Rebecca Fjelde Farstad

Cross-border Shopping and Inflation

An analysis of Norwegian cross-border shopping in Sweden

Bachelor's thesis in Economics
Supervisor: Doriane Mignon
May 2023

Bachelor's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Economics and Management
Department of Economics

NTNU
Norwegian University of
Science and Technology

Rebecca Fjelde Farstad

Cross-border Shopping and Inflation

An analysis of Norwegian cross-border shopping in Sweden

Bachelor's thesis in Economics
Supervisor: Doriane Mignon
May 2023

Norwegian University of Science and Technology
Faculty of Economics and Management
Department of Economics

NTNU
Norwegian University of Science and Technology

Preface

This research paper is written as a part of my bachelor's thesis in the Department of Economics at the Norwegian University of Science and Technology. I would like to thank my supervisor, Doriane Mignon, for the great feedback and good discussions on how to improve my thesis. Her assistance, answers, and advice have been crucial for the final result of this paper.

May 2023
Rebecca Fjelde Farstad

Abstract

Following the increased inflation in Sweden in 2023, Swedes started cross-border shopping in Norway. Norwegians are known for cross-border shopping in Sweden, but the effect of inflation on how much Norwegians spend on cross-border shopping in Sweden has not been investigated. Thus, this paper aims to answer the question: "Does inflation affect cross-border shopping from Norway to Sweden?". To capture the effect of inflation in both countries, differential inflation was used. A multiple linear regression model with the ordinary least squares assumptions was implemented to model the effect of the differential inflation on the sum spent in Sweden between 2004 and 2019. Hypothesis testing was utilized to test for statistical significance. The conclusion was that no statistically significant effect was found of differential inflation on the sum spent in Sweden. This could be a result of there being no effect on the population, however, there are possibilities for improving the model which could give a different conclusion.

Key Words: Inflation, Differential Inflation, Cross-Border Shopping, Norwegian Cross-Border Shopping in Sweden

Content

Tables

Figures

1 - Introduction

“Harryhandel” is the phenomenon of Norwegians travelling abroad, mostly to Sweden, to buy goods they have in their home country for a lower price (NAOB, n.d.). The act got its own nickname after cross-border shopping has increased significantly since the early 2000s, and now Norwegians go on a “harrytur” to do “harryhandel” (Remix Consulting, 2021). Norway and Sweden share a long border in which all citizens can travel freely due to the Nordic Border Union. This in combination with a long border, Sweden's considerably lower price level and lower taxes on goods like tobacco and alcohol compared to Norway, leads to Norwegians travelling to Sweden to cross-border shop (Steen et al., 2020).

Norwegians shopping in Sweden is far from the only instance of cross-border shopping recorded. Swedes shop in Denmark, Danes and Swiss shop in Germany, Canadians shop in the US, etc. This type of shopping is the most simple form of cross-border trade; travelling abroad, usually by car, buying goods and bringing them home. While this might be a good opportunity for consumers to save money, the state and local businesses can lose great amounts of money on people's cross-border shopping. In 2019, the grocery stores in Norway lost over 9 million NOK as an effect of people travelling abroad to buy goods (NTB, 2020). In addition, the state loses tax money on goods they could have sold in Norway but that were bought abroad (Asplund, Friiberg and Wilander, 2006). Thus, economists have looked into what affects cross-border shopping, and in some instances researched how to minimise it.

However, in 2023 one effect that was previously not included in this research became a part of news headlines in Norway. The inflation in Sweden had increased to such a level that Swedes now travelled to Norwegian stores to buy certain foods and products, despite the historically higher price level of Norway compared to Sweden (Kampestuen et al., 2023). This made me wonder whether or not inflation levels in Norway and Sweden affect cross-border shopping from Norway to Sweden, ad what the effect might be. Thus, this paper aims to answer the question:

“Does inflation affect Norwegian cross-border shopping in Sweden?”

To investigate this, STATA17 is utilized to create a regression model investigating how inflation affects how much money Norwegians choose to spend on cross-border shopping in Sweden. The model is based on the theoretical framework presented, and available data from Norway and Sweden. The analysis is done using a multiple linear regression model, which utilizes the ordinary least squares assumptions. An explanation and discussion of the model and these assumptions are presented. Hypothesis testing is used to determine whether or not the effect is statistically significant. Lastly, a part identifying possible deficiencies of the model, criticism of the model and a conclusion of the findings are presented.

2 - Theoretical framework and method

This part of the paper examines the theoretical framework of cross-border shopping and inflation. The first part looks into why people choose to cross-border shop and how inflation is measured. The second part explains the method used for the model and analysis. The Ordinary Least Squares (OLS) assumptions, goodness of fit and hypothesis testing are introduced and explained.

2.1 - Theoretical framework

2.11 - Cross-border shopping

SSB defines cross-border shopping as “Norwegians' physical trade abroad on trips abroad without accommodation” (SSB, 2023). The act of cross-border shopping is predominantly motivated by buying goods you could get in the country you live in for a lower price in a different country (Remix Consulting, 2021 and Asplund et al., 2001, 142). According to this, the price elasticity of demand would be the main factor influencing the change in levels of cross-border shopping. However, there are other factors in addition to price that affect whether or not a consumer chose to cross-border shop, and what goods they decide to buy.

Amzi (2015) divides the characteristics that affect cross-border shopping into consumer characteristics and market characteristics. Consumer characteristics refer to the demographic and include age, marital status, gender, education level, occupation and average income per month (p. 29). The main market characteristics are cheaper prices and a higher variety of goods but also include lower taxes, transportation, communication, social status and opening hours (p. 30-31). In addition to this, Ferris (2001) models the effect of exchange rates on cross-border shopping. The results show that when a country’s currency appreciates compared to a neighbouring country’s currency, it increases the number of day trips people do (p. 807). This suggests that the exchange rate also is a key factor in the change in cross-border shopping.

In his paper, Ferris (2001) also models the decision-making of a consumer that decides to cross-border shop and he uses the context of Canadians cross-border shopping tendencies between 1972-1997. He creates a utility model in which maximisation of utility given several factors that affect cross-border shopping. The consumers' time and income restraints are used in the most basic equations in his model, and he further expands this by including exchange rate and tax collections by the government (p. 803-804). He concludes that no one factor alone can explain a high rise and fall in cross-border shopping, but that there are many variables that may provide an insight into the fluctuations. His conclusion on what external factors led to the increased cross-border shopping from Canada to the US include a fall in the exchange rate, making the Canadian dollar worth less, an increase in Canadian unemployment rates, heavier taxes on goods in Canada compared to the US, and an increase in the price of finished goods in Canada (Ferris, 2001, p. 822).

