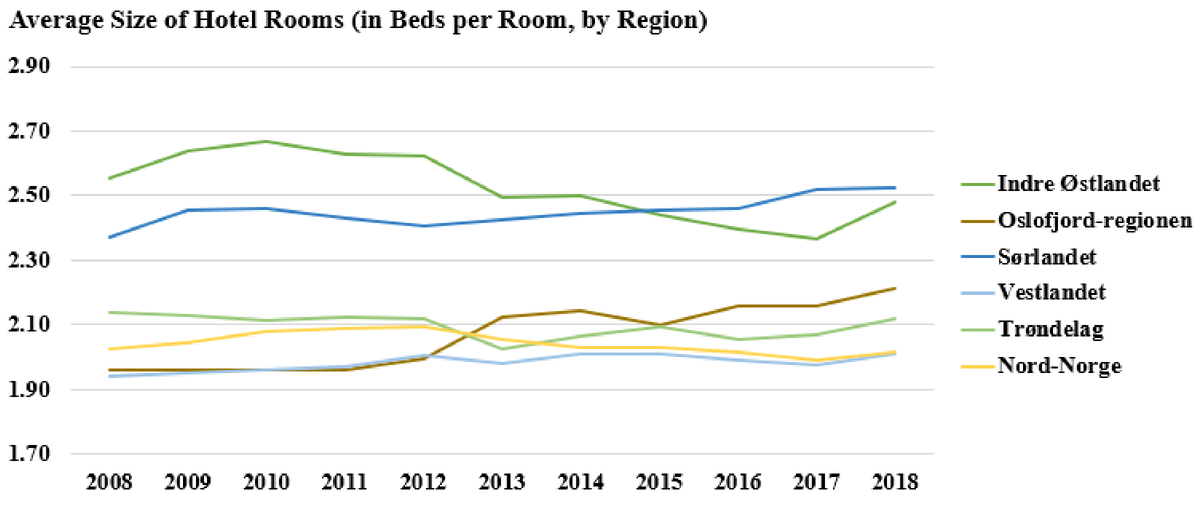


Figure 4.2.13



4.3 Multiple Regression Models

Model 4.3.1 - RevPAR

The log-transformed autoregressive model with a lag length of m=1 was used to model the relationship between the performance metric RevPAR at time t, and its preceding values at time t -1. Relations for both county and municipality levels were estimated through pooled OLS regression of the following proposed model:

$$ln(RevPAR_t) = a + b\, ln(RevPAR_{t-1}) + \epsilon_t \qquad \textbf{(equation 4.3.1)}$$

Table 4.3.1A – RevPAR yearly time lag (county)

Regression Statistics				
Multiple R	*R Square*	*Adjusted R Square*	*Observations*	*Significance F*
0.951025222	0.904448973	0.903803358	150	2.37135E-77

	Coefficients	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.413958887	2.742236352	0.006855603	0.115650002	0.712267772
RevPAR t-1	0.935512119	37.4287404	2.37135E-77	0.886119954	0.984904285

Table 4.3.1B – RevPAR yearly time lag (municipality)

Regression Statistics				
Multiple R	*R Square*	*Adjusted R Square*	*Observations*	*Significance F*
0.838454486	0.703005925	0.702814439	1553	0

	Coefficients	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	53.98236456	9.55802709	4.50698E-21	42.90412432	65.0606048
RevPAR t-1	0.856618017	60.59147134	0	0.828887189	0.884348845

Both estimates are found statistically significant at a 99% confidence level, using the F value of the regression, and feature R square values of 0.90 and 0.70, for county and municipality respectively.

Model 4.3.2 – Demand

Two iterations of the log-transformed multiple regression model were used to model the relationship between the "demand"-group variables and the hotel capacity, measured as available rooms (AR). Relations for both county and municipality levels were estimated through pooled OLS regression of the following proposed models:

(equation 4.3.2.1)

$$ln(AR) = a + b_1\,ln(DV) + b_2\,ln(IV) + b_3\,ln(AvP) + b_4\,ln(SD) + b_5\,ln(SI) + \epsilon$$

$$ln(AR) = a + b_1\,ln(AvP) + b_2\,ln(SD) + b_3\,ln(SI) + \epsilon$$ **(equation 4.3.2.2)**

Results presented in table 4.3.2 A and B are estimates of equation 4.3.2.1 (part 1), and results presented in table 4.3.2 C and D are estimates of equation 4.3.2.2 (part 2).

