Analysis of Macroeconomic Variables Affecting International Tourism Consumption in Sweden

JUN HO LEE

NOEL MATTAR
Analysis of Macroeconomic Variables Affecting International Tourism Consumption in Sweden

JUN HO LEE

NOEL MATTAR
Abstract

There is an evident trend of growing tourism in the world. Tourism in Sweden is gaining more economic and social attention. The main purpose of this thesis is to discover what macroeconomic variables contribute to the growth of annual international tourism income in Sweden. A multiple linear regression approach over a time period of 1978-2017 is used for the analysis.

The final results show that GDP is the major macroeconomic factor that drives the annual international tourism income in Sweden across all time periods. NOK-SEK exchange rate seem to another relevant variable in the long term from 1978-2017, but not in shorter periods of time. USD-SEK exchange rate and unemployment rate hold no significant relevance to the international tourism consumption in Sweden for all time. The devaluation of Swedish krona in 1992 did not change the relationship between these variables and the response variable. However, these results can be unstable due to the limited number of observations used in the analysis, and therefore, we recommend other regression approaches, such as panel data regression, for this subject.
Abstrakt


Contents

1 Introduction 1
  1.1 Background ................................................. 1
  1.2 Purpose .................................................. 1
  1.3 Research Question ................................. 2

2 Statistical Framework 3
  2.1 Multiple Linear Regression Analysis ........... 3
    2.1.1 Ordinary Least Squares ......................... 4
    2.1.2 Assumptions ........................................ 4
  2.2 Error Validation ................................. 5
    2.2.1 Multicollinearity ............................. 5
    2.2.2 Micronumerosity ............................. 5
    2.2.3 Heteroscedasticity ............................ 5
    2.2.4 Endogeneity ..................................... 6
    2.2.5 $R^2$ and Adjusted $R^2$ ......................... 6
    2.2.6 Akaike Information Criterion ................... 7
    2.2.7 Breusch-Pagan Test ............................. 8
    2.2.8 Shapiro - Wilk Normality Test ................ 8

3 Economic Framework 9
  3.1 Macroeconomics and Tourism .................. 9
  3.2 Overview of Tourism in Sweden ................ 12

4 Methodology 16
  4.1 Previous Studies and Software ............... 16
  4.2 Selection of Variables ............................. 16
    4.2.1 Response Variable .............................. 16
    4.2.2 Explanatory Variable .......................... 16
  4.3 Data Collection ...................................... 18
  4.4 Limitations ........................................ 19

5 Results 20
  5.1 Initial Models ...................................... 20
5.2 Model Reduction .......................................................... 26
5.3 Final Models ................................................................. 27
  5.3.1 Testing the Final Models ............................................. 27
  5.3.2 Presentation of Final Models ....................................... 32

6 Discussions ................................................................. 33
  6.1 Interpretation of the Final Result .................................... 33
    6.1.1 GDP ................................................................ 33
    6.1.2 Exchange Rates ..................................................... 33
    6.1.3 Removed Variables ................................................... 34
  6.2 Review of the Research Approach .................................... 35
    6.2.1 Model ................................................................ 35
    6.2.2 Data ................................................................ 35
    6.2.3 Methodology ........................................................... 35

7 Conclusion ................................................................. 37

8 References ................................................................. 39
1 Introduction

1.1 Background

Tourism has been continuously growing over last decades and become one of the fastest growing economic sectors in the world according to the UNWTO, Tourism Highlights, 2018 Edition. Modern tourism is closely linked to a wide range of service industries which include, but not limited to, accommodation, transportation, national parks, shopping facilities, entertainment facilities, food and beverage establishments. Tourism in industrialized and developed countries has produced economic and employment benefits in related business sectors (UNWTO, 2018). The business volume of tourism in the world equals or even surpasses that of oil exports, food products or automobiles and has generated 1.6 trillion USD in 2017 in its export earnings (UNWTO, 2018).

Sweden has also experienced a growth in tourism during the last decade (Tillväxtverket, 2017). The international tourism consumption in Sweden has shown positive growths since the global financial crisis in 2008 (Tillväxtverket, 2017). The total consumption by foreign tourism in Sweden has increased by 11.4 percent during 2017, and 229 percent in between 2000 and 2017 as published by Tillväxtverket, Fakta om svensk turism, 2017.

However, there is no clear answer to why there is such a significant growth in tourism in Sweden. There are countlessly many socio-economic factors on the individual level that could affect tourists’ decision-making processes in their destination choice for travel. In this paper, we therefore limit our scope to a macroeconomics to explain the growth of international tourism consumption in Sweden by foreign visitors using a multiple linear regression approach.

1.2 Purpose

The main purpose of this paper is to combine mathematical statistics with industrial economics to identify which macroeconomic variables are affecting the growth of international tourism in Sweden. The paper uses exchange rates, GDP, trade openness, and unemployment rate as explanatory variables, and the net international tourism
consumption (SEK) as the response variable. This paper aims to identify macroeconomic factors that are affecting the tourism consumption in Sweden during a time period of from 1978 to 2017.

1.3 Research Question

The main research question of this paper is:

- Which macroeconomic variables play a statistically significant role in the growth of international tourism consumption Sweden?
2 Statistical Framework

2.1 Multiple Linear Regression Analysis

A model that is constructed from analyzing the relation between variables in a data set is called a regression model. There are multiple regression model which includes several covariates (more than one). The following formula is a multiple regression model including n observations:

\[ y_i = \sum_{j=0}^{n} x_{ij} \beta_j + \varepsilon_i \quad i = 1, \ldots, n \]  

(1)

\( y_i \) corresponds to the response variable and \( x_j \) are called covariates. The coefficients of the covariates are called \( \beta_j \) and the error term is denoted by \( \varepsilon_i \). The different variables can be expressed in matrix notations:

\[ Y = X\beta + \varepsilon \]  

(2)

(Lang, 2016)

Where the matrices and vectors can be written as:

\[
Y = \begin{bmatrix}
    y_1 \\
y_2 \\
    \vdots \\
y_n
\end{bmatrix}, \quad 
\beta = \begin{bmatrix}
    \beta_1 \\
    \beta_2 \\
    \vdots \\
    \beta_k
\end{bmatrix}, \quad 
\varepsilon = \begin{bmatrix}
    \epsilon_1 \\
    \epsilon_2 \\
    \vdots \\
    \epsilon_n
\end{bmatrix}
\]

\[
X = \begin{bmatrix}
    1 & x_{11} & x_{12} & \cdots & x_{1k} \\
    1 & x_{21} & x_{22} & \cdots & x_{2k} \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    1 & x_{n1} & x_{n2} & \cdots & x_{nk}
\end{bmatrix}
\]  

(3)
2.1.1 **Ordinary Least Squares**

In multiple linear regression models, Ordinary Least Squares (OLS) are applied as an estimation of the relationship between a dependent variable and a set of regressors. More specifically, the OLS method is used to estimate the \( \beta_j \) coefficients in a regression model by minimizing the following least-square function:

\[
S(\beta) = \sum_{i=1}^{n} \epsilon_i^2 = (y - X\beta)'(y - X\beta)
\]  

(4)

Taking the derivative of \( S(\beta) \) with respect to \( \beta \) and putting it equal to zero will turn the least-square normal equation into the following:

\[
X'X\hat{\beta} = X'y
\]  

(5)

\[
\hat{\beta} = (X'X)^{-1}X'y
\]  

(6)

Solving this normal equation will provide the least-square estimators \( \hat{\beta} \).