While these papers provide several insights into what affects cross-border shopping, one effect has proved that it might be significant in the Nordic countries; inflation. In 2023, Swedes travelled to Norway to shop for goods despite Norway's higher price level and increased inflation (Lilleås, 2023 and SSB, 2023). When interviewing the Swedes that shop in Norway, the high inflation, and following high price increases for several goods, was the main reason for this shift (Kampestuen et al., 2023). While it is clear that price is a major effect in determining levels of cross-border shopping, few studies investigate if inflation has a direct effect on spending habits across the border.

2.12 - Inflation

Inflation is defined as an increase in prices over time. While there are several ways of measuring inflation, both Norway and Sweden measure it as the change in the consumer price index (CPI) (SSB, 2023). The CPI measures the general price level in an economy and is estimated by comparing the prices of a set "basket" of goods over time. These goods are rarely changed, and the price changes are used to estimate the price change of an entire economy.

The theory of low and steady inflation has presented itself as the most popular way of using inflation to benefit the economy (European Central Bank, 2023). However, shocks in the economy, like a natural disaster or sudden increase in activity, can lead to inflation levels increasing at a rapid pace. In 2023 the low-interest rates of 2020-2022 led to increased activity in both Norway and Sweden. Following this came higher inflation levels and increased prices. However, looking at the differential inflation, Sweden's inflation level increased more than Norway's (Kampestuen et al., 2023). Sweden's CPI in March of 2023 was estimated to be 10.6% compared to the same period in 2022 (SCB, 2023). In Norway, the CPI for March 2023 was at 6.4% (SSB, 2023). Thus, it was the high differential inflation for Sweden that led to a shift in cross-border shopping (Kampestuen et al., 2023).

Based on the theoretical data, I expect the differential inflation to have a positive effect on the sum of money Norwegians spend in Sweden. This means that when the inflation in Norway is higher than the inflation in Sweden, I expect Norwegians to shop more in Sweden as Norwegian goods are relatively more expensive compared to the same Swedish goods.

2.2 - Method

This part of the paper first explains how the values of the model have been estimated. Following, is an introduction to the regression model using ordinary least squares (OLS) and multiple linear regression (MLR). It will also explain how we will use the goodness of fit to determine how good the model is, and hypothesis testing to find what results are statistically significant.

2.1 - OLS assumptions

MLR.1 - Linearity in parameters

Linearity in parameters assumes that there is a linear relationship between the dependent and independent variables. This is expressed as $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \dots \beta_k x_k + u$ (Wooldridge, 2018, p. 80). Non-linear variables can be included in the model, as long as the parameters are included linearly. In addition to the variables, an error term u must be included to capture any disturbance or unobserved factors that are not captured by any of the independent variables in the model. This error term also serves as a way to indicate that the model is an estimate of a population found by using a sample, and a model of the population.

MLR.2: Random sampling

Following MLR.1, MLR.2 states that the sample must be randomly drawn where each variable has the same probability of being drawn. This is expressed as $\{(x_{i1}, x_{i2}, \dots, x_{ik}, y_i): i = 1, 2, \dots, n\}$ (Wooldridge, 2018, p.80). Random sampling is important to get the best possible unbiased estimate of a population.

MLR.3: No perfect collinearity

MLR.3 states that in the sample, and thus in the population, no variables are constant. As an effect of this, there is no perfectly linear relationship between any of the independent variables. Some correlation between the variables is accepted, but a high correlation will lead to a biased result (Wooldridge, 2018, p. 80-81).

MLR.4: Zero conditional mean

This assumption states that the mean value of the error term *u* is expected to be zero across all independent variables in a sample. This is expressed as $E(u|x_1, x_2, \dots, x_k) = 0$ (Wooldridge, 2018, p. 82). MLR.4 is an assumption that is often discussed as there are several ways it can be violated. If MLR.1-MLR.4 are not violated, we assume the model to be unbiased and $E(\widehat{\beta}) = \beta$.

MLR.5: Homoscedasticity

MLR.5 states that the error term *u* has the same variance across all of the explanatory, or independent, variables. This assumption is written as $Var(u|x_1, x_2, \dots, x_k) = \sigma^2$ (Wooldridge, 2018, p.88). It means that the variance of the independent variables is the same across the whole sample, and we use this to find the standard errors of the variables.

MLR.6: Normality

Following MLR.4 and MLR.5, MLR.6 states that the population error *u* is independent of the explanatory variables and normally distributed with a zero mean and variance σ^2. For this to hold, both the zero conditional mean and homoscedasticity assumptions must not be broken. MLR.6 is written as $u \sim normal(0, \sigma^2)$ (Wooldridge, 2018, p. 118).

Assumptions MLR.1-MLR.5 are called the Gauss Markov Assumptions. When these are not violated we assume to have the Best Linear Unbiased Estimators (BLUEs). This implies that we have an unbiased model and that there is no other linear estimator that has a sample variance lower than the current estimators. In addition to this, MLR.1-MLR.6 are referred to as the Classical Linear Model Assumptions (CLM) (Wooldridge, 2019, p. 118)

2.2 - Goodness of fit

R-squared (R^2) is a value used in regression analysis that explains how much of the total variation is explained by the model compared to the total population variation (Wooldridge, 2018, p. 35). R^2 is measured by using the total sum of squared (SST), explained sum of squares (SSE) and the residual sum of squared (SSR). The total sum of squared consists of the explained and residual sum of squares, while the explained sum of squares is the variation of the independent variables and the residual sum of squares is the variation found in the error term u. R^2 can be expressed as:

$$R^2 = \frac{SSE}{SST} = \frac{(1-SSR)}{SST} = \frac{SST-SSR}{SST}$$

The R^2 is used to determine how well a model fits with the population and takes a value between 0 and 1. An R^2 of 1 means that the model explains all variation in the population, and thus is a perfect representation of the population. R^2 always increases when a new variable is added to the model. Therefore, it can be a poor tool to use when deciding whether or not the model is good and if the variables included are significant to the estimation and model (Wooldridge, 2018, p. 77).

