1 Introduction

Start of Introduction

1. Our problem: “Predicting the pricing of salmon using similar commodities and macroeconomic factors”

Why is our problem important and what is its relevance?

1.1 Fish as a commodity

”As the largest traded food commodity in the world, seafood provides sustenance to billions of people world wide. More then 3 billion people in the world rely on wild-caught and farmed seafood.” (WWF, 2019) Seafood has been a traded commodity for hundreds of years, and for most of this time has been one of Norway’s biggest exports. ”Norwegian clipfish has been exported to Southern Europe and beyond since the early 1700s.” (Norway, 2023) Fish world wide has been an important source of protein. Facing an ever larger growing population and the sustainability issues that comes with it, the seafood industry plays a major role in solving these issues.

1.2 Why salmon?

”The salmon industry is one of the biggest industries in Norway.” (Johansen et al., 2019) The companies in the industry impact the rest of the Norwegian economy and society as a whole through labor and culture. This in turn means that the Norwegian society is indirectly effected by the the price of salmon on the open market. Salmon export makes up about 2/3 of the Norwegian seafood export by value, amounting to a total of more than 105 billion NOK in 2022. This made salmon export the third largest exported commodity by value in Norway, below oil and gas. (Meisingset, 2023) (Sjømatråd, n.d.)

1.3 Fish farming

Most of the salmon that is exported comes from fish farms. Salmon farming is a Norwegian lead global industry with four of the 5 biggest companies being Norwegian. (Berge, 2020) These companies all rely heavily on the price of salmon. We hope that by creating predictive models on the price of salmon short term we will gain insight into the short term future of these major companies. This in turn could give us an idea of the impact they might have on the Norwegian economy and society as a whole in the short term.

1.4 Is this relevant from a business analytics standpoint?

This thesis is relevant in Business analytics as we will be creating a model using methods and logic form previous courses we have had. Its highly relevant in the economy today for multiple reasons. For one the Industry is a cornerstone in the Norwegian economy and has been touted as one of the ways for Norway to prosper after the oil is gone. The market and industry surrounding salmon has also been a subject of debates as of recently. This debate comes as a result of massive growth, profits and old tax-related incentives in the industry. This has caused a lot of turbulence in the price of the salmon farming companies. In this thesis we will not be tackling the issues of the salmon tax debates and its impact on the industry. But we might have to make adjustments in our data set because of it.

1.5 Hva ønsker vi å oppnå med thesisen?

Skrive litt om hva vi ønsker å finne ut av eller å finne ut? Hva som gjør det vanskelig å finne ut av? Noen statistiske faktorer med datasettet vårt som random sampling osv.

Kilde 2: https://www.worldwildlife.org/industries/sustainable-seafood Kilde 3: https://fromnorway.com/sea from-norway/clipfish/ Kilde 4: https://e24.no/norsk-oekonomi/i/xgOd7X/skyhoeye-gasspriser-ga-historisk-hoey-eksport-i-2022 (ikke direkte sitert i avsnittet, men) Kilde 5: https://nokkeltall.seafood.no (ikke direkte sitert i avsnittet, men) Kilde 6: https://ilaks.no/dette-er-verdens-20-storste-lakseoppdrettere-2/

Alle disse lagt inn under references

2 Theory and literature

Relevant theory will be: We envision using economic theories such as market equilibrium, Pareto efficiency and consumer choice theory. Statistical theory, such as the presumptions for regression to evaluate whether our data sources meet the requirements, and our model is robust. Neural networks, especially LSTM compared to other predictive models

2.1 ARIMA — SARIMAX

Formula ARIMA:

$$\hat{y'}_t = c + \phi_1 y'_{t-1} + \phi_2 y'_{t-2} + ... + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

The ARIMA-model is one of the more popular and useful approaches to time series forecasting. The name is an acronym that stands for AutoRegressive Integrated Moving Average and the model utilizes these in order to predict future values solely on earlier values, it is therefore an univariate model. The SARIMAX-model is an extension of the ARIMA-model that also takes external factors and seasonality into account in order to better predict future values. SARIMAX can therefore be a multivariate model. (Hyndman and Athanasopoulos, 2021)

2.1.1 Seasonality

2.1.2 Auto Regressive (AR)

The first part of the ARIMA acronym is the Auto Regressive part. Auto comes from the greek word autos and mean self, in this context it means that the model is regressing on itself. This part of the model can be written as follows:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t \tag{1}$$

Where c is a constant, p is the number of lag observations or autoregressive terms, ϕ are the AR coefficients and ε_t is the error term. y_t is the data on which the AR-model is applied on. (Oracle, 2023) This model is a “pure” AR-model and relies therefore solely on its own lags. If p is set to 1, the model looks at the previous value and tries to predict the next value. If p is set to 2, the model looks at the previous two values and tries to predict the next value, and so on. (Artley, 2022)