Table 4.3.2A – Demand effect on AR (county)

Regression Statistics

Multiple R	*R Square*	*Adjusted R Square*	*Observations*	*Significance F*
0.984076099	0.968405768	0.965634344	63	2.01922E-41

	Coefficients	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-3.906913409	-8.889752418	2.33838E-12	-4.786967069	-3.026859749
DV	0.636197476	9.17999792	7.85667E-13	0.497421468	0.774973485
IV	0.124138402	5.124219281	3.70058E-06	0.075627039	0.172649765
AvP	0.206845259	3.017324959	0.003807291	0.069571186	0.344119331
SD	0.006136079	0.94546887	0.348410361	-0.006859892	0.019132051
SI	-0.026029289	-2.833981678	0.006347763	-0.044421346	-0.007637232

Table 4.3.2B – Demand effect on AR (municipality)

Regression Statistics

Multiple R	*R Square*	*Adjusted R Square*	*Observations*	*Significance F*
0.983621067	0.967510403	0.965869514	105	5.55604E-72

	Coefficients	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-3.376536101	-13.5586895	2.67368E-24	-3.870667939	-2.882404263
DV	0.674698326	17.19712735	1.70127E-31	0.596851146	0.752545505
IV	0.137521924	5.343403175	5.84325E-07	0.086454594	0.188589254
AvP	0.09264886	3.331304918	0.001215939	0.037464648	0.147833073
SD	-0.014552128	-1.301483957	0.196113406	-0.036738018	0.007633761
SI	0.004319462	0.694551037	0.488963494	-0.008020524	0.016659448

Table 4.3.2C – Demand effect on AR (county)

Regression Statistics

Multiple R	*R Square*	*Adjusted R Square*	*Observations*	*Significance F*
0.898054487	0.806501861	0.796826954	64	2.24731E-21

	Coefficients	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-4.737988032	-4.788384161	1.1399E-05	-6.717233394	-2.75874267
AvP	1.06545103	11.4532794	9.79764E-17	0.879371629	1.251530432
SD	0.068625897	5.389002194	1.25219E-06	0.043153235	0.094098559
SI	-0.096256069	-4.854181824	8.98898E-06	-0.135921003	-0.056591134

Table 4.3.2D – Demand effect on AR (municipality)

Regression Statistics

Multiple R	*R Square*	*Adjusted R Square*	*Observations*	*Significance F*
0.885364933	0.783871064	0.777451392	105	1.81379E-33

	Coefficients	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-0.012691884	-0.025221015	0.979928439	-1.010958007	0.985574238
AvP	0.431323615	7.106300447	1.7245E-10	0.310919189	0.551728041
SD	0.136362959	5.996806292	3.14716E-08	0.091254377	0.181471542
SI	0.045216114	3.000398262	0.003396406	0.015321214	0.075111014

All four estimates are found statistically significant at a 99% confidence level, using the F value of the regression, and feature R square values from 0.77 to 0.96. However, from the individual t-testing, not all variables were shown to have a statistically significant (at least 95% confidence level) effect different from the trivial relation (b =0).

Model 4.3.3 – Financial and Economic

The log-transformed multiple regression model was used to model the relationship between the “financial and economic”-group variables and the hotel capacity, measured as available rooms (AR). Relations for both county and municipality levels were estimated through pooled OLS regression of the following proposed model:

(equation 4.3.3)

$$ln(AR) = a + b_1\, ln(GNPB) + b_2\, ln(CW) + b_3\, ln(BP) + b_4\, ln(KPR) + \epsilon$$

Table 4.3.3A – Financial effect on AR (county)

Regression Statistics

Multiple R	*R Square*	*Adjusted R Square*	*Observations*	*Significance F*
0.848121823	0.719310627	0.713241667	190	6.17219E-50

	Coefficients	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-1.080131338	-1.154747006	0.249683437	-2.925521405	0.765258729
GNPB	0.691535909	20.4749592	1.9893E-49	0.624902789	0.758169028
CW	0.383218577	2.922808792	0.003901751	0.124549445	0.641887709
BP	0.094824377	0.719687581	0.472625577	-0.165116421	0.354765175
KPR	0.07610744	1.815867194	0.071009724	-0.00658037	0.158795249

Table 4.3.3B – Financial effect on AR (municipality)

Regression Statistics

Multiple R	*R Square*	*Adjusted R Square*	*Observations*	*Significance F*
0.242947914	0.059023689	0.056435034	1459	2.71815E-18

	Coefficients	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.907612571	0.874720319	0.381870562	-1.127746548	2.942971689
GNPB	0.368655157	8.322453162	1.96154E-16	0.281763373	0.455546941
CW	0.306590878	1.980174375	0.047872197	0.00287636	0.610305397
BP	0.048165637	0.33770604	0.73563347	-0.231608937	0.327940211
KPR	0.020709146	0.447792029	0.654370004	-0.070009296	0.111427587

Both estimates are found statistically significant at a 99% confidence level, using the F value of the regression, and feature R square values of 0.71 and 0.05, for county and municipality respectively. However, from the individual t-testing, not all variables were shown to have a statistically significant (at least 95% confidence level) effect different from the trivial relation (b =0).

