Norwegian University of Life Sciences
School of Economic and Business

Philosophiae Doctor (PhD)
2023:72

Essays in Salmon Commodity Markets

Essayer i Laksemarkeder

Mikaella Zitti

Essays in Salmon Markets

Essayer i Laksemarkeder

Philosophiae Doctor (PhD) Thesis

Mikaella Zitti

Norwegian University of Life Sciences
School of Economic and Business

Ås (2023)

Thesis number 2023:72
ISSN 1894-6402
ISBN 978-82-575-2103-5

Acknowledgments

Getting this PhD has been a wild ride and I couldn't have done it without the support of many amazing people. First, my parents, Christiana and Constantinos Zittis. You've always had my back and encouraged me to follow my dreams - I can't thank you enough for that. To my boyfriend, Clemens Knoppe, thank you for your unwavering support and for being there through the best and the toughest times. Big thank you to my brother, Demetris Zittis. Those endless phone calls and your unintentional inspiration have meant the world to me.

Gratitude is also extended to my friends who have been there for me during this time. Zeta Rizikianou, for the comfort and encouragement she always offered. Marilena Papadopoulou, my long-standing friend, for lending a patient ear. My fellow researcher and friend, Torgunn Aslaug Skjerve, for stimulating philosophical research discussions. The office companions who also became my friends, Wuddasie Dereje Bekele and Irene Namugenyi. My friends in Kiel, Tràn Lê Nga and Mariia Okuneva. Sarah Elpida Tione, for the support during my initial paper writing. My friends who kept faith in me, Ioanna Savva, Alexandra Katsarou, and Arnau Picón Martínez. And my other friends in Norway, Martin Hansen, Niels Oliver Nagel, Yi-kuang Chen, Ingvild Solhjem for all the fun memories.

Special thanks to my main supervisor, Atle G. Guttormsen, for granting me complete intellectual independence. My co-supervisor, Christian Oliver Ewald, who always extended a helping hand even when it was beyond his responsibilities. My appreciation also extends to the management and staff of the School of Economics and Business (HH). In particular, to Frode Alfnes for his honest guidance. And not to forget Kirsti Pettersen and Reidun Aasheim. Acknowledgement is also due to the Erasmus office, especially Tonje Nore for facilitating my exchange semester at the University of Kiel.

Lastly, I want to acknowledge my co-authors, Mariia Okuneva and Clemens Knoppe, for a memorable research experience and making the final part of this journey enjoyable. We indeed had our fair share of long hours in the office, debating sentiment and debugging codes.

I want to wrap this up by saying that getting this PhD has been an amazing experience. The people I've met and the things I've learned have truly shaped me. I'm ready to take on the next chapter knowing that I have the best support network anyone could ask for.

Kiel, June 2023

Mkaella Zitti

Contents

List of Research Articles

This thesis is based on the following research articles:

1. Zitti, M., & Guttormsen, A. G. (2022). Climate risk and financial disclosure in salmon aquaculture. *Aquaculture Economics & Management*. https://doi.org/10.1080/13657305.2022.2143934

2. Zitti, M. (2023). *Forecasting Salmon Market Volatility using Long Short-term Memory (LSTM)*. Working Paper. Ås, Norway.

3. Zitti, M. (2023). *The Influence of Market Adaptability on Forecasting Salmon Spot Prices: A Comparison of Hybrid Deep Learning and Traditional Models*. Working Paper. Ås, Norway

4. Knoppe, C., Okuneva, M. & Zitti, M. (2023). *Salmon price movements around salmon market news*. Working Paper. Kiel, Germany.

Abstract

This thesis examines key aspects of the salmon industry, including financial effects of climate change, market volatility, price predictions, and the role of news on market prices. The work consists of four research articles that offer practical insights to various stakeholders in the salmon market, thus adding to the growing body of literature in the field and enhancing the understanding of the salmon industry.

The first article underlines the importance of climate-related financial disclosures in the salmon industry. It showcases a trend of increased transparency, encouraged by organizations such as the TCFD and CDP, leading to better practices in addressing climate-related risks. This shift benefits not only the companies themselves but also investors and policymakers, enabling sustainable decision-making.

The second study investigates the prediction of salmon market volatility using a deep learning technique known as LSTM. The findings reveal that the LSTM model did not outperform the benchmark ARMA model in forecasting accuracy, suggesting that the salmon market volatility may not exhibit complex temporal patterns that can be effectively captured by LSTM.

In the third article, the effectiveness of a hybrid VAR-LSTM model is tested against a traditional VAR model for predicting salmon spot prices. The study suggests that the hybrid model does not significantly improve salmon price forecasting, suggesting that salmon price series do not exhibit (exploitable) non-linear patterns. This result implies the efficiency of the salmon market, where new information is swiftly identified and incorporated by investors.

The final study investigates how news affects the stock prices of major salmon com-

panies. This paper brings attention to the need for industry-specific sentiment analysis tools ("lexicons"). In addition, it addresses the competition within the global salmon market and discusses factors contributing to market volatility, with Covid-19 identified as a major influence.

Overall, this research contributes to the existing knowledge by enhancing the understanding of climate-related disclosures, improving prediction models for market volatility and prices, and offering insights into the effects of news on the market. These findings have significant implications for the salmon industry and could assist policymakers, companies, and investors in making informed decisions, promoting the industry's resilience and sustainability. Furthermore, these findings provide insights into potential future trends and price fluctuations in the salmon industry, which are crucial for efficient market operation and strategic planning.

Norsk sammendrag

Denne avhandlingen undersøker nøkkelaspekter ved lakseindustrien, inkludert de finansielle effektene av klimaendringer, markedsvolatilitet, prisprognoser og nyheters rolle på markedsprisene. Arbeidet består av fire forskningsartikler som tilbyr praktiske innsikter til ulike interessenter i laksemarkedet.

Den første artikkelen understreker betydningen av klimarelaterte finansielle avsløringer i lakseindustrien. Den viser en trend med økt transparens, oppmuntret av organisasjoner som TCFD og CDP, noe som fører til bedre praksis i håndtering av klimarelaterte risikoer. Denne endringen gagner ikke bare selskapene selv, men også investorer og beslutningstakere, ved å muliggjøre bærekraftig beslutningstaking.

Den andre studien fokuserer på å forutsi laksemarkedets volatilitet ved hjelp av tradisjonelle tidsseriemodeller og en dyp læringsmetode kalt LSTM. Funnene avslører at LSTM presterer bedre, spesielt for langsiktig prognose, noe som gjør det mulig for markedsdeltakerne å ta informerte beslutninger om risikostyring og produksjonsplanlegging.

I den tredje artikkelen testes effektiviteten av en hybrid VAR-LSTM-modell mot en tradisjonell VAR-modell for å forutsi lakse spotpriser. Studien antyder at den hybride modellen ikke forbedrer laks prisprognosen betydelig, noe som antyder at laksepriser ikke viser ikke-lineære trender. Dette resultatet antyder effektiviteten i laksemarkedet, der prisendringer raskt blir identifisert og innarbeidet av investorer.

Den endelige studien undersøker hvordan nyheter påvirker aksjekursene til de største lakseselskapene. Denne artikkelen retter oppmerksomheten mot behovet for bransjespesifikke verktøy for sentimentanalyse ("leksikoner"). I tillegg tar den opp konkurransen i det globale laksemarkedet og diskuterer faktorer som bidrar til markedsvolatilitet, med

Covid-19 identifisert som en hovedpåvirker.

Denne forskningen bidrar til eksisterende kunnskap ved å forbedre forståelsen av klimarelaterte avsløringer, forbedre prognosemodeller for markeds volatilitet og priser, og tilby innsikt i effekten av nyheter på markedet. Disse funnene kan hjelpe politikere, selskaper og investorer med å ta informerte beslutninger og bidra til bransjens motstandsdyktighet og bærekraft. Videre kan resultatene gi veiledning til markedsdeltakere i håndtering av potensielle fremtidige trender og prisfluktuasjoner, og fremme mer effektive praksis innen lakseoppdrettsindustrien.

Synopsis

5.1 Introduction

The purpose of this thesis is to provide a comprehensive analysis of the critical features that constitute the salmon markets. These characteristics are of significant interest to a wide range of market participants, including farmers, processors, and investors. In today's era, where sustainable food production is a major concern, investors are increasingly aware of the economic implications of climate change for salmon farming companies. Therefore, their demand for climate-related financial disclosure has surged. Additionally, salmon is a volatile commodity, and the increasing demand for salmon over the past few decades has led to high salmon prices and subsequently heightened volatility, particularly since the mid-2000s. This volatility poses a significant concern for many market participants, such as farmers, processors, traders, and hedgers, who must prepare for future uncertainty. Therefore, forecasting salmon market volatility is crucial for reducing future uncertainty for all salmon market participants. Furthermore, predicting salmon spot prices using a variety of factors that provide predictive information on the future development of the salmon spot price is essential for salmon market participants. Finally, news related to the salmon market can influence trading behavior in the market. Studying the drivers of the salmon market and investor sentiment using news articles enables policymakers to gain a better understanding of what drives the market.

This thesis comprises two co-authored papers: "Climate risk and financial disclosure in salmon aquaculture" and "Salmon price movements and trading behaviors around salmon market news", and two single-authored papers: "Forecasting Salmon Market Volatility using Long Short-term Memory (LSTM)" and "Multi-step ahead forecasting of salmon spot prices using a hybrid deep learning model". These research articles offer a comprehensive understanding of the salmon market features by using qualitative information to understand investors' sentiment around potential financial impacts of climate change on salmon farming producing companies, forecasting salmon market volatility and prices using traditional time-series models and deep learning techniques, and examining the factors driving salmon markets as well as the investors' sentiment using textual analysis. These articles will provide a better understanding of the salmon market factors for all market participants as well as policy makers and regulators. Furthermore, they will encourage aquaculture economists to explore the potential of machine and deep learning techniques within the field of financial economics for salmon markets.

The remainder of the introductory section presents the background of the research, detailed problems encountered during the research construction, the data used, the methods used in the research articles, a summary of the research articles, and the contributions and limitations of the research.

5.2 Overview of the salmon farming industry

The farming of Atlantic salmon had its roots in experimental ventures during the 1960s, which eventually evolved into an established industry in Norway during the 1980s (Mowi, 2020). In the subsequent decade, Atlantic salmon farming underwent a process of commercialization (Y. Liu et al., 2011), ultimately becoming a commercially viable industry during the 1980s (Asche & Bjorndal, 2011). The industry grew rapidly and continuously, such that by the late 1990s, farmed salmon surpassed wild capture in terms of production volume. Figure 5.1 shows that since 2012, global production of farmed salmon has exceeded two million metric tonnes. Despite stagnation in export volume of other species, the volume of farmed salmon has continued to increase, such that by the end of 2021, salmon represented approximately 40% of the quantity of all seafood species exported from Norway and 60% of the value (see Figure 5.1).

The initial introduction of farmed salmon into commercial markets resulted in con-

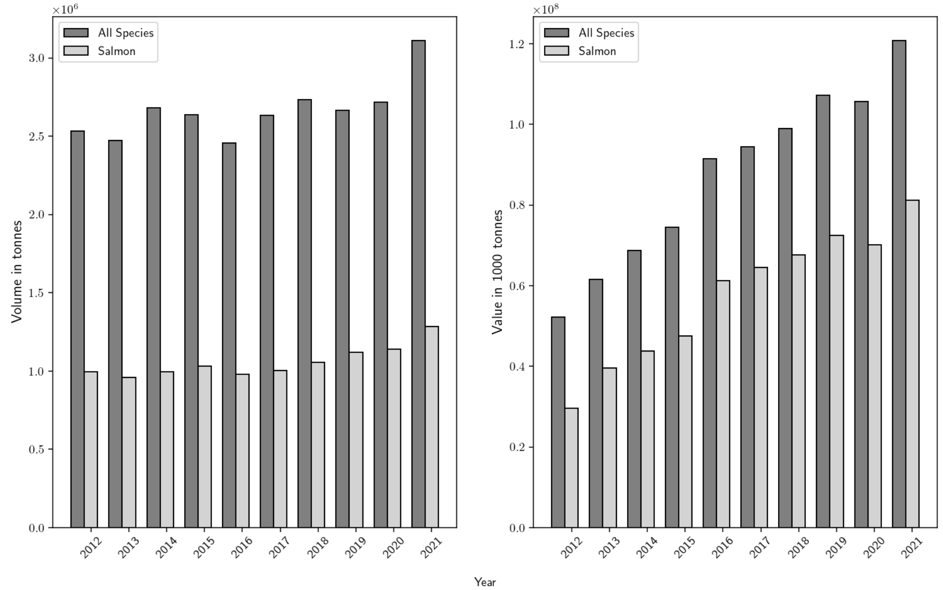


Figure 5.1: Comparison of the Volume in Metric Tonnes and Value in 1000 Tonnes of salmon and all species exported from Norway from 2012 to 2021. Source: Norwegian Seafood Export Council (2022)

siderably higher prices than wild salmon. Nevertheless, technological advancements have facilitated a gradual reduction in the price of farmed salmon from the 1980s through the early 2000s. However, since then, prices have been rising slowly due to a variety of factors such as limited availability of production sites (Hvas et al., 2021), slower productivity growth (Oglend, 2013; Vassdal & Sørensen Holst, 2011), stricter industry regulation (Asche & Bjorndal, 2011), increasing feed costs (Iversen et al., 2020), and a surge in demand (Brækkan et al., 2018). Figure 5.2 provides an overview of annual Norwegian farmed salmon export prices and production costs from 2008 until 2020. It is evident that the export price of salmon in 2020 has more than doubled compared to the price in 2008. Although production costs are also increasing, they remain significantly below the levels of the export prices. This trend supports the rapid growth of the industry in recent years.

The Norwegian aquaculture industry has become one of the most significant export industries of Norway, second only to oil and gas, since its inception. The salmon farming sector is a significant constituent of the Norwegian aquaculture industry (Asche, 2008),

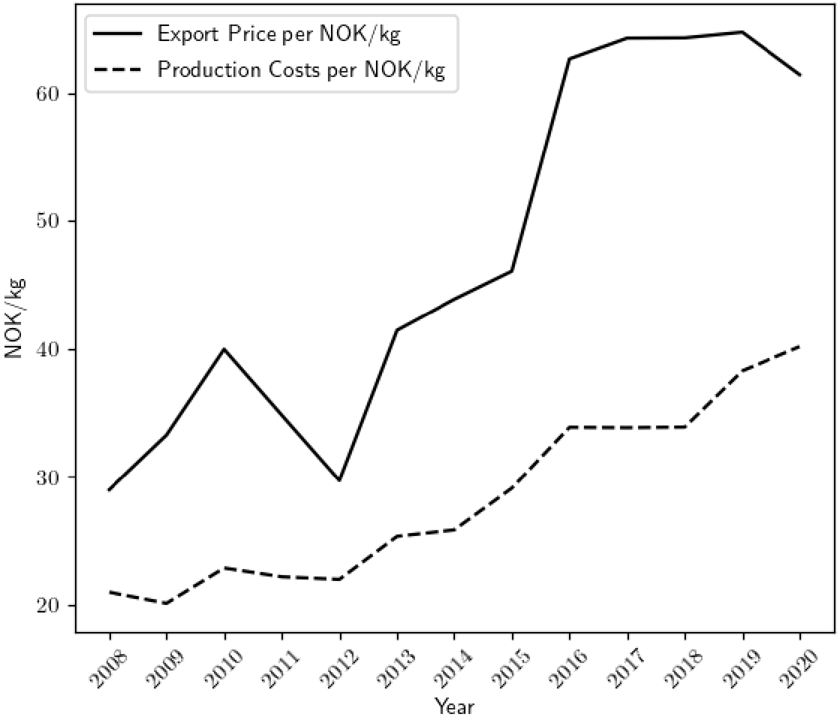


Figure 5.2: Salmon Export Price and Production Costs (including delivery costs) measured in NOK/kg. Source: Norwegian Directorate of Fisheries (2022)

leading Norway to become the world's largest salmon producer, responsible for over 50% of the total production (Hersoug, 2021). While other countries such as Chile, Scotland, and Canada have also been involved in salmon farming to a lesser extent, the natural conditions have placed severe limitations on the geographical distribution of salmon producers worldwide, due to factors such as sea lice and disease problems. As depicted in Figure 5.3, the global production volume distribution of salmon as of 2019 highlights this constrained distribution. With the growing demand for salmon, Bjørndal and Tusvik (2019) suggest that technological advancements such as land-based aquaculture could play a significant role in shaping the dynamics of the salmon market.

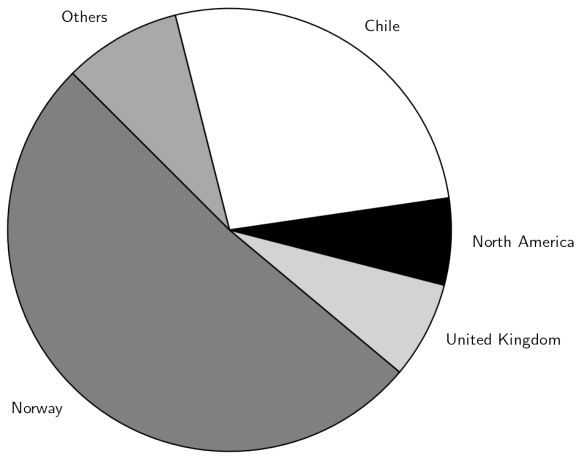


Figure 5.3: Distribution of farmed Atlantic salmon production volume across countries, 2019. Source: Marine Harvest (2020)

5.2.1 Supply and Demand for salmon

The supply of farmed salmon is subject to various factors with varying time horizons. Long-term factors, such as the availability of production sites and industry regulations, as well as medium- and short-term factors, such as the amount of young fish put into seawater for growing, the pace of growth and mortality, and the farmer's slaughtering schedule, all play a crucial role in determining the supply of farmed salmon. Fish growth is influenced by numerous factors, including seawater temperature, feeding intensity and frequency, and feed properties. Although the seawater temperature is beyond the farmer's control, the farmer can regulate the speed of growth by adjusting feeding practices. Disease precautions, such as vaccination and treatments against parasites, are also applied at different stages of production to manage production risks, which may include outbreaks of infectious diseases, unexpected water temperature changes, storms, and damage to farming facilities that may lead to fish escapes.

Price is a key determinant of salmon supply, and farmers aim to harvest when it is most profitable for them. Specifically, when the salmon spot price is high, the farmers tend to be more eager to harvest to make a profit. However, considering that the salmon market is highly volatile (Bloznelis, 2016; Guttormsen, 1999; Oglend, 2013; Oglend & Sikveland, 2008), it is difficult for farmers to plan the production process early enough to avoid causing price fluctuations (Asche, 2008). Furthermore, the salmon spot price is not the only price that affects supply, as the expected future price can also impact supply, given that farmers are interested in how the salmon price will look in the future. Several studies have analyzed the effects of price on salmon supply, with Landazuri-Tveteraas et al. (2018) finding evidence of price causality from export to retail level, and Asche et al. (2014) indicating that the supply chain for fresh salmon fillets can be characterized by a high degree of price transmission. However, Asche, Cojocaru, et al. (2018) showed that the salmon supply chain remains semi-automated, with less control over production processes.

Numerous factors also affect the demand for salmon, including the current price and lagged price effects, the price of substitutes such as wild salmon and trout, consumer income, and regulations. For instance, partial or complete trade restrictions for Norwegian salmon have been introduced and subsequently revoked repeatedly over the years. Media coverage may also affect demand by altering consumer tastes, with recurrent attacks on pollution caused by the industry's practices and spread of parasites from salmon farms (P. Liu et al., 2016), as well as fish escapes from farms into the wilderness increasing wild salmon mortality (Jonsson & Jonsson, 2006). On the other hand, salmon has long been recognized as a healthy food by dieticians, which boosts demand. Many studies have analyzed the global and country-level demand for salmon, including Asche (1996), Asche et al. (1998), Xie et al. (2009), Xie and Myrland (2011), and Asche et al. (2019).

5.2.2 Salmon Markets

Numerous stakeholders within the salmon market rely on the evolution of salmon prices to inform their future operations. However, due to the inherent volatility of salmon as a commodity (Guttormsen, 1999), it presents challenges for market participants, including farmers, processors, traders, and hedgers, to effectively strategize. Specifically, farmers, who are responsible for salmon production, seek to maximize profits by harvesting at optimal times (Asche, 2008). High salmon spot prices often incentivize farmers to increase harvesting activities, thus generating profits. Nonetheless, given the pronounced volatility of the salmon market (Bloznelis, 2016; Guttormsen, 1999; Oglend, 2013; Oglend & Straume, 2019), formulating long-term production plans to mitigate price fluctuations proves arduous.

Harvesting decisions by farmers can influence market volatility, subsequently affecting other market participants' operations. For instance, processors' operational plans hinge upon anticipated input costs, dictated by salmon spot prices. Lacking control over production, processors remain entirely reliant on the trajectory of salmon spot prices. Furthermore, traders strive to minimize salmon acquisition costs while preserving profitability, though heightened market volatility can jeopardize their operations. Other participants, such as hedgers, similarly depend on salmon price forecasts. Consequently, the elevated volatility of the salmon market exacerbates uncertainty surrounding future price trends.