(Lang, 2016)

2.1.2 **Assumptions**

In order for the linear regression model to be stable, five assumptions have to be met.

- A linear relationship between the predictor variables and the response variable.
- Error terms have a mean of zero.
- All the error terms are normally distributed.
- The variance of the error term is constant.
- There exists an uncorrelated relationship between the errors.

These assumptions are required for the regression analysis to be feasible (Montgomery, 2012).
2.2 Error Validation

2.2.1 Multicollinearity

When regressor variables are correlated then the model is said to be affected by multicollinearity. Multicollinearity is undesirable because it increases the variance of regression coefficient estimates and leads to coefficients with less precision.

Detection of multicollinearity in a regression model can be done in various ways. However, variance inflation factors (VIF) is used in this thesis. VIF explains how large a coefficients variance can grow due to the correlation between the regressors. If the value of the VIF equals one it implies that there is no correlation at all between the regressors.

The VIF of the j:th regression coefficient estimate can be expressed as the following:

\[
VIF_j = \frac{1}{1 - R_j^2}
\]  

(7)

Where \( R_j^2 \) is the coefficient of determination that can be obtained from regressing \( x_j \) on the other regressors in the model (Montgomery, 2012).

2.2.2 Micronumerosity

Micronumerosity arises when the sample taken in the linear regression is not enough. This causes the coefficients of the regressor variables from the linear regression model to be inaccurate. It could result in a wide difference when comparing the fitness of the estimated model to the actual reality. Limiting the number of regressor variables used in the regression analysis is one way of responding to micronumerosity when the model is suffering from a low number of observations (Montgomery, 2012).

2.2.3 Heteroscedasticity

Heteroscedasticity occurs when the error terms have non-constant variance. It goes against the previously mentioned assumption that the variance of the error term is constant for
the linear regression model. One can detect heteroscedasticity by analyzing a so called scale-location plot and the dispersion of the points over the predicted values. If the model has a constant variance then the plot should be showing a horizontal band of points. In this case one say that heteroscedasticity doesn’t exist and the model is called to be of the type homoscedastic. If the plot shows a non-straight band of points then the model is heteroscedastic (Lang, 2016).

2.2.4 Endogeneity

When an error term is correlated with at least one of the independent variables, it is said that the endogeneity exist in the regression model. The estimated coefficients $\beta_j$ will not be unbiased when endogeneity occurs.

$$E(\beta_j) \neq 0$$  \hspace{1cm} (8)

Three general sources of endogeneity are listed below.

1. Simultaneity : When the regressors are dependent on the response variable.

2. Error in measuring the explanatory variables : When there exist measurement errors in the data.

3. Model miss-specification : When an essential regressor variable is excluded from the model.

(Lang, 2016)

2.2.5 $R^2$ and Adjusted $R^2$

$R^2$ or so called the coefficient of determination is defined as:

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_{Res}}{SS_T}$$  \hspace{1cm} (9)

and

$$SS_{Res} = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)$$  \hspace{1cm} (10)
\[ SS_R = \sum_{i=1}^{n} (\hat{y}_i - \bar{y}) \] \hspace{1cm} (11)

\[ SS_T = SS_R + SS_{Res} \] \hspace{1cm} (12)

\( R^2 \) is a good tool to compare various models with each other, it is also a statistic that describes the variance proportion explained by the explanatory variables. For an example, if \( R^2 = 0.73 \) for a regression model, it means that 73% of the variance is described by the model. It should be noted that \( R^2 \) has a non-decreasing in \( x_j \). In other words, adding a new regressor variable will always increase the \( R^2 \) value. This could result in adding unnecessary regressor variables to the model.

Therefore, adjusted \( R^2 \) that penalizes its value by the number of regressor variables used in the model is often used in place of \( R^2 \) when comparing the models. Adjusted \( R^2 \) can be mathematically expressed by:

\[ R^2_{adj} = 1 - \frac{(n - 1)(1 - R^2)}{n - k - 1} \] \hspace{1cm} (13)

Where \( n \) refers to the number of total observations and \( k \) refers to the number of regressor variables used in the model. (Lang, 2016)

### 2.2.6 Akaike Information Criterion

Akaike information criterion (AIC) is an estimator of the relative quality of statistical models for a given set of data. Its values are determined by following equation:

\[ AIC = n \ln \left( \frac{SS_{Res}}{n} \right) + 2p \] \hspace{1cm} (14)

\( p \) represents the number of covariates and \( n \) corresponds to the number of observations.

AIC value alone does not provide any information about the quality of the regression model, it must be used as a comparison to another model that is reduced or have different
set of regressor variables.

\[ \Delta AIC = AIC_{FULL} - AIC_{REDUCED} \]  

(15)

The reduced model is more preferred when \( \Delta AIC > 0 \) since it implies less loss of information. A model with lowest AIC value is preferred. (Montgomery, 2012)

### 2.2.7 Breusch-Pagan Test

A test for heteroscedasticity is called Breusch-Pagan Test. The test checks the null hypothesis that the error variances are all equal and the error variance does not change as a function of one or more variables (Breusch and Pagan, 1979).

### 2.2.8 Shapiro - Wilk Normality Test

The Shapiro–Wilk test is for testing the normality. The test checks the null hypothesis that the data is normally distributed. If p-value is smaller then the alpha level then the null hypothesis must be rejected (Razali and Wah, 2011).
3 Economic Framework

3.1 Macroeconomics and Tourism

Macroeconomics is the study of the behaviour of a national economy as a whole (Chappelow, 2019). It uses aggregated measures such as GDP and price indices to understand how the economy functions on a larger scale. These macroeconomic measures could be used to determine the factors that drive the long-term economic growth within a specific business sector (Chappelow, 2019).

"Sustainable tourism: Contribution to economic growth and sustainable development” published by United Nations (UN) in 2013 states that tourism encompasses various sectors within an economy. Tourism stimulates other economic industries through direct and indirect interaction (UN, 2013). Moreover, in "Questionnaire on Tourism and Employment: Overview of Results” conducted by UNWTO in 2010 emphasis on the relationship between tourism and employment that is often overlooked by governments. Smeral (1988) claims that "economic growth influences tourism demand through mechanisms caused by the interdependence of certain elements of the socioeconomic system we live in”. International tourism can be perceived as having a positive effect on the nation’s economy, and it can be argued that there is a positive correlation between tourism and economy.

This led economists to support the idea of tourism-led growth hypothesis (TLGH) (Samimi and Sadeghi, 2011). The hypothesis claims that the international tourism is a strong driver for the growth of economy (Samimi and Sadeghi, 2011). It originates from export-led growth hypothesis (ELGH) (Panagiotidis and Mussoni, 2012). ELGH implies that the growth in economy is not only dependent upon the increasing size of labor and capital of the economy, but also on the increasing size of exports (Balassa, 1978). International tourism can be interpreted as one of subcategories of export (Anggraeni, 2017), and therefore TLGH can be understood as an altered form of ELGH with a narrower focus.

However, TLGH has a flaw in its argument; it is theoretically debatable whether a
relationship between tourism development and economy growth exists or not (Anggraeni, 2017). Many empirical researches have been conducted in this field of economics in order to confirm the tourism-led growth hypothesis. Few of these empirical researches are discussed below.