Model 4.3.4 - Urbanization

The log-transformed multiple regression model was used to model the relationship between the "urbanization"-group variables and the hotel capacity, measured as available rooms (AR). Only relationships for county level were estimated through pooled OLS regression of the following proposed model:

$$ln(AR) = a + b_1\, ln(PUA) + b_2\, ln(GNP_{pc}) + \epsilon \qquad \textbf{(equation 4.3.4)}$$

Table 4.3.4 – Effect of urbanization on AR (county)

Regression Statistics

Multiple R	*R Square*	*Adjusted R Square*	*Observations*	*Significance F*
0.760939289	0.579028602	0.574017038	171	2.73664E-32

	Coefficients	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-13.77329491	-7.942624605	2.71088E-13	-17.19672686	-10.34986296
PUA	0.299493631	5.766392389	3.79719E-08	0.196958831	0.402028431
GNPpc	1.42101959	9.290442861	7.82644E-17	1.119058115	1.722981066

Model 4.3.5 – Urbanization differenced

The log-differenced multiple regression model was used to model the relationship between the "urbanization"-group variables and the hotel capacity, measured as available rooms (AR). Only relationships for county level were estimated through pooled OLS regression of the following proposed model:

$$\Delta AR_{\mathrm{t}} = a + b_1 \Delta PUA_{\mathrm{t}} + b_2 \Delta GNP_{pc,\mathrm{t}} + \epsilon \qquad \textbf{(equation 4.3.5)}$$

Table 4.3.5 – Effect of differenced urbanization on AR (county)

Regression Statistics

Multiple R	*R Square*	*Adjusted R Square*	*Observations*	*Significance F*
0.081125604	0.006581364	-0.008702	133	0.651028043

	Coefficients	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.012502735	1.315744994	0.190575257	-0.006296626	0.031302095
dPUA	0.344062048	0.67023136	0.503899178	-0.671535931	1.359660028
dGNPpc	-0.102137366	-0.656593673	0.512602824	-0.409887155	0.205612423

This estimate is not found statistically significant (at least 95% confidence level) and should be rejected.

Model 4.3.6 – Demand differenced

Two iterations of the log-differenced multiple regression model were used to model the relationship between the "demand"-group variables and the hotel capacity, measured as available rooms (AR). Relations for both county and municipality levels were estimated through pooled OLS regression of the following proposed models:

(equation 4.3.6.1)

$$\Delta AR = a + b_1\Delta DV + b_2\Delta IV + b_3\Delta AvP + b_4\Delta SD + b_5\Delta SI + \epsilon$$

$$\Delta AR = a + b_3\Delta AvP + b_4\Delta SD + b_5\Delta SI + \epsilon \qquad \textbf{(equation 4.3.6.2)}$$

Results presented in table 4.3.6 A and B are estimates of equation 4.3.6.1, and results presented in table 4.3.6 C and D are estimates of equation 4.3.6.2.

Table 4.3.6A – Differenced demand effect on AR (county)

Regression Statistics

Multiple R	*R Square*	*Adjusted R Square*	*Observations*	*Significance F*
0.370435387	0.137222376	0.034510754	48	0.267919445

	Coefficients	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.016188417	1.010852438	0.317875407	-0.016130393	0.048507226
dDV	-0.077534645	-0.65007045	0.519186838	-0.318233555	0.163164265
dIV	0.037181458	1.456699381	0.152635368	-0.014328978	0.088691893
dAvP	1.040842639	0.604690639	0.548637689	-2.432843493	4.514528772
dSD	0.009205837	0.517298712	0.607661509	-0.026707903	0.045119578
dSI	-0.021145722	-0.570832422	0.571155053	-0.095902853	0.053611409

Table 4.3.6B – Differenced demand effect on AR (municipality)

Regression Statistics

Multiple R	*R Square*	*Adjusted R Square*	*Observations*	*Significance F*
0.529013136	0.279854898	0.226903052	74	0.000368705

	Coefficients	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.0296476	1.768906136	0.081393908	-0.003797286	0.063092486
dDV	0.52214055	4.144605083	9.64106E-05	0.270749833	0.773531268
dIV	0.09969542	2.059265725	0.043300776	0.0030886	0.196302239
dAvP	-2.755032797	-2.132970363	0.03653953	-5.332463055	-0.177602539
dSD	0.005770633	0.187195177	0.852065024	-0.055743339	0.067284605
dSI	-0.011032204	-0.670557226	0.504773787	-0.043862245	0.021797836

Table 4.3.6C – Differenced demand effect on AR (county)

Regression Statistics

Multiple R	*R Square*	*Adjusted R Square*	*Observations*	*Significance F*
0.112057698	0.012556928	-0.05327261	48	0.902155177

	Coefficients	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.016896979	1.046696152	0.300829506	-0.015617005	0.049410963
dAvP	0.844871332	0.479936191	0.63359744	-2.700721281	4.390463944
dSD	0.005239318	0.2856334	0.776469623	-0.031704992	0.042183629
dSI	-0.022324837	-0.583087935	0.562745022	-0.099439329	0.054789655

Table 4.3.6D – Differenced demand effect on AR (municipality)

Regression Statistics

Multiple R	*R Square*	*Adjusted R Square*	*Observations*	*Significance F*
0.312629629	0.097737285	0.059068883	74	0.064337978

	Coefficients	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.056696334	3.356305818	0.001279535	0.023005342	0.090387326
dAvP	-3.642519799	-2.592380013	0.011594938	-6.444877721	-0.840161876
dSD	0.024422036	0.725921372	0.470307183	-0.042676436	0.091520507
dSI	0.014320543	0.87493653	0.384600631	-0.01832345	0.046964536

Out of the four estimates, only 4.3.6B was found statistically significant at a confidence level of at least 95%, with a confidence level of 99%. However, from the individual t-testing, not all variables were shown to have a statistically significant effect different from the trivial relation (b =0). The results for 4.3.6D are significant at 90%, but the other estimates are found insignificant and should be rejected.