In response to these concerns, Fish Pool, a futures exchange dedicated to salmon, was founded in 2006 in Bergen, Norway, with the aim of offering hedging opportunities within the salmon market. Futures contracts are available for each month, extending up to 60 months into the future. It is important to note that contracts with maturities exceeding twelve months are rarely traded; in fact, even contracts of similarly long maturity exhibit low trading volumes. It is plausible that the market's low liquidity is partially attributable to the wide array of available maturities, resulting in a lack of concentration that deters potential buyers and sellers. Contracts are settled against a monthly benchmark index known as the Fish Pool Index. This monthly index is calculated as an equally-weighted average of the relevant weekly index values, with each trading month consisting of either four or five complete weeks, as specified in the trading rules (Fish Pool, 2017a). The weekly index is derived from several price indices that should represent the spot price of 3-6 kg salmon. While the weights of the various component indices have historically fluctuated, the dominant component has consistently been the spot price by NOS, later substituted by NASDAQ. The current composition of the weekly index is available on the Fish Pool website (Fish Pool, 2017b).

Fish Pool's inception has sparked significant debate among aquaculture economists. Since its establishment, Fish Pool has enabled trading in financial derivatives tied to the price of Norwegian farmed Atlantic salmon. However, many aquaculture economists have questioned Fish Pool's effectiveness (Bloznelis, 2018b; Oglend, 2013), with research showing that salmon price volatility has doubled since its introduction and remains persistently high. Several studies have investigated the efficiency and role of the salmon futures market (Ankamah-Yeboah et al., 2017; Ewald et al., 2017; Larsen & Asche, 2011; Misund & Asche, 2016; Solibakke, 2012), particularly in relation to the spot market (Asche et al., 2016a, 2016b; Chen & Scholtens, 2019). There have been concerns raised about the market's low liquidity and infrequent trading (Andersen & de Lange, 2021; Bloznelis, 2018a; Dahl et al., 2021; Ewald et al., 2022). Nonetheless, a study suggested that the majority of salmon price fluctuations share common factors, thus justifying the presence of a salmon price index (Oglend & Straume, 2019). Consequently, while Fish Pool offers futures contracts with the intention of providing hedging opportunities for market participants, the persisting concerns about market efficiency, liquidity, and price volatility suggest that the benefits and effectiveness of these contracts in managing risk exposure remain a matter of debate.

5.2.3 Sustainability of the Salmon Industry

Climate change poses substantial fiscal challenges for the aquaculture sector, particularly salmon cultivation, due to its unpredictable nature and the potential for significant economic repercussions (Asche, 2008; Asche et al., 2022; Asche et al., 2017; Misund, 2017). While some research has explored price fluctuations resulting from climate change (Asche et al., 2019; Asche et al., 2017), inquiries into the financial consequences of climate-related hazards, categorized as physical or transition risks (TCFD, 2017), are sparse.

Climate-related financial risks are generally divided into physical or transition risks (TCFD, 2017). Physical risks encompass threats associated with economic damages originating from the direct impacts of climate change (Bovari et al., 2018; Dafermos et al., 2017, 2018; Dietz et al., 2016; TCFD, 2017). Conversely, transition risks relate to the challenges companies encounter due to market disruptions and policy implications stemming from the shift towards a low-carbon economy (Battiston et al., 2017; Dafermos et al., 2018; Leaton, 2011; Stolbova et al., 2018; TCFD, 2017).

Examples of physical risks include the following. Salmon farming, conducted in open cages, faces climate-related hazards such as widespread disease outbreaks, feed waste pollution, and algal blooms (Abolofia et al., 2017; Asche, 2008; Asche et al., 2021; Asche et al., 1999; Asche, Sikveland, et al., 2018; Fischer et al., 2017; Torrissen et al., 2013). Climate change may intensify these risks through consequences such as elevated water temperatures, increased precipitation, rising sea levels, and extreme weather events (De Silva & Soto, 2009), leading to financial losses (Olaussen, 2018; Pincinato et al., 2021).

Persistent physical risks demand climate adaptation from industry firms (De Silva & Soto, 2009). Temperature-dependent dynamics of salmon lice and the danger of algal blooms can lead to disease outbreaks and augmented production costs (Abolofia et al., 2017; Asche et al., 1999; Jansen et al., 2012; Torrissen et al., 2013). As a result, salmon aquaculture must address these climate-related hazards to ensure financial sustainability.

Transition risks can influence the industry through elevated carbon taxes, stringent regulations, and challenges in feeding practices. High carbon taxes can affect air cargo, exacerbate price volatility, and ultimately impact profitability (Asche et al., 2019). Regulatory measures can shape the industry's growth and price fluctuations (Asche et al., 2017; Hersoug et al., 2019). Furthermore, soy-based feeding practices raise environmental concerns, such as deforestation in the Brazilian Amazon (Dou et al., 2018; Sun et al., 2018). Therefore, transitioning to a low-carbon economy presents multiple challenges for the salmon aquaculture industry, and it is essential for companies to address potential climate-related financial impacts to maintain their profitability and reputation (TCFD, 2017).

5.3 Research Objectives and Contributions

Examining the volatility of salmon prices, the impact of salmon-related news on price trends, and the potential financial ramifications of climate change on the salmon farming industry are critical research domains within the field of salmon production and market. Precise understanding of these subjects is vital for various market participants, including farmers, processors, traders, hedgers, investors, as well as policy makers and regulators.

In particular, anticipating future salmon spot prices is crucial for strategic planning among stakeholders in the salmon market, such as farmers, processors, traders, and hedgers (Bloznelis, 2018b; Oglend & Sikveland, 2008). Spot prices exhibit greater volatility and are more challenging to predict than production or consumption volume, thereby contributing substantially to uncertainty in future revenues and costs for market participants. Consequently, developing a more comprehensive understanding of price volatility and enhancing forecasting models for future salmon spot prices is imperative.

Prior studies have utilized standard econometric models to forecast salmon price volatility (Bloznelis, 2016; Oglend, 2013; Oglend & Sikveland, 2008). However, to the best of my knowledge, no study has yet attempted to employ deep learning techniques, such as neural networks, to estimate and forecast salmon price volatility. Research from various fields has applied machine learning and deep learning techniques for forecasting volatility and prices of other commodities as well as stocks, yielding promising results (Kim & Won, 2018; Parot et al., 2019; Ramyar & Kianfar, 2019). Therefore, the second paper of this thesis aims to forecast salmon market volatility using a Recurrent Neural Network (RNN) variant, the Long Short-term Memory (LSTM).

Similarly, Zitti (2023a) investigates the forecasting performance of a hybrid architecture model that integrates a conventional multivariate forecasting model, the Vector Autoregressive (VAR), with the LSTM deep learning technique. This approach seeks to address the limitations of traditional statistical methods, which are unable to detect complexities and non-linearity in financial data series. Employing a hybrid architecture technique can be advantageous when forecasting over longer horizons, as deep learning techniques like LSTM can identify long-term patterns and handle long minimal time lags (Hochreiter & Schmidhuber, 1997).

Nonetheless, forecasting salmon prices and volatility is not the sole concern of salmon industry participants, such as investors and farmers, as well as policymakers. Examining climate-related risks specific to the salmon industry and exploring the extent to which some publicly listed firms have implemented climate-related financial disclosures is of significant importance to market participants and regulators. Initiatives like the Task Force for Financial Disclosures (TCFD) and the Carbon Disclosure Project (CDP) aid companies in comprehending the external effects of their unsustainable practices, thereby maintaining transparency for their investors. The growing demand for climate-related financial disclosures and the industry firms' responses are discussed in Zitti and Guttormsen (2022).

The responses of salmon markets to future price expectations and climate-related expectations (at the firm level) do not solely determine salmon market behavior. While a substantial portion of the salmon market literature concentrates on analyzing market volatility (Dahl & Oglend, 2014; Dahl & Yahya, 2019; Oglend, 2013; Oglend & Sikveland, 2008), the fourth paper of this thesis expands upon this body of work by investigating which types of news influence the market and the mechanisms underlying these effects (Knoppe et al., 2023). Given that investor behavior is shaped by news surrounding their invested securities, the paper examines news articles pertinent to salmon production and their impacts on stock prices. This study employs text mining techniques in conjunction with the Vector Autoregressive (VAR) model to explore the relationship between investors' trading behavior and salmon market-related financial news. The focus on stock prices is informed by the findings of Dahl et al. (2021), which suggest that the salmon stock market exhibits greater reactivity than commodity prices.

The primary objectives of this thesis are to underscore the significance of understanding salmon price volatility, the influence of salmon-related news on price trends, and the potential financial implications of climate change on the salmon farming industry. The goals of the studies discussed entail the development of enhanced forecasting models for future salmon spot prices, the application of deep learning techniques such as LSTM for market volatility prediction, the examination of climate-related risks and financial disclosures within the industry, and the exploration of the relationship between investors' trading behavior and salmon market-related financial news, employing text mining techniques.

5.4 Data

5.4.1 Data Sources

Data sources relevant to analyzing climate-related financial disclosures, modeling salmon prices and their volatility, as well as examining the impact of salmon-related news, are plentiful and generally of high quality. Table 5.1 displays the data series and their respective sources. The columns "Range" and "Frequency" indicate the specific ranges and frequencies employed in the four research papers, although they do not necessarily encompass all available ranges and frequencies at the sources. The majority of the datasets utilized in this thesis are publicly and freely accessible, with a few exceptions (refer to Table 5.1). The papers focused on predicting volatility and prices primarily use weekly data, and most of the data obtained from these sources are in weekly frequency. The text data employed in the fourth paper of this thesis is daily, as stock prices are available on a daily basis and the two datasets are made consistent. The data procured for evaluating the responsiveness of salmon firms to their climate-related risks is obtained annually, as these reports are intended for annual publication.

5.4.2 Data Limitations

The collected data sets offer crucial insights into the salmon market, but it is essential to acknowledge certain data limitations in some research papers. Specifically, in Zitti (2023a), both NOS and NASDAQ Salmon Index data were utilized, yet there are incon-

Table 5.1: Data sources

Object	Range	Frequency	Source	Access
Status categories & Scores	2016 - 2021	Annual	CDP	https://www.cdp.net/en
Salmon spot price (NOS price)	2007W27: 2013W13	Weekly	NOS	Upon request from: https://salmonprice.nasdaqomxtrader.com
Salmon spot price (NASDAQ Salmon Index)[a]	2013W14: 2020W53	Weekly	NASDAQ	Upon request from: https://salmonprice.nasdaqomxtrader.com
Salmon export volume (SSB volume)	2013W14: 2020W53	Weekly	Statistics Norway (SSB)	https://www.ssb.no/en/statbank
Salmon futures prices	2013-01-01 – 2020-12-31	Daily	Fish Pool	https://www.fishpool.eu
Soybean prices (Chicago Board of Trade (CBoT))	2013W14: 2020W53	Weekly	Refinitiv	Private Access
EUR/NOK exchange rate	2013W14: 2020W53	Weekly	Norges Bank	https://www.norges-bank.no/en/topics/
Shares Prices[b]	2013W1: 2022W53	Daily/ Weekly	Refinitiv	Private Access
News Articles	2012-11-28 – 2022-07-08	Daily	IntraFish	Private Access from: https://www.intrafish.com/

Note: "W" denotes week, e.g. 2013W14 denotes 2013 week 14; The range for the frequencies are as used in the four research papers; wider ranges and/or extra frequencies may also be available at the sources.
[a]The range is up to 2020W53 as the same data series was used for the second and third research article.
[b]Shares Prices are utilised for constructing the Share Price Index (SPI) used in research articles 3 and 4 where the frequency of the data series is daily and weekly, respectively.

sistencies in their definitions. NOS represents the prices paid by exporters to salmon farmers, while the NASDAQ displays prices received by salmon exporters from international buyers, leading to a variation in the exporter's margin. Moreover, the NASDAQ survey's inclusion criteria are more flexible compared to the NOS survey, which could cause NASDAQ to encompass a larger spot market share than NOS—although this may not always be true in practice. Nonetheless, in Zitti (2023a), a structural break-point was identified from 2012 week 18 to 2019 week 51, and the study's sample begins from 2013W1, based on similar findings from Bloznelis (2016). Consequently, the discrepancy between NOS and NASDAQ spot prices was not an issue.

Additionally, in Zitti (2023a), a univariate-based forecasting study was conducted, favoring the use of spot prices over futures prices to predict volatility. This preference is due to the salmon futures market's low liquidity, which trades less than 10% of the physical market volume (Fish Pool, 2020), and infrequent trades. Although this method may constrain the capture of forward-looking information, the salmon futures market remains thin due to the absence of speculative traders and sporadic trades. Consequently, salmon spot prices were chosen over futures contract prices, under the assumption that their lagged transformations carry more predictive information than salmon forward prices.

In Zitti (2023b), three distinct models were employed, with each containing varying predictor variables. In this multivariate-based approach, multiple data sets were incorporated. Similar to Zitti (2023a), the salmon spot price series served as the predictive variable, while export volume, futures contract prices, soybean prices, EUR/NOK exchange rates, and share prices acted as predictor variables. Given the multivariate predictive study's structure, futures contract prices were included as a predictive variable since they carry significant information regarding the future development of salmon spot prices. However, the study's aim was to forecast weekly salmon spot prices for various forecasting horizons, with each horizon using a distinct model format. For example, if the forecasting horizon is 26 weeks (approximately 6 months), the futures contracts' explanatory variable is represented by contracts maturing 6 months after their issue date. This method's drawback is that the futures contract price reflects the anticipated spot price for an entire month rather than a week, potentially decreasing forecast accuracy for predicting weekly salmon spot prices due to the approximated number of weeks within a month, half year, and one year. Despite this limitation, futures contract price series were still utilized as predictor variables, as they offer valuable information for the future development of the spot price series.

In Knoppe et al. (2023), the text data is confined to news articles published on IntraFish between 2016 and 2022, containing specific keywords related to salmon, prices, and finance. This constraint may exclude crucial information about the salmon market not covered by IntraFish or not encompassed by the chosen keywords. Furthermore, eliminating non-relevant information, such as journalist names and quantitative data, can decrease the information available for analysis. Another limitation when using text data is that the data pre-processing methods are restricted. Despite cleaning the data to the best of our ability, some noise might still persist, potentially impacting topic modeling. However, these limitations can be easily addressed when identified by the authors.

While data limitations are a significant concern for most articles presented in this thesis, for Zitti and Guttormsen (2022), the data collection does not hinder the article's results, as the hypothesis question can be addressed even with a limited data set. Nonetheless, since the article relies on publicly available data on companies' climate-related financial reporting habits for various reports, it is essential to consider that the available data for some reports may be insufficient throughout the years. If more data were available, the research conclusions would have been more robust.

Lastly, in all the research articles, the data set is restricted to a specific time frame. Although the chosen time period is suitable for the research questions raised in each paper, the results may not be generalizable to other time periods. Moreover, the data sources may have been altered or updated since the data collection period ended, which could affect the relevance of the results for current market conditions and the implementation of new policies and regulations.

5.5 Methodology

In this section, I will discuss some of the key methodologies employed throughout the technical articles of this thesis. This discussion will not include any information about Zitti and Guttormsen (2022), as this article is a qualitative study that does not incorporate econometric or statistical analysis. I will begin by addressing some of the data preprocessing methodologies applied in Zitti (2023a), Zitti (2023b), and Knoppe et al. (2023). Following that, I will outline the models that have been utilized to address the research questions of each of these research article.

5.5.1 Data pre-processing requirements

Seasonal Adjustment

Seasonal adjustments were implemented in research articles Zitti (2023a) and Zitti (2023b) due to the seasonality characteristic of salmon production caused by supply and demand factors. Seasonality in supply and demand creates patterns in salmon price and production volume. Modelling seasonality in weekly time series is complex, so these articles employ a technique introduced by Hyndman and Athanasopoulos (2018) and applied by Bloznelis (2018b) that uses regression with autoregressive moving average (ARMA) errors, incorporating Fourier terms as regressors.

The regression with ARMA error is given by the following equations:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \cdots + \beta_K x_{K,t} + u_t \tag{5.1}$$

$$u_t = \phi_1 u_{t-1} + \cdots + \phi_p u_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \tag{5.2}$$

where, y represents the dependent variable, x_1 through x_K are independent variables, and ϵ is an independent and identically distributed error. When implementing the model in the articles, $x_{i,t}$, where $i = 1, \ldots, K$, takes the form of Fourier terms and dummy variables that represent seasonal effects. The number of Fourier terms, up to 26 pairs for weekly data, was selected by minimizing the Akaike information criterion (AIC), as was the order of the ARIMA model (Hyndman & Athanasopoulos, 2018).

In the salmon spot price and export volume series, four dummy variables were introduced for the weeks before and after Christmas (including Christmas week) and four more for the same weeks around Easter (including Easter week). For the soybean series, 17 dummy variables were constructed to specify weeks from May to August. The seasonal component, composed of the Fourier terms and the seasonal dummy variables, is subsequently subtracted from the original series to generate the seasonally-adjusted version. Seasonal dummy variables are not employed for adjusting futures price variables, as the underlying of a future contract is the average FPI recorded over a month, not a week.

Structural Changes

In Zitti (2023a), the analysis starts from 2007 week 27, as indicated by Bloznelis (2016). However, a change in the variability of the series was observed from mid-2012 onwards. To test for a potential structural breakpoint, the sample was split into two periods: 2007 week 27 to 2012 week 17 and 2012 week 18 to 2019 week 51. An $F-$test was used to examine whether the two sub-samples have equal variance. After testing for normal distribution using the Shapiro-Wilk normality test, it was confirmed that the two sub-samples met the required assumptions. If the samples had not been confirmed to be normally distributed, employing an $F-$test would not have been appropriate. The results strongly reject equal variances with p-values well below 0.01. However, since the focus is on forecasting using an in-sample (train sample) and a hold-out sample (test sample), forecasting with two sub-samples before and after the breakpoint is not feasible. Thus, the sample from 2012 week 18 to 2019 week 51 is used for forecasting.

Text Data Pre-processing

In Knoppe et al. (2023), conventional cleaning of the text data is performed to remove HTML tags, images, videos, tables, graphs, tweets, white spaces, and regular text patterns (e.g., “Click here for. . . ”). Quantitative information and articles with 20 or fewer words are also removed to reduce noise, and all articles’ publication times are converted to

Greenwich Mean Time (GMT). The final news data format includes title, post time, and content.

To align with daily log returns data, articles are sorted according to their impact on stock returns. Articles published after 2:20 pm GMT on weekdays or during weekends and public holidays are considered to impact the stock market on the next trading day.

After cleaning and organizing the textual data, text pre-processing steps are performed to prepare the dataset for further analysis. This includes feature extraction using collocations, converting uppercase letters to lowercase, splitting contractions, tokenization, removing non-alphabetic characters and stop words, removing journalist names, and stemming tokens using the Porter Stemmer for English. To reduce dimensionality, the Term Frequency-Inverse Document Frequency (TF-IDF) method is used, as shown in the following formula for each token v:

$$log\left(1 + N_v\right) \times log\left(\frac{D}{D_v}\right) \tag{5.3}$$

where N_v is the token index, D_v is the number of documents containing term v, and D is the total number of documents. Tokens with the lowest TF-IDF scores are discarded before proceeding with the topic modeling approach.

5.5.2 Dimensionality Reduction

Least Absolute Shrinkage and Selection Operator (LASSO) regression

A variable selection method such as Least Absolute Shrinkage and Selection Operator (LASSO) regression is essential to make forecasts more accurate and reduce dimensionality. LASSO is a shrinkage and variable selection method for linear regression models, which minimizes the usual sum of squared errors with a bound on the sum of the absolute values of the coefficients.

The objective function of LASSO can be formally defined as:

$$\min_{\beta}\left\{\frac{1}{2n}\|y - X\beta\|_2^2 + \lambda\|\beta\|_1\right\} \tag{5.4}$$

where the term $|y - X\beta|_2^2$ represents the sum of squared residuals, reflecting how well the model fits the data. On the other hand, $|\beta|_1$ represents the absolute value of the coefficients, and $\lambda \geq 0$ is a tuning parameter that controls the level of shrinkage. In simple terms, LASSO tries to find a balance between fitting the data well and keeping the coefficients as small as possible, with the exact balance determined by λ. The ability of the LASSO to shrink some of the coefficients to zero, allows it to perform a variable selection.

In Zitti (2023b), the LASSO method is employed to ensure that the VAR model only include the most relevant predictors, reducing dimensionality, improving interpretability, and potentially enhancing the accuracy of our forecasts.