2.23 - Hypothesis testing

A hypothesis test is a test conducted on a value to find if its effect is true or false given a significance level. The test states that there are two hypotheses, H_0 and H_1, which are mutually exclusive. H_0 is the null-hypothesis and is the hypothesis that is examined to be proven correct or incorrect given the specified significance level. H_1 is the alternative hypothesis, which will be true if H_0 is false. If H_0 is incorrect we say that we reject the null-hypothesis. To do this test finding a p-value, or a critical value, denoted as c and a test statistic, denoted as TS is needed (Wooldridge, 2018, p. 121). TS is found using the estimate and standard error of the value of interest and can be expressed as:

$$TS_{\widehat{\beta}_j} = \frac{\widehat{\beta}_j}{se(\widehat{\beta}_j)}$$

This equation is the estimated value divided by the standard error of the estimated value (Wooldridge, 2018, p. 128). It shows how many estimated standard deviations the estimated value is from the hypothesised value. The TS is distributed given t_{n-k-1}, and is found using the number of observations, n, minus the number of independent variables k, minus the constant equal to one.

The critical value, c, is the threshold level used to decide if the null-hypothesis should be rejected. The c depends on the chosen significance level and the test statistic distribution. It will also differ with the type of test conducted, a one-tailed or a two-tailed test. A one-tailed test is used to determine whether a value is significantly bigger or significantly smaller than a chosen value (Wooldridge, 2018, p. 122), while a two-tailed test is used to determine if a value is significantly different from a chosen value (Wooldridge, 2018, p. 126). As this paper aims to investigate the effect of one independent variable x on a dependent variable y without specifying whether this effect should be positive or negative, a two-tailed test will be conducted to test for statistical significance.

With both c and TS obtained, they are compared to see whether or not we reject the null-hypothesis. If we use a 5% significance level and we see that c is lower than the absolute value of TS, ($|TS| > |c|$)we say that we can reject the null-hypothesis with 95% confidence. This means that we reject H_0 if the TS is larger than the upper critical value, or smaller than the lower critical value (Wooldridge, 2018, p. 737).

3 - Data

The data used for the regression model in this paper is gathered from three different sources; the Norwegian statistical database, SSB, the Swedish statistical database, SCB, the website for Sweden's central bank, the Riksbank, and the World Bank's online database.

3.1 - Variables

The dependent variable in this model is the sum of money Norwegians spend in Sweden, measured in millions of NOK. The independent variable of interest is the differential inflation between Norway and Sweden. Instead of using the inflation levels of both countries as two variables, the differential inflation shows how the changes in both inflation levels affect the sum of money spent in Sweden. The other independent variables, the control variables, included are the exchange rate between the countries, the change in real income for Norwegians and the change in gas prices in Norway.

The differential inflation is measured by taking the Norwegian inflation rate and subtracting the Swedish inflation rate. This means that a positive relationship shows that when there are higher inflation in Norway than in Sweden, Norwegians cross-border shop more, and vice versa. The exchange rate is measured as how many SEK one can buy for 100 NOK. The change in real income and the change in gas prices are both measured to be a % change. All variables have 64 observations. In addition, there are some dummy variables for the time periods 2004-2007, 2008-2011, 2012-2015 and 2016-2019 that are included to show the time-fixed effect of differential inflation on the sum.

3.2 - Estimating data

The dependent variable

The Norwegian statistical database began publishing records on Norwegian cross-border shopping in 2004. However, the database has not singled out how much Norwegians buy specifically in Sweden in their quarterly reports, only in the yearly reports. If only the yearly values were included in the model, it would have 16 observations. This number of observations is not enough to assume a normal distribution for a time series (Wooldridge, 2018, p. 119). Therefore, an estimation of the quarterly sum of money Norwegians spend in Sweden is used.

This estimation is based on the quarterly total sum of money Norwegians spend on all cross-border shopping, the yearly total sum, and the yearly sum used only in Sweden. First, the percentage of the total yearly sum spent in Sweden was calculated. This percentage was then applied to all the total quarterly sums of that year. By doing this we assume that the change in the total sum and the change in the sum of money spent in Sweden are perfectly correlated.

The independent variables

While the sum spent in Sweden had accurate yearly data, the variables inflation and gas prices had monthly data. As inflation is measured as the change in CPI in both Norway and Sweden, that is how it has been estimated in this model. An average of the intervals Jan-March, April-June, July-Sept and Oct-Dec was calculated, and this value estimates the quarterly values. Thus, the differential inflation is the difference between the quarterly change in CPI in Noway and the quarterly change in CPI in Sweden. Data on the quarterly real income in Norway existed, and the change in real income and change in gas prices was found by calculating the percentage change from one quarter to the next.

3.3 - Descriptive statistics

Using central tendency and dispersion, we can present the raw statistics obtained in their most basic form. Here, the number of observations, mean, standard deviation, minimum value and maximum value are presented. The standard deviation explains how much the values in the data obtained vary from the mean value.

VARIABLES	N	mean	sd.	min	max
sum	64	2,708	751.2	1,485	4,751
diffinf	64	0.843	1.359	-1.783	3.601
chincome	64	0.664	2.712	-9.100	8.300
exchrate	64	111.8	6.235	97.87	122.8
chgas	64	0.970	4.243	-14.60	9.007

Table 1: Descriptive statistics

The dependent variable *sum* has a range between 1485,492 million NOK and 4751,096 million NOK. It has a mean value of 2707.658 million NOK and a standard deviation of 751,2183. The differential inflation in percentage, *diffinf*, has a range from -1,78% to 3,6%. The mean value is 0.843% and the standard deviation of 1.3589. The positive mean value shows that Norway historically has higher levels of inflation than Sweden. This is also shown in *Figure 1*. It is worth mentioning that in some years, Sweden has had deflation rather than inflation. This explains why the maximum value of *diffinf* is over 3,5% while Norway rarely reaches such high levels of inflation. The percentage change in real income for Norway, *chincome*, has a range between -9.1% and 8.3%, the mean value is 0.664% and the standard deviation is 2,7115. Lastly, the change in gas prices, *chgas*, has a minimum value of -14.6% and a maximum value of 9.007%. The mean is 0.9697%, and the standard deviation is 4.243.