Model 4.3.7 – Financial and Economic differenced

The log-differenced multiple regression model was used to model the relationship between the "financial and economic"-group variables and the hotel capacity, measured as available rooms (AR). Only relationships for county level were estimated through pooled OLS regression of the following proposed model:

(equation 4.3.7)

$$\Delta AR = a + b_1\Delta GNPB + b_2\Delta CW + b_3\Delta BP + b_4\Delta KPR + \epsilon$$

Table 4.3.7A – Differenced financial effect on AR (county)

Regression Statistics

Multiple R	*R Square*	*Adjusted R Square*	*Observations*	*Significance F*
0.220423809	0.048586656	0.025661033	171	0.080620815

	Coefficients	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.012145005	1.329874674	0.185383975	-0.005885705	0.030175715
dGNPB	0.105935628	0.748459223	0.455242344	-0.173511566	0.385382823
dCW	-0.144254446	-2.089137262	0.038220634	-0.280583394	-0.007925498
dBP	-0.032448276	-0.937170106	0.350032537	-0.100807805	0.035911252
dKPR	-0.007016623	-0.397310836	0.691648651	-0.041884348	0.027851102

Table 4.3.7B – Differenced financial effect on AR (municipality)

Regression Statistics

Multiple R	*R Square*	*Adjusted R Square*	*Observations*	*Significance F*
0.078394272	0.006145662	0.00305676	1292	0.093799822

	Coefficients	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-0.012184057	-0.690532067	0.489984163	-0.046799091	0.022430978
dGNPB	0.050615413	0.184884956	0.8533484	-0.486463282	0.587694109
dCW	-0.169312264	-1.293184304	0.196179394	-0.426165301	0.087540772
dBP	0.041521515	0.638508371	0.523256409	-0.086052819	0.16909585
dKPR	-0.044430086	-1.300540501	0.193648611	-0.111450945	0.022590772

Both estimates are found to not be statistically significant (at least 95% confidence level) and should be rejected.

Chapter 5: Discussion and Conclusion

The descriptive statistics of table 4.1.2 shows that the data exhibits a lot of variation in terms of standard deviations and range. Although computed from the pooled data for all municipalities and counties, the descriptive statistics table may still carry some implications of high dispersion for the individual time series data sets. This claim is supported by the time series graphs (figure 4.2.3 through 4.2.7), showing that most of the monthly data is characterized by high seasonality and volatility. The time series graphs also show that the yearly aggregates for the data tend to be much more stable, and better represent the overall trend and development of the data. From the tables 4.1.3 through 4.1.6 (as well as figures 4.2.1 and 4.2.2), showing the paired correlations, an interesting observation is that the relations present in the monthly data is not only also present when looking at correlations of yearly data, but more prevalent in most cases. Thus, the yearly data carry the same relational data, with a lesser degree of uncertainty than that of the monthly data, and so the monthly data was left out of further studies, in favor of the yearly data.

The regression model (model 4.3.1) explores the relation between RevPAR and its past values, and concludes at a 99% confidence level that a 10% increase in $RevPAR_{t-1}$ will cause the $RevPAR_t$ to increase with 9.0% on county level and 8.6% on municipality level. It was expected that RevPAR for a following year, as tend to be true for time series data of economic variables, would be dependent on the value from the preceding year. From the autocorrelation tables 4.1.7 and 4.1.8 it was also expected that the relationship would be stronger on county level, but also that the relations should be fairly similar. The regression results are in line with these intuitions.

The first main regression model (4.3.2) explores the relation between the “demand”-group variables and hotel capacity, over two parts. In the first part (table 4.3.2 A and B), the relationship between visitors, both domestic and international, and hotel capacity is found to be statistically significant at a 99% confidence level. This is in line with Newell and Seabrook’s (2006) previous research on driving factors for hotel investment. However, it is likely that including measures for numbers of visitors in this regression model also introduces the issue of endogeneity. While the number of available hotel rooms certainly depends on the amount of hotel visitors, it is also quite possible that the number of visitors is affected by the number of hotel rooms. There is also, undoubtably, some information contained in the number of visitors that coincide with that of the number of flight passengers. Intuitively, this would be especially true for international hotel visitors, who are most likely to have arrived in Norway by airplane. The amount of flight passengers, both domestic and international, will be closely linked to the average population as well. These claims are supported by the high correlation between the dependent variables presented in tables 4.1.5 and 4.1.6. This introduces a degree of collinearity to the model, which will further lower the validity of the part 1 results. Lastly, it is possible that the inflation of the slope coefficients variances, because of collinearity, is causing the coefficient for domestic flight passengers to appear as statistically insignificant. To minimize the effect of these issues, part 2 of the regression results (table 4.3.2 C and D) excludes the visitor-variables from the model. Thus, these findings are presented as more reliable.