Principal Components Analysis

Principal Components Analysis (PCA) is a widely used statistical method for dimensionality reduction, which aims to transform a set of correlated variables into a smaller set of uncorrelated variables called principal components (PCs). This method identifies the linear combinations of the original variables that capture the largest amount of variance in the data. Mathematically, let X be an $n \times p$ data matrix, where n is the number of observations and p is the number of variables. The first principal component PC_1 is the linear combination of the variables that maximizes the variance:

$$PC_1 = w_{11}x_1 + w_{21}x_2 + \cdots + w_{p1}x_p \tag{5.5}$$

where w_{i1} are the coefficients (loadings) of the first principal component. The second principal component PC_2 is the linear combination of the variables that is orthogonal to PC_1 and has the largest variance. This process continues for subsequent principal components, each orthogonal to the previous ones and capturing the maximum remaining variance.

In Knoppe et al. (2023), PCA is used to identify the most influential topics driving market news. By applying PCA to a large dataset of market news articles, the authors were able to reduce the dimensionality of the data and focus on the most important topics that captured the majority of the variance in the data. The principal components can be considered as more easily understood representations of the original variables, capturing the essential structure and patterns in the market news.

5.5.3 Benchmark Models

Volatility Measure

In Zitti (2023a), a volatility proxy measure is needed to compare the predicted output of each model during the volatility forecasting analysis. The volatility is estimated using the sample standard deviation of logarithmic returns of the salmon spot price series computed over 4-week intervals using a rolling approach, as shown in the formula:

$$V_t = \frac{1}{T} \sum_{j=0}^{T-1} (r_{t-j} - \bar{r}_t)^2, \tag{5.6}$$

where r_j is the logarithmic return at time t, and $\bar{r}_j$ is the average of the logarithmic returns for the same period. The volatility is calculated using a rolling-window approach, which reduces the length of the logarithmic returns to $(k-4)$, where k is the length of the logarithmic return series.

One limitation of the employed volatility proxy is its smoothing effect on the logarithmic returns series, which results in persistence within the measure. Although an ideal volatility proxy is not clearly defined, the use of spot prices instead of forward prices in Zitti (2023a) (see Section 5.4.2) makes the implied volatility measure inapplicable. Consequently, this volatility measure, despite its drawbacks, is deemed sufficient as a benchmark to support the paper's objectives.

Autoregressive Moving Average (ARMA Model)

The Autoregressive Moving Average (ARMA) model is a standard tool in time-series forecasting. An ARMA(p, q) process is formulated as follows:

$$V_t = \mu + \phi_1(V_{t-1} - \mu) + \phi_2(V_{t-2} - \mu) + ... + \phi_p(V_{t-p} - \mu) + \epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} - ... - \theta_q\epsilon_{t-q}, \quad (5.7)$$

In this formulation, V_t is the actual value at time t, and ϵ_t is the random error at the same time point. The parameter μ represents the intercept, while $\phi_i : (i = 1, 2, ..., p)$ and $\theta_j : (j = 1, 2, ..., q)$ are the model parameters. The quantity p denotes the number of autoregressive terms, and q denotes the number of moving average terms.

In Zitti (2023a), the estimation of the ARMA model parameters is performed using the method of maximum likelihood, which seeks the parameters that maximize the probability of the observed data. The lag order values p and q are selected using the Akaike Information Criterion (AIC), a common method for model selection.

Zitti (2023a) employs the ARMA model as a benchmark for forecasting salmon market volatility and finds that salmon spot price volatility is best represented by an AR(5) model.

Vector Autoregressive Model

Vector Autoregressive (VAR) models are widely used for multivariate time series analysis, modeling linear dependencies among multiple time-evolving features. A reduced form of a Gaussian error VAR model of order p is:

$$\mathbf{y}_t = \nu + \mathbf{A}_1\mathbf{y}_{t-1} + \cdots + \mathbf{A}_p\mathbf{y}_{t-p} + \epsilon_t;\ t = 1, .., T \quad (5.8)$$

where $\mathbf{y}_t$ is an $n \times 1$ vector of estimated endogenous variables, ν is a $n \times 1$ vector of intercepts, $\mathbf{A}_i$ $(i = 1, ..., p)$ are $n \times n$ matrices of coefficients, ϵ_t is a $n \times 1$ error term vector, T is the time series length, and p determines the model's lag order. Error term distribution is Gaussian, without autocorrelation, zero mean, and constant variance.

In Zitti (2023b) and Knoppe et al. (2023), the AIC is used for selecting the optimal VAR order, and the OLS estimator is employed to estimate each VAR equation separately. Zitti (2023b) uses the generated $\mathbf{y}t$ series to forecast long-term salmon prices through the conditional expectation of $\mathbf{y}T + h$ given $\mathbf{y}t$, $t \leq T$:

$$\mathbf{y}_{T+h|T} = E\left(\mathbf{y}_{T+h}|\mathbf{y}_T, \mathbf{y}_{T-1}, \dots\right) = \nu + \mathbf{A}_1\mathbf{y}_{T+h-1|T} + \dots + \mathbf{A}_p\mathbf{y}_{T+h-p|T}, \tag{5.9}$$

where $\mathbf{y}_{T+j|T} = \mathbf{y}_{T+j}$ for $j \leq 0$. The optimal parameter selection for the first VAR model using front-month futures contracts as a predictor is $p = 4$ for forecasting $h = 1$ and 4 weeks ahead. For the remaining two VAR models using six-month and twelve-month futures contracts to forecast 26 and 52 weeks ahead, the optimal order selection is $p = 2$.

In Knoppe et al. (2023), impulse responses of the y_t series to one standard deviation innovations are found using 68 percent confidence bands, computed with bootstrap standard errors, using 1000 replications.

5.5.4 Deep-learning forecasting techniques

Long Short-term Memory (LSTM) is a type of Recurrent Neural Network (RNN) capable of maintaining information for extended periods (Hochreiter & Schmidhuber, 1997). Utilizing a memory cell and gates, it can store information effectively. The cell state (C_t) represents the network's memory, with forget (f_t), input (i_t), and output (o_t) gates filtering information via activation functions. The gates, cell state, and hidden state (h_t) are defined by the following equations (Eqs. 5.10 - 5.15):

$$f_t = \sigma\Big[W_f h_{t-1} + U_f x_t + b_f\Big], \tag{5.10}$$

$$i_t = \sigma\Big[W_i h_{t-1} + U_i x_t + b_i\Big], \tag{5.11}$$

$$\tilde{C}t = \tanh\Big[Wch_{t-1} + U_c x_t + b_c\Big], \tag{5.12}$$

$$C_t = f_t C_{t-1} + i_t \tilde{C}t, \tag{5.13}$$

$$ot = \sigma\Big[W_o h_{t-1} + U_o x_t + b_o\Big], \tag{5.14}$$

$$h_t = o_t \tanh[C_t]. \tag{5.15}$$

Fig.5.4 shows a hidden LSTM layer, where the forget gate (f_t), input gate (i_t), and

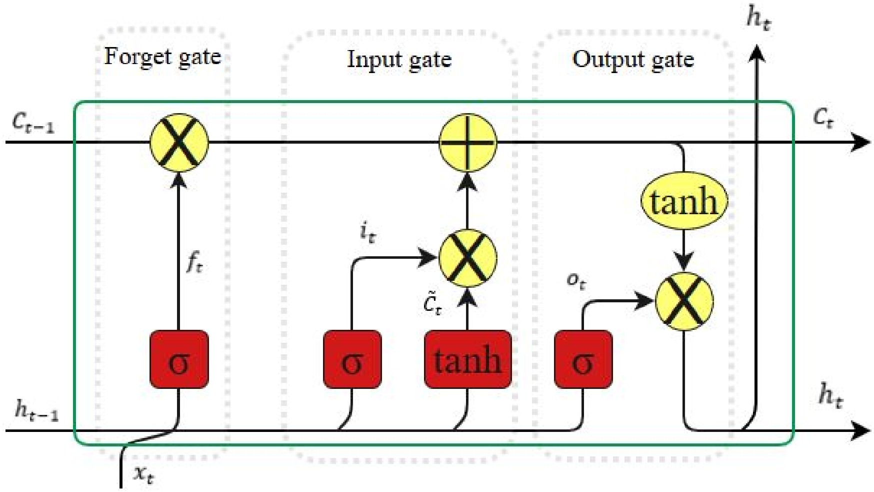


Figure 5.4: Structure of a long short-term memory (LSTM) layer with a forget gate as introduced by Gers et al. (2000).

output gate (o_t) process inputs (h_{t-1}) and (x_t), updating the cell state (C_t) and generating the new hidden state.

The LSTM model is utilized in both Zitti (2023a) and Zitti (2023b) under different circumstances. In Zitti (2023a), a single LSTM network is employed to forecast volatility using various input variables. These input series are transformed into supervised learning, data is scaled and split into train and test samples, and hyperparameters are tuned using the Random Search technique. Once the LSTM model is trained, it is used to forecast salmon market volatility. The resulting forecasts are then compared with the actual volatility measure (see Eq. 5.6), as well as the forecasts generated from the ARMA model that serves as a benchmark for forecasting in this study. In Zitti (2023b), a hybrid architecture combining VAR and LSTM is developed to test for non-linear relationships in the residuals of the VAR model. This hybrid approach takes advantage of the strengths of both models, as demonstrated by Eqs. 5.16, 5.17, 5.18, and 5.19, which illustrate the process of combining the linear component's forecast with the LSTM model's forecast as follows:

$$y_{1,t} = L_{1,t} + N_{1,t}. \tag{5.16}$$

$$\hat{\epsilon}_{1,t} = y_{1,t} - \hat{L}_{1,t}. \tag{5.17}$$

$$\hat{\epsilon}_{1,t} = f\left(\hat{\epsilon}_{1,t}, \hat{\epsilon}_{1,t-1}, \hat{\epsilon}_{1,t-2}, \ldots, \hat{\epsilon}_{1,t-k}\right) + u_{1,t}. \tag{5.18}$$

$$\hat{y}_{1,t} = \hat{L}_{1,t} + \hat{N}_{1,t}. \tag{5.19}$$

After generating the forecasts, the performance of the hybrid VAR-LSTM model is evaluated against the benchmark VAR model and examined under various multi-step ahead forecast horizons.

A disadvantage of incorporating deep learning techniques for forecasting is their computational intensity, requiring more processing time compared to traditional econometric models. Additionally, these methods are challenging to interpret, as they do not stem from a theoretical framework but are engineering-oriented instead. Consequently, many economists are hesitant to adopt them.

5.5.5 Text Mining Techniques

Latent Dirichlet allocation (LDA) is an unsupervised generative probabilistic model introduced by Blei et al. (2003), which represents documents as random mixtures of latent topics. Each topic is characterized by a distribution of words. LDA enables documents to belong to multiple topics, providing a mixed-membership model. It estimates parameters by generating topic distributions for each document and the distribution between topics and vocabulary.

In Knoppe et al. (2023), an extended version of LDA is used, placing Dirichlet priors on the probability vectors for smoother estimation. Collapsed Gibbs sampling, a Markov Chain Monte Carlo (MCMC) algorithm, is employed to estimate the LDA parameters. To select the optimal number of topics, 10-fold cross-validation is conducted, and the final perplexity is calculated as the average over 10 folds, and the optimal number of topics is determined to be 100.

In Knoppe et al. (2023), sentiment is also considered essential for explaining salmon market stock returns, as it offers various information types. A lexicon-based approach is utilized for sentiment quantification, which is relatively simple, efficient, and transparent. The Loughran-McDonald (LM) dictionary, a finance industry-specific lexicon, is used in the study. However, the dictionary's insensitivity to industry-specific language requires expansion to include terms relevant to salmon production. Consequently, Knoppe et al. (2023) expands the LM dictionary to encompass terms related to technology, diseases and natural disasters, and market-specific expressions. One limitation of this approach is that while incorporating industry-specific words, these terms are based on the available data set, which has a limited range.

5.6 Summary of the Research Articles

Research Article I: Climate risk and financial disclosure in salmon aquaculture.

This research paper investigates the financial impacts of climate change on the salmon aquaculture industry. It highlights the potential physical and transition risks associated with climate change and the increasing importance of climate-related financial disclosures. Organizations such as the Task Force for Financial Disclosures (TCFD) and the Carbon Disclosure Project (CDP) have been crucial in promoting transparency and accuracy in these disclosures. The study finds a sharp increase in the number of firms reporting their climate, forest, and water-security CDP reports between 2016 and 2021, indicating that salmon aquaculture companies are actively addressing environmental concerns and meeting investors' demands regarding transparency.

The implications of the study extend to salmon aquaculture companies, investors, and policymakers. As climate change becomes a more pressing issue, investors are increasingly expecting companies to adapt and respond to their demands for disclosure of climate-related risks. The establishment of organizations like the CDP and TCFD has played a significant role in raising awareness within firms, leading financial risk managers to develop strategies for addressing climate-related financial impacts and integrating them into the company's risk management processes. Consequently, firms are setting more ambitious goals to improve their climate-related risk disclosure, maintain climate resilience, and reduce their environmental footprints.

Research Article II: Forecasting Salmon Market Volatility using Long Short-term Memory (LSTM)

This study investigates the forecasting performance of deep-learning techniques, specifically the Long Short-term Memory (LSTM) model, in predicting salmon market volatility. The research aims to examine whether deep learning can accurately forecast salmon market volatility and help overcome challenges stemming from the illiquidity of the salmon futures market.

The paper utilizes the LSTM model to forecast salmon market volatility and assesses its performance against the actual volatility series and a benchmark forecasting model, represented by the ARMA model. The findings demonstrate that the ARMA model outperforms the LSTM in short-term salmon market volatility forecasting, while both models show similar capabilities for longer forecasting horizons. However, a significant discrepancy exists between the forecasts generated by both models and the actual volatility values, especially for mid-term forecasting horizons.

The results of this study provide valuable insights into the development of salmon price volatility, suggesting that there are no complex patterns that can be exploited by an LSTM model in the short term. The forecasting framework introduced in this paper could be applied to other commodities and expanded to include multivariate predictive analyses accounting for factors driving salmon market volatility.

Research Article III: The Influence of Market Adaptability on Forecasting Salmon Spot Prices: A Comparison of Hybrid Deep Learning and Traditional Models

Salmon price prediction is crucial for various market participants, such as farmers, processors, traders, and hedgers. Due to the highly volatile nature of the salmon market, accurate price forecasts are essential to reduce risks and support market participants.

This study explores the forecasting performance of a hybrid model that combines the traditional vector autoregressive (VAR) model with a deep learning technique, the long short-term memory (LSTM). We hypothesize that this hybrid VAR-LSTM model, by capturing both linear and potentially complex non-linear dependencies, could enhance the precision of salmon spot price predictions, thus improving the accuracy of salmon spot price forecasts compared to the performance of a traditional VAR model

However, the results revealed that both the hybrid VAR-LSTM model and the traditional VAR model performed similarly, suggesting that salmon prices might not fluctuate in complex, non-linear ways as assumed. This points to the salmon market's efficiency - price changes are quickly recognized and integrated by market players. The VAR model was also tested against a benchmark forecasting model, the random walk with a drift model, and outperformed it across different prediction periods indicating that salmon spot prices, although potentially exhibiting random walk properties, are not purely random and do contain predictable dynamics that can be captured effectively by VAR. Even though VAR and VAR-LSTM showed similar accuracy, the simpler VAR model demonstrated a modest advantage, highlighting the balance between complexity and accuracy in forecasting. These findings offer useful insights for industry participants, providing valuable perspectives on market behavior and price prediction strategies.

Research Article IV: Salmon price movements around salmon market news

This study examines the impact of news on stock returns of major salmon-producing companies listed on the Oslo Stock Exchange by analyzing articles related to salmon production and markets. Latent Dirichlet Allocation (LDA) is employed to generate topics, and a dictionary approach is used for sentiment analysis.

Initially, the study evaluated the impact of news topics on the salmon market by grouping them into components and applying Vector Autoregression (VAR) analyses to examine the relationship between these components and logarithmic stock returns. The analysis was constrained to absolute returns due to the undefined direction of news coverage effects. While news about the Covid-19 pandemic predominantly shaped market repercussions, components associated with corporate news also elicited significant market reactions.

Upon evaluating topics combined with sentiment, it was found that sentiment dictionaries, such as the Loughran-McDonald dictionary, were not sufficiently adaptable to domains beyond their original design. As a result, the dictionary was expanded with industry-specific terms. To assess the relevance of the expanded lexicon, the effects of industry-specific topics, when multiplied by the extended sentiment index, on returns were analyzed, including factors such as algal blooms, diseases, and R&D, while accounting for competition among producers. However, the extended dictionary was unable to account for the Covid-19 topic, which represented an unexpected exogenous shock to the market.

In conclusion, this study emphasizes the importance of accounting for competition and market structure in the salmon industry but recognizes a trade-off between focusing on firm-specific articles and retaining crucial market news. The results could be applied in similar studies within aquaculture economics, beyond the salmon industry. Future research could consider expanding the time series range, incorporating additional news sources, broadening the dictionary to include competitive seafood markets, and exploring enhanced data filtering techniques to preserve general market news and competitor information while removing unrelated noise.

5.7 Conclusions and Contributions

The thesis consists of four research articles that examine critical aspects of the salmon aquaculture industry, including climate-related financial disclosures, market volatility and price forecasting, and price movements in response to salmon market news. These studies provide valuable insights for salmon aquaculture companies, investors, policymakers, and other market participants, emphasizing the necessity of transparent climate-related financial disclosures, reliable forecasting models to predict market volatility and prices, and identifying the underlying market drivers.

The first article contributes to the growing body of literature on climate risk and financial disclosures, emphasizing the importance of transparency and the role of organizations like TCFD and CDP in raising awareness and promoting better climate-related risk practices. This study highlights the importance of increased disclosure, which benefits salmon aquaculture companies, investors, and policymakers by promoting more sustainable and resilient business practices.

The second article evaluates the forecasting performance of deep-learning techniques, such as LSTM, in predicting salmon market volatility, compared against the benchmark ARMA model. The ARMA model exhibits better forecasting ability than the LSTM model, indicating that there are no complex patterns in salmon market volatility that can be exploited by an LSTM. Nevertheless, this study encourages future research into the use of deep learning techniques and their potential applications in the field of aquaculture finance.

The third article investigates the efficacy of a hybrid VAR-LSTM model for multi-step ahead forecasting of salmon spot prices. While the hybrid model does not significantly improve salmon price forecasting compared to the traditional VAR model, suggesting that salmon prices do not depict non-linear trends, it still points to the salmon market's efficiency where price changes are quickly recognized and integrated by investors.

Finally, the fourth article explores the impact of news on stock returns of major salmon-producing companies, highlighting the importance of accounting for competition and market structure. The study recognizes the limitations of existing sentiment dictionaries, suggesting the need for domain-specific lexicons and improvements in data filtering techniques to better understand the effects of news on the salmon industry.

The cumulative insights from these research articles significantly contribute to advancing climate-related financial disclosure within the industry, developing more sophisticated forecasting models to support salmon market participants, and enhancing a comprehensive understanding of the factors driving the salmon market. These findings can assist policymakers in comprehending the complexities of the salmon industry and the market, considering the challenges of the market participants before implementing new policies and regulations.

Moreover, these results can help producing companies in identifying the areas to prioritize in their future development, such as sustainability, market risk management, and internal research and development. By doing so, they can make more informed decisions that contribute to their long-term success. Additionally, these conclusions can help market participants to better anticipate future trends, enabling them to avoid planning decisions that may lead to price fluctuations and increased market volatility. Ultimately, the findings serve to enhance decision-making processes across the salmon aquaculture industry, promoting more efficient, sustainable, and resilient practices.

5.8 Future Research Directions

There are several opportunities for future research that build upon the findings of this thesis. In the context of climate risk and financial disclosure, further research could explore the effectiveness of current disclosure practices and investigate potential improvements to enhance the quality of reporting. This could involve analyzing the impact of new regulatory frameworks or assessing the role of novel technologies in promoting transparency and accountability in the industry.

For price forecasting and volatility analysis, advancements in deep-learning techniques and alternative models, such as Convolutional Neural Networks (CNNs), Graph Neural Networks (GNNs), and emerging deep learning techniques, such as Transformers can be explored to improve forecasting performance. Additionally, incorporating external factors, such as economic factors, into the forecasting models may provide a more comprehensive understanding of price dynamics and volatility in the salmon aquaculture industry.

In terms of news impact on stock returns, future research can focus on expanding the time series range and incorporating additional news sources, including social media and alternative data streams, to gain deeper insights into the relationship between news and stock market price development. Researchers could also explore a supervised approach instead of a lexicon-based approach.

Moreover, the findings of this thesis can be applied to other sectors within aquaculture economics or the broader seafood market, enabling comparative analyses and providing a more comprehensive understanding of the challenges and complexities faced by different industry participants. By exploring these future research directions, researchers can continue to advance theirs and the industry's knowledge, ultimately supporting more sustainable, efficient, and resilient business practices in the face of evolving market conditions and global challenges.

References

Abolofia, J., Asche, F., & Wilen, J. E. (2017). The cost of lice: Quantifying the impacts of parasitic sea lice on farmed salmon. *Marine Resource Economics*, *32*(3), 329–349.

Andersen, B. P., & de Lange, P. E. (2021). Efficiency in the atlantic salmon futures market. *Journal of Futures Markets*, *41*(6), 949–984.