Akan, Arslan and İş k (2009) looks at the growth of economy and tourism in Turkey over a time period of 1985 to 2007. They conclude that there is a two-way causality between tourism and economy. In other words, growth in economy within a nation will lead to growth in tourism and vice versa (Akan, Arslan and İş k, 2009). For instance, a growth in tourism will bring more foreign exchanges into the country which can be used to import capital goods to produce more goods and services, hence leading to a growth in economy (Akan, Arslan and İş k, 2009). Growth in economy, on the other hand, will produce more goods and services for tourists in the country and stimulate the growth of tourism (Akan, Arslan and İş k, 2009). Lee and Chien (2008) also conducted a research on the relationship between GDP and tourism development in Taiwan over the time period of 1959 to 2003. Their study also confirmed that there is a bi-directional causality between tourism and economic growth in Taiwan.

On the other hand, Tang and Abosedra (2012) studies the contribution of tourism to the economic growth of Lebanon over a time period of 1995 to 2010. They found that there is a one-way causality between tourism and economy. Increase in tourism contributed to the growth of economy for the case of Lebanon (Tang and Abosedra, 2012). Akinboade and Braimoh (2010) demonstrated an analysis to determine the direction of causality between economy (GDP) and international tourism income in South Africa over a time period of 1980 to 2005. They claim that there is a uni-direction causality between tourism and economy, and the growth in international tourism income induces economic growth in South Africa (Akinboade and Braimoh, 2010).

Payne and Mervar (2010) examines and tests the tourism-led growth hypothesis in Croatia from 2001 to 2008. They also confirmed that there is a one-way causality, but in an opposite direction. Payne and Mervar (2010) observed a positive unidirectional causality from real GDP to tourism in Croatia, where growth in economy stimulated the growth of tourism.
Khandaker and Islam (2017) investigates correlation between international tourism demand and macroeconomic factors in countries with the largest, emerging and developed economies in the world, which includes India, Japan, Germany, the UK and the USA as their sample countries over a time period from 2007 to 2013. Their study shows that GDP and foreign direct investments are statistically insignificant to the growth of tourism. In doing so, they confirmed that tourism-led growth hypothesis do not hold for these countries.

To categorize the different outcomes of the empirical studies conducted by researchers, Tugcu (2013) proposes four classification of tourism-led growth hypothesis. Growth hypothesis, that suggests there is a one-way causality from economy to tourism. Conservation hypothesis, that argues that there is a one-way causality from tourism to economy. Feedback hypothesis, that claim that there is a two-way causality between economy and tourism. Finally, neutrality hypothesis, that indicates that there is no causality between economy and tourism.

In case of Sweden, there have been few researches conducted to check TLGH in the country. Ajvaz (2015) analyzes 21 counties in Sweden using a panel regression to check the tourism-led growth hypothesis in Sweden over a time period of 2003 to 2013. In his research, Ajvaz (2015) confirms that TLGH holds true for Sweden and claims that there exists a one-way causality from economy to tourism. Ajvaz (2015) observed that the growth of economy in Sweden led to the the development in tourism industry.

Bohlin, Brandt and Elbe (2014) studies the development of Swedish tourism public policy from 1930 to 2010. They claim that the developments in Swedish tourism policy and the general regimes, which encompasses social, economic and technological regimes, show a very strong connection (Bohlin, Brandt and Elbe, 2014). Their analysis shows that the economy, among other general regimes, can be identified as the major driver of development of tourism policy for Sweden. The Swedish tourism policy can be sorted and understood along with the development of the Swedish economy after 19th century (Bohlin, Brandt and Elbe, 2014). Their study is not aimed at finding the relationship between economic growth and tourism development in Sweden. However, it
still provides a valuable insight that supports the results obtained by Ajvaz (2015) and supports the growth hypothesis of TLGH suggested by Tugcu (2013) in the case of Sweden.

Outside tourism-led growth hypothesis, there are few empirical studies conducted by researchers. Culiuc (2014) looks at over 50 countries in the world over the time period of 1999 to 2009. He claims that economic ties have a large impact on tourism arrivals and its growth. In the report, Culiuc (2014) concludes that trades and real exchange rate are strong factors that affect tourism demand, and OECD countries tend to be more sensitive to these factors. Khandaker and Islam (2017) argue in their research that exchange rates and political stability has a strong correlation to the growth of tourism in a country. Smeral (1988) states that based on empirical studies conducted by Menges (1958), Schulmeister (1975), Smeral, Kramer, and Walterskirchen (1984) the demands for tourism depends greatly on relative prices of tourism and consumer goods, energy costs, and transportation costs, along with other social factors such as educational backgrounds and availability of leisure times of the visitors.

UN (2013) claims that the relationship between economy and tourism could be different across countries depending on their infrastructure and policies. UNWTO in "Questionnaire on Tourism and Employment" in 2010 concluded that there are uneven distribution of tourism activities and related employments across geographical, social, and industrial variables. It is difficult to acknowledge that there is a shared pattern between the tourism and employment across all countries (UNWTO, 2010). These suggests that the empirical studies conducted by researchers could be country-specific, and cannot easily be generalized for other countries without through examinations at the country’s infrastructure, geography and economy.

3.2 Overview of Tourism in Sweden

The statistical data and its trends in Sweden are archived and analysed by an official government agency named Swedish Agency for Economic and Regional Growth/Statistics Sweden (Tillväxtverket/SCB). Their statistical data and annual reports are published to the public. We have therefore relied a large part of our overview of tourism in Sweden on the
annual reports provided by Tillväxtverket/SCB.

The growth of tourism in Sweden is not a recent phenomena. Sweden has experienced a continuous growth in tourism back in 1980s because of its relative strengths in the well developed infrastructure and natural environments compared to other travelling destinations (Swedish Institute, 1995). This trend broke when there was a tax reform and Persian Gulf war that raised the price of goods and service provided in hotels and restaurants in 1990 (Swedish Institute, 1995). However, the devaluation of the Swedish krona in November 1992 made Sweden a more affordable travelling destination according to ”Tourism in Sweden” published by Swedish Institute in 1995.

Tillväxtverket points out in ”Tourism in Sweden” published in 2010 that the foreign tourism in Sweden measured in number of foreign nights spent has increased by over 120 percent from 1991 to 2010. This growth is higher than the approximate increase in the number of international travel in the world (110 percent) and in Europe (80 percent) during the same period of time which are provided by UN World Tourism Organization (UNWTO) (Tillväxtverket, 2010). This trend is illustrated in the figure 3.2 below.

Figure 3.1: Index for overnight stays spent by foreign visitors. (Tillväxtverket/SCB, ”Tourism in Sweden”, 2010.)

Tourism in Sweden has grown very rapidly during the last decade, and the international tourism consumption in Sweden has increased by 229 percent (Tillväxtverket, 2017). The international tourism consumption in Sweden has grown in relation to Sweden’s GDP, total exports, and employment (Tillväxtverket, 2017). Its contribution to Sweden’s total GDP, however, has remained at a constant level of 2.6 to 2.9 percent from 2000 to 2017 according to ”Tourism in Sweden” published in 2017 by Tillväxtverket. These numbers
are presented in the figure below.