2.1.3 Integrated (I)

The second part of the ARIMA acronym is the Integrated part. This part of the model is used to make the time series stationary. In the ARIMA-equation it is represented by the letter d and is the number of differencing required to make the time series stationary. Usually the optimal amount of differencing is the least amount needed to make the data fluctuate around a well defined mean. (Nau, 2019)

2.1.4 Moving Average (MA)

$$\hat{y} = c + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$$

2.2 Exploratory analysis

Exploratory data analysis (EDA is a method used to analyse datasets and summarize the main characteristics. It helps determine how best to manipulate datasources to get the answers needed. This makes it easier to discover patterns, anomalies, test hypotheses or checking assumptions (https://www.youtube.com/watch?v=QiqZliDXCCg).

2.3 Regression

Regression analysis is a reliable method of identifying which variables have impact on a topic of interest. Regression analysis consists of two types of variables; dependent and independent. The dependent variable is the main variable or factor that aasdasd is trying to predict or understand. The independent variables are the factors that dasdasd hypothesize to have an impact on the chosen dependent variable.

Linear regression models often use a ”least-squares approach” to determine a line of best fit (regression line) to a given dataset. This line is the yellow line in the figure below. A square is the squared distance between a datapoint and the regression line. These values needs to be squared in order to not counteract each other.

BILDE

When the process above has been completed a regression model is constructed. The genereal form of a **multiple linear regression model** is:

- Regression, https://www.alchemer.com/resources/blog/regression-analysis/

https://www.investopedia.com/terms/r/regression.asp

2.4 Neural networks

2.4.1 Recurrent Neural Network

2.4.2 Long Short-Term Memory

3 Methodology

We intend to utilize the following methods:

- Exploratory analysis
- Regression, https://www.alchemer.com/resources/blog/regression-analysis/

- ARIMA, https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/
- LSTM — Tensorflow

3.1 Data gathering

In order to analyze data connected to salmon price, we first need to gather this data. The main data point is the price of salmon. There are several sources for this data, but we utilized the data from the NASDAQ salmon exchange. The reason for this being a combination of the accessibility of the data, and the fact that the NASDAQ salmon exchange (NQSALMON) uses a wighted average for the salmon price, gathered from a spectrum of salmon exporters and it is therefore the best source of meaningful data. Another reason for using the NASDAQ salmon exchange is that the data is updated weekly with no missing values for the entire time frame. We downloaded data from March 2013 through December 2022, for a total of 507 data points. This was our base for the independent factors.

The next step was to gather data from the other relevant factors for our analysis.

3.2 Exploratory analysis

3.3 ARIMA and SARIMAX

3.4 LSTM — Tensorflow

4 Results and discussion

5 Conclusion

Works Cited

Artley, B., 2022. *Time series forecasting with arima , sarima and sarimax*, May. Available from: https://towardsdatascience.com/time-series-forecasting-with-arima-sarima-and-sarimax-ee61099e78f6 [Accessed 22 March 2023].

Berge, A., 2020. *Dette er verdens 20 største lakseoppdrettere*, July. Available from: https://ilaks.no/dette-er-verdens-20-storste-lakseoppdrettere-2/ [Accessed 21 March 2023].

Hyndman, R.J. and Athanasopoulos, G., 2021. *Forecasting: principles and practice*. Melbourne: Otexts.

Johansen, U., Bull-Berg, H., Vik, L.H., Stokka, A.M., Richardsen, R. and Winther, U., 2019. The norwegian seafood industry – importance for the national economy. *Marine policy* [Online], 110, p.103561. Available from: https://doi.org/10.1016/j.marpol.2019.103561.

Meisingset, S., 2023. *Skyhøye gasspriser ga historisk høy eksport i 2022*, January. Available from: https://e24.no/norsk-oekonomi/i/xgOd7X/skyhoeye-gasspriser-ga-historisk-hoey-eksport-i-2022 [Accessed 21 March 2023].

Nau, R., 2019. *Introduction to arima: nonseasonal models*. Duke University. Available from: https://people.duke.edu/~rnau/411arim.htm [Accessed 22 March 2023].

Norway, S. from, 2023. *Clipfish*. Available from: https://fromnorway.com/seafood-from-norway/clipfish/ [Accessed 21 March 2023].

Oracle, 2023. *Oracle® crystal ball reference and examples guide*. Available from: https://docs.oracle.com/cd/E57185_01/CBREG/ch06s03s04s01.html [Accessed 22 March 2023].

Sjømatråd, N., n.d. *Nøkkeltall*. Available from: https://nokkeltall.seafood.no/ [Accessed 21 March 2023].

WWF, 2019. *Sustainable seafood — industries — wwf*. Available from: https://www.worldwildlife.org/industries/sustainable-seafood [Accessed 21 March 2023].