In part 2 of the demand model (4.3.2), all the coefficients were found to be statistically significant, for both county and municipality level. The sign of the effect from the average population and the flight passengers (municipality level only, for international flight passengers), are consistent with general understanding. If the population grows, naturally, the demand for accommodation services would also increase. Similarly, if the number of flight passengers increase the demand for accommodation is also expected to increase. In contrast, the sign for international flight passengers on county level is negative. This implies that as the number of international flights increase, the number of available hotel rooms decrease, which does not compile with general intuition. As the correlation table 4.1.5 and 4.1.6 shows, the correlation between population and flight data is much greater for county level, and so this discrepancy may be a result of the high collinearity in the model.

In the second main regression model (4.3.3) the relation between the "financial and economic"-group variables and hotel capacity is explored. While there are definitely some implications of causality, between the different financial and economic variables, computed correlations between them are still low. As previously mentioned, Norwegian economy is sensitive to developments in the price of oil, and the key policy rate is the central bank's most important tool to help stabilize the price of currency and development in the Norwegian economy. However, these effects are not instant, and takes time to manifest. Therefore, these dependencies are not reflected in the descriptive statistics, and so, collinearity in the model is also expected to be low. The resulting estimates show a significant relationship between hotel capacity and GNPB, weighted currency, and key policy rate* (* 90% confidence level) for county level with a fairly high R square of 0.71. On municipality level, only GNPB and weighted currency was significant. Intuitively, the sign of key policy rate should be negative, as that would imply that investments in hotel capacity would decline as the cost of capital increased. This may be in part a result of the correlation between the dependent variables. Both the coefficient and the values for KPR are very small, so collinearity inflation of variation could cause the KPR to falsely appear positive.

The third, and last main regression model (4.3.4) explores the relation between the "urbanization"-group variables and hotel capacity. As this model included people living in urban areas (PUA) as a variable, which was only available on county level and could not be copied down to municipalities, the model was only estimated for county level. As Luo and Lam (2017) suggested, these variables indicate the level of urbanization, and the model estimates shows a statistically significant effect on hotel capacity for both PUA and GNP per capita. The results show that a 10% increase in PUA and GNP per capita, causes an increase in hotel capacity by 3% and 14%, respectively.

For all non-differenced regression cases, the R square value is higher for county level estimates than for municipality level. This may indicate more volatility, and hence more variance in the municipality data set. Looking at the correlation tables (4.1.3 through 4.1.6), the same remark can also be made for the pairs of correlations, as relationships are generally stronger for yearly county level data than for yearly municipality level data. This would also imply that yearly county level data carries the least uncertainty. As figure 4.2.8 through 4.2.13 helps to show visually, the regional/county development overall has been much more stable than that of each municipality.

The regression estimates of the time differenced (4.3.5 to 4.3.7) models repeatedly yield statistically insignificant results. Except for the model showing the effect of municipality level market demand on hotel capacity, all the differenced estimates fail to reject the null hypothesis for a 95% confidence level. Although these estimates are significant at a 90% confidence level, the individual testing of the coefficients shows that there are, primarily, no non-trivial linear relationship between the dependent variable and the independent variables. As pointed out by Newell and Seabrook (2006), hotel investors were shown to place greater importance on location attributes that they can specifically identify themselves. Therefore, one could expect that they are able to react quicker to variables that are more tangible to them, like visitors and region population, as opposed to macroeconomic impacts. It also makes intuitive sense that the effects of changes in macroeconomic aspects (e.g. GNPB, CW, KPR) would be slower to manifests itself in society than the more immediate effects of changes in microeconomic aspects (e.g. RevPAR, number of visitors). Thus, some reasoning for the poor statistical significance of these differenced estimates could be that the time-frame restriction is too limiting, in that it excludes too much information. While the regression estimates on level form accounts for long term effects, the differenced models have more emphasis on the implications of short-term change. And so, the regression estimates of the differenced models that rely on macroeconomic variables, have poor explanatory power.

5.1 Limitations and Basis for Future Studies

This research has several limitations that deserve further investigations. The differenced approach to achieve time series stationarity was too limiting, causing too much information to get lost in the conversion, leading to poor statistical significance and explanatory power in the results. More advanced regression methods, like Feasible Least Squares and other iterations of GLS, should be able to account for the skewness of the data and provide more accurate results. Another interesting aspect of this assignment to explore further could be a more detailed case study of Stavanger, or larger regions like Vestlandet, where the oil industry is most essential. Lastly, while addressed in other ways, the present study makes no effort to explicitly express how the effect of the uncertainty and volatility of accommodation demand impacts capacity investments.