Ankamah-Yeboah, I., Nielsen, M., & Nielsen, R. (2017). Price formation of the salmon aquaculture futures market. *Aquaculture Economics & Management*, *21*(3), 376–399.

Asche, F. (1996). A system approach to the demand for salmon in the european union. *Applied Economics*, *28*(1), 97–101.

Asche, F. (2008). Farming the sea. *Marine Resource Economics*, *23*(4), 527–547.

Asche, F., Anderson, J. L., Botta, R., Kumar, G., Abrahamsen, E. B., Nguyen, L. T., & Valderrama, D. (2021). The economics of shrimp disease. *Journal of invertebrate pathology*, *186*, 107397.

Asche, F., & Bjorndal, T. (2011). *The economics of salmon aquaculture*. John Wiley & Sons.

Asche, F., Bjørndal, T., & Salvanes, K. G. (1998). The demand for salmon in the european union: The importance of product form and origin. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*, *46*(1), 69–81.

Asche, F., Cojocaru, A. L., & Roth, B. (2018). The development of large scale aquaculture production: A comparison of the supply chains for chicken and salmon. *Aquaculture*, *493*, 446–455.

Asche, F., Dahl, R. E., Valderrama, D., & Zhang, D. (2014). Price transmission in new supply chains—the case of salmon in france. *Aquaculture Economics & Management*, *18*(2), 205–219.

Asche, F., Eggert, H., Oglend, A., Roheim, C. A., & Smith, M. D. (2022). Aquaculture: Externalities and policy options. *Review of Environmental Economics and Policy*, *16*(2), 282–305.

Asche, F., Guttormsen, A. G., & Tveterås, R. (1999). Environmental problems, productivity and innovations in Norwegian salmon aquaculture. *Aquaculture Economics & Management*, *3*(1), 19–29.

Asche, F., Misund, B., & Oglend, A. (2016a). Determinants of the atlantic salmon futures risk premium. *Journal of Commodity Markets*, *2*(1), 6–17.

Asche, F., Misund, B., & Oglend, A. (2016b). The spot-forward relationship in the atlantic salmon market. *Aquaculture Economics & Management*, *20*(2), 222–234.

Asche, F., Misund, B., & Oglend, A. (2019). The case and cause of salmon price volatility. *Marine Resource Economics*, *34*(1), 23–38.

Asche, F., Oglend, A., & Selland Kleppe, T. (2017). Price dynamics in biological production processes exposed to environmental shocks. *American Journal of Agricultural Economics*, *99*(5), 1246–1264.

Asche, F., Sikveland, M., & Zhang, D. (2018). Profitability in norwegian salmon farming: The impact of firm size and price variability. *Aquaculture economics & management*, *22*(3), 306–317.

Battiston, S., Mandel, A., Monasterolo, I., Schütze, F., & Visentin, G. (2017). A climate stress-test of the financial system. *Nature Climate Change*, *7*(4), 283–288.

Bjørndal, T., & Tusvik, A. (2019). Economic analysis of land based farming of salmon. *Aquaculture Economics & Management*, *23*(4), 449–475.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993–1022.

Bloznelis, D. (2016). Salmon price volatility: A weight-class-specific multivariate approach. *Aquaculture Economics & Management*, *20*(1), 24–53.

Bloznelis, D. (2018a). Hedging salmon price risk. *Aquaculture Economics & Management*, *22*(2), 168–191.

Bloznelis, D. (2018b). Short-term salmon price forecasting. *Journal of Forecasting*, *37*(2), 151–169.

Bovari, E., Giraud, G., & Mc Isaac, F. (2018). Coping with collapse: A stock-flow consistent monetary macrodynamics of global warming. *Ecological Economics*, *147*, 383–398.

Brækkan, E. H., Thyholdt, S. B., Asche, F., & Myrland, Ø. (2018). The demands they are a-changin'. *European Review of Agricultural Economics*, *45*(4), 531–552.

Chen, X., & Scholtens, B. (2019). The spot-forward relationship in the atlantic salmon market. *Reviews in Fisheries Science & Aquaculture*, *27*(2), 142–151.

Dafermos, Y., Nikolaidi, M., & Galanis, G. (2017). A stock-flow-fund ecological macroeconomic model. *Ecological Economics*, *131*, 191–207.

Dafermos, Y., Nikolaidi, M., & Galanis, G. (2018). Climate change, financial stability and monetary policy. *Ecological Economics*, *152*, 219–234.

Dahl, R. E., & Oglend, A. (2014). Fish price volatility. *Marine Resource Economics*, *29*(4), 305–322.

Dahl, R. E., Oglend, A., & Yahya, M. (2021). Salmon stock market prices revealing salmon price information. *Marine Resource Economics*, *36*(2), 173–190.

Dahl, R. E., & Yahya, M. (2019). Price volatility dynamics in aquaculture fish markets. *Aquaculture Economics & Management*, *23*(3), 321–340.

De Silva, S. S., & Soto, D. (2009). Climate change and aquaculture: Potential impacts, adaptation and mitigation. *Climate change implications for fisheries and aquaculture: overview of current scientific knowledge. FAO Fisheries and Aquaculture Technical Paper*, *530*, 151–212.

Dietz, S., Bowen, A., Dixon, C., & Gradwell, P. (2016). 'Climate value at risk' of global financial assets. *Nature Climate Change*, *6*(7), 676–679.

Dou, Y., da SILVA, R. F. B., Yang, H., & Liu, J. (2018). Spillover effect offsets the conservation effort in the amazon. *Journal of Geographical Sciences*, *28*(11), 1715–1732.

Ewald, C.-O., Haugom, E., Kanthan, L., Lien, G., Salehi, P., & Størdal, S. (2022). Salmon futures and the fish pool market in the context of the capm and a three-factor model. *Aquaculture Economics & Management*, *26*(2), 171–191.

Ewald, C.-O., Ouyang, R., & Siu, T. K. (2017). On the market-consistent valuation of fish farms: Using the real option approach and salmon futures. *American Journal of Agricultural Economics*, *99*(1), 207–224.

Fischer, C., Guttormsen, A. G., & Smith, M. D. (2017). Disease risk and market structure in salmon aquaculture. *Water Economics and Policy*, *3*(02), 1650015.

Fish Pool. (2017a). Fish pool index. https://fishpool.eu/wp-content/uploads/2022/03/Trade-membership-agreement-2018.pdf

Fish Pool. (2017b). Fish pool index. https://fishpool.eu/fish-pool-index/

Fish Pool, A. (2020). Annual report 2020.

Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with lstm. *Neural Computation*, *12*(10), 2451–2471.

Guttormsen, A. G. (1999). Forecasting weekly salmon prices: Risk management in fish farming. *Aquaculture Economics & Management*, *3*(2), 159–166.

Hersoug, B. (2021). Why and how to regulate norwegian salmon production?–the history of maximum allowable biomass (mab). *Aquaculture*, *545*, 737144.

Hersoug, B., Mikkelsen, E., & Karlsen, K. M. (2019). "Great expectations" – Allocating licenses with special requirements in norwegian salmon farming. *Marine Policy*, *100*, 152–162.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Hvas, M., Folkedal, O., & Oppedal, F. (2021). Fish welfare in offshore salmon aquaculture. *Reviews in Aquaculture*, *13*(2), 836–852.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice*. OTexts.

Iversen, A., Asche, F., Hermansen, Ø., & Nystøyl, R. (2020). Production cost and competitiveness in major salmon farming countries 2003–2018. *Aquaculture*, *522*, 735089.

Jansen, P. A., Kristoffersen, A. B., Viljugrein, H., Jimenez, D., Aldrin, M., & Stien, A. (2012). Sea lice as a density-dependent constraint to salmonid farming. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1737), 2330–2338.

Jonsson, B., & Jonsson, N. (2006). Cultured atlantic salmon in nature: A review of their ecology and interaction with wild fish. *ICES Journal of Marine Science*, *63*(7), 1162–1181.

Kim, H. Y., & Won, C. H. (2018). Forecasting the volatility of stock price index: A hybrid model integrating lstm with multiple garch-type models. *Expert Systems with Applications*, *103*, 25–37.

Knoppe, C., Okuneva, M., & Zitti, M. (2023). Salmon price movements and trading behaviors around salmon market news. *Working Paper*.

Landazuri-Tveteraas, U., Asche, F., Gordon, D. V., & Tveteraas, S. L. (2018). Farmed fish to supermarket: Testing for price leadership and price transmission in the salmon supply chain. *Aquaculture Economics & Management*, *22*(1), 131–149.

Larsen, T. A., & Asche, F. (2011). Contracts in the salmon aquaculture industry: An analysis of norwegian salmon exports. *Marine Resource Economics*, *26*(2), 141–150.

Leaton, J. (2011). Unburnable carbon—are the world's financial markets carrying a carbon bubble? *Carbon Tracker Initiative*, 6–7.

Liu, P., Lien, K., & Asche, F. (2016). The impact of media coverage and demographics on the demand for norwegian salmon. *Aquaculture Economics & Management*, *20*(4), 342–356.

Liu, Y., Olaussen, J. O., & Skonhoft, A. (2011). Wild and farmed salmon in norway—a review. *Marine policy*, *35*(3), 413–418.

Misund, B. (2017). Financial ratios and prediction on corporate bankruptcy in the Atlantic salmon industry. *Aquaculture Economics & Management*, *21*(2), 241–260.

Misund, B., & Asche, F. (2016). Hedging efficiency of atlantic salmon futures. *Aquaculture Economics & Management*, *20*(4), 368–381.

Mowi, A. (2020). Salmon farming industry handbook 2020.

Oglend, A. (2013). Recent trends in salmon price volatility. *Aquaculture Economics & Management*, *17*(3), 281–299.

Oglend, A., & Sikveland, M. (2008). The behaviour of salmon price volatility. *Marine Resource Economics*, *23*(4), 507–526.

Oglend, A., & Straume, H.-M. (2019). Pricing efficiency across destination markets for norwegian salmon exports. *Aquaculture Economics & Management*, *23*(2), 188–203.

Olaussen, J. O. (2018). Environmental problems and regulation in the aquaculture industry. insights from norway. *Marine Policy*, *98*, 158–163.

Parot, A., Michell, K., & Kristjanpoller, W. D. (2019). Using artificial neural networks to forecast exchange rate, including var-vecm residual analysis and prediction linear combination. *Intelligent Systems in Accounting, Finance and Management*, *26*(1), 3–15.

Pincinato, R. B. M., Asche, F., & Roll, K. H. (2021). Escapees in salmon aquaculture: A multi-output approach. *Land Economics*, *97*(2), 425–435.

Ramyar, S., & Kianfar, F. (2019). Forecasting crude oil prices: A comparison between artificial neural networks and vector autoregressive models. *Computational Economics*, *53*(2), 743–761.

Solibakke, P. B. (2012). Scientific stochastic volatility models for the salmon forward market: Forecasting (un-) conditional moments. *Aquaculture Economics & Management*, *16*(3), 222–249.

Stolbova, V., Monasterolo, I., & Battiston, S. (2018). A financial macro-network approach to climate policy evaluation. *Ecological Economics*, *149*, 239–253.

Sun, J., Mooney, H., Wu, W., Tang, H., Tong, Y., Xu, Z., Huang, B., Cheng, Y., Yang, X., Wei, D., et al. (2018). Importing food damages domestic environment: Evidence from global soybean trade. *Proceedings of the National Academy of Sciences*, *115*(21), 5415–5419.

TCFD. (2017). Recommendations of the task force on climate-related financial disclosures.

Torrissen, O., Jones, S., Asche, F., Guttormsen, A., Skilbrei, O. T., Nilsen, F., Horsberg, T. E., & Jackson, D. (2013). Salmon lice–impact on wild salmonids and salmon aquaculture. *Journal of fish diseases*, *36*(3), 171–194.

Vassdal, T., & Sørensen Holst, H. M. (2011). Technical progress and regress in norwegian salmon farming: A malmquist index approach. *Marine Resource Economics*, *26*(4), 329–341.

Xie, J., Kinnucan, H. W., & Myrland, Ø. (2009). Demand elasticities for farmed salmon in world trade. *European Review of Agricultural Economics*, *36*(3), 425–445.

Xie, J., & Myrland, Ø. (2011). Consistent aggregation in fish demand: A study of french salmon demand. *Marine Resource Economics*, *26*(4), 267–280.

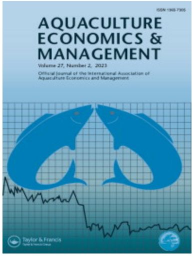
Zitti, M. (2023a). Forecasting salmon market volatility using long short-term memory (lstm). *Working Paper*.

Zitti, M. (2023b). Multi-step ahead forecasting of salmon spot prices using a hybrid deep learning model. *Working Paper*.

Zitti, M., & Guttormsen, A. G. (2022). Climate risk and financial disclosure in salmon aquaculture. *Aquaculture Economics & Management*, 1–27.

Research Articles

Research Article I



Aquaculture Economics & Management

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uaqm20

Climate risk and financial disclosure in salmon aquaculture

Mikaella Zitti & Atle G. Guttormsen

To cite this article: Mikaella Zitti & Atle G. Guttormsen (2022): Climate risk and financial disclosure in salmon aquaculture, Aquaculture Economics & Management, DOI: 10.1080/13657305.2022.2143934

To link to this article: https://doi.org/10.1080/13657305.2022.2143934

Published online: 17 Nov 2022.

AQUACULTURE ECONOMICS & MANAGEMENT
https://doi.org/10.1080/13657305.2022.2143934

Climate risk and financial disclosure in salmon aquaculture

Mikaella Zitti and Atle G. Guttormsen

Department of Economics and Resource Management, Norwegian University of Life Sciences, Norway

ABSTRACT
The growth of the salmon aquaculture industry has attracted an increasing number of investors. Investors are conscious of economic consequences of climate change for the salmon farming companies, hence their demand for climate-related financial disclosure has increased. This study discusses potential climate-related financial impacts imposed on the salmon aquaculture production as identified by the Task Force on Climate-related Financial Disclosure (TCFD). We use data from 2016 to 2021 available on the Carbon Disclosure Project (CDP) platform, and we show to what extent salmon aquaculture companies disclose their climate-related risks. We find that when the demand of investors for climate-related financial disclosure increases, more firms tend to comply with their requests, while based on the CDP's evaluation system, the firms perform better in minimizing their carbon impact. We argue that when salmon aquaculture companies publish their climate-related financial disclosure, they ensure transparency for their investors and secure a smooth transition into a low carbon economy.

KEYWORDS
CDP; Climate risk; financial disclosure; Salmon; TCFD

Introduction

Over the past decades, aquaculture has been the world's fastest-growing food production industry (Smith et al., 2010), with an annual growth of 7% (FAO, 2016). Global production has increased from around 14.9 million tonnes in 1995 to 82.1 million tonnes in 2018 (FAO, 2020). The increase in production has been possible due to substantial technological innovation, which led to productivity growth and lower production costs. As a result, aquaculture has become an economically competitive food production (Asche, 2008; Asche, Roll, Sandvold et al., 2013; Bergesen & Tveterås, 2019;

CONTACT Mikaella Zitti mikaella.zitti@nmbu.no School of Economics and Business, Norwegian University of Life Sciences, P.O. Box 5003, Ås 1430, Norway.

Kumar & Engle, 2016). Overall, modern aquaculture has developed into a technologically advanced and profitable industry.

As fish farming technology has improved, the salmon-producing industry has matured and consolidated. The increased use of advanced technology has made the industry more capital intensive. This drove companies to seek for external financing, hence many were publicly listed. Today 26 companies engaged with salmon farming are traded on the Oslo Stock Exchange (OSE), one of the major seafood stock exchanges. The market value of these companies is around 368 billion Norwegian Kroner (€36.8 billion). As Sikveland et al. (2021) showed, return on assets (ROA) has increased for publicly listed companies between 2001 and 2014. As a result, they become more attractive to investors. Thus, they are also responsible to meet investors' demands to keep them satisfied.

Aquaculture production is a biological production that is subject to severe economic consequences due to the uncertain nature of climate-change (Asche, 2008; Asche et al., 2017, 2022). The success of the industry has led to a number of studies investigating price volatility as a consequence of climate change (Asche et al., 2017, 2019), but so far, little attention has been paid to the potential financial impacts from climate change. Misund (2017) argues that common and industry-specific risk factors can impact aquaculture firms' returns. Listed salmon aquaculture firms have a responsibility toward their investors to tackle such risks and cannot ignore them.

Financial climate-related risks are usually identified as either physical or transition risk (TCFD, 2017). Physical risks are defined as the risks related to economic damages that come as a result of the effect of climate change (Bovari et al., 2018; Dafermos et al., 2017, 2018; Dietz et al., 2016; TCFD, 2017). For instance, farmed salmon is kept in pens and if pens are destroyed as a result of extreme weather, salmon can escape, causing major financial consequences. Transition risks are linked to the risk of the firms as a result of market shocks and policy implications related to transitioning into a low-carbon economy (Battiston et al., 2017; Dafermos et al., 2018; Leaton, 2011; Stolbova et al., 2018; TCFD, 2017). In salmon aquaculture, transition risks are mainly related to policies and regulations. A large amount of the supplies of the Norwegian salmon aquaculture industry is exported to a diverse market (Asche et al., 2013; Straume, 2017), hence a substantial increase in the carbon tax can impact air cargo and increase price volatility (Asche et al., 2019).

Climate-related risks are expected to cause turmoil not only for the salmon aquaculture sector, but also across various sectors outside the food production industry. Companies across sectors are becoming aware of the financial risks associated with climate change and are looking to become climate resilient to maintain future profitability. Investors, lenders, and other stakeholders also become more aware of climate-related risks and their demands for climate resilience and adaptation increase. Climate-related financial disclosures are key in reassuring investors and other stakeholders regarding climate risks. Therefore, over the recent years organizations such as the Carbon Disclosure Project (CDP) and the Task Force for Financial Disclosures (TCFD) were created to promote the importance of reporting on climate-related risks. Such disclosures assist investors, stakeholders, and other market participants to gain a better understanding of the industry's challenges related to climate change as well as support their future investment decisions. A standardized framework, such as the one provided by the TCFD, can also assist financial risk managers to understand and identify the risks their companies face. It is however important to note that disclosures are crucial to maintain transparency, but they are not to be seen as an indicator of good financial performance per se. They are an indication that the underlying company is tackling climate-related financial risks while considering stakeholders' demands.

In the literature on business studies and finance there are theories that are often used to explain the relationship between companies, their stakeholders, and society. This study is compatible with theories such as the *legitimacy theory*, *signaling theory*, and *stakeholder theory*. In line with the legitimacy theory, organizations must ensure that they carry out activities in accordance with societal boundaries and norms (Deegan et al., 2002). Based on Patten (2002) and their definition of legitimacy theory, if there is a large scale environmental disaster incident in a salmon-producing company (e.g., disease outbreak) then other companies also respond by increasing the amount of environmental disclosures in their annual reports even though the incident itself was directly related to one company. There are also significant information asymmetries concerning climate-related risks between corporate insiders and directors and other stakeholders, such as investors. Disclosing climate-related financial risk can reduce such asymmetries in the access to information, as discussed in the literature on signaling theory (e.g. Connelly et al., 2011). Nygård (2020) found that sustainability reports contribute to reduce the information gap between managers and shareholders. By gaining access to information on the company they are invested in, investors and other stakeholders can impact decisions regarding the operations and finances of that company. According to the stakeholder theory, a company is responsible to meet all the demands from their stakeholders (Parmar et al., 2010). Therefore, if the investors' demands for climate-related financial disclosure increase, the company is responsible to disclose its exposure on climate-risk.

In this study, we discuss climate-related risks specific to the salmon industry and explore to what extent some of the publicly listed firms have been implementing climate-related financial disclosures. Next, we investigate the role of TCFD and CDP in increasing responses from the companies. There are other initiatives that also attempt to quantify sustainability matters such as the Global Reporting Initiative (GRI),[1] and specific to the aquaculture industry, the Aquaculture Stewardship Council (ASC).[2] These initiatives help companies understand the outward impacts of their unsustainable practices which in turn helps them become more sustainable and impacts their willingness to report on their climate risks. We will not explain further the specifications of these initiatives as this study focuses only on the TCFD and CDP. Furthermore, we will only focus on investors' demands regarding climate-related financial disclosures and not on other stakeholders. This is because we collect publicly available data from the CDP platform which is an investor-driven global disclosure system. The outcome of this study is applicable to salmon aquaculture firms, as well as their investors and other stakeholders, but also policy-makers and regulators.

In Section "Climate-related risks conceptualization" we divide the climate-related risks into physical and transition, and discuss the ones relevant to the Norwegian salmon aquaculture industry. In Section "Climate-related financial disclosure," we explain the role of TCFD and CDP and the specifics of climate-risk financial disclosures for salmon aquaculture firms, as well as the main potential financial impacts they can face. In Section "Results" we analyze the publicly available reports obtained from the CDP platform. Section "Discussion" presents a discussion of the outcomes.