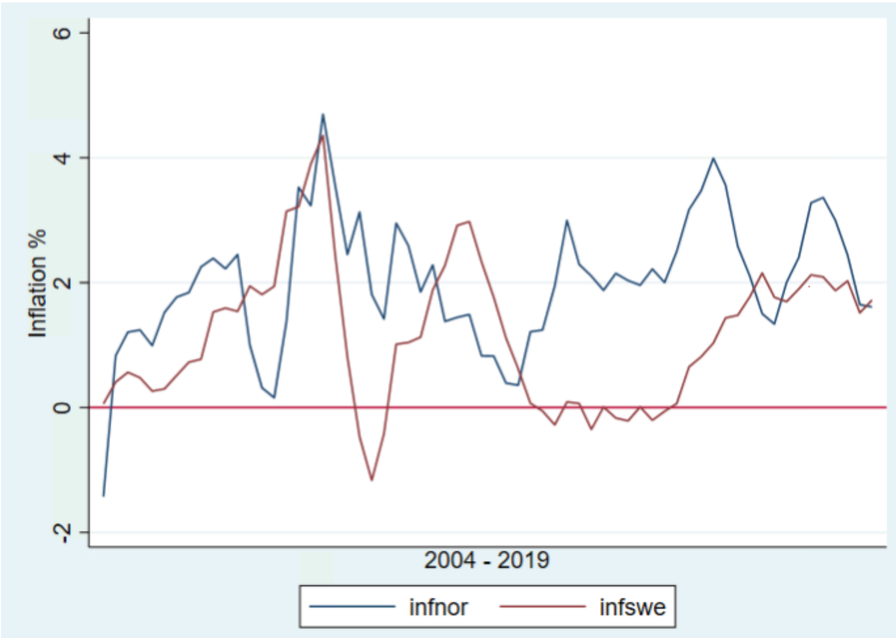

Figure 1: Inflation % for Norway and Sweden 2004-2019

In addition to these values, the dummy variables of 2008-2011, 2012-2015 and 2016-2019 are included in the model. These values capture the time-fixed effect and obtain a value of 0 or 1. The reference period is the effect between the years of 2004-2007, and the other dummy variables explain the change we see in the other time periods compared to the reference period. These dummies do not explain the singled-out effect of differential inflation on the sum, but the effect all the dependent variables over time have on the sum. Since the dummies all have the same amount of observations and only obtain the values 0 or 1, they have the same mean value and standard deviation of 0.25 and 0.436 respectively.

4 - Regression analysis

The regression analysis is done using the Ordinary Least Squares model using STATA 17.0. This model makes it possible to investigate any relationship between the differential inflation and the sum of money spent in Sweden, and whether this is positive or negative.

4.1 - Model

The following models are estimated:

(1) $\widehat{Sum} = \beta_0 + \beta_1 diffinf + u$

(2) $\widehat{Sum} = \beta_0 + \beta_1 diffinf + \beta_2 period2 + \beta_3 period3 + \beta_4 period4 + u$

(3) $\widehat{Sum} = \beta_0 + \beta_1 diffinf + \beta_5 exchrate + \beta_6 chincome + \beta_7 chgas + u$

(4) $\widehat{Sum} = \beta_0 + \beta_1 diffinf + \beta_2 period2 + \beta_3 period3 + \beta_4 period4 + \beta_5 exchrate + \beta_6 chincome + \beta_7 chgas + u$

VARIABLES	(1) sum	(2) sum	(3) sum	(4) sum
diffinf	117.7*	-10.43	64.40	-7.902
	(68.60)	(46.83)	(63.17)	(46.80)
2008-2011		448.6***		363.8*
		(166.2)		(182.0)
2012-2015		1,097***		1,094***
		(177.9)		(182.8)
2016-2019		1,569***		1,686***
		(168.8)		(213.8)
exchrate			-54.88***	13.85
			(13.91)	(15.70)
chincome			22.75	28.53
			(31.03)	(21.64)
chgas			-14.14	-20.63
			(20.18)	(14.36)
Constant	2,608***	1,938***	8,786***	381.5
	(109.1)	(116.6)	(1,564)	(1,784)
Observations	64	64	64	64
R-squared	0.045	0.640	0.264	0.664

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2: Model

Table 2 shows the models of equations (1) to (4). From column (1) we can interpret that there is a positive relationship between the differential inflation and the sum of money spent in Sweden. A one percentage point increase in the differential inflation is expected to increase the sum spent by 117,7 million NOK, all else equal. However, looking at the time-fixed effect in column (2), the relationship goes from positive to negative. There is a reduction in the constant in column (2), from 2608 to 1938. In addition, there is a significant increase in r^2 between the models in columns (1) and (2). The increase from the model explaining 4.5% to explaining 64% of the variance in the sum suggest that time is an important factor.

Column (3) shows the non-time-fixed model when the control variables are included. There is a positive relationship between differential inflation and the sum Norwegians spend in Sweden, though it has been reduced from 117,7 in column (1) to 64,39 million NOK per percentage point increase in the differential inflation. For the change in income and change in gas prices, the results are as expected. When the Norwegian real income increase by one percentage point, the sum of money spent in Sweden increases by 72,75 million NOK, all else equal. This means that when Norwegian households increase their disposable income, they spend more money on goods in Sweden. When there is a one percentage point increase in gas prices in Norway, the sum of money spent in Sweden on cross-border shopping decreases by a little over 14 million NOK. As the cost of travelling to Sweden increases, Norwegians are less willing to travel, and the sum of money spent in Sweden decreases.

However, for the exchange rate, there is a negative relationship. For each percentage point increase in the exchange rate from NOK to SEK, Norwegians spend 54,88 million NOK less on cross-border shopping in Sweden, all else equal. One explanation of this could be that when the value of the NOK increases they have to spend less NOK to buy the same amount of goods in Sweden. However, there might also be some effect of an excluded independent variable that is captured by the exchange rate, leading to the unexpected effect.

Column (4) shows the full, time–fixed, model. As in column (2), there is a negative relationship between the differential inflation and the sum of money spent in the first time period. In addition, we can observe that r^2, again, increases significantly when the time-fixed effect is included. This strengthens the indication of time being an important factor in explaining the changes in the independent variable.

In column (4) there is also a reduction in how much of the model is explained by the differential inflation and the constant. The value of the constant reduces when comparing the constants in columns (1) and (3) to the constants in the time-fixed models in columns (2) and (4). This, again, indicate that time itself is a factor affecting the sum Norwegians spend in Sweden on cross-border shopping. In column (4) there is also an increase in how much of the model is explained by the change in income and change in gas prices from column (3). The change in income increased from 22,75 to 28,5 million NOK per percentage point and the change in gas prices increased from -14 to -20,6 million NOK per percentage point. Interestingly, the effect of the exchange rate has changed from a negative to a positive relationship in the full model.

4.11 Hypothesis tests

As there is no clear explanation of why the effect of the differential inflation changes from positive to negative, it could indicate that one or both of the values are statistically insignificant. To test this, a hypothesis test will be done on both the non-time-fixed model, column (3), and the time-fixed model, column (4). All tests will have a confidence interval of 95%.