References

Avinor. (2017). *The Avinor group – About the company*. Retrieved from https://avinor.no/en/corporate/about-us/the-avinor-group/about-the-company

Avinor. (2019). *Passengers monthly 2014 – 2018* [table]. Retrieved from https://avinor.no/en/corporate/about-us/statistics/archive

Berglihn, H. (2018). Stavanger-hotellene på bunn. *Dagens næringsliv*. Retrieved from https://www.dn.no

Billington, L. (2015). Finanskrise. *Store Norske Leksikon*. Retrieved from https://snl.no/finanskrise

Bjørshol, E. (2017). Hotell-Norge går så det suser. *Hotell & Restaurantmagasinet*. Retrieved from https://www.hotellmagasinet.no/

Cappelen, Å., Eika, T., Prestmo, J. B. (2014). Virkninger på norsk økonomi av et kraftig fall i oljeprisen. *SSB: Økonomiske analyser*, Vol 3, pp 31-41. https://www.ssb.no/nasjonalregnskap-og-konjunkturer/artikler-og-publikasjoner/_attachment/180823?_ts=14662dc53a8

Chen C. & Lin Y. (2013). *The influence of uncertain demand on hotel capacity*. International Journal of Hospitality Management, Vol 34, pp 462-465, https://doi.org/10.1016/j.ijhm.2012.11.008

Hayes, D. K. & Miller A. (2011). *Revenue Management for the Hospitality Industry*. John Wiley & Sons.

IndexMundi (2013). *About IndexMundi*. Retrieved from https://www.indexmundi.com/about.html

IndexMundi (2019). *Crude Oil (petroleum); Dated Brent Monthly Price - Norwegian Krone per Barrel* [table]. Retrieved from https://www.indexmundi.com/commodities/?commodity=crude-oil-brent&months=300&currency=nok

Juel, S. (2017). *Urbanisering og økonomisk utvikling – en konflikt mellom sentrum og periferi?* Civita, issue 19.

Kennedy, P. (2003). A guide to econometrics (5th ed.). Cambridge, MA: The MIT Press.

Koop, G. (2000). *Analysis of economic data* (2nd ed.). University of Leicester: Wiley.

Luo J. M. & Lam C. F.(2017). *Urbanization effects on hotel performance: A case study in China.* Cogent Business & Management, Vol 4. https://doi.org/10.1080/23311975.2017.1412873

Neuman, W.L. (2007). *Basics of social research: Qualitative and quantitative approaches*. (2nd ed.). Pearson Education

Newell G. & Seabrook R., (2006). *Factors influencing hotel investment decision making*. Journal of Property Investment & Finance, Vol 24 Issue: 4, pp 279-294, https://doi.org/10.1108/14635780610674499

Norges Bank (2017). *About the bank*. Retrieved from https://www.norges-bank.no/en/topics/about/

Norges Bank (2019). *Alle valutaer – månedsgjennomsnitt – siste 20 år* [table]. Retrieved from https://www.norges-bank.no/tema/Statistikk/Valutakurser/

Pallant, J. (2010). *SPSS survival manual*. McGraw Hill

Statistics Norway (2014). *This is Statistics Norway – statistics and analyses that benefit Society* [table]. Retrieved from https://www.ssb.no/en/omssb/om-oss/vaar-virksomhet/_attachment/198854?_ts=148c59d56c0

Statistics Norway (2019). 03614: *Hotel og liknande overnattingsbedrifter. Opne bedrifter, senger og rom (F) 1985M01 – 2019M03* [table]. Retrieved from https://www.ssb.no/statbank/table/03614/

Statistics Norway (2019). *10638: Turistkonsum I Norge, etter konsumentgruppe og Reiselivsnæring (mill. kr) 2007 – 2017* [table]. Retrieved from https://www.ssb.no/statbank/table/10638/

Statistics Norway (2019). *11342: Areal og befolkning I komuner, fylker og hele landet (K) 2007 – 2019* [table]. Retrieved from https://www.ssb.no/statbank/table/11342/

Statistics Norway (2019). *01222: Endringar I befolkninga I løpet av kvartalet, for Kommunar, fylke og heile landet (K) 1997K4 – 2019K1* [Table]. Retrieved from https://www.ssb.no/statbank/table/01222/

Statistics Norway (2019). *09391*: *Hovedtall fylkesfordelt nasjonalregnskap (F) 2008 – 2017* [table]. Retrieved from https://www.ssb.no/statbank/table/09391

Statistics Norway (2019). 08402: *Hotell og liknande overnattingsbedrifter. Overnattingar, etter gjestene sitt bustadland (F) 1985M01 - 2019M03* [table]. Retrieved from https://www.ssb.no/statbank/table/08402/

Statistikknett (2015). *Om statistikknett*. Retrieved from https://www.statistikknett.no/reiseliv/OmProfilen/meny_omprofilen.aspx

Statistikknett (2019). *Hotellenes losjiøkonomi* [table]. Retrieved from https://www.statistikknett.no/reiseliv/mnd/profil/profil_ok.aspx

Trochim, W. M. K. (2001). *The research methods knowledge base* (2nd ed.). Thomson learning.