Climate-related risks conceptualization

Salmon farming production is a biological practice that interacts with its surrounding environment. It can therefore be vulnerable to potential financial impacts as a consequence of climate-related risks. A set of studies raising the environmental challenges of the salmon farming industry were selected and will be discussed (Asche et al., 1999, 2009; Smith et al., 2010). To facilitate the discussion around climate-related risks, the studies raising the environmental challenges of the aquaculture industry will be categorized according to physical and transition risks.

Physical risks

There are physical and biological risks associated with the salmon farming industry's practices. Since salmon is farmed in open cages it can be exposed to a number of climate-related risks (Asche et al., 2018). The major ones are large-scale disease outbreaks (Abolofia et al., 2017; Asche et al., 2021; Fischer et al., 2017; Torrissen et al., 2013), pollution through feed waste (Asche, 2008; Asche et al., 1999), and algal blooms. Climate change can potentially worsen the impacts of these risks. The change in water temperature, increasing precipitation, raise in sea-levels, frequent storm surges and other extreme weather events are some of the climate-related risks that can financially impact the salmon aquaculture industry.

Sea cages, where salmon is kept, are an open system of stock cultivation. If storms are stronger than expected, they can destroy the cages the salmon is kept in, and it can escape, resulting to unexpected losses for the farmers (Olaussen, 2018). Salmon escaping from the sea cages is not something unfamiliar to the farmers and the relevant authorities and therefore they manage to keep the escapes under control without them causing large losses (Pincinato et al., 2021). Nevertheless, the farmers and the relevant authorities must consider the added risk from climate change. It is possible that salmon escaping becomes more frequent and the farmers are no longer able to control its potential outcome. Storm surges can also destroy materials and installations exposing the industry's firms to severe economic consequences. Given that the installations are a valuable asset for the salmon farming practice, it being destroyed can significantly decrease stockholders' equity.

Chronic physical risks, e.g. increase in water temperature, can impose long-term changes that will demand climate adaptation from the industry's underlying firms (De Silva & Soto, 2009). Specifically, the dynamics of salmon lice are depending on the water temperature, and they are characterized by annual oscillations in parasite abundance (Jansen et al., 2012). For instance, salmon thrives between 9 and 14 degrees, hence instability in water temperature, can cause a disease outbreak, potentially impacting the mortality rate of salmon as well as increasing production costs as a result of necessary treatment processes, e.g. medicine and vaccines (Asche et al., 1999; Torrissen et al., 2013). Abolofia et al. (2017) found that lice parasitism alone produced 436 million US dollars in damages to the Norwegian industry in 2011. Algal bloom which is linked to an increased concentration of nutrients in the sea, known as eutrophication is also an existing threat to the salmon industry, since it can be the cause of diseases spreading (Asche et al., 1999). It is a biological risk for the salmon aquaculture production and as for storm surges, such events can become more frequent with the imposed risk of climate change. Norwegian coastal waters are particularly vulnerable to an increase in the frequency of toxic algal bloom (Edwards et al., 2006). As a consequence of diseases spreading, the industry's firms must invest sufficient funds into research (Asche, 2008). To be able to do this, the firms must hold a required capital amount. All physical risks related to the salmon aquaculture industry can potentially increase costs and expenses. Therefore, a so-called sustainable food production such as salmon aquaculture industry cannot afford ignoring these risks and their potential financial impacts.

Transition risks

Climate change and its impacts are receiving increased political attention. It is then sensible to expect that the political scene regarding matters related to climate change is changing. The impacts from climate change are enormous and the scientific world has been urging politicians to act immediately on the climate (IPCC, 2022). However, the long inactivity from politics has now emphasized the urgency for "new" relevant policy implementations that ensure a smooth transition into a low-carbon economy. What if a "smooth transition" is no longer possible? To be more specific, the urgency to save our planet can result in ill-considered decisions. Abate et al. (2016) argued that stricter environmental regulations in developed countries have contributed to lower growth rates. For instance, in Norway, environmental concerns regarding the exponential growth of the Norwegian aquaculture production drove the authorities to restrict the licenses for new production (Asche et al., 2017; Hersoug et al., 2019). Asche et al. (2019) argued that such restrictive regulations can result in the stagnation of the industry's production growth, causing an increase in salmon price volatility. This can have severe consequences for the producers (farmers) and other market participants such as investors, whose aim is to maintain profitability at a minimum cost. Implementing policies without consulting the relevant industry can have undesirable outcomes. Operating based on a holistic approach is vital for creating policies and regulations that consider the industry's environmental challenges without harming the farmers and jeopardizing the production (Osmundsen et al., 2017).

The key for the industry (or any industry) to maintain its profitability is the trust from its stockholders. In return, it must ensure the protection of the stockholders' equity. Therefore, it is vital for the industry's firms to be prepared for transitioning into a low-carbon economy. Otherwise, the costs will increase, potentially resulting in high losses, so that many shareholders will pull out if they have not yet. If stockholders sell a large number of shares from a publicly traded firm, it is likely to cause the value of the firm's stock to fall. Listed Norwegian aquaculture firms have high liquidity buffers, but using it as a risk reduction measure is costly (Sikveland et al., 2021; Sikveland & Zhang, 2020), which can have adverse effects on returns. This in turn can reduce the stocks' value. Moreover, considering that the salmon market offers hedging opportunities for investors through the Fish Pool futures exchange for salmon, Ewald et al. (2022) found a correlation between the shares prices and longer salmon futures contracts. This implies that the salmon stock market reflects on the salmon market risk, hence, climate-related risks associated with salmon aquaculture companies are likely to impact salmon prices. To avoid these, the salmon industry firms have a responsibility toward their stakeholders to maintain their growth and reputation. The TCFD (2017) recommendations argue that the key to achieve this are the financial risk managers. They are required to have sufficient knowledge about climate-related risks associated with the transition into a low-carbon economy. This knowledge is important for developing trustworthy climate-related financial disclosures as well as managing the exposure into climate-related risks (TCFD, 2017).

Transitioning to a low-carbon economy does not appear to be a smooth transition for the salmon aquaculture industry. The transition risks associated with the industry are various and challenging. An increase in the carbon tax, bringing farms inland or demand closed containment to improve sustainability, importing soybeans from South America despite the procedure causing deforestation, are some of the challenges the salmon industry is facing and must overcome to attain a smooth transition.

An increase in the carbon tax can severely impact the salmon aquaculture industry. Figure 1 shows that at the end of 2021, salmon represented approximately 40% of the quantity of all seafood species exported from

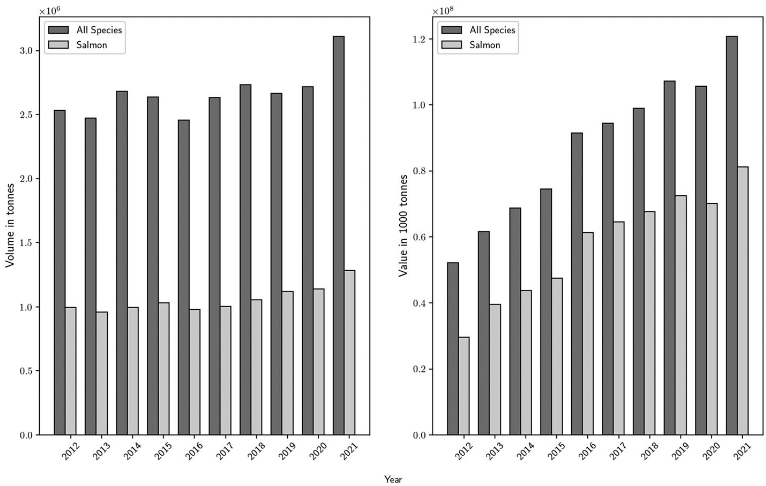


Figure 1. Comparison of the Volume in Metric Tonnes and Value in 1000 Tonnes of salmon and all species exported from Norway from 2012 to 2021. *Source:* Norwegian Seafood Council (2022).

Norway and 60% of the value. The increasing trend of the export volume of salmon verifies that it is a significant species for the export growth of the Norwegian seafood industry. Around 20% of salmon exported from Norway is shipped with air carriers; hence, if the Norwegian authorities are raising concerns relating to the carbon emissions of air transport, an increase in the carbon tax on air cargo is a likely policy measure. If the cost of delivering their product to their consumers rises, salmon producers are forced to either raise prices and face lower demand, or accept lower profit margins. Figure 2 shows the development of the annual export prices and production costs for salmon[3] from 2008 to 2021. It is evident that both the annual export prices and production costs have been increasing over this period. Even though the production costs for Norwegian farmed Atlantic salmon are increasing, they are low compared to other competitor countries (e.g., Chile, Faroe Islands) (Iversen et al., 2020). Climate-related risks can increase production costs, causing an increase in prices and volatility which in turn creates distress for the investors (Oglend & Sikveland, 2008).

One of the biggest critiques of the salmon farming industry is the feeding practice. In the early years of salmon aquaculture production the farmers used fish meal and fish oil to feed the salmon. Later, it became clear that

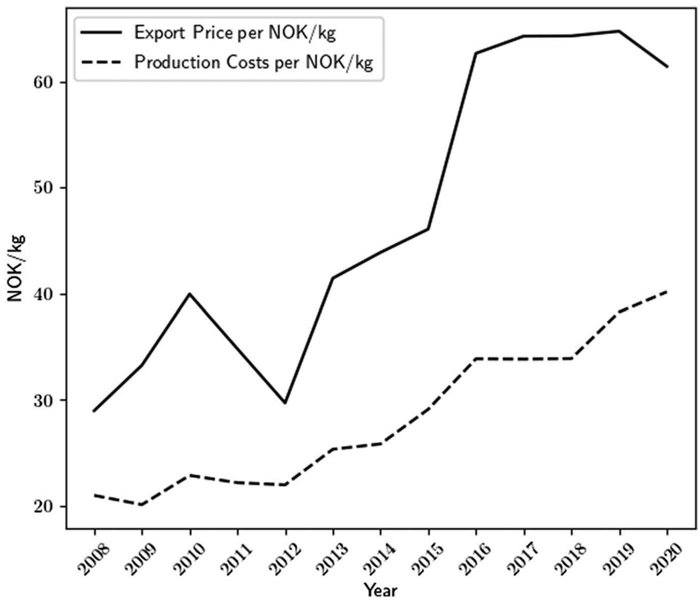


Figure 2. Salmon Export Price and Production Costs (including delivery costs) measured in NOK/kg. *Source:* Norwegian Directorate of Fisheries (2022).

this feed practice was not sustainable mainly because a lot of the fish meal and fish oil came from wild-caught fish. Because of the rapid growth of the salmon aquaculture industry, more fish meal and oil were required to feed the salmon and therefore more wild-caught fish was required which then led to over-fishing wild fish populations (Tveteras & Asche, 2008). Researchers in collaboration with salmon fish feed manufacturers substituted fish meal and fish oil with plant-based protein sources, mainly soy (Egerton et al., 2020). Therefore, most of the ingredients used for feeding salmon are now plant-based. Although this transition into plant-based feed is seen as an improvement in terms of implementing a sustainable feeding practice, it also comes with a number of environmental concerns. Given that soy is mainly grown in South America, having to grow more of it raises concerns relating to the unprecedented deforestation in the Brazilian Amazon forest (Dou et al., 2018; Sun et al., 2018). These concerns are valid, especially considering that importing countries (e.g., Norway) are gaining environmental benefits as well as maintaining production growth, while the exporting countries (e.g., Brazil) suffer environmental losses. As a response, certification programs for sustainable aquaculture have been developed with a focus on restricting the global trading of soybeans (Luthman et al., 2019). The Norwegian salmon aquaculture industry set a goal together with their soy suppliers in Brazil to become 100% deforestation and conversion-free. Their goal was achieved in February 2022 and has set an important example for other food production industries. On the other hand, the demand for growing soy has been increasing not only for feeding salmon but also for feeding humans. Solberg et al. (2021) argued that yeast produced from nonfood resources such as wood can serve as a high-quality protein source for farmed fish. They found that there is indeed potential for use in commercial production but the present costs of producing yeast from lignocellulosic biomass may still be too high, and there is a need to develop more efficient processes for economic utilization (Solberg et al., 2021). Such findings emphasize on the importance of research and innovation developments within the salmon aquaculture industry. Research and development are capital intensive and therefore the salmon aquaculture industry must be financially prepared to invest in projects that promote a more sustainable food production.

Government regulations and policies can also drive the need for investing in research and innovation (Asche & Smith, 2018). Given the binding government regulations and environmental challenges in sea-based salmon aquaculture, the rapid growth once observed has now been limited (Abate et al., 2016; Bjørndal & Tusvik, 2019). Technological developments on land-based farming have changed the potentials of the aquaculture industry as well as the cost of production. Bjørndal and Tusvik (2019) found that although land-based salmon farming is early in development, if successful, it can potentially have an important impact on the dynamics of the salmon market. Investing in land-based salmon farming projects is important to further investigate to which extent it can support the salmon aquaculture production. An additional reason for dedicating part of the salmon aquaculture research to land based farming, is that several political parties, e.g., in Denmark and North America, propose to forbid net pen farming. Such proposals aim to prevent salmon escape incidents that spread parasites and pathogens to wild stocks (Bailey & Eggereide, 2020; Olaussen, 2018). The salmon aquaculture production must be prepared to implement new developments to remain a successful food production industry. Attitudes toward innovation, levels of investment and social norms influence adoption of technological, organizational and informational practices (Lebel et al., 2021). Moreover, policies that explicitly ban or limit the adoption of new technologies could undermine aquaculture's green potential (Asche et al., 2022).

It has become clear that a "good" reputation for a food production industry is closely linked to sustainability. Being prepared to transition into a low-carbon economy is vital for being considered a successful industry. Salmon aquaculture is heavily dependent on changes in regulations and therefore it must be prepared to tackle climate-related risks or its successful reputation will be jeopardized. Assessing climate-related risks should be set as the main priority for the industry. Therefore, the question to be asked is: are the salmon aquaculture industry's underlying companies addressing potential climate-related financial impacts?

Climate-related financial disclosure

Task force on climate-related financial disclosures (TCFD)

The Financial Stability Board (FSB)[4] established the TCFD to assess climate-related risks and opportunities that assist investors, lenders, insurance writers, and other stakeholders. The role of the TCFD is to provide a standardized framework on climate-related financial disclosure so that it helps market participants understand climate-related risks. The prospect of creating a climate-related financial disclosure framework is to be singular instead of a regime, and accessible to various organizations across sectors.

The framework is based on a set of recommendations. These recommendations aim to support a company to develop a regular procedure when disclosing climate-related risks. They are based on four key features: First, the framework must be adaptable by organizations in different industries (e.g., financial services, agriculture). Second, it ought to be recorded in the files of every organization that follows the TCFD recommendations. Next, it is designed based on a “Scenario Analysis” approach. Specifically, potential hypothetical climate-related scenarios are set to happen in the future and the aim is to assess whether the underlying organization is resilient to these scenarios. Last, the framework focuses on risks and opportunities related to the transition into a low-carbon economy. The climate-related opportunities fall outside the scope of this study and therefore will not be thoroughly discussed but only mentioned.

The structure of the TCFD (2017) recommendations is based on four thematic areas, that represent the core elements of how firms operate: governance, strategy, risk management, and metrics and targets. These recommendations integrated into the financial disclosure framework, create the information that will assist investors and other stakeholders to understand how to assess climate-related risks and opportunities. Since these recommendations are applicable to organizations across sectors and jurisdictions, the TCFD (2017) provided supplemental guidance in developing these disclosures for both the financial and non-financial sectors. In this study, we will only focus on the Non-Financial sector recommendations that are relevant for the salmon aquaculture industry.

Financial impacts

The first step in assessing the potential climate-related financial impacts is to acknowledge and understand the physical and transition risks related to the salmon aquaculture industry. To derive the financial impacts associated with these risks, the TCFD (2017) identified four major climate-related financial impact categories: Revenues, Expenditures, Assets and Liabilities, and Capital and Financing. All four categories are highly relevant for the salmon aquaculture production.

It is evident from Tables 1 and 2 that there are various financial impacts that could potentially affect the financial stability of the salmon aquaculture industry. Physical risks, presented in Table 1, can mainly impact the revenues of a company, and destroy its installations which in turn will increase insurance premiums and costs. Table 2 describes in detail possible transition risks of the salmon aquaculture industry and their potential financial impacts. Policy implementations associated with transition risks can also have a major impact on the revenues of a company, increase its costs and cause changes in supply and demand. Requirements for technological developments and innovations can also impact the demand for salmon, highlight the industry’s research and development requirements and emphasize on the importance of climate adaptation to minimize costs. Climate change brings instability to the market and that can reduce the demand and the revenues. Gaining a reputation of being a non- sustainable

Table 1. Examples of physical climate-related risks and Potential Impacts concerning aquaculture.

	Climate-related risks	Potential financial impacts
Physical risks	- Storm surge	- Reduced revenue from decreased production capacity (e.g., destroy cages, high water, impacts mortality rate)
	- Rising mean water temperature	
	- Algal bloom	
	- Storms surge	- Write-offs and early retirement of existing assets (e.g., destroy installations and plants)
	- Rising sea levels	
	- Storms surge	- Increased capital costs (e.g. damage to installations, plants, land-based facilities)
	- Rising sea levels	
	- Rising mean water temperature	- Reduced revenue from lower sales/output (e.g. fish stock escaping, mortality rate, diseases)
	- Algal bloom	
	- Storm surge	
	- Rising sea levels	
	- Storm surge	- Increased insurance premiums and insurance costs (e.g. damage to facilities)
	- Rising sea levels	

food production, or a food production that is not willing to transition and adapt to a low-carbon economy, can decrease the demand for salmon which in turn can reduce a company's revenues.

The response to the climate-related risks is depending on the industry's firms' cost structure (TCFD, 2017). A firm with lower production costs is more resilient to changes that can impact its cost structure, unlike a firm with a high-cost structure. By disclosing its cost structure, a firm better informs its investors about their investment potential. With the discussion around climate-related risks strengthening, investors will start demanding disclosure of capital expenditure plans and the level of debt or equity required for funding these plans. Disclosing these plans allows investors to understand how flexible the salmon aquaculture firms to re-invest their capital, as well as how willing the capital markets are to fund firms that are significantly exposed to climate-related risks. The salmon aquaculture production is an industry that can be severely exposed to climate-related risks because it is a food production practice highly dependent on its surrounding environment. Debt and equity structures should also be disclosed as they can also be impacted from climate-related risks. A firm's ability to raise new debt or refinance existing debt is important to maintain investors' trust. It shows certain flexibility to handle climate-related issues. Operating losses, asset write-downs, and the need to raise new equity for

Table 2. Examples of Transition climate-related risks and Potential Impacts concerning aquaculture.

	Climate-related risks	Potential financial impacts
	Policy and Legal	
Transition risks	- Increase in carbon tax for air cargo	- Increased operating costs (e.g., higher transportation costs)
	- Mandates on and regulations of global trade (e.g., soybeans imports)	- Increased costs and/or reduced demand for products and services
	- Regulation of production growth	
	Technology and Innovation Developments	
	- Disease outbreak imposes research and innovation developments	- Reduced demand for products and services (e.g., disease outbreak impacts demand)
		- Research and development (R&D) expenditures in new and alternative technologies and medicine
	- Substituting soy with more sustainable options	- Capital investments in technology development (e.g., to inform about the water temperature rise)
		- Costs to adopt/deploy new practices and processes
	Market	
	- Changing customer behavior	- Reduced demand for goods and services due to a shift in consumer preferences (e.g. veganism, ASC label)
	- Uncertainty in market signals	- Reduced revenues
	- Increase in raw materials costs (e.g., soy)	- Reduced profits
	Reputation	
	- Stigmatization of sector (e.g., media scrutiny regarding animal welfare and sustainability goals	- Reduced revenue from decreased demand for goods/services
		- Reduced revenue from decreased production capacity (e.g., supply chain interruptions)
		- Reduction in capital availability (e.g., to fund the repercussions from a shift in consumers preferences

required investments, can cause changes in capital and reserves. The transparency that comes with climate-related financial disclosure is the key for the industry's firms to prove their resilience to their stockholders.

TCFD recommendations

Investors' demand for better climate-related financial disclosures has increased (Hahn et al., 2015). The number of firms across sectors that are pursuing financial disclosures upon investors' requests is on the rise (Reid & Toffel, 2009). The TCFD (2017) suggests that the fear of potential climate-related financial impacts can drive companies to conduct regular climate-related financial disclosures. Task Force's aim is to provide a standardize framework with sufficient information and instructions that will assist financial risk managers in constructing better climate-related financial disclosures. Table 3 presents the major TCFD (2017) recommendations to be followed when creating climate-related financial disclosures. These recommendations are created with the purpose to be easily adaptable across various sectors including the salmon aquaculture industry. They can be seen as suggested steps the financial risk managers of a company should follow when creating climate-related financial disclosures. The recommendations aim to provide a holistic approach for constructing climate-related financial disclosures as well as emphasize the benefits of such disclosures for investors, and other stakeholders.