Non-time-fixed model

The hypothesis is formulated as:

H_0: $Variable\ is\ = 0 \Leftrightarrow H_0: \beta_{diffinf} = 0$

H_1: $Variable\ is\ \neq 0 \Leftrightarrow H_1: \beta_{diffinf} \neq 0$

The test statistic $TS_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} = \frac{64.3966}{63.16601} = 1.02$. The degrees of freedom are 59, and from this, we find that the critical value is 2.001. The null-hypothesis is rejected if the critical value is bigger than the absolute value of the test statistic, $c > |TS|$. For the non-time fixed model, we find that 2.001 > 1.02, which means that the null-hypotheis cannot be rejected. We can, with 95% certainty, state that the value is equal to zero, and thus not statistically significant. It cannot be determined that differential inflation has an effect on the sum of money spent in Sweden.

Time-fixed model

The hypothesis is formulated as:

H_0: $Variable\ is\ =\ 0\ \Leftrightarrow\ H_0$: $\beta_{diffinf}\ =\ 0$

H_1: $Variable\ is\ \neq\ 0\ \Leftrightarrow\ H_1$: $\beta_{diffinf}\ \neq\ 0$

The test statistic $TS_{\widehat{\beta}_j} = \frac{\widehat{\beta}_j}{se(\widehat{\beta}_j)} = \frac{-7.902131}{46.80307} = -0.17$. The degrees of freedom is 56, which is lower than for the non-time-fixed model as there are more independent variables. The critical value is 2.003 and null-hypothesis is still rejected if $c > |TS|$. For the time-fixed model, 2.003 > 0.17, the null hypothesis cannot be rejected. As for the non-time-fixed model, we can with 95% certainty state that the value is equal to zero, and thus not statistically significant. It cannot be determined that differential inflation has an effect on the sum of money spent in Sweden.

4.2 - Reviews of OLS assumptions

This section goes evaluates the OLS assumptions MLR.1 - MLR.6 to check whether or not they hold for the model. Hypothesis testing using two-tailed tests and chi² testing, tests of correlation and modelling in STATA.17.0 is used.

MLR.1 - Linearity

MLR.1 states that all parameters must be linear to be able to use the OLS-model. This can be tested by including the quadratic terms of the independent variables in the model. A hypothesis test is done to investigate if there is a statistically significant effect between the non-linear independent variable and the dependent variable (Grambsch & O'Brien, 1991).

The hypothesis is formulated as:

H_0 : *Variable is linear*

H_1 : *Variable is non − linear*

$H_0: \beta1\ =\ 0, \beta2\ =\ 0,\ \beta3\ =\ 0, \beta4\ =\ 0$

$H_1: \beta1 \neq 0,\ \beta2 \neq 0,\ \beta3 \neq 0, \beta4 \neq 0$

VARIABLES	(1) Linearity test
diffinf	129.1
	(86.56)
exchrate	625.2
	(486.1)
chincome	3.932
	(31.31)
chgas	-3.630
	(20.66)
diffinf2	-0.119
	(43.07)
exchrate2	-3.014
	(2.182)
chincome2	-15.23**
	(5.776)
chhgas2	-3.046
	(2.586)
Constant	-29,341
	(27,023)
Observations	64
R-squared	0.375

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3: Test for linearity

For differential inflation, exchange rate and the change in gas prices we can conclude that there is linearity as there are no stars indicating significance on the squared terms. However, the squared term of the change in income appears to be statistically significant. Usually, when this is the case, the variable will be included in the model, however, as it is not the variable of interest it has been omitted. We can conclude that the assumption of linearity partially holds.

MLR.2 - Random Sampling

Assuming random sampling means assuming that each observation, *i*, in the sample has the same probability of being selected. For this assumption to hold true, the sample size must be sufficiently large. Anderson (2017) explains through the central limit theorem that as long as there are more than 30 observations, $n \geq 30$, random sampling holds. As there are 64 observations in this data set and $64 \geq 30$, random sampling can be assumed to hold.

However, time-series data is data from several observations done on the same subject over time. Because of this, we cannot assume random sampling as the subject and time have been selected and not randomly drawn. The observations cannot be independent of each other, but rather they depend on the earlier observations of the variable. (Wooldridge, 2019, p. 343). An example of this is that the inflation levels of 2006 were dependent on the inflation in 2005, which was dependent on the year before and so on.

This indicates that MLR.2 does not hold. Still, doing some corrections can make it possible to use time-series data and still have MLR.2 fulfilled, including achieving stationery. Stationary means that there is the same distribution for all independent variables and that the correlation between the terms is the same across all time periods. (Wooldridge, 2019, p. 367). As it can be difficult to achieve stationary, the results of the analysis should be interpreted with the knowledge that MLR.2 might not hold.

MLR.3:

MLR.3 states that there must be enough variation in the independent variables and no perfect collinearity. Collinearity is the correlation between the variables used in an analysis. If there are perfectly correlated variables, MLR.3 does not hold, though some correlation is acceptable. If two variables perfectly correlate, one variable could be written as a direct expression of the other, and the coefficients would work in the same way (Wooldridge, 2018, p.81). This can be measured in two ways: in a collinearity matrix or using the Variance Inflation Factor (VIF).

Variables	(1) sum	(2) diffinf	(3) exchrate	(4) chincome	(5) chgas	(6) 2008-2011	(7) 2012-2015	(8) 2016-2019
(1) sum	1.000							
(2) diffinf	0.213	1.000						
(3) exchrate	-0.488	-0.196	1.000					
(4) chincome	0.048	-0.024	0.059	1.000				
(5) chgas	-0.169	-0.115	0.178	0.058	1.000			
(6) 2008-2011	-0.254	-0.083	0.603	0.048	0.022	1.000		
(7) 2012-2015	0.240	0.325	-0.138	0.044	-0.152	-0.333	1.000	
(8) 2016-2019	0.611	0.051	-0.644	-0.096	0.011	-0.333	-0.333	1.000

Table 4: Collinearity matrix

If two variables have a correlation value above 0.9, we assume a very high correlation to the point where it could pose a problem in the model (Wooldridge, 2018, p. 92). None of the values seen in the matrix are close to 0.9, and MLR.3 is assumed to hold.

	VIF	1/VIF
exchrate	2.827	.354
2016-2019	2.569	.389
2012-2015	1.877	.533
2008-2011	1.862	.537
diffinf	1.193	.838
chgas	1.095	.913
chincome	1.016	.985
Mean VIF	1.777	.