In "Tourism in Sweden" published in 2017 by Tillväxtverket shows that shopping and good purchases are the largest and fastest growing spending activity in Sweden which accounts for approximately 30 percent of the total expenditures by visitors from abroad and grew by 11 percent from 2016 to 2017. Transportation accounts for 21 percent, and it is the second largest and third fastest growing spending activity with 5 percent growth (Tillväxtverket, 2017). Accommodation accounts for approximately 20 percent, and it is the third largest source of expenditures by international tourists in Sweden, and second fastest growing spending activity with 9 percent growth from last year (Tillväxtverket, 2017). There have been a great change in the ratio of spending activities for shopping and transportation compared to 2010. "Tourism in Sweden" published in 2010 by Tillväxtverket show that shopping and good purchases accounts for 48 percent, accommodation takes 22 percent, and transportation shares 12 percent in 2010 for international tourists in Sweden. Although there is no official clarification for this change in the percentages for shopping and good purchases and transportation, it could be due to development Swedish economy. As discussed earlier, Ajvaz (2015) showed that there is an one-way causality from economy to tourism, and it could be possible that the growth of Swedish economy has led to the development of other touristic activities other than shopping and good purchases.

Table 3.2: Key figures of tourism incomes in current prices, million Swedish krona. (Tillväxtverket/SCB, "Tourism in Sweden", 2017.)

<table>
<thead>
<tr>
<th>Year</th>
<th>Total tourism income (ISK)</th>
<th>Transport (ISK)</th>
<th>Shopping (ISK)</th>
<th>Accommodation (ISK)</th>
<th>Restaurant (ISK)</th>
<th>Hotel rooms (ISK)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020</td>
<td>686,312</td>
<td>186,312</td>
<td>186,312</td>
<td>186,312</td>
<td>186,312</td>
<td>186,312</td>
</tr>
<tr>
<td>2019</td>
<td>686,312</td>
<td>186,312</td>
<td>186,312</td>
<td>186,312</td>
<td>186,312</td>
<td>186,312</td>
</tr>
<tr>
<td>2018</td>
<td>686,312</td>
<td>186,312</td>
<td>186,312</td>
<td>186,312</td>
<td>186,312</td>
<td>186,312</td>
</tr>
<tr>
<td>2017</td>
<td>686,312</td>
<td>186,312</td>
<td>186,312</td>
<td>186,312</td>
<td>186,312</td>
<td>186,312</td>
</tr>
<tr>
<td>2016</td>
<td>686,312</td>
<td>186,312</td>
<td>186,312</td>
<td>186,312</td>
<td>186,312</td>
<td>186,312</td>
</tr>
</tbody>
</table>

"Tourism in Sweden" published in 2010 shows that approximately 46 percent of the total number of nights spent by visitors in Sweden are from other Nordic countries, 41 percent
from European countries excluding Nordic, 13 percent from countries outside Europe. The trend continues in 2017, where “Tourism in Sweden” published in 2017 shows that Norway is the largest market for Swedish tourism, where 3.4 millions visited Sweden in 2017, Germany is the second largest, with around 3 millions visitors, Denmark is the third largest, with around 1.1 million tourists during the same year. USA is the fastest growing market where the number of nights spent increased by 45 percent during 2017 compared to 2016 from 549 000 to 795 000, and China was the second fastest growing market with 13 percent growth in during the same period of time from 323 000 to 365 000.
4 Methodology

4.1 Previous Studies and Software

*Introduction to Linear Regression Analysis* by D. C. Montgomery, G. Geoffrey Vining, and Elizabeth A. Peck, 2015 is used as the main reference literature for the mathematical part of the thesis. Rstudio, LaTex (overleaf) and Microsoft excel are software used to perform regression analysis, write and store data respectively.

A thorough review of previous studies examining the relationship between macroeconomics and tourism is conducted prior to the collection of data. These previous studies are obtained from various online databases, journals, and financial databases, such as International Monetary Fund (IMF) and International Journal of Economics and Financial Issues.

4.2 Selection of Variables

4.2.1 Response Variable

Annual international tourism income in Sweden is the response variable used in the regression analysis in this paper. Million Swedish krona (mil SEK) is the base unit used to measure the response variable. Some data points represented in million United States dollar (mil USD) are converted into SEK by multiplying its value by yearly average exchange rate from USD to SEK of the same year.

4.2.2 Explanatory Variable

Explanatory variables are chosen based on the discussions made in *Economic Framework* section. Below are the explanation of each variables used in the project.

- **Gross Domestic Products (GDP)**

  Real GDP is the inflation-adjusted measure of the market value of all the final goods and services produced in a period of time (Chappelow, 2019). It is a common way
of expressing the size of an economy within a nation with respect to inflation rate (Chappelow, 2019). By using this explanatory variable in the regression analysis, it will be able to tell whether the growth in the economic size of Sweden does or does not affect tourism in Sweden. Annual Real GDP will be used in this paper. The base unit used for Real GDP in the paper is billion SEK.

- **Exchange Rates**

Exchange rates could have a direct impact on tourism according Culiuc (2014) and Khandaker and Islam (2017). Also, the annual reports published by Tillväxtverket show that visitors from abroad in Sweden spends approximately 30 percent of their total expenditures only in shopping and good purchases, and it is the largest spending activity by foreign tourists in Sweden. This indicates that exchange rate could be an important factor for the international tourism in Sweden.

In this paper, the annual exchange rates between USD-SEK and NOK-SEK (Norwegeian krone) are used as explanatory variables. USD is chosen because it is the fastest growing market, and NOK is chosen because it is the largest market for Swedish tourism (Tillväxtverket, 2017). Exchange rates are noted in SEK value as in per 100 of the specified currency. In other words, the annual SEK value of 100 NOK and 100 USD are used as base units.

- **Trade Openness**

Trade openness is the ratio of imports and exports to a nation’s GDP (Vijil, Huchet-Bourdon and Le Mouël, 2018). It is a measure of how open a nation is for trading. As the extent of globalization and trade liberalization within a nation have a large impact on the size of trades, it is a macroeconomic variable that encompasses a nation’s political aspects as well (Keman, 2013).

Culiuc (2014) argues that the number of economic ties, businesses, and trades between countries have a direct impact on the number of tourists between countries. Trading increases the number of touchpoints between people in different countries (Culiuc, 2014). It also stimulates transfer of skills and labours flowing in and out of the country (Culiuc,
2014). These altogether accelerates the growth of tourism in a country.

• **Unemployment Rate**

Unemployment rate is the ratio of labor forces in the country that are jobless. In this paper, annual unemployment rates of Swedish population between ages of 15 to 74 are used. UNWTO argue that there is a relationship between unemployment rate and tourism, although this could vary greatly depending on the geographical, infrastructural and political aspects of the country (UNWTO, 2010).

• **Nature Reserve**

Nature reserve is one of the major tourism resources in Nordic countries according to ”Tourism, nature and sustainability” review published by Nordic Council of Ministers in 2018. The touristic attractiveness of nature areas could be an important variable driving the growth of tourism in Sweden (Øian, et al., 2018). It is therefore taken as a relevant explanatory variable in this paper. Nature reserve is measured in hectares and only refers to land surface area.

### 4.3 Data Collection

All data are collected from various official government websites, and financial institutes, which includes, but not limited to, OECD database, Statistics Sweden, Riksbanken, and European Central Bank. The time range of data is from 1978 to 2017. More detailed sources of data are listed below.

- Annual international tourism income : SCB
- Gross Domestic Products (GDP) : Ekonomifakta
- Exchange rates (USD): Riksbanken
- Exchange rates (NOK): Riksbanken
- Trade Openness: The Global Economy
• Unemployment rate: European Central Bank

• Nature reserve: SCB

4.4 Limitations

To avoid the varying values of response variable due to the seasonality, we collected annual data with a time range from 1978 to 2017. Although it is a long time period, we are still limited by the number of observations. It is possible that our regression model could be unstable and have high errors due to this limitation. Moreover, we are missing annual international tourism income data from 1997 to 1999 from SCB. This further reduces the number of observations included in our regression model.