Wooldridge, J. M. (2001). *Econometric analysis of cross section and panel data*. The MIT Press

Wooldridge, J. M. (2014). *Introduction to econometrics*. Cengage Learning EMEA

Appendix

Table A1 – part 1

Region	County	Sub-Region	Sub-Region Description and Municipality Number(s)
Indre Østlandet	Buskerud	Drammen	K-0602 Drammen
		Kongsberg	K-0604 Kongsberg
		Ringerike	K-0605 Ringerike
		Hole	K-0612 Hole
		Nes	K-0616 Nes (Buskerud)
		Gol	K-0617 Gol
		Hemsedal	K-0618 Hemsedal
		Ål	K-0619 Ål
		Hol	K-0620 Hol
		Nore og Uvdal	K-0633 Nore og Uvdal
		Midt-Buskerud	K621,622,623
		Ringerike/Hole	K605,612
		Drammen-region	K602,623,624,625,626,627,628,631
		Kongsberg-region	K604,631,632,633
		Hallingdal	K615,616,617,618,619,620
	Hedmark	Hedmark	K402,403,412,415,417,418,419, K420,423,425,426,427,428,429,437
	Oppland	Lillehammer	K-0501 Lillehammer
		Gjøvik	K-0502 Gjøvik
		Dovre	K-0511 Dovre
		Lesja	K-0512 Lesja
		Skjåk	K-0513 Skjåk
		Lom	K-0514 Lom
		Vågå	K-0515 Vågå
		Nord-Fron	K-0516 Nord-Fron
		Sel	K-0517 Sel
		Sør-Fron	K-0519 Sør-Fron
		Ringebu	K-0520 Ringebu
		Øyer	K-0521 Øyer
		Gausdal	K-0522 Gausdal
		Nord-Aurdal	K-0542 Nord-Aurdal
		Øystre Slidre	K-0544 Øystre Slidre
		Vang	K-0545 Vang
		Skjåk/Vågå	K513,515
		Sør-Gudbrandsdal	K501,521,522
		Gjøvik-region	K502,528,529,536,538
		Midt-Gudbrandsdal	K516,519,520
		Nord-Gudbrandsdal	K511,512,513,514,515,517
		Hadeland	K532,533,534
		Valdres	K540,541,542,543,544,545
	Telemark	Skien	K-0806 Skien
		Notodden	K-0807 Notodden
		Bamble	K-0814 Bamble
		Kragerø	K-0815 Kragerø
		Bø	K-0821 Bø (Telemark)
		Tinn	K-0826 Tinn
		Seljord	K-0828 Seljord
		Kviteseid	K-0829 Kviteseid
		Nissedal	K-0830 Nissedal
		Vinje	K-0834 Vinje
		Grenland-Kragerø	K805,806,811,814,815
		Vest-Telemark	K817,828,829,830,831,833
		Midt/Øst-Telemark	K807,819,821,822,827
		Fjell-Telemark	K826,834
		Nome-Sauherad-Hjartdal	K819,822,827

Table A1 – part 2

Oslofjord-regionen	Akershus	**Bærum**	K-0219 Bærum
		Asker	K-0220 Asker
		Skedsmo	K-0231 Skedsmo
		Ullensaker	K-0235 Ullensaker
		Romerike	K221,226,227,228,230,231,233, K234,235,236,237,238,239
		Follo-region	K211,213,214,215,216,217,229
		Asker/bærum	K219,220
	Oslo	**Oslo**	K-0301 Oslo
	Østfold	**Halden**	K-0101 Halden
		Moss	K-0104 Moss
		Sarpsborg	K-0105 Sarpsborg
		Fredrikstad	K-0106 Fredrikstad
		Indre Østfold	K111,121,124,125,136
	Vestfold	**Horten**	K-0701 Horten
		Tønsberg	K-0704 Tønsberg
		Sandefjord	K-0710 Sandefjord
		Larvik	K-0712 Larvik
		Vestfold-Rest	K711,713,715,716
Sørlandet	Aust-Agder	**Grimstad**	K-0904 Grimstad
		Arendal	K-0906 Arendal
		Bykle	K-0941 Bykle
		Valle-Bygland-Evje	K938,940,941
		Risør-Tvedestrand	K901,911,912,914,929
		Setesdal	K937,938,940,941
	Vest-Agder	**Kristiansand-region**	K926,928,935,1001,1014,1017,1018
		Lister-region	K1003,1004,1032,1037,1046
		Mandal-region	K1002,1021,1026,1027,1029,