Table 5: VIF factor

Using the same logic as for the collinearity matrix, if a VIF factor has a value of 10 we can assume that there is a problem with collinearity in the model (Wooldridge, 2018, p. 92). As none of the values has a VIF close to 10, the conclusion of no collinearity is strengthened.

MLR.4

MLR.4 is one of the most discussed assumptions when it comes to whether or not a regression is unbiased. This assumption states that there is zero conditional mean, or that the error term has an expected value of zero given any values of the independent variables (Wooldridge, 2018, p. 82). This assumption is hard to satisfy as it is often violated due to omitted variables. When there is a factor in the error term that is correlated with the dependent variable and one or more independent variable(s), we cannot assume that MLR.4 holds (Wooldridge, 2018, p. 87).

For the model of *diffinf* on *sum*, it has already been mentioned that some possibly relevant factors have been omitted due to the lack of data. How much Norwegians spend in Sweden on cross-border shopping and the differential inflation are big variables that have many other factors affecting them. This can be a problem, so it will be important to know that the results of the analysis should be looked at with a critical eye regarding MLR.4. However, the assumption has been considered to the best degree possible, it is assumed to hold for the model and analysis in this paper.

MLR.5

MLR.5 is the assumption of homoscedasticity and relates to the variance of the error term, u. It states that the error is constant over the observed factors, e.g. that there is no increase or decrease in the variance of u over time (Wooldridge, 2018, p. 262). Whether or not this holds can be tested in two ways: using a scatter plot of the variance of the error term, and doing a Breuch-Pagan hypothesis test (Wooldridge, 2018, p. 270).

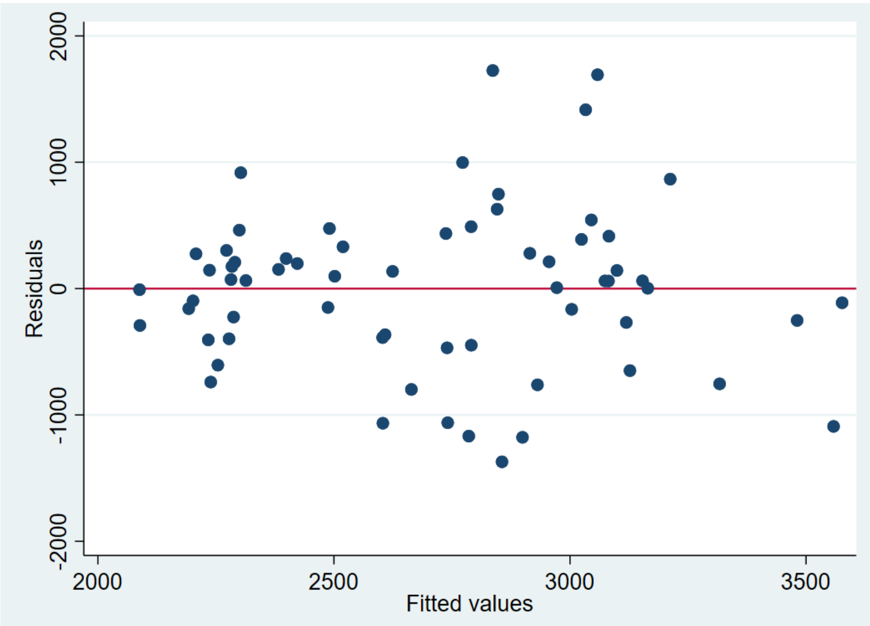

Figure 2: Test for homoscedasticity

From *Figure 2* we can see that the variance in u is not perfectly constant across the variables in the model. To check whether or not the variance is a problem for the model, the Breuch-Pagan hypothesis test is done.

The following hypothesis is:

H_0: *Homoscedasticity is present*

H_1: *Heteroscedasticity is present*

$chi^2(1) \;=\; 7.73$

$Prob \;>\; chi2 \;=\; 0.0054$

For this test, a 5% significance level is used. The p-value is found to be 0.0054. This means that the null-hypothesis can be rejected as the p-value is not greater than 0.05. Thus, there is assumed to be heteroscedasticity and that the residuals are not distributed with equal variance. We conclude that MLR.5 does not hold.

MLR.6

MLR.6 states that the population error, *u*, is independent of the explanatory variables and is normally distributed with zero mean and variance (Wooldridge, 2018, p. 118). This assumption rarely holds as it states that the error term is independent of all variables. Like with MLR.5, we can test whether or not this assumption holds by graphing the kernel density estimate or doing a Jarque-Bera test.

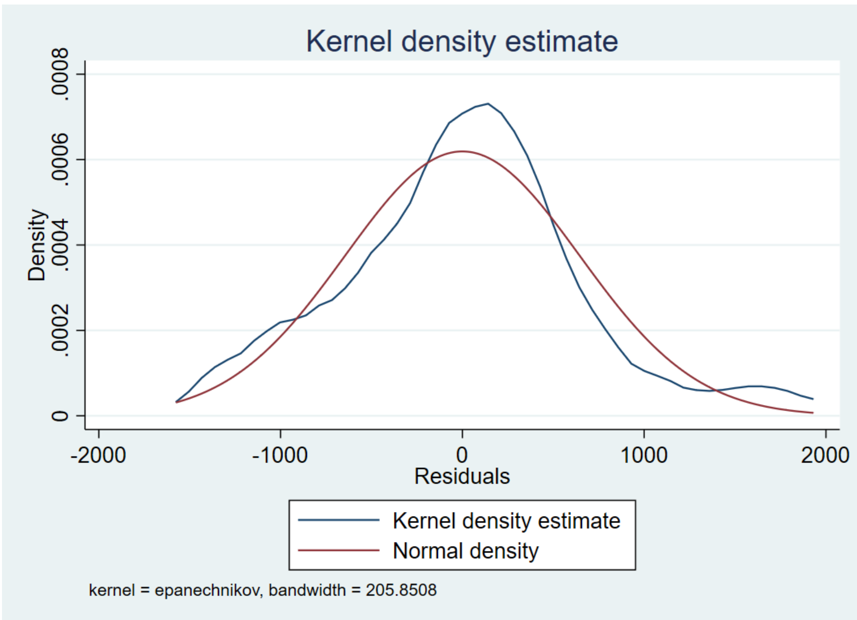

Figure 3: Test for normality

Looking at *Figure 3*, we see that the model is close to normally distributed. However, as it is not a perfect normal distribution, doing a hypothesis test will show whether or not MLR.6 holds. The hypothesis done to check goodness-of-fit for normality is called the Jarque-Bera test (Jarque, 2014, p. 701).