Some of the data in SCB annual reports have variations in their numbers. For instance, the annual international tourism income for 1987 stated in "Statistisk årsbok för Sverige 1987" is 2030 million USD whereas in "Statistisk årsbok för Sverige 1992" it is 2033 million USD for the same year. We assume that this could be due to correction of old data, and not due to the change in exchange rate. It is because the values of only specific years are affected. We will use the data from the more recent annual report whenever there is a mismatch in data between annual reports provided by SCB.

Some of the relevant variables mentioned in the economic framework section are not available for the time period from 1978 to 2017. For instance, political stability and political risk indices were not measured in Sweden before 2000. Some economic variables, such as EUR-SEK exchange rate, did not exist before 1999. Furthermore, having too many variables happen to make models very unstable. Some variables, such as consumer price index and CO\textsubscript{2} emissions, are excluded after testing the model for normality, heteroscedasticity and multicollinearity before diving in for a deeper analysis.
5 Results

Three models are made to check if there is any difference in models in the time periods of: 1978 to 2017 (full time period), 1978 to 1992 (before the devaluation of Swedish krona), and 1993 to 2017 (after the devaluation of Swedish krona).

5.1 Initial Models

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std. error</th>
<th>p-value</th>
<th>Signif. code</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-1.733e+05</td>
<td>3.447e+04</td>
<td>2.15e-05</td>
<td>***</td>
</tr>
<tr>
<td>Exchange rate (USD)</td>
<td>1.391e+01</td>
<td>8.232e+00</td>
<td>0.101517</td>
<td>.</td>
</tr>
<tr>
<td>Exchange rate (NOK)</td>
<td>4.416e+02</td>
<td>2.375e+02</td>
<td>0.072853</td>
<td>.</td>
</tr>
<tr>
<td>GDP</td>
<td>6.285e+01</td>
<td>4.969e+00</td>
<td>1.48e-13</td>
<td>***</td>
</tr>
<tr>
<td>Trade openness</td>
<td>-9.604e+01</td>
<td>2.291e+02</td>
<td>0.678072</td>
<td>.</td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>9.724e+02</td>
<td>5.717e+02</td>
<td>0.099334</td>
<td>.</td>
</tr>
<tr>
<td>Nature reserve</td>
<td>-1.376e-02</td>
<td>3.284e-03</td>
<td>0.000226</td>
<td>***</td>
</tr>
</tbody>
</table>

***** < 0.001, **'** < 0.01, *'** < 0.05, '.' < 0.1, <' '< 1

Table 5.1: Estimations of the model covering the full time period.

Initial values are shown in the table above for the whole period from 1978 to 2017. GDP has an extremely low p-value compared to other variables. It indicates that the model depends very much upon this variable to predict its values. It is also possible to observe that variables with high p-values generally have very high standard error value, which makes them unfeasible for the model.

<table>
<thead>
<tr>
<th>Observation</th>
<th>R²</th>
<th>Adjusted R²</th>
<th>Df</th>
</tr>
</thead>
<tbody>
<tr>
<td>37</td>
<td>0.9819307</td>
<td>0.9783168</td>
<td>30</td>
</tr>
</tbody>
</table>

Table 5.2: Statistics of the model covering the full time period.

Adjusted R² of the full period model is 0.9783168. This implies that 97.83168% of the variation in the response variable can be explained by the regressors in the model.
Table 5.3: Estimations of the model covering a time period of 1978-1992.

Similar to the full period model, GDP has the lowest p-value during 1978 to 1992.

Table 5.4: Statistics of the model covering a time period of 1978-1992.

Adjusted $R^2$ of the model is slightly higher compared to the full period model.

Table 5.5: Estimations of the model covering a time period of 1993-2017.

Similar trend continues in this model covering a time period of 1993-2017. However, the low p-value of nature reserve indicates that it could have an impact on this model.

Table 5.6: Statistics of the model covering a time period of 1993-2017.
The $R^2$ and adjusted $R^2$ of the model is slightly worse compared to other models. However, the values of $R^2$ and adjusted $R^2$ are still greater than 0.96, which implies that the model is highly precise.

- **Normality**

The normal quantile-quantile plot for each time periods are shown below.

![Normal Q-Q plot](image)

(a) 1978-2017  
(b) 1978-1992  
(c) 1993-2017

Figure 5.1: Normal Q-Q plots of models

However, it is very difficult to determine whether the plots are actually normal or not by looking at the normal Q-Q plots because there are not many data points plotted on the graph. We therefore used Shapiro–Wilk test to check the normality assumptions of each model. Using R, Shapiro–Wilk test generates a p-value of the null-hypothesis that the data tested are normally distributed. According to Royston (1995) a model is adequate if the p-value is less than 0.1. Below are the results of Shapiro–Wilk test on the models.

<table>
<thead>
<tr>
<th>Model period</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1978-2017</td>
<td>0.005136</td>
</tr>
<tr>
<td>1978-1992</td>
<td>0.0502</td>
</tr>
<tr>
<td>1993-2017</td>
<td>0.6251</td>
</tr>
</tbody>
</table>

Table 5.7: Shapiro–Wilk test for the models.

The model covering the time period from 1993 to 2017 has a p-value that is greater than 0.1 and the normality assumption is rejected. We reduced the time period from 1993 to 2014 to make the model feasible for the multiple linear regression. Below are the normal Q-Q plot and the result of the Shapiro–Wilk test on the new model.
The normal q-q plot and the p-value of Shapiro–Wilk test became significantly better by reducing the time period. Below are the new estimations and statistics of the new model.

Table 5.9: Estimations of the model covering a time period of 1993-2014.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std. error</th>
<th>p-value</th>
<th>Signif. code</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-5.414e+04</td>
<td>5.434e+04</td>
<td>0.338719</td>
<td></td>
</tr>
<tr>
<td>Exchange rate (USD)</td>
<td>-5.617e+00</td>
<td>1.755e+01</td>
<td>0.754446</td>
<td></td>
</tr>
<tr>
<td>Exchange rate (NOK)</td>
<td>-2.047e+02</td>
<td>3.328e+02</td>
<td>0.549996</td>
<td></td>
</tr>
<tr>
<td>GDP</td>
<td>4.986e+01</td>
<td>1.151e+01</td>
<td>0.000973 ***</td>
<td></td>
</tr>
<tr>
<td>Trade openness</td>
<td>-2.829e+02</td>
<td>3.845e+02</td>
<td>0.475988</td>
<td></td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>-3.447e+02</td>
<td>1.616e+03</td>
<td>0.834629</td>
<td></td>
</tr>
<tr>
<td>Nature reserve</td>
<td>-6.302e-03</td>
<td>9.817e-03</td>
<td>0.532982</td>
<td></td>
</tr>
</tbody>
</table>

'***' < 0.001, '**' < 0.01, '*' < 0.05, '.' < 0.1, '<' < 1

Table 5.10: Statistics of the model covering a time period of 1993-2014.