Table A1 – part 3

Vestlandet	Hordaland	**Bergen**	K-1201 Bergen
		Stord	K-1221 Stord
		Kvinnherad	K-1224 Kvinnherad
		Odda	K-1228 Odda
		Ullensvang	K-1231 Ullensvang
		Eidfjord	K-1232 Eidfjord
		Ulvik	K-1233 Ulvik
		Voss	K-1235 Voss
		Kvam	K-1238 Kvam
		Lindås	K-1263 Lindås
		Hardanger	K1227,1228,1231,1232,1233,1234,1238,
		Sunnhordland	K1211,1216,1219,1221,1222,1223,1224,1244
		Nordhordland	K1256,1260,1263,1264,1265,1266
		Osterfjord-Bjørnefjord	K1241,1242,1251,1252,1253
		Sotra-Øygarden	K1245,1246,1247,1259
		Bergen-region	Bergen, Osterfjord-Bjørnefjord, Sotra-Øygarden
	Møre og Romsdal	**Molde**	K-1502 Molde
		Ålesund	K-1504 Ålesund
		Kristiansund	K-1505 Kristiansund
		Molde-region	K1502,1535,1539,1543,1545,1546,1547, K1548,1551,1557
		Ålesund-region	K1504,1511,1514,1515,1516,1517,1519, K1520,1523,1526,1528,1529,1531,1532,1534
		Kristiansund-region	K1505,1554,1560,1563,1566,1571,1573,1576
		Geiranger-trollstigen	K1525,1526
		Nordmøre-Romsdal	Molde-region, Kristiansund-region
	Rogaland	**Sandnes**	K-1102 Sandnes
		Stavanger	K-1103 Stavanger
		Haugesund	K-1106 Haugesund
		Sola	K-1124 Sola
		Stavanger-Jæren	K1102,1103,1119,1120,1121,1122,1124,1127
		Ryfylke	K1129,1130,1133,1134,1135,1141,1142,1144
		Nord-rogaland	K1106,1145,1146,1149,1151,1160
		Dalane	K1101,1111,1112,1114
	Sogn og Fjordane	**Sogndal**	K-1420 Sogndal
		Årdal	K-1424 Årdal
		Luster	K-1426 Luster
		Stryn	K-1449 Stryn
		Sognefjord	K1417,1418,1419,1420,1421,1422,1424,1426
		Sunnfjord	K1416,1430,1431,1432,1433
		Fjordkysten	K1401,1411,1412,1413,1428,1429,1438,
		Vik-Balestrand-Leikanger	K1417,1418,1419
		Aurland-Lærdal	K1421,1422
		VisitSognefjord	Vik-Balestrand-Leikanger, Aurland-Lærdal
		Nordfjord	K1439,1441,1443,1444,1445,1449

Table A1 – part 4

Trøndelag	**Trøndelag**	**Trondheim**	K-5001 Trondheim
		Steinkjer	K-5004 Steinkjer
		Namsos	K-5005 Namsos
		Oppdal	K-5021 Oppdal
		Røros	K-5025 Røros
		Stjørdal	K-5035 Stjørdal
		Verdal	K-5038 Verdal
		Grong	K-5045 Grong
		Orkdal-region	K5016,5022,5023,5024,5029
		Fosen	K5015,5017,5018,5019,5020,5054
		Hitra-Frøya	K5011,5012,5013,5014
		Namdal	K5005,5040,5042,5043,5044,5045,5046, K5047,5048,5049,5050,5051,5052
		Innherred	K5004,5037,5038,5039,5041,5053
		Værnes-region	K5031,5032,5033,5034,5035,5036
Nord-Norge	**Finnmark**	**Alta**	K-2012 Alta
		Nordkapp	K-2019 Nordkapp
		Porsanger	K-2020 Porsanger - Porsángu - Porsanki
		Sør-Varanger	K-2030 Sør-Varanger
		Alta-Hammerfest-Kvalsund	K2004,2012,2017
		Nordkapp/Kysten	K2014,2015,2018,2019,2022,2023,2024,2028
		Indre Finnmark	K2011,2020,2021
		Varanger-Region	K2002,2003,2025,2027,2030
	Nordland	**Bodø**	K-1804 Bodø
		Narvik	K-1805 Narvik
		Rana	K-1833 Rana
		Fauske	K-1841 Fauske - Fuosko
		Vågan	K-1865 Vågan
		Bodø-region	K1804,1837,1838,1839,1848
		Fauske-region	K1840,1841,1845,1849,1850
		Mo-Nesna-Sandnessjøen	K1818,1820,1822,1827,1828,1832, K1833,1834,1835,1836
		Helgeland	Mo-Nesna-Sandnessjøen, Brønnøysund-region, Mosjøen-region
		Salten	Bodø-region, Fauske-region
		Narvik-region	K1805,1852,1853,1854
		Brønnøysund-region	K1811,1812,1813,1815,1816
		Mosjøen-region	K1824,1825,1826
		Lofoten	K1856,1857,1859,1860,1865,1874
		Vesterålen	K1851,1866,1867,1868,1870,1871
	Troms	**Tromsø**	K-1902 Tromsø
		Harstad	K-1903 Harstad - Hárstták
		Tromsø-region	K1902,1933,1936,1938,1939,1940
		Sør-Troms	K1903,1911,1913,1917,1919,1920,1923
		Indre Troms	K1922,1924,1925,1926
		Senja-region	K1927,1928,1929,1931
		Nord-Troms	K1941,1942,1943

Table A2

Currency	Country
GBP	Storbritannia
SEK	Sverige
DKK	Danmark
EUR	Finland Andorra Belgia Estland Frankrike Hellas Irland Italia Kosovo Latvia Litauen Luxembourg Malta Monaco Montenegro Nederland Portugal San Marino Slovakia Slovenia Spania Tyskland Østerrike Vatikanstaten Kypros
USD	USA