The following hypothesis is:

H_0: *The data follows a normal distribution*

H_1: *The data does not follow a normal distribution*

$chi^2(2) = 0.56$

$Prob > chi2 = 0.7541$

This test is also done with a 5% significance level, and the p-value obtained is 0.7541. As the p-value is greater than 0.05, 0.7541 > 0.05, we state that the data follow a normal distribution with 95% confidence. We thus assume MLR.6 to hold.

5 - Discussion

The result of whether or not there is a relationship between differential inflation and the sum of money Norwegians spend in Sweden came back inconclusive. The null-hypothesis stating that the coefficient of differential inflation was equal to zero could not be rejected, and the results were found to be insignificant. However, no significance in this model does not mean that there is no relationship between the variables in the population. Therefore we cannot conclude that there is no relationship between the sum of money and the differential inflation, though we conclude that no statistically significant relationship was found using this model.

Both Amzi (2015) and Ferris (2001) stated that price is a major factor in how consumers chose to cross-border shop. However, they pinpointed several other factors, including some that can change prices aside from inflation. Ferris (2001) explained that the increased tax levels of certain goods in Canada led to more Canadians shopping across the border, and Amzi (2015) came to the same general conclusion. For future investigations on the topic of cross-border shopping and inflation, a variable capturing the tax difference should be included to account for another reason for changes in price that are not related to CPI.

While including this, it could also be interesting to look into whether there is a statistically significant relationship between differential inflation and the sum of money spent on certain groups of goods. The Norwegian and Swedish governments tax goods like tobacco, alcohol and sweets very differently from food and non-alcoholic drinks (Skatteetaten, n.d. and Skatteverket, n.d.). Creating a model investigating whether or not the differential inflation has an effect on how much Norwegians buy of different groups of goods could give a clearer result on if there is a relationship, and whether different taxes alter that relationship.

Another factor Amzi (2015) touched on is how far a consumer live from the border. The model in this paper investigates how differential inflation affects Norwegian cross-border shopping in Sweden on a national level. Gathering data on where the consumer travel from and including this in the model could give an indication of whether or not the relationship between differential inflation and the sum is affected by region and distance. As Amzi (2015) explained, there is more leisure for a consumer that has to travel far, and thus it could indicate that the price elasticity of demand is higher for people living further from the border.

6 - Criticism

Estimation of variables

In section 2.1 it was explained how some of the variables are estimated rather than found. For the *sum*, the variables are based on two different tables, the total sum spent by Norwegians on cross-border shopping per quarter, and the total sum spent by Norwegians in Sweden per year. It is assumed that the percentage of all cross-border shopping happening in Sweden stayed constant for all quarters, and thus estimated the *sum*. There is no possibility of proving or denying whether or not this assumption holds. The annual percentage of cross-border shopping done by Norwegians in Sweden is between 82% and 95% across the years in this study. Therefore, the estimation is most likely not far from the truth as Norwegians mostly cross-border shop in Sweden. Still, the reader should know that there might be some seasonal changes in the sum that is not captured in the model.

Omitted variables

As mentioned in section 3.1, several variables could have been included in the model to make it more explanatory and robust, however, they were omitted due to the lack of data. Some of the control variables that could have been included are the demographics of the shoppers and the tax levels in Norway and Sweden. The demographic could have given results on how the differential inflation affected people of different age groups, genders and especially closeness to the border as these are factors we know to determine how people chose to cross-border shop (Ferris, 2001, 813). For tax levels, the theory shows how many cross-border shoppers chose to shop as the taxes and fees are lower on certain goods (Amzi, 2015, p.31). The squared term of the change in income could also be included as it proved to be statistically significant. This should be investigated further in future analyses.

Random Sampling

In part 4.2, it was mentioned that there are some issues regarding time-series datasets and the random sampling assumption. As time-series data is a set of observations with temporal ordering, random sampling will not hold for the population in question. This has been taken into account as best as possible, but the reader should be aware that it is a weakness of all OLS regression models done using time-series data.

Goodness of fit

As explained in section 2.22, the r^2 of a model shows how much of the effect on the dependent variable is captured by the independent variables. We want this variable to be as close as possible to 1, while still having no multicollinearity. For the full, time-fixed, model in column (4), r^2 is 0,66, which is a satisfactory value. It shows that the model only explains 66% of the variance in the *sum*, however, r^2 values above 0,5 is usually considered good when investigating human behaviour. A low r^2 might not be a reason for concern as it also means lower variance. It is still worth mentioning that the model does not explain all of the variance in the *sum* so the reader can take it into account when reading this paper.

OLS-assumptions

Two of the OLS-assumptions were concluded not to hold; linearity and homoscedasticity. For linearity, the squared term of the change in income proved to be statistically significant. Including the squared term in the model could lead to more reliable results. For homoscedasticity, it was concluded that it does not hold across the model. This can lead to unreliable confidence intervals. In addition, it is assumed that the zero-conditional mean assumption holds, though it cannot be tested. Due to these factors, the model is not a BLUEs model - it is not the best linear unbiased estimate. For future analysis, the squared term of the change in income should be evaluated and possibly included in the model. The heteroscedasticity and zero-conditional mean assumption should be considered when studying the results in this paper.

7 - Conclusion

This research paper aimed to investigate if inflation levels have an effect on cross-border shopping. This topic became of interest following the increased inflation in Norway and Sweden in 2023, leading to increased cross-border shopping from Sweden to Norway despite the historically higher price level in Norway. Due to this, the expectation was that when inflation increases in Norway, Norwegians would buy more goods in Sweden. A time-fixed multiple linear regression model was set up to investigate this question:

"Does inflation affect Norwegian cross-border shopping in Sweden?"

To capture the effect of changes in inflation in Sweden, differential inflation was utilized. Using hypothesis testing to test the statistical significance of the effect, the results came back as insignificant. Thus, we conclude that this model finds no significant effect of differential inflation on the sum of money Norwegians spend in Sweden. However, it appeared that time was an important factor in explaining the changes in the sum.

The insignificant results in the model could be a result of there being no effect of differential inflation on the sum spent in Sweden. Yet, the results could show different results if the model was changed and improved in regards to excluded variables and strengthening the hold of the multiple linear regression assumptions. To increase the robustness of a model examining inflation and cross in future research, these factors should be taken into consideration.