<table>
<thead>
<tr>
<th>Observation</th>
<th>R²</th>
<th>Adjusted R²</th>
<th>Df</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>0.9723544</td>
<td>0.9585316</td>
<td>12</td>
</tr>
</tbody>
</table>

This model covering the time period from 1993 to 2014 will be used instead of the previous model using a time period from 1993 to 2017.
• **Heteroscedasticity**

![Figure 5.3: Residual versus fitted values of the models.](image)

There are very few points. It is difficult to tell whether there is potential heteroscedasticity hidden in the model or not with only graphical interpretations. Therefore, Breusch–Pagan test is used to determine the presence of heteroscedasticity of the models.

In R, the *Non-constant Variance Score Test* is equivalent to the Breusch–Pagan test. It checks the null hypothesis that the error variances are all equal and the error variance does not change as a function of one or more variables (Breusch and Pagan, 1979). A $\chi^2$ value that is greater than 3.841 for one degree of freedom paired with a low p-value signals a presence of heteroscedasticity in the model.

Below are the results of the Breusch–Pagan test for each model.

<table>
<thead>
<tr>
<th>Model period</th>
<th>$\chi^2$</th>
<th>Df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1978-2017</td>
<td>3.635391</td>
<td>1</td>
<td>0.056563</td>
</tr>
<tr>
<td>1978-1992</td>
<td>2.580595</td>
<td>1</td>
<td>0.10818</td>
</tr>
<tr>
<td>1993-2014</td>
<td>0.8659187</td>
<td>1</td>
<td>0.35209</td>
</tr>
</tbody>
</table>

Table 5.11: Breusch-Pagan test for the models.

The results show that there is no strong evidence of heteroscedasticity in the models.
• **Multicollinearity**

<table>
<thead>
<tr>
<th>Variable</th>
<th>VIF-value (Full)</th>
<th>VIF-value (1978-1992)</th>
<th>VIF-value (1993-2014)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exchange rate (USD)</td>
<td>1.82</td>
<td>5.91</td>
<td>3.31</td>
</tr>
<tr>
<td>Exchange rate (NOK)</td>
<td>7.59</td>
<td>8.25</td>
<td>1.45</td>
</tr>
<tr>
<td>GDP</td>
<td>22.10</td>
<td>31.41</td>
<td>33.77</td>
</tr>
<tr>
<td>Trade openness</td>
<td>11.85</td>
<td>6.88</td>
<td>10.22</td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>2.66</td>
<td>2.27</td>
<td>5.10</td>
</tr>
<tr>
<td>Nature reserve</td>
<td>18.79</td>
<td>28.86</td>
<td>29.62</td>
</tr>
</tbody>
</table>

Table 5.12: VIF values of models.

VIF values that are over 10 implies the presence of multicollinearity. In the table above, GDP and Nature reserve seem to be highly correlated. One way of dealing with multicollinearity is removing one of the highly correlated variable from the model. Removing GDP, however, is not good for the model because as shown earlier in the section 5.1, all three models depends strongly on GDP. Removing GDP might result in breaking the models.

Therefore, we decided to remove Nature reserve and Trade openness variables from the model, which have the highest VIF values.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Exchange rate (USD)</td>
<td>1.72</td>
<td>2.79</td>
<td>2.15</td>
</tr>
<tr>
<td>Exchange rate (NOK)</td>
<td>3.77</td>
<td>2.72</td>
<td>1.40</td>
</tr>
<tr>
<td>GDP</td>
<td>1.93</td>
<td>1.37</td>
<td>2.79</td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>2.27</td>
<td>1.26</td>
<td>2.79</td>
</tr>
</tbody>
</table>

Table 5.13: VIF values of models after the removal.

<table>
<thead>
<tr>
<th>Model period</th>
<th>Adjusted R² (Before)</th>
<th>Adjusted R² (After)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1978-2017</td>
<td>0.9783168</td>
<td>0.9677553</td>
</tr>
<tr>
<td>1978-1992</td>
<td>0.9802357</td>
<td>0.9786700</td>
</tr>
<tr>
<td>1993-2014</td>
<td>0.9585316</td>
<td>0.9618757</td>
</tr>
</tbody>
</table>

Table 5.14: Adjusted R² values before and after the removal.

Removing the variables from the model greatly reduces VIF for GDP. Adjusted R² values in the models are not affected greatly by the removal. We decide to keep this change and
continue to the model reduction.

5.2 Model Reduction

Models are reduced using two methods: all possible regression and backward elimination. All possible regression method, also known as best subset regression, retrieves the best subset of variables for each number of predictors using an exhaustive method (testing all combinations) by comparing the residual sum of squares, and then compares the results using various criteria (Montgomery 2012). In this paper, we will use Adjusted $R^2$ as the criterion. Backward elimination in R, on the other hand, is another exhaustive method that compares AIC value of removing a variable and reduces the model until removing a variable no longer gives a lower AIC value.

Below are the summaries of the model reduction for all models.

<table>
<thead>
<tr>
<th>Removed variables</th>
<th>$R^2$</th>
<th>Adjusted $R^2$</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>(None)</td>
<td>0.9713</td>
<td>0.9678</td>
<td>654.60</td>
</tr>
<tr>
<td>Exchange rate (USD) + Unemployment rate</td>
<td>0.9698</td>
<td>0.9680</td>
<td>652.59</td>
</tr>
</tbody>
</table>

Table 5.15: Comparison of the models with removed variables (1978-2017).

<table>
<thead>
<tr>
<th>Removed variables</th>
<th>$R^2$</th>
<th>Adjusted $R^2$</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>(None)</td>
<td>0.9848</td>
<td>0.9787</td>
<td>205.32</td>
</tr>
<tr>
<td>Exchange rate (NOK) + Unemployment rate</td>
<td>0.9830</td>
<td>0.9802</td>
<td>202.93</td>
</tr>
<tr>
<td>Exchange rate (USD) + Exchange rate (NOK) + Unemployment rate</td>
<td>0.9813</td>
<td>0.9798</td>
<td>202.42</td>
</tr>
</tbody>
</table>

Table 5.16: Comparison of the models with removed variables (1978-1992).
The result of all possible regression (highest adjusted $R^2$ value) and backward elimination (lowest AIC value) are same for the model with the time period from 1978 to 2017. For the other models that have different results, we compared the values between AIC and adjusted $R^2$. In the both model, we decide to choose a model reduction with lower AIC value, because the change in AIC are greater than the change in adjusted $R^2$.

## 5.3 Final Models

### 5.3.1 Testing the Final Models

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std. error</th>
<th>p-value</th>
<th>Signif. code</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-1.586e+05</td>
<td>1.824e+04</td>
<td>3.71e-10</td>
<td>***</td>
</tr>
<tr>
<td>Exchange rate (NOK)</td>
<td>5.540e+02</td>
<td>1.441e+02</td>
<td>0.000506</td>
<td>***</td>
</tr>
<tr>
<td>GDP</td>
<td>4.678e+01</td>
<td>1.767e+00</td>
<td>&lt;2e-16</td>
<td>***</td>
</tr>
</tbody>
</table>

$*** < 0.001$, $** < 0.01$, $* < 0.05$, $.$ $< 0.1$, $<$ $< 1$

Table 5.18: Estimations of the final model covering the full time period.

<table>
<thead>
<tr>
<th>Observation</th>
<th>$R^2$</th>
<th>Adjusted $R^2$</th>
<th>Df</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.9697554</td>
<td>0.9679764</td>
<td>34</td>
</tr>
</tbody>
</table>

Table 5.19: Statistics of the final model covering the full time period.