Carbon Disclosure project (CDP)

The Carbon Disclosure Project (CDP) was established in 2003 as an investor-led nonprofit initiative, and it began surveying large firms regarding their carbon-related risks and strategies. The main objectives of the CDP are twofold: to inform risk managers about the concerns of investors over climate-related risks and to inform investors about the climate-related risks the firms are exposed to (Stanny & Ely, 2008). To maintain the first objective, the CDP developed a questionnaire by translating the TCFD recommendations into actual disclosure questions and a standardized annual format. The CDP provides investors with a unique platform where the TCFD framework can be brought into real-world practice. The questionnaire is distributed to a range of firms across sectors with high market value. To achieve the second objective, the CDP decided to make the responses of individual firms publicly available online and produce reports with accumulated responses.

Considering that answering the CDP questionnaire is voluntary, the response rate has been rising gradually. Several studies indicate that an increasing number of firms submit the CDP questionnaire (Lewis et al., 2014; Luo et al., 2012; Stanny & Ely, 2008; Wegener et al., 2013). The number of Global 500 firms responding increased from 221 in 2003 to 403 in 2013 (Depoers et al., 2016) with the respondent firms accumulating more than 10% of the total global emissions (EU Technical Expert Group, 2019). As a result, the CDP holds the largest database on GHG emissions in the world (Reid & Toffel, 2009). The success of the CDP is driven by the investors who support it.

The companies' decision to implement a regular practice on their reporting has enabled the CDP to influence corporate governance and climate-related financial disclosure. The CDP has obtained critical mass over the

Table 3. Task force recommendations.

Recommendations			
Governance	Strategy	Risk management	Metrics and targets
- Describe the board's oversight of climate-related risks and opportunities.	- Describe the climate-related risks (see Section "Climate-related risks conceptualization").	- How does the salmon industry identify and assess the climate-related risks?	- Disclose the metrics used to assess the climate-related risks in line with its strategy and risk management process.
- Describe management's role in assessing and managing climate-related risks.	- Describe the financial impacts of climate-related risks (see Tables 1 and 2)	- How is the industry managing these risks?	- Disclose Scope 1, Scope 2, and Scope 3[a] Greenhouse Gas (GHG) emissions.
	- Describe the resilience of the industry's strategy by considering different climate-related scenarios relevant to the aquaculture industry (i.e., increase in average water temperature, maximum sea levels rise, storms surge frequency, carbon price increase).	- How these processes are integrated into the industry's overall risk management?	- What are the targets of the industry to manage climate-related risks and performance against targets?

[a]Scope 1 emissions are equivalent to direct emissions, and indirect emissions are divided into Scope 2 and Scope 3 (WRI, 2011).
Source: TCFD (2017).

last years and it is considered the leading reporting initiative for firms worldwide. It signals a positive reputation for the firms who choose to report on the CDP platform. The CDP created a "Status Report" where they disclose the status of the reporting for each company publicly on their platform. They also developed an internal score system. It measures the comprehensiveness of disclosure, awareness and management of climate-related risks and best practices associated with environmental leadership, such as setting ambitious and meaningful targets. The details of the CDP Scores are presented on Table 4. The CDP internal scoring system is used to drive investment decisions toward a low-carbon and resilient economy. Even for low-scoring companies, choosing to report shows respect to investors' demands as well as commitment to lower their GHG emissions. The companies that do not report on the CDP platform despite investors' requests, indicate weaknesses in their governance and risk management strategies (Sullivan & Gouldson, 2012).

The main report on the CDP platform is the climate change report. This is in accordance with the TCFD (2017) recommendations. In addition to this report, the CDP added the forest and the water security reports. The

Table 4. Explanation of the CDP scores.

CDP Scores	Explanation
A and A-	Leadership level
B and B-	Management level
C and C-	Awareness level
D and D-	Disclosure level
F	Failure to provide sufficient information to be evaluates
Not requested	The company was not requested to disclose by investors or its customers
See Another	The company's data has been covered by its parent company's response
Not available	The score is private to the company and any requesting customers the response has been submitted to
Forthcoming	The score has not yet been released

Source: Carbon Disclosure Project (2022).

forest report is targeting firms that are inclined to measure and manage forest-related risks. Forest-related risks are relevant for organizations and industries, whose practices can cause deforestation and forest degradation. The main objective of this report is for companies to show their commitment to restore the forests. There are four forest-related reports: Cattle Products, Palm Oil, Soy, and Timber. Palm Oil and Soy are the main relevant ones for the salmon aquaculture industry and the ones we will focus on in this study. The aim of the water security report is to disclose whether firms are doing enough to tackle water pollution. The Carbon Disclosure Project (2017) reported that many firms underestimate the risks related to water pollution with only 28% of the disclosing firms acknowledging any water-related risks in their practice. Climate-related financial disclosure allows stakeholders to access more information on strategies that mitigate climate change as well as assist companies to improve the disclosure of their actions that contribute to the reduction of GHG emissions (de Faria et al., 2018).

Climate-related financial disclosures for the salmon industry

The physical and transition climate-related risks presented in Tables 1 and 2 are specific to the salmon aquaculture industry and can be seen as guidance for creating climate-related financial disclosures for individual companies. In this study, we discuss industry specific climate-related financial disclosures without focusing on individual companies. This is because the TCFD recommendations (see Table 3) are constructed based on sectors (e.g. Agricultural, Food and Forest Products Group in the case of salmon producers) and not for individual companies. However, the climate-related financial disclosures are to be submitted by individual companies. By reporting in a common framework, i.e. the CDP's industry-specific questionnaire, the differences in sustainability between the companies can be captured by the internal scoring system of the CDP (see Table 4).

Assessing the impacts of climate-related risks for firms in the salmon aquaculture industry involves a number of interactions and tradeoffs among the climate-related aspects of chemical and biological pollution, disease outbreaks, unsustainable feeds and competition for coastal space (Carballeira Brana et al., 2021), complicated by maintaining production sufficient to meet the rising demand for blue food (Naylor et al., 2021). Disclosures in the salmon aquaculture industry should focus on qualitative and quantitative information related to both the industry's policy and market risks in the areas of GHG emissions, as well as the industry's opportunities around increasing food. The salmon-producing companies should provide evidence of their efforts to reduce GHG emissions, pollution, disease outbreaks, high fish mortality rates and unsustainable feeds. Also they must provide information on how they improve sustainability through adequate environmental monitoring, location of farms, reduction and exploitation of wastes as well as how the chemicals are being used to ensure the growth and continuity of salmon aquaculture production. The companies must also disclose the impacts that physical and transition risks had on salmon production. They should also report opportunities that capture shifts in both business and consumer trends toward consuming salmon, as well as processes and services that produce lower emissions, are less chemical-intensive, and use sustainable feeding while maintaining adequate food security.

Results

In this study, we utilized the public information available on the CDP platform. We collected publicly available data on all three types of CDP reports: climate change, forest, and water security, for seven Norwegian salmon aquaculture companies with the highest market cap. The analysis is in line with the content and converge of the firms' responses to the CDP questionnaires.

It is evident from Figure 3 that all seven firms have been requested to disclose their climate-change reports from 2016 to 2021. This could be because physical and transition risks of the salmon aquaculture industry have been receiving increasing attention from investors. Their demands for climate-related financial disclosures and transparency have increased. From the results demonstrated in Figure 3, the firms are not obliged to respond to the investors' requests and hence some firms "Decline to participate." In 2016, 3 out of 7 firms chose not to publicly disclose their climate change

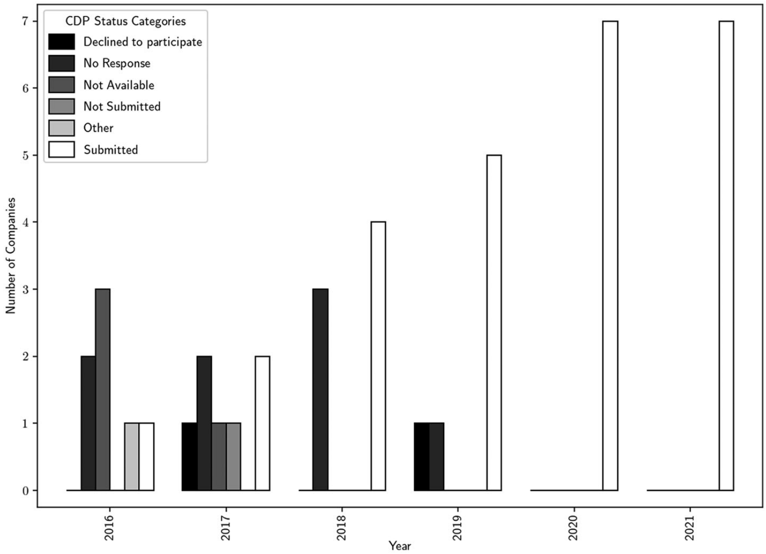


Figure 3. Salmon Aquaculture Firms CDP Status Results for the Climate Change report from 2016 to 2021. *Source:* Carbon Disclosure Project (2022).

reports but only make them available for the investors who have requested them (see Table 4 for explanation). Over the years the number of companies submitting their climate change reports on the CDP platform has increased. In fact, in 2020 and 2021 all seven companies have submitted their climate change reports. This can be because discussions about sustainability and sustainable food practices increased, raising awareness for consumers, investors, and other stakeholders. Therefore, it is important for the salmon aquaculture companies to publish their climate-related financial disclosures, because it emphasizes their willingness to respond to investors' demands, tackle climate-related risks, and lower their GHG emissions.

Figure 4 presents the corresponding *CDP Scores* of the companies that submitted the CDP climate change report from 2016 to 2021. Given that number of companies submitting their climate change reports have increased over the period from 2016 to 2021 (see Figure 3), the corresponding CDP scores have on average also become better. The first A was granted in 2019 which is an improvement from 2018 where the average score was C. In 2020, when all the companies submitted their reports, they on average managed to seal an "Awareness Level" about the requirements of lowering GHG emissions, with two of them taking this awareness up to the "Management level," and the last two up to the "Leadership level" (see Table 4). All the companies made their CDP scores publicly available in

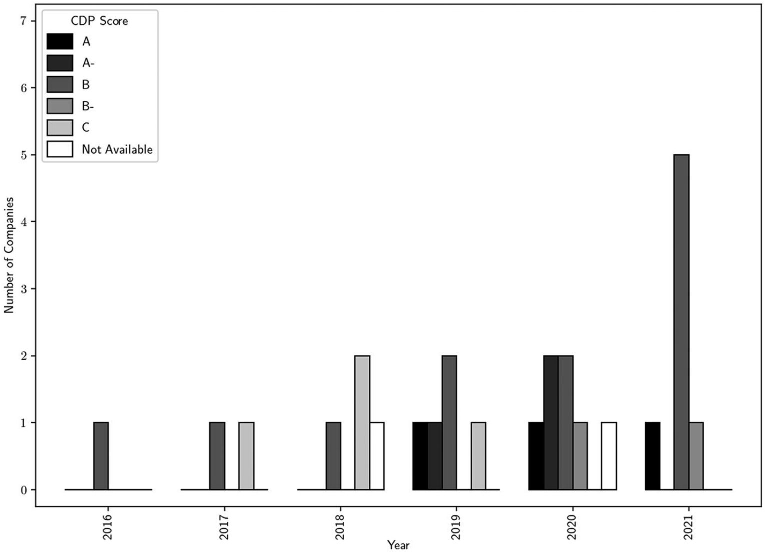


Figure 4. Salmon Aquaculture Firms CDP Scores from 2016 to 2021. *Source:* Carbon Disclosure Project (2022).

Table 5. Salmon Aquaculture Firms CDP Status on Forest Report from 2016 to 2021.

		Firms		
Year	Number of Firms requested to participate	No response	Submitted	Declined to Participate
2016	1	1	–	–
2017	3	3	–	–
2018	5	4	1	–
2019	7	4	1	2
2020	7	4	3	–
2021	6	3	3	–

Source: Carbon Disclosure Project (2022).

2021 with B as the average score. It is the first year that all companies submitted the climate change report and also maintained a sufficient score. These results are encouraging for the future of the salmon aquaculture industry. They can be seen as proof that the increased demands from investors to submit climate-related financial disclosures, lower GHG emissions, and adapt more sustainable practices, have a positive impact on the industry's companies.

In Table 5 we see that from 2016 to 2021, institutional investors have increasingly requested the salmon aquaculture firms to disclose the CDP forest report. The number of companies submitting the report has also increased from 2016 to 2021. In 2020 when all seven companies were requested to submit the forest report, three of them responded. In 2021

Table 6. Salmon aquaculture firms CDP results on forest report.

		Scores	
Year	Firm	Palm Oil	Soy
2018	1	Not Scored	**A-**
2019	1	Not Scored	Not Scored
2020	1	Not Scored	Not Scored
	2	Not Scored	Not Scored
	3	**A-**	**A-**
2021	1	Not Scored	**B**
	2	Not Scored	**A-**
	3	**B**	**B**

Source: Carbon Disclosure Project (2022).

Table 7. Salmon aquaculture firms CDP results on water security report.

		Firms		
Year	Number of Firms requested to participate	No response	Submitted	Declined to participate
2016	–	–	–	–
2017	–	–	–	–
2018	1	1	–	–
2019	2	1	–	1
2020	3	2	1	–
2021	6	4	1	1

Source: Carbon Disclosure Project (2022).

also three companies submitted the forest report, but six were requested to do so. Table 6 presents the scores for the two types of forest-related reports we focus in this study: Palm Oil and Soy. It is evident from the results that the Soy forest report is the one mostly reported since it is quite relevant for the salmon aquaculture industry. For the Palm Oil report we notice that most companies responded after the deadline and therefore have the status of "Not Scored." In 2021, all three companies which submitted their forest reports, were given an average score of B. The link between the salmon aquaculture's feeding practice and soy production is what mainly drives investors to request disclosure of forest-related risks.

Table 7 presents the number of companies which were requested to disclose the CDP Water Security Report and their responses. It is evident that the water security report is not requested as much as the climate and forest reports from investors. However, it appears that in 2021, 6 out of 7 companies were requested to disclose the CDP water security report with only one of them submitting it. This shows that the relevance of the water security report is increasing for the salmon aquaculture companies and the investors appear to believe that it does. The responses of the companies in the upcoming years will clarify investors view on the industry's responsibility on water pollution.

Discussion

Physical and transition climate-related risks have been receiving increasing attention from investors over the last years. While the impacts of climate change worsen, investors' expectations will continue to increase. A fast-growing food production industry such as salmon aquaculture has a responsibility to identify and tackle climate-related risks to protect its growth, secure transparency for its investors and other stakeholders, and smoothly transition into a low-carbon economy. To secure these, organizations such as the TCFD and CDP have been founded to encourage climate-related financial disclosures.

We discussed the potential climate-related risks, physical and transition, that are likely to financially impact the salmon aquaculture companies. Policies and regulations play a significant role in maintaining and supporting the sustainability of the industry but are not always successful. We argue that the policy makers should implement a holistic approach by consulting the salmon producers when constructing policies and regulations. Companies on the other hand, need to be prepared for the possibility of new regulation concerning climate-related risks. For instance, the growing discussions on climate-related financial disclosures have resulted in some countries, such as the United Kingdom (UK),[5] to implement a new legislation that will require some companies to disclose climate-related financial information. This, or other regulatory changes could be implemented in Norway too, so Norwegian salmon producers must be prepared to make the necessary adjustments.

The results from the publicly available CDP reports are compatible with the increased investors' demands for transparent and accurate climate-related financial disclosure. The number of firms disclosing their climate, forest, and water-security CDP reports has sharply increased between 2016 and 2021. The most important, and largely reported one is the climate report. We found that in 2020 and 2021 all seven companies submitted the CDP climate report, and were granted a B score on average. This shows that the salmon aquaculture companies are working to tackle their carbon and climate risks, while responding to the investors' requests. Although this cannot necessarily be interpreted as good financial performance for the companies, it is a step forward toward not only transparency, but also climate change awareness and willingness to do better in terms of their own environmental impacts.

The second most important report is the CDP forest report. This report is divided into four forest-related risks, of which we only focus on the ones relevant to the salmon aquaculture industry: Palm Oil and Soy. We found that in 2021, out of six companies that were requested to submit the report, three of them submitted and were given a B score on average for the soy forest report. This was the highest number of companies reporting on their soy forest report within the predefined time requirements of the CDP. Soy is the main source of feeding for the salmon aquaculture production and it has been accused of causing deforestation. It is also a plant-based raw material that can be consumed by humans. This makes the choice of soy an unsustainable feeding practice and causes major criticism for the salmon aquaculture industry. We argue that this attracts investors' attention and increases their demands for disclosure and as long as it remains controversial, the companies have an ethical responsibility to be transparent about it.

The third report is the CDP water security report that aims to decouple growth from depletion of water resources and capitalize a water secure economy. While this report was introduced back in 2010, it was only in 2018 that the first salmon aquaculture company was requested to submit it. The number of salmon aquaculture companies requested to report increased since 2018 with the number significantly rising in 2021 when six out of seven companies were asked to report. However, only one company submitted. This indicates that the salmon aquaculture industry does not consider that its practices cause water pollution. Considering that 6 out of 7 companies were requested by their investors to submit the water-security report, it is likely that more companies will submit this report in the future. This expectation stems from the fact that disclosing climate- and forest-related risks improved substantially upon investors' demands.

The results from this study are relevant for salmon aquaculture companies, investors and other market participants, as well as policy makers. As climate change is increasingly discussed in politics, business, media and academia, investors have also become more aware of the issue. We verify that a large and growing number of investors are increasingly expecting salmon producers to respond to their demands and adapt to climate change. Therefore, they impose pressure on the companies to disclose the climate-related risks and financial impacts. The salmon-producing companies have a responsibility toward their investors, society and the environment to report on their climate-related risks and environmental impacts. The establishment of the CDP and the TCFD played a significant role in promoting climate-related financial disclosures. They increased awareness within the firms and thus lead financial risk managers to develop dynamic strategies to tackle climate-related financial impacts and integrate these strategies into the risk management process of the company. As a result, firms set more ambitious goals to become more competent in disclosing climate-related risks, as well as maintain their climate resilience and lower environmental footprints.

Notes

1. An independent, international organization established in 1997 that helps businesses take responsibility for their climate-related impacts, by providing them with the standards for sustainable reporting.

2. An independent, international non-profit organization established in 2010 that promotes certified responsibly farmed seafood and aims to reduce the environmental impacts of aquaculture.
3. The production costs are per kg produced salmon and rainbow trout because companies produce both species and it is not possible for the Norwegian Directorate of Fisheries to separate the costs.
4. The Financial Stability Board (FSB) is an international body that monitors and makes recommendations about the global financial system.
5. https://www.gov.uk/government.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Mikaella Zitti http://orcid.org/0000-0002-7918-9613

References

Abate, T. G., Nielsen, R., & Tveterås, R. (2016). Stringency of environmental regulation and aquaculture growth: A cross-country analysis. *Aquaculture Economics & Management*, *20*(2), 201–221. https://doi.org/10.1080/13657305.2016.1156191

Abolofia, J., Asche, F., & Wilen, J. E. (2017). The cost of lice: Quantifying the impacts of parasitic sea lice on farmed salmon. *Marine Resource Economics*, *32*(3), 329–349. https://doi.org/10.1086/691981

Asche, F. (2008). Farming the sea. *Marine Resource Economics*, *23*(4), 527–547. https://doi.org/10.1086/mre.23.4.42629678

Asche, F., Anderson, J. L., Botta, R., Kumar, G., Abrahamsen, E. B., Nguyen, L. T., & Valderrama, D. (2021). The economics of shrimp disease. *Journal of Invertebrate Pathology*, *186*, 107397. https://doi.org/10.1016/j.jip.2020.107397

Asche, F., Eggert, H., Oglend, A., Roheim, C. A., & Smith, M. D. (2022). Aquaculture: Externalities and policy options. *Review of Environmental Economics and Policy*, *16*(2), 282–305. https://doi.org/10.1086/721055

Asche, F., Guttormsen, A. G., & Nielsen, R. (2013). Future challenges for the maturing Norwegian salmon aquaculture industry: An analysis of total factor productivity change from 1996 to 2008. *Aquaculture*, *396–399*, 43–50. https://doi.org/10.1016/j.aquaculture.2013.02.015

Asche, F., Guttormsen, A. G., & Tveterås, R. (1999). Environmental problems, productivity and innovations in Norwegian salmon aquaculture. *Aquaculture Economics & Management*, *3*(1), 19–29. https://doi.org/10.1080/13657309909380230

Asche, F., Misund, B., & Oglend, A. (2019). The case and cause of salmon price volatility. *Marine Resource Economics*, *34*(1), 23–38. https://doi.org/10.1086/701195

Asche, F., Oglend, A., & Selland Kleppe, T. (2017). Price dynamics in biological production processes exposed to environmental shocks. *American Journal of Agricultural Economics*, *99*(5), 1246–1264. https://doi.org/10.1093/ajae/aax048

Asche, F., Roll, K. H., Sandvold, H. N., Sørvig, A., & Zhang, D. (2013). Salmon aquaculture: Larger companies and increased production. *Aquaculture Economics & Management*, *17*(3), 322–339. https://doi.org/10.1080/13657305.2013.812156

Asche, F., Roll, K. H., & Tveteras, R. (2009). Economic inefficiency and environmental impact: An application to aquaculture production. *Journal of Environmental Economics and Management*, *58*(1), 93–105. https://doi.org/10.1016/j.jeem.2008.10.003

Asche, F., Sikveland, M., & Zhang, D. (2018). Profitability in Norwegian salmon farming: The impact of firm size and price variability. *Aquaculture Economics & Management*, *22*(3), 306–317. https://doi.org/10.1080/13657305.2018.1385659

Asche, F., & Smith, M. D. (2018). Induced innovation in fisheries and aquaculture. *Food Policy*. *76*, 1–7. https://doi.org/10.1016/j.foodpol.2018.02.002

Bailey, J. L., & Eggereide, S. S. (2020). Indicating sustainable salmon farming: The case of the new norwegian aquaculture management scheme. *Marine Policy*, *117*, 103925. https://doi.org/10.1016/j.marpol.2020.103925

Battiston, S., Mandel, A., Monasterolo, I., Schütze, F., & Visentin, G. (2017). A climate stress-test of the financial system. *Nature Climate Change*, *7*(4), 283–288. https://doi.org/10.1038/nclimate3255

Bergesen, O., & Tveterås, R. (2019). Innovation in seafood value chains: The case of Norway. *Aquaculture Economics & Management*, *23*(3), 292–320. https://doi.org/10.1080/13657305.2019.1632391

Bjørndal, T., & Tusvik, A. (2019). Economic analysis of land based farming of salmon. *Aquaculture Economics & Management*, *23*(4), 449–475. https://doi.org/10.1080/13657305.2019.1654558

Bovari, E., Giraud, G., & Mc Isaac, F. (2018). Coping with collapse: A stock-flow consistent monetary macrodynamics of global warming. *Ecological Economics*, *147*, 383–398. https://doi.org/10.1016/j.ecolecon.2018.01.034

Carballeira Brana, C. B., Cerbule, K., Senff, P., & Stolz, I. K. (2021). Towards environmental sustainability in marine finfish aquaculture. *Frontiers in Marine Science*, *343*, 62. https://doi.org/10.3389/fmars.2021.666662

Carbon Disclosure Project. (2017). *Harnessing the power of purchasing for a sustainable future*. Technical Report.