8 - References

Amzi, A. (2015, December). Factors influencing cross-border shopping. *ESTEEM article journal*, *11*(2), 27-34.

Asplund, M., Friberg, R., & Wilander, F. (2001, February). Demand and distance: Evidence on cross-border shopping. *Journal of Public Economics*, *91*(1-2), 141-157. https://doi.org/10.1016/j.jpubeco.2006.05.006

Cochran, J. J., Williams, T. A., Sweeney, D. J., Anderson, D. R., & Camm, J. D. (2017). *Essentials of Modern Business Statistics with Microsoft Office Excel* (7th ed.). Cengage Learning.

European Central Bank. (2023). *Benefits of price stability*. European Central Bank. Retrieved April, 2023, from https://www.ecb.europa.eu/mopo/intro/benefits/html/index.en.html

Ferris, J. (2001, February). The Determinants of Cross Border Shopping: Implications for Tax Revenues and Institutional Change. *National Tax Journal*, *53*(4), 801-824. https://www.researchgate.net/publication/228128864

Grambsch, P. M., & O'Brien, P. C. (1991, May). The effects of transformations and preliminary tests for non-linearity in regression. *10*(Statistics in Medicine), 697-709. https://doi.org/10.1002/sim.4780100504

Jarque, C. M. (2014, December). Jarque-Bera test. *International Encyclopedia of Statistical Science*, 701-702. https://doi.org/10.1007/978-3-642-04898-2_319

Kampestuen, K. Å., Haagesen, V. W., & Meskau, F. (2023, March 18). Svenske kunder strømmer over grensa og fyller handlekurvene med matvarer. *NRK*. Retrieved March, 2023, from https://www.nrk.no/innlandet/svenske-kunder-strommer-over-grensa-og-fyller-handlekurvene-med-matvarer-1.16339550

Lilleås, H. S. (2023, March 6). Sverige, Svenskehandel | Ekstremt prishopp på mat i Sverige. *Nettavisen*. Retrieved March, 2023, from https://www.nettavisen.no/okonomi/ekstremt-prishopp-pa-mat-i-sverige/s/5-95-955107

NAOB. (n.d.). *harryhandel - Det Norske Akademis ordbok*. NAOB. Retri eved April, 2023, from https://naob.no/ordbok/harryhandel

NTB. (2020, May 14). Ny rapport: Norsk dagligvarebransje taper 9 milliarder kroner på grensehandelen. *E24*. Retrieved March, 2023, from https://e24.no/naeringsliv/i/opq2vB/ny-rapport-norsk-dagligvarebransje-taper-9-milliarder-kroner-paa-grensehandelen

Remix Consulting. (2021, September 6). *Historisk grensehandel*. Remix Consulting AS. Retrieved May 11, 2023, from https://remixconsulting.com/historisk-grensehandel/

SCB. (2023). *Consumer Price Index (CPI)*. Statistics Sweden. Retrieved May, 2023, from https://www.scb.se/en/finding-statistics/statistics-by-subject-area/prices-and-consumption/consumer-price-index/consumer-price-index-cpi/

SCB. (2023). *Consumer Price Index (CPI), total 1980=100. Month 1980M01 - 2023M03* [Data set used for inflation in Sweden]. Statistics Sweden. Retrieved March, 2023, from https://www.statistikdatabasen.scb.se/pxweb/en/ssd/START__PR__PR0101__PR0101A/KPItotM/table/tableViewLayout1/

Skatteetaten. (n.d.). *Særavgifter*. Skatteetaten. Retrieved May 13, 2023, from https://www.skatteetaten.no/bedrift-og-organisasjon/avgifter/saravgifter/

Skatteverket. (n.d.). *Punktskatter*. Skatteverket. Retrieved May 13, 2023, from https://www.skatteverket.se/foretag/skatterochavdrag/punktskatter.4.71004e4c133e23bf6db800057013.html

SSB. (2023). *05678: Cross border trade, by contents, destination and year* [Data set used for the sum of money spent in Sweden by Norwegians,]. SSB. Retrieved March, 2023, from https://www.ssb.no/en/statbank/table/05678/tableViewLayout1/

SSB. (2023). *08460: Cross border trade, by quarter and contents* [Data set used for the sum of money spent in Sweden by Norwegians,]. Statistics Norway. Retrieved March, 2023, from https://www.ssb.no/en/statbank/table/08460/tableViewLayout1/

SSB. (2023). *08981: Consumer Price Index (2015=100), by contents, month and year* [Data set used for inflation in Norway]. Statistics Norway. Retrieved March, 2023, from https://www.ssb.no/en/statbank/table/08981/tableViewLayout1/

SSB. (2023). *09654: Prices on engine fuel (NOK per litres), by petroleum products, month and contents* [Data used for changes in gas prices in Norway]. Statistics Norway. Retrieved March, 2023, from https://www.ssb.no/en/statbank/table/09654/tableViewLayout1/

SSB. (2023). *11020: Quarterly non-financial sector accounts. Income, expenditure and saving, by sector, transaction, quarter and contents* [Data used for changes in disposable income in Norway]. Statistics Norway. Retrieved March, 2023, from https://www.ssb.no/en/statbank/table/11020/tableViewLayout1/

SSB. (2023, February 24). *Cross border trade*. Statistics Norway. Retrieved May, 2023, from https://www.ssb.no/en/varehandel-og-tjenesteyting/varehandel/statistikk/grensehandel

SSB. (2023, April 11). *Consumer price index*. Statistics Norway. Retrieved April, 2023, from https://www.ssb.no/en/priser-og-prisindekser/konsumpriser/statistikk/konsumprisindeksen

Steen, F., Ulsaker, S., & Friberg, R. (2020, December 9). Svenskehandelen vil fortsette på tross av billigere alkohol, brus og snus. *NHH*. Retrieved April, 2023, from https://www.nhh.no/nhh-bulletin/artikkelarkiv/2020/desember/svenskehandelen-vil-fortsette-pa-tross-av-billigere-alkohol-brus-og-snus/

Sveriges Riksbank. (2023). *Annual average exchange rates (aggregate)* [Data used for the exchange rate between Norway and Sweden]. Riksbank. Retrieved March, 2023, from https://www.riksbank.se/en-gb/statistics/search-interest--exchange-rates/annual-average-exchange-rates/

Wooldridge, J. M. (2018). *Introductory Econometrics: A Modern Approach* (7th ed.). Cengage Learning.

NTNU

Norwegian University of
Science and Technology