The model covering the full time period from 1978 to 2017 has been reduced into two variables, both with p-values lower than 0.05. There has been a slight decrease in $R^2$ and adjusted $R^2$ values compared to the initial model, but they still maintain values greater than 0.96.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std. error</th>
<th>p-value</th>
<th>Signif. code</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-4.915e+04</td>
<td>2.288e+03</td>
<td>1.54e-11</td>
<td>***</td>
</tr>
<tr>
<td>GDP</td>
<td>2.536e+01</td>
<td>9.722e-01</td>
<td>1.30e-12</td>
<td>***</td>
</tr>
</tbody>
</table>

*** < 0.001, ** < 0.01, * < 0.05, . < 0.1, < 1

Table 5.20: Estimations of the final model covering a time period of 1978-1992.

<table>
<thead>
<tr>
<th>Observation</th>
<th>R²</th>
<th>Adjusted R²</th>
<th>Df</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>0.9812579</td>
<td>0.9798162</td>
<td>13</td>
</tr>
</tbody>
</table>


The model covering the time period from 1978 to 1992 has been reduced into one variables. There has been a slight decrease in R² and adjusted R² values compared to the initial model, but the changes are less than 0.01 and can be neglected.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std. error</th>
<th>p-value</th>
<th>Signif. code</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-1.036e+05</td>
<td>1.255e+04</td>
<td>3.69e-07</td>
<td>***</td>
</tr>
<tr>
<td>GDP</td>
<td>4.223e+01</td>
<td>2.201e+00</td>
<td>1.82e-12</td>
<td>***</td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>1.089e+03</td>
<td>7.954e+02</td>
<td>0.19</td>
<td></td>
</tr>
</tbody>
</table>

*** < 0.001, ** < 0.01, * < 0.05, . < 0.1, < 1

Table 5.22: Estimations of the model covering a time period of 1993-2014.

The model covering the time period from 1993 to 2014 has been reduced into two variables. However, unemployment rate has p-value that is significantly higher than 0.05, which implies that this variable does not contribute to the model. We will remove this variable from the final model.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std. error</th>
<th>p-value</th>
<th>Signif. code</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-8.894e+04</td>
<td>6.744e+03</td>
<td>2.34e-10</td>
<td>***</td>
</tr>
<tr>
<td>GDP</td>
<td>4.053e+01</td>
<td>1.867e+00</td>
<td>7.78e-14</td>
<td>***</td>
</tr>
</tbody>
</table>

*** < 0.001, ** < 0.01, * < 0.05, . < 0.1, < 1

Table 5.23: Estimations of the final model covering a time period of 1993-2014.

<table>
<thead>
<tr>
<th>Observation</th>
<th>R²</th>
<th>Adjusted R²</th>
<th>Df</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>0.9651968</td>
<td>0.9631496</td>
<td>17</td>
</tr>
</tbody>
</table>

Table 5.24: Statistics of the final model covering a time period of 1993-2014.
All variables except one is removed from the initial model. R² and adjusted R² values still have values over 0.96. The change in these values are minimal.

**Heteroscedasticity**

<table>
<thead>
<tr>
<th>Model period</th>
<th>χ²</th>
<th>Df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1978-2017</td>
<td>6.93114</td>
<td>1</td>
<td>0.0084708</td>
</tr>
<tr>
<td>1978-1992</td>
<td>2.570955</td>
<td>1</td>
<td>0.108840</td>
</tr>
<tr>
<td>1993-2014</td>
<td>0.6941045</td>
<td>1</td>
<td>0.404770</td>
</tr>
</tbody>
</table>

Table 5.25: Breusch-Pagan test for the final models.

As shown in the results of Breusch-Pagan tests for the models above, the final model covering the full time period from 1978 to 2017 has a χ² value that is larger than 3.841, and the corresponding p-value is also very low. This implies that there is heteroscedasticity in the final model.

One of the common correction of the heteroscedasticity in a model is rescaling the explanatory variables using natural logarithms. We plotted a graph of each explanatory variable against the response variable for the model below.

![Graphs of GDP and NOK exchange rate](image)

(a) GDP  
(b) NOK exchange rate

Figure 5.4: Explanatory variable against the response variable.

We see that there is a clear exponential relationship between GDP and the response variable, whereas NOK exchange rate does not show any clear relationship with the response variable. We decided to rescale GDP variable with the natural logarithm. The results of rescaling are summarized below.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std. error</th>
<th>p-value</th>
<th>Signif. code</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-1.259e+06</td>
<td>8.099e+04</td>
<td>&lt;2e-16</td>
<td>***</td>
</tr>
<tr>
<td>Exchange rate (NOK)</td>
<td>8.183e+02</td>
<td>2.105e+02</td>
<td>0.000446</td>
<td>***</td>
</tr>
<tr>
<td>log(GDP)</td>
<td>1.524e+05</td>
<td>8.098e+03</td>
<td>&lt;2e-16</td>
<td>***</td>
</tr>
</tbody>
</table>

Table 5.26: Estimations of the final model covering the full time period with a natural log on GDP.

<table>
<thead>
<tr>
<th>Observation</th>
<th>R(^2)</th>
<th>Adjusted R(^2)</th>
<th>Df</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.9427235</td>
<td>0.9393543</td>
<td>34</td>
</tr>
</tbody>
</table>

Table 5.27: Statistics of the final model covering the full time period with a natural log on GDP.

<table>
<thead>
<tr>
<th>Model period</th>
<th>(\chi^2)</th>
<th>Df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1978-2017</td>
<td>2.766374</td>
<td>1</td>
<td>0.096264</td>
</tr>
</tbody>
</table>

Table 5.28: Breusch-Pagan test for the final model covering the full time period with a natural log on GDP.

The model estimations and p-values do not greatly change from the logarithmic rescaling. \(R^2\) and adjusted \(R^2\) values have deteriorated, but they still maintain values close to 0.94. The Breusch-Pagan test now shows that there is no presence of heteroscedasticity in the final model.

Figure 5.5: Residual versus fitted values of the models.

Looking at the residual versus the fitted values of the final models, the graph for period corresponding to 1978 to 2017 shows a U-shaped pattern. This suggests that there could be a non-linear relationship between the explanatory variables and the response
variable. However, the linear model still has a high precision with the adjusted R² value greater than 0.94, which implies that more than 94% of the variances in the response variable. We decided to continue to use this linear model in this paper without any further transformation.

• Normality

The normal quantile quantile plot for each time periods are shown below.

![Normal Q-Q plots of final models](image)

**Figure 5.6: Normal Q-Q plots of final models**

<table>
<thead>
<tr>
<th>Model period</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1978-2017</td>
<td>0.03230</td>
</tr>
<tr>
<td>1978-1992</td>
<td>0.09682</td>
</tr>
<tr>
<td>1993-2014</td>
<td>0.05099</td>
</tr>
</tbody>
</table>

Table 5.29: Shapiro–Wilk test for the models.

The Shapiro–Wilk test for testing the normality of the final models shows that all models have p-value less than 0.1. The normality assumption is not violated for all final models.

• Multicollinearity

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Exchange rate (NOK)</td>
<td>2.13</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GDP</td>
<td>2.13</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5.30: VIF values of the final models.

The VIF values of the final models does not exceed 10 for all variables in all final models. There is no evidence of multicollinearity in the final models.
5.3.2 Presentation of Final Models

Below are the models with their respective covariates and beta estimates expressed in equation forms.