Carbon Disclosure Project. (2022). *Scores*. https://www.cdp.net/en/scores

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of Management*, *37*(1), 39–67. https://doi.org/10.1177/0149206310388419

Dafermos, Y., Nikolaidi, M., & Galanis, G. (2017). A stock-flow-fund ecological macroeconomic model. *Ecological Economics*, *131*, 191–207. https://doi.org/10.1016/j.ecolecon.2016.08.013

Dafermos, Y., Nikolaidi, M., & Galanis, G. (2018). Climate change, financial stability and monetary policy. *Ecological Economics*, *152*, 219–234. https://doi.org/10.1016/j.ecolecon.2018.05.011

De Silva, S. S., & Soto, D. (2009). Climate change and aquaculture: Potential impacts, adaptation and mitigation. *Climate change implications for fisheries and aquaculture: Overview of current scientific knowledge. FAO Fisheries and Aquaculture Technical Paper*, *530*, 151–212.

Deegan, C., Rankin, M., & Tobin, J. (2002). An examination of the corporate social and environmental disclosures of bhp from 1983–1997: A test of legitimacy theory. *Accounting, Auditing & Accountability Journal*, *15*(3), 312–343. https://doi.org/10.1108/09513570210435861

Depoers, F., Jeanjean, T., & Jérôme, T. (2016). Voluntary disclosure of greenhouse gas emissions: Contrasting the carbon disclosure project and corporate reports. *Journal of Business Ethics*, *134*(3), 445–461. https://doi.org/10.1007/s10551-014-2432-0

Dietz, S., Bowen, A., Dixon, C., & Gradwell, P. (2016). 'Climate value at risk' of global financial assets. *Nature Climate Change*, *6*(7), 676–679. https://doi.org/10.1038/nclimate2972

Dou, Y., Silva, R. F. B. d., Yang, H., & Liu, J. (2018). Spillover effect offsets the conservation effort in the amazon. *Journal of Geographical Sciences*, *28*(11), 1715–1732. https://doi.org/10.1007/s11442-018-1539-0

Edwards, M., Johns, D., Leterme, S., Svendsen, E., & Richardson, A. (2006). Regional climate change and harmful algal blooms in the Northeast Atlantic. *Limnology and Oceanography*, *51*(2), 820–829. https://doi.org/10.4319/lo.2006.51.2.0820

Egerton, S., Wan, A., Murphy, K., Collins, F., Ahern, G., Sugrue, I., Busca, K., Egan, F., Muller, N., Whooley, J., McGinnity, P., Culloty, S., Ross, R. P., & Stanton, C. (2020). Replacing fishmeal with plant protein in Atlantic salmon (salmo salar) diets by supplementation with fish protein hydrolysate. *Scientific Reports*, *10*(1), 1–16. https://doi.org/10.1038/s41598-020-60325-7

EU Technical Expert Group. (2019). *Report on Climate-related disclosures.*

Ewald, C.-O., Haugom, E., Kanthan, L., Lien, G., Salehi, P., & Størdal, S. (2022). Salmon futures and the fish pool market in the context of the capm and a three-factor model. *Aquaculture Economics & Management*, *26*(2), 171–191. https://doi.org/10.1080/13657305.2021.1958105

FAO. (2016). The state of world fisheries and aquaculture 2016. *Contributing to food security and nutrition for all, 200.* FAO.

FAO. (2020). *The state of world fisheries and aquaculture*. FAO.

Faria, J. A. d., Andrade, J. C. S., & Silva Gomes, S. M. d (2018). The determinants mostly disclosed by companies that are members of the carbon disclosure project. *Mitigation and Adaptation Strategies for Global Change*, *23*(7), 995–1018. https://doi.org/10.1007/s11027-018-9785-0

Fischer, C., Guttormsen, A. G., & Smith, M. D. (2017). Disease risk and market structure in salmon aquaculture. *Water Economics and Policy*, *03*(02), 1650015. https://doi.org/10.1142/S2382624X16500156

Hahn, R., Reimsbach, D., & Schiemann, F. (2015). Organizations, climate change, and transparency: Reviewing the literature on carbon disclosure. *Organization & Environment*, *28*(1), 80–102. https://doi.org/10.1177/1086026615575542

Hersoug, B., Mikkelsen, E., & Karlsen, K. M. (2019). "Great expectations" – Allocating licenses with special requirements in Norwegian salmon farming. *Marine Policy*, *100*, 152–162. https://doi.org/10.1016/j.marpol.2018.11.019

IPCC. (2022). *Climate Change 2022: Impacts, Adaptation, and Vulnerability*. Cambridge University Press.

Iversen, A., Asche, F., Hermansen, Ø., & Nystøyl, R. (2020). Production cost and competitiveness in major salmon farming countries 2003–2018. *Aquaculture*, *522*, 735089. https://doi.org/10.1016/j.aquaculture.2020.735089

Jansen, P. A., Kristoffersen, A. B., Viljugrein, H., Jimenez, D., Aldrin, M., & Stien, A. (2012). (1737). Sea lice as a density-dependent constraint to salmonid farming. *Proceedings of the Royal Society B*, *279*(1737), 2330–2338. https://doi.org/10.1098/rspb.2012.0084

Kumar, G., & Engle, C. R. (2016). Technological advances that led to growth of shrimp, salmon, and tilapia farming. *Reviews in Fisheries Science & Aquaculture*, *24*(2), 136–152. https://doi.org/10.1080/23308249.2015.1112357

Leaton, J. (2011). Unburnable carbon–Are the world's financial markets carrying a carbon bubble? *Carbon Tracker Initiative*, 6–7. http://www.carbontracker.org/wp-content/uploads/2014/09/Unburnable-Carbon-Full-rev2-1.pdf (Investor Watch, 2011).

Lebel, L., Jutagate, T., Thanh Phuong, N., Akester, M. J., Rangsiwiwat, A., Lebel, P., Phousavanh, P., Navy, H., Navy, H., Soe, K. M., & Lebel, P. (2021). Climate risk management practices of fish and shrimp farmers in the Mekong region. *Aquaculture Economics & Management*, *25*(4), 1–23. https://doi.org/10.1080/13657305.2021.1917727

Lewis, B. W., Walls, J. L., & Dowell, G. W. (2014). Difference in degrees: CEO characteristics and firm environmental disclosure. *Strategic Management Journal*, *35*(5), 712–722. https://doi.org/10.1002/smj.2127

Luo, L., Lan, Y.-C., & Tang, Q. (2012). Corporate incentives to disclose carbon information: Evidence from the CDP Global 500 report. *Journal of International Financial Management & Accounting*, *23*(2), 93–120. https://doi.org/10.1111/j.1467-646X.2012.01055.x

Luthman, O., Jonell, M., & Troell, M. (2019). Governing the salmon farming industry: Comparison between national regulations and the ASC salmon standard. *Marine Policy*, *106*, 103534. https://doi.org/10.1016/j.marpol.2019.103534

Misund, B. (2017). Financial ratios and prediction on corporate bankruptcy in the Atlantic salmon industry. *Aquaculture Economics & Management*, *21*(2), 241–260. https://doi.org/10.1080/13657305.2016.1180646

Naylor, R. L., Kishore, A., Sumaila, U. R., Issifu, I., Hunter, B. P., Belton, B., Bush, S. R., Cao, L., Gelcich, S., Gephart, J. A., Golden, C. D., Jonell, M., Koehn, J. Z., Little, D. C., Thilsted, S. H., Tigchelaar, M., & Crona, B. (2021). Blue food demand across geographic and temporal scales. *Nature Communications*, *12*(1), 1–14. https://doi.org/10.1038/s41467-021-25516-4

Norwegian Directorate of Fisheries. (2022). *Statistikk fra akvakultur [Aquaculture statistics]*. https://www.fiskeridir.no/English/Aquaculture/Statistics

Norwegian Seafood Council. (2022). *Open reports [Åpne rapporter]*. https://seafood.no/markedsinnsikt/apne-rapporter/

Nygård, R. (2020). Trends in environmental CSR at the oslo seafood index: A market value approach. *Aquaculture Economics & Management*, *24*(2), 194–211. https://doi.org/10.1080/13657305.2019.1708996

Oglend, A., & Sikveland, M. (2008). The behaviour of salmon price volatility. *Marine Resource Economics*, *23*(4), 507–526. https://doi.org/10.1086/mre.23.4.42629677

Olaussen, J. O. (2018). Environmental problems and regulation in the aquaculture industry. insights from norway. *Marine Policy*, *98*, 158–163. https://doi.org/10.1016/j.marpol.2018.08.005

Osmundsen, T. C., Almklov, P., & Tveterås, R. (2017). Fish farmers and regulators coping with the wickedness of aquaculture. *Aquaculture Economics & Management*, *21*(1), 163–183. https://doi.org/10.1080/13657305.2017.1262476

Parmar, B. L., Freeman, R. E., Harrison, J. S., Wicks, A. C., Purnell, L., & De Colle, S. (2010). Stakeholder theory: The state of the art. *Academy of Management Annals*, *4*(1), 403–445. https://doi.org/10.5465/19416520.2010.495581

Patten, D. M. (2002). The relation between environmental performance and environmental disclosure: A research note. *Accounting, Organizations and Society*, *27*(8), 763–773. https://doi.org/10.1016/S0361-3682(02)00028-4

Pincinato, R. B. M., Asche, F., & Roll, K. H. (2021). Escapees in salmon aquaculture: A multi-output approach. *Land Economics*, *97*(2), 425–435. https://doi.org/10.3368/le.97.2.425

Reid, E. M., & Toffel, M. W. (2009). Responding to public and private politics: Corporate disclosure of climate change strategies. *Strategic Management Journal*, *30*(11), 1157–1178. https://doi.org/10.1002/smj.796

Sikveland, M., Tveterås, R., & Zhang, D. (2021). Profitability differences between public and private firms: The case of Norwegian salmon aquaculture. *Aquaculture Economics & Management*, *2021*, 1–25. https://doi.org/10.1080/13657305.2021.1970856

Sikveland, M., & Zhang, D. (2020). Determinants of capital structure in the norwegian salmon aquaculture industry. *Marine Policy*, *119*, 104061. https://doi.org/10.1016/j.marpol.2020.104061

Smith, M. D., Roheim, C. A., Crowder, L. B., Halpern, B. S., Turnipseed, M., Anderson, J. L., Asche, F., Bourillón, L., Guttormsen, A. G., Khan, A., Liguori, L. A., McNevin, A., O'Connor, M. I., Squires, D., Tyedmers, P., Brownstein, C., Carden, K., Klinger, D. H., Sagarin, R., & Selkoe, K. A. (2010). Sustainability and global seafood. *Science*, *327*(5967), 784–786.

Solberg, B., Moiseyev, A., Hansen, J. Ø., Horn, S. J., & Øverland, M. (2021). Wood for food: Economic impacts of sustainable use of forest biomass for salmon feed production in Norway. *Forest Policy and Economics*, *122*, 102337. https://doi.org/10.1016/j.forpol.2020.102337

Stanny, E., & Ely, K. (2008). Corporate environmental disclosures about the effects of climate change. *Corporate Social Responsibility and Environmental Management*, *15*(6), 338–348. https://doi.org/10.1002/csr.175

Stolbova, V., Monasterolo, I., & Battiston, S. (2018). A financial macro-network approach to climate policy evaluation. *Ecological Economics*, *149*, 239–253. https://doi.org/10.1016/j.ecolecon.2018.03.013

Straume, H.-M. (2017). Here today, gone tomorrow: The duration of Norwegian salmon exports. *Aquaculture Economics & Management*, *21*(1), 88–104. https://doi.org/10.1080/13657305.2017.1262477

Sullivan, R., & Gouldson, A. (2012). Does voluntary carbon reporting meet investors' needs? *Journal of Cleaner Production*, *36*, 60–67. https://doi.org/10.1016/j.jclepro.2012.02.020

Sun, J., Mooney, H., Wu, W., Tang, H., Tong, Y., Xu, Z., Huang, B., Cheng, Y., Yang, X., Wei, D., Zhang, F., & Liu, J. (2018). Importing food damages domestic environment: Evidence from global soybean trade. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(21), 5415–5419. https://doi.org/10.1073/pnas.1718153115

TCFD. (2017). Recommendations of the task force on climate-related financial disclosures.

Torrissen, O., Jones, S., Asche, F., Guttormsen, A., Skilbrei, O. T., Nilsen, F., Horsberg, T. E., & Jackson, D. (2013). Salmon lice–impact on wild salmonids and salmon aquaculture. *Journal of Fish Diseases*, *36*(3), 171–194. https://doi.org/10.1111/jfd.12061

Tveteras, S., & Asche, F. (2008). International fish trade and exchange rates: An application to the trade with salmon and fishmeal. *Applied Economics*, *40*(13), 1745–1755. https://doi.org/10.1080/00036840600905134

Wegener, M., Elayan, F. A., Felton, S., & Li, J. (2013). Factors influencing corporate environmental disclosures. *Accounting Perspectives*, *12*(1), 53–73. https://doi.org/10.1111/1911-3838.12007

Wri, W. (2011). *Greenhouse Gas Protocol Corporate Value Chain (Scope 3) accounting and reporting standard*. World Resources Institute and World Business Council for Sustainable Development.

Research Article II

Forecasting Salmon Market Volatility using Long Short-term Memory (LSTM)

Mikaella Zitti[1]

[1]Department of Economics and Resource Management, Norwegian University of Life Sciences, Ås, Norway.

Abstract

Forecasting salmon market volatility is crucial for reducing future uncertainty for market participants. This study explores the efficacy of the Long Short-term Memory (LSTM) network, a deep learning technique, in forecasting multi-step ahead salmon market volatility. The performance of the LSTM is assessed against a constructed volatility proxy and the Autoregressive Moving Average (ARMA) model, a traditional benchmark in time-series analysis. Evaluation is performed across various forecasting horizons using different forecast error measures. Our findings indicate that the ARMA model outperforms the LSTM in predicting salmon market volatility, suggesting that any non-linear patterns in the salmon market volatility might be too insignificant for an LSTM model to exploit effectively. However, we observed a significant discrepancy between the actual volatility values and the forecasts obtained by both models, indicating the complexity of accurately predicting salmon market volatility.

Keywords— Salmon Market; Volatility Forecasting; Volatility Models; GARCH; LSTM; Neural Network;

1 Introduction

Salmon is a volatile commodity, with the increased demand over recent decades resulting in high salmon prices and increased volatility, especially since the mid-2000s (Asche et al., 2019; Bloznelis, 2016; Guttormsen, 1999; Oglend, 2013). This high price volatility presents challenges for all market participants, including farmers, processors, traders, and intermediaries. Farmers, who control production, can adjust harvesting to maximize their profits or to meet biomass limitations (Asche, 2008; Forsberg & Guttormsen, 2006; Guttormsen, 2008), causing fluctuations in the salmon spot price and, in turn, increased volatility. Processors face thin profit margins and customers demanding lower prices, operating based on their expectations of the future salmon spot price (Bergfjord, 2007; Kvaløy & Tveterås, 2008).

There are hedging opportunities available in the salmon market to mitigate this volatility, however, they remain thin due to a lack of speculative traders (Andersen & de Lange, 2021; Asche et al., 2016; Ewald et al., 2022). Fish Pool, a futures exchange of salmon, was established in 2006 to provide these hedging opportunities. However, its role has been controversial among aquaculture economists, with some arguing that its launch increased salmon price volatility (Bloznelis, 2016). Further studies found that shorter futures contracts contribute to more volatility (Ankamah-Yeboah et al., 2017), futures prices are efficient in the long-run but not in the short-run (Andersen & de Lange, 2021), and Fish Pool futures can be considered as a hedging instrument but not an investment asset (Ewald et al., 2022). In support of Fish Pool, other researchers found that the contract settlement price used is representative of salmon transaction prices (Oglend & Straume, 2019), and stock prices reflect salmon price information earlier than the Fish Pool Index (Dahl et al., 2021). Despite these debates, the salmon futures market is characterized by low liquidity, with infrequent trades that account for less than 10

Given these challenges, this study aims to fill a critical gap in the literature by implementing a volatility forecasting model that could significantly aid salmon market participants and provide valuable insights for both the academic field and market participants. We will examine whether further exploration of neural networks for forecasting salmon market volatility is necessary or if reliance on traditional time-series forecasting models suffices. Despite several studies analyzing salmon price volatility, there has been a limited effort towards forecasting this volatility (Asche et al., 2015; Asche et al., 2019; Asche & Oglend, 2016; Bloznelis, 2016; Dahl & Oglend, 2014; Dahl & Jonsson, 2018; Oglend, 2013; Oglend & Sikveland, 2008; Solibakke, 2012).

Deep learning techniques such as neural networks have shown great potential in forecasting financial data, including commodity prices (Hamid & Iqbal, 2004; Manogna & Mishra, 2021; Verma, 2021; Xu & Zhang, 2021, 2022). Among these techniques, Recurrent Neural Networks (RNNs) have been found particularly suitable for predicting financial market volatility due to their ability to learn temporal dependencies of time-series data (Selvin et al., 2017). Specifically, Long Short-Term Memory (LSTM) networks, a type of RNN capable of capturing long-term dependencies, have yielded promising results in various forecasting tasks (Kim & Won, 2018; Nelson, 1991). Despite these promising results, the application of LSTM networks in forecasting salmon market volatility remains unexplored.

In this study, we implement LSTM networks to forecast salmon market volatility, filling both a gap in the literature and a practical need for robust forecasting models in the salmon industry. The approach we undertake involves creating a volatility proxy based on the standard deviation of the logarithmic returns rolling over 4 weeks. This proxy, although a mere estimation of volatility, is forecasted using the ARMA model, with the forecasting ability of the LSTM assessed against this. Each model's forecasting performance is evaluated under different, multi-step ahead forecast horizons using various forecast error measures, with the model yielding the lowest errors deemed the best performing one. The expected value of the forecast losses generated by each model is compared using the Diebold-Mariano (DM) test to assess robustness, and the forecasts obtained by LSTM are examined for significant difference against actual volatility values.

Our study is structured as follows: Section 2 discusses the specifications of neural networks and the LSTM model. Section 3 demonstrates seasonality and structural changes in the salmon spot price series. Section 4 outlines the statistical metrics and tests used to evaluate the forecasting performance of the LSTM model compared to a benchmark model. Section 5 describes the application of the proposed models and Section 6 reports their results. Finally, Section 7 concludes and discusses potential future research directions.

2 Methodology

2.1 Neural Networks (NN)

NNs are linear and polynomial methods that connect a set of input variables $\{x_t^i\}$ with $i = 1, .., n$, n being the number of inputs connected to an output $\{\tilde{y}_t\}$. They consist of three different types of layers. The input, the hidden, and the output layers. The *input layer* includes the input nodes, and each node represents a different variable. When applied on univariate time-series data, a lagged version of the data is used, and the nodes correspond to each lag. The *output layer* is usually formed with one node which represents the output of the NN. The *hidden layer(s)*, $\{h_t^i\}$ where $i = 1, ..., m$ is the number of nodes that separate the input from the output layer and define the amount of complexity the model is capable of fitting. The number of hidden layers and the number of nodes in each layer are based on the complexity of the model under study and a trial-and-error approach. An illustration of a type of NN is presented in Figure 1.