Annual international tourism income \(_{1978–2017} = -1259000 + 818.3 \cdot (\text{Exchange rate (NOK)}) + 152400 \cdot \log (\text{GDP})\)

Annual international tourism income \(_{1978–1992} = -41060 + 25.36 \cdot (\text{GDP})\)

Annual international tourism income \(_{1993–2014} = -88940 + 40.53 \cdot (\text{GDP})\)
6 Discussions

6.1 Interpretation of the Final Result

6.1.1 GDP

The results shows that the GDP is the major explanatory variables of the annual international tourism income in Sweden for all time periods. It shows a strong positive correlation to the growth of the international tourism income. This is an expected result, where the increase in GDP will eventually lead to producing more goods and service in the country for tourism as discussed earlier in the economic framework section. This results aligns with previous researches conducted in Sweden by Ajvaz (2015) and Bohlin, Brandt and Elbe (2014).

However, unlike Ajvaz (2015) and Bohlin, Brandt and Elbe (2014) where they suggested an uni-directional causality from economy to tourism, our result shows an evidence of bi-directional causality between economy and tourism. It is possible to see this evidence of two-way relationship between tourism and economy in Figure 5.4 where GDP grows exponentially with the growth of the international tourism income. However, this could be due to the difference in the approach, where we used multiple linear regression, but Ajvaz (2015) used a panel data regression approach. Also, there is a difference in data samples because time periods and the size of the geographical focus were not same.

6.1.2 Exchange Rates

The NOK-SEK exchange rate is positively correlated to the response variable during the time period of 1978 to 2017. This is an expected result because Norway historically has been the largest market for Swedish tourism as discussed in section 3.1. If the NOK exchange rises, the valuation of NOK becomes higher. Since shopping and good purchases are the largest spending activities in Sweden, this will positively affect the purchase power of the tourists from Norway and increase the international tourism income in Sweden.
In shorter time periods, on the other hand, the NOK-SEK exchange rate appears to be not correlated to the response variable. This is an odd result because the tourism consumption should benefit from the devaluation of Swedish krona in 1992. We assume that tourists from Norway is not as price sensitive in a shorter period of time and the devaluation of Swedish krona did not affect the tourism in Sweden to a larger extent.

USD-SEK exchange is not correlated to the response variable regardless of the time periods. This result does not aligns itself with the previous studies, where exchange rates and international tourism income and growth are positively correlated. However, this could be due to the touristic nature of Sweden where its main market is Europe. As previously mentioned in section 3.2, more than 85% of the tourists coming to Sweden are from other European countries. Although USA is growing very rapidly as a new tourism market for Sweden, the absolute number of tourists from USA is incomparably small.

### 6.1.3 Removed Variables

Trade openness and nature reserves appeared to be collinear to other variables and had to be removed from the regression model. It appeared that VIF values for GDP drops significantly after removing these two variables. The collinearity between GDP and trade openness could be explained by the definition of GDP which is the sum of consumption, government spending, investment, and net exports. Increase in GDP could imply increase in net exports and result in increased trade openness, which causes collinearity. These removed variables could be related to the response variable, as shown in Table 5.1 nature reserves shows a low p-value. However, it is difficult to determine their relationship with the response variable when they are collinear with GDP, which is our most important explanatory variable.

Unemployment rate was insignificant for all models, implying that there is no observable relationship between this variable and the tourism consumption for Sweden. As previously discussed in section 3.3, UN and UNWTO both note that there may be variations in the macroeconomic variables depending on the infrastructure and policy of the country. In other words, a variable that could be relevant for other countries may not be relevant to
6.2  Review of the Research Approach

6.2.1  Model

Several different measures and analyses are conducted to ensure that no assumptions of the linear regression are violated. All of our models yield robust results with high R\(^2\) and low p-values. The R\(^2\) and adjusted R\(^2\) values varied between 0.9-0.96 for all time periods even after the model reduction, which indicates that the chosen explanatory variables adequately explain the response variable.

6.2.2  Data

The data collected are not from the same database. This could incur possible measurement errors depending on how the measurements are conducted by statistical warehouses. Our final models could be suffering from these errors.

The data are assembled on the annual base and this lead to a very few number of observations. Having too many explanatory variable on very few observation often violated the fundamental assumptions of the linear regression. It put a limit on the number of explanatory variables to fit in our linear regression models.

The limitation on the number of observations also raises the question of the feasibility of our final models. The models have high adjusted R\(^2\) values. However, the standard errors tend to be large due to the micronumerosity.

6.2.3  Methodology

There are several detectable outliers in the models that does not align with other data points, but we decided to not eliminated these outlier data points in our analysis. There are not a lot of data points used in the analysis. Moreover, by taking these outliers from the model, we could also risk ourselves getting a biased result by overfitting the model.
The final regression model over the time period of 1978 to 2017 with logarithmic scaling shows a U-shaped pattern in the residual versus fitted value graph presented under result section. This implies that other approach than a multiple linear regression would be more suitable for the analysis. For instance, many of the modern researchers in this field uses panel data regression. Panel data analysis is an increasingly popular form of longitudinal data analysis among social and behavioral science researcher (Yaffee, 2003). This approach is more fit for small sample of data (Yaffee, 2003).
7 Conclusion

The purpose of this report is to identify the macroeconomic variables that are affecting the tourism consumption in Sweden. This thesis combines mathematical theories and macroeconomic backgrounds to derive a tolerable results. We stay within the boundaries of ordinary least square regression, performing three multiple linear regressions over different time periods of 1978-2017, 1978-1992, and 1993-2014.

GDP seem to be the major macroeconomic variable that drives the tourism consumption in Sweden. Our results shows that GDP is very closely correlated to the international tourism consumption in Sweden. A growth in GDP leads directly to the growth of tourism income. This aligns with previous studies conducted by Ajvaz (2015) and Bohlin, Brandt and Elbe (2014) that have already suggested that there is a one-way causality from economy to tourism in Sweden.

NOK-SEK exchange rate appeared to have a significant impact on the model over the time period of 1978-2017. However, it was insignificant when we divided the time period into before and after the devaluation of Swedish krona in 1992. USD-SEK exchange rate, on the other hand, was insignificant for all time periods.

Unemployment rate appeared to have no impact on the international tourism income for Sweden. Nature reserves and trade openness happen to be coincide with the GDP and had to be removed from the model due to the multicollinearity.

To summarize, the results of the multiple regression analysis show that GDP is a major driving variable for the international tourism consumption in Sweden regardless of the time. NOK-SEK exchange rate is a notable long-term variable for the international tourism consumption, but has no significance in a shorter time periods. USD-SEK exchange rate and unemployment appeared to have no impact on the international tourism consumption across all time periods.

The results of the should be taken as a general overview of the macroeconomic variables and international tourism income in Sweden during a specific time period. Although the
models have relatively high $R^2$ and adjusted $R^2$ values, the models themselves are unstable because the analysis is conducted on very numbers of few observations. Furthermore, in the graph of the residual versus fitted values for the model covering the time period of 1978-2017 shows a possible non-linear relationship between the regressor and the dependent variables. For these reasons, we recommend that a further research in this field of knowledge should be conducted using a different approach other than the multiple linear regression, preferably a panel data regression.
8 References


Statistical Modeling and Analytics. 2 (1), 21-33.


Tugcu, Can Tansel. (2014). Tourism and economic growth nexus revisited: A panel
causality analysis for the case of the Mediterranean Region. *Tourism Management*. 42. 207-212.