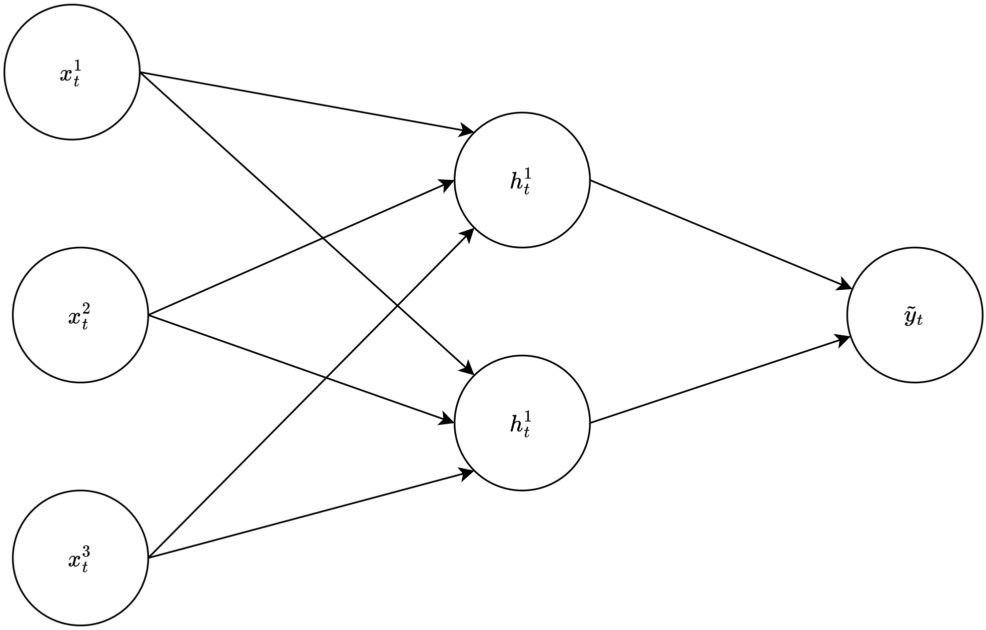


Figure 1: A feed-forward Artificial Neural Network (ANN) with three inputs and one hidden layer with two hidden nodes.

2.1.1 Recurrent Neural Network (RNN)

RNN is also known as the Elman network (Elman, 1990), a class of ANNs that incorporates a recurrent hidden state, that consists of hidden layers, whose activation after each iteration depends on previous states and the current input. The RNN's feature is that it incorporates a *'memory'* mechanism that is saving a copy of the previous values of the layer containing the recurrent nodes and using them as an additional input for the next step (Makridakis et al., 2018). The weights of an RNN determines how much significance to give to the present input and the past hidden state. The weights are adjusted via backpropagation until the error function (SSE) is minimised. The network's 'memory' feature allows them to exhibit dynamic temporal behaviour. The illustration in Figure 2 portrays this procedure.

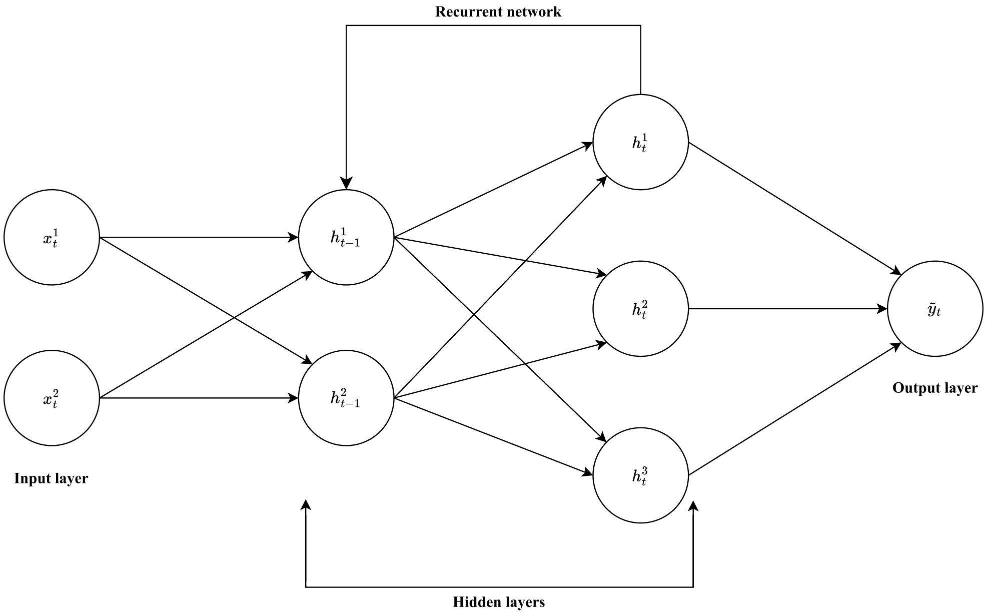


Figure 2: A RNN with one input layer that consists of two neurons, two hidden layers consisting of two and three neurons, respectively, and an output layer.

2.1.2 Long Short-term Memory(LSTM)

The training process of RNNs suffers from the *vanishing gradient problems* (Bengio et al., 1994). RNNs are only able to store short-term information and they have difficulties carrying information for longer periods. Hochreiter and Schmidhuber (1997) developed the Long Short-term Memory (LSTM) as a solution. It is an advanced type of recurrent neural network and is applied in a number of different areas (e.g. handwriting recognition, speech recognition, see (Graves et al., 2013). LSTM is a network architecture that in combination with an appropriate gradient-based algorithm can use memory cells and gates to store information for long periods.

Gers et al. (2000) specify that the cell state is the core of the LSTM model because the cell state represents the 'memory' feature. Simpler, the cell state behaves like a 'transport line' that captures and stores information. The information passing through the cell state is filtered by the gates. The gates have the power to add or remove information to and from the cell state.

There are three gates: the forget gate, the input gate, and the output gate. The *forget gate* takes as inputs information carried from the last hidden state, h_{t-1}, and the current input, x_t. It passes these inputs via a sigmoid function that returns values between zero and one, where zero means "nothing goes through" and one "everything goes through". The output of the first gate f_t is:

$$f_t = \sigma[W_f(h_{t-1}, x_t) + b_f]. \tag{1}$$

The *input gate* decides which information to store in the cell state. This gate has two parts. First, it also receives inputs h_{t-1} and x_t and passes them via a sigmoid function as follows:

$$i_t = \sigma[W_i(h_{t-1}, x_t) + b_i]. \tag{2}$$

Next, it passes the same inputs via a hyperbolic tangent function that creates a vector of new information:

$$\tilde{C}_t = tanh[W_c(h_{t-1}, x_t) + b_c]. \tag{3}$$

The two outputs, i_t and $\tilde{C}_t$ are combined and added to the cell state. The current period cell state is created using the outputs from the first and second gates as follows:

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t, \tag{4}$$

The forget gate's output, f_t, is multiplied with the information carried from the previous period cell state, C_{t-1}. Then, the product from the input gate, $i_t\tilde{C}_t$, is added and the new cell state, C_t is created.

The *output gate* also receives inputs h_{t-1} and x_t and passes them via a sigmoid activation. The output, o_t, is denoted as follows:

$$o_t = \sigma[W_o(h_{t-1}, x_t) + b_o]. \tag{5}$$

In the meantime, the cell state, C_t, passes through a hyperbolic tangent function and the output is multiplied with o_t, to decide which information the new hidden state, h_t, should carry to the next period. This is described as follows:

$$h_t = o_t tanh[C_t]. \tag{6}$$

The structure of the LSTM layer is shown in Figure 3. It is evident from the figure that the three sigmoid functions, σ, and the hyperbolic tangent function, $tanh$, are controlling the three gates. Scalar multiplication is denoted as $\times$ and addition is denoted as $+$. Overall, the LSTM updates the cell state, from C_{t-1} to C_t, filters important and non-important information via the three gates, and generates h_t.

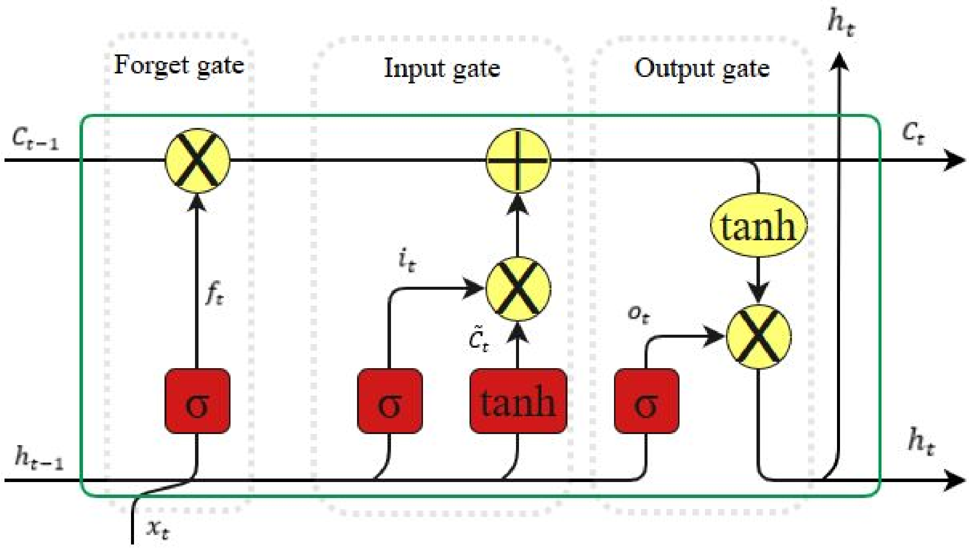


Figure 3: Structure of a long short-term memory (LSTM) layer with a forget gate as introduced by Gers et al. (2000).

3 Data

Our analysis employs weekly salmon spot prices sourced from the NASDAQ Salmon Index[1]. This data extends from week 27 of 2007 until week 51 of 2019. The prices included represent averages for all weight classes and are given in Norwegian Krone (NOK) per kilogram (Kg).

The decision to use spot prices as opposed to futures contracts prices was informed by the relatively low liquidity and infrequent trades of the salmon futures market (Bergfjord, 2007; Bloznelis, 2018a). Further supporting this choice, Asche et al. (2016) discovered that innovations in the spot price impact futures prices, suggesting that futures prices do not provide a price discovery function. This is an expected finding considering the immaturity of the salmon futures

[1]https://salmonprice.nasdaqomxtrader.com.

market and provides a compelling rationale for choosing the salmon spot prices over the futures prices in our volatility forecasting study.

An important note on the choice of our dataset timeframe is necessary here as we have intentionally chosen not to include data from 2020 onwards. The primary reason for this decision is the onset of the COVID-19 pandemic, which would introduce a structural breakpoint in the data and significantly disrupted market dynamics. The pandemic represents an outlier event that, while undeniably impactful, may not be indicative of typical market behavior. Including data from this period could skew our model and undermine the reliability of the forecasts by potentially overfitting to the exceptional conditions brought on by the pandemic. Consequently, limiting our analysis to pre-2020 data allows us to avoid this issue and provide forecasts that are grounded in more standard market conditions, therefore offering a more unbiased and reliable prediction of typical salmon market volatility.

3.1 Seasonality

Before obtaining the logarithmic return series, we considered the seasonal patterns in salmon spot prices. As a result of factors related to supply and demand a key characteristic of salmon production is seasonality. Seasonality in supply does not match the seasonality in demand and that generates seasonal patterns in salmon price. Modelling seasonality when having weekly time-series is complicated. In the existing literature, the most common technique is the Fourier series; that is, sums of trigonometric functions (Bloznelis, 2016, 2018b; Oglend, 2013). Here, we follow a technique introduced by Hyndman (2014), and also applied by Bloznelis (2018b), that uses a regression with ARMA errors, having Fourier terms as regressors. The number of Fourier terms could be up to 26 pairs for weekly data. However, the number of Fourier terms for the fitted model was selected by minimising the Akaike information criterion (AIC) and choosing between none to 26 pairs, while the same applies for selecting the order of the ARMA model (Hyndman & Athanasopoulos, 2018).

The salmon spot price exhibits large movements around Christmas and Easter as shown by Bloznelis (2018b). Therefore, we also incorporate the possibility of deterministic seasonality occurring, by adding four dummy variables to specify the weeks before and after Christmas (including Christmas week) and four more to specify the weeks before and after Easter (including Easter week). The deterministic seasonality is expressed by the means of these eight seasonal dummy variables. The combination of the Fourier terms and these eight seasonal dummy variables exhibit the seasonal component, which is subtracted from the data before proceeding. The seasonally adjusted version of the salmon spot price will be used in place of the original series without explicit reference. The development of the seasonally adjusted salmon spot price series is depicted in Figure 5.

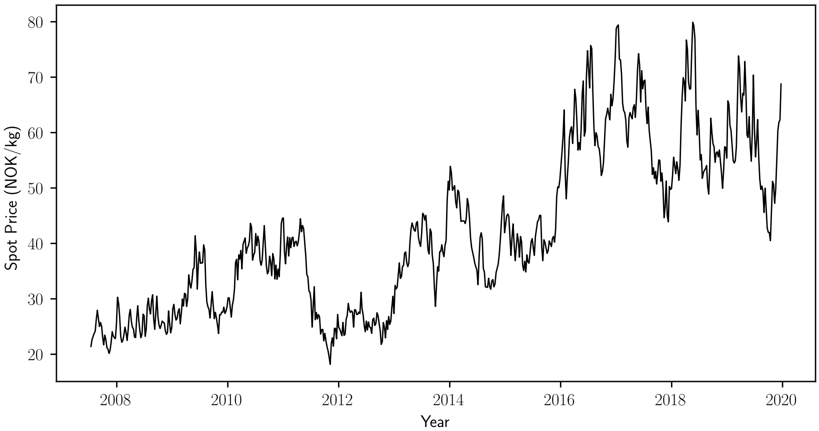


Figure 4: Original time-series of weekly salmon spot prices in NOK/kg.

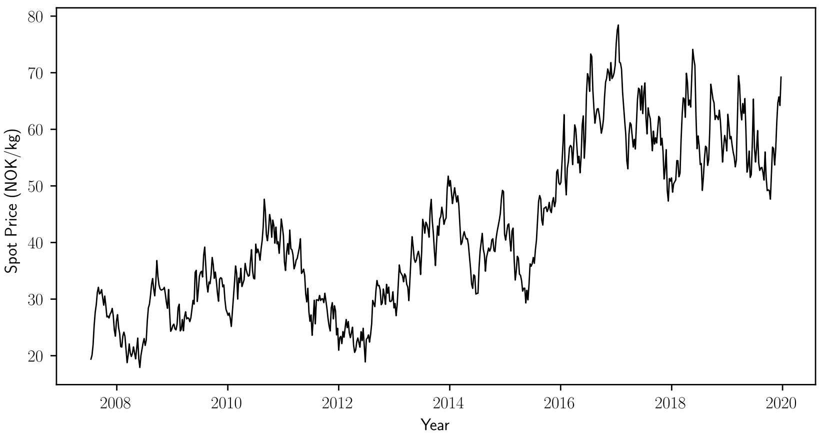


Figure 5: Seasonally adjusted time-series of weekly salmon spot prices in NOK/kg.

Given that our aim is to forecast salmon market volatility, the parametric elimination of seasonality prior to forecasting proves unfeasible. This stems from the fact that, based on the information available at the time of forecasting, we can't predict the precise timings and occurrences of future seasonal patterns, making the removal of seasonality to use the seasonally adjusted series for forecasting inappropriate. Ideally, seasonality should be considered only until the point of initiating forecasting, specifically, it should be controlled only within the in-sample (training) data and not extended to the out-of-sample (testing) data. Nevertheless, upon noting the minimal influence of the seasonal component (refer to Appendix A.1), we found the original series has a standard deviation of 14.703, while the seasonally adjusted series indicates a 14.339 standard deviation. Since we determined the seasonal component to be minimal, we opted to disregard the seasonal adjustment and conduct our volatility analysis and forecasting on the original, unadjusted series. These minor differences of the two series can be observed in Figures 4 and 5.

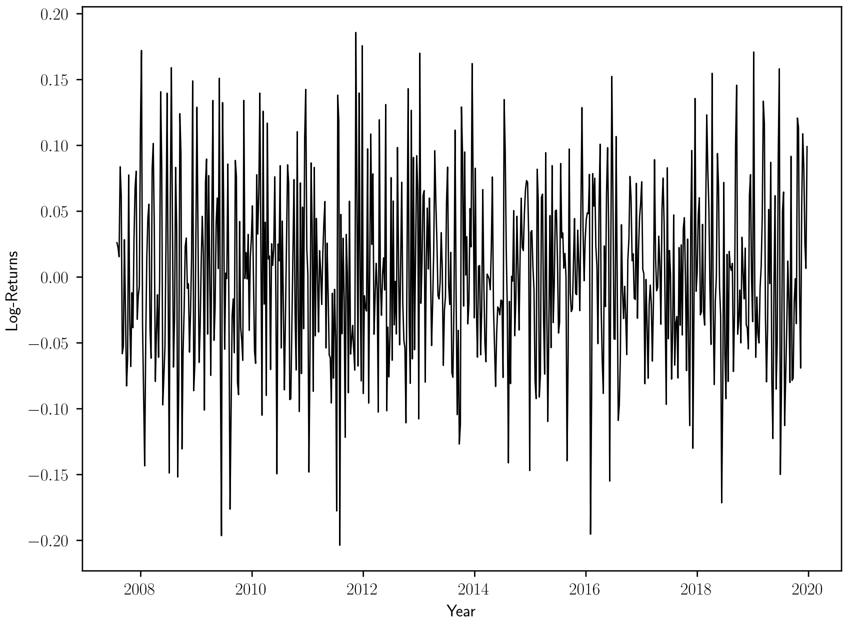


Figure 6: Salmon spot price logarithmic returns.

Moreover, we apply a logarithmic transformation on the original spot price series as it is often used in volatility analysis and forecasting. The return from week to week is denoted as $Y = W_t/W_{t-1}$, where W_t is the spot price at time t (or current week's price) and W_{t-1} is the spot price at time $t-1$ (or previous week's price). To consider for proportional changes in the returns, we apply a logarithmic transformation of the first price difference $r_t = ln(W_t/W_{t-1})$. Figure 6 presents the development of the logarithmic returns.

3.2 Structural Changes

We choose to analyse the sample period starting from 2007 week 27 based on the findings of Bloznelis (2016), who estimated the salmon spot price volatility of different salmon weight classes, and found a structural break-point in the logarithmic returns series from 1996 week 1 until 2005 week 45, and from 2007 week 27 until 2013 week 13. From Figure 6 there is a noticeable change in the variability of the logarithmic returns from middle 2012 and onward. Therefore, to test whether the variability of the logarithmic returns is different before middle 2012, we split the sample into two periods; one from 2007 week 27 until 2012 week 17 and one from 2012 week 18 until 2019 week 51. We use an $F-$test to examine whether the two sub-samples have equal variance. The $F-$test assumes that the two sub-samples are independent of each other and hence independent across time. For simplicity of the analysis we ignore any potential time dependence at this point, assuming that it is not strong enough to invalidate the results. The second main assumption of the $F-$test is that the two sub-samples must be normally distributed. We test for normality of each of the two sub-samples using the Shapiro-Wilk normality test and we find that they are likely normally distributed, hence we are not violating the normality assumption of the $F-$test. The results strongly reject that the two sub-samples have equal variances with p-values well below 0.01. However, as we are interested in forecasting using an in-sample (train sample) and a hold-out sample (test sample), forecasting using two sub-samples, before and after the break-point, is not feasible. Therefore, for forecasting salmon returns volatility we use the sample from 2012 week 18 until 2019 week 51. This is the main sample which we refer to throughout the remainder of the article.

3.3 Descriptive Statistics

Table 1 shows descriptive statistics such as mean, standard deviation, skewness, and kurtosis of the salmon spot prices and their logarithmic returns, as well as the results of the augmented Dickey-Fuller (ADF) test, a unit-root test, and the Shapiro-Wilk test, a normality test. The ADF test was used to confirm the stability of the time-series (Dickey & Fuller, 1979). For the ADF test, a negative value, whose absolute value exceeds the critical value, suggests stationarity. From the ADF statistic in table 1, we interpret that the spot price series contains a unit-root at all significant levels, e.g., 1%, 5%, 10%, implying it is non-stationary, while the logarithmic returns series does not contain a unit root, thus, it is stationary. Figure 6, illustrates the logarithmic returns, where the stationarity of the series can be observed.

	Mean	Standard Deviation	Min	Max	Skewness	Kurtosis	Shapiro-Wilk	ADF
spot prices	49.23	13.634	21.78	79.91	-0.14	-0.78	0.98***	-2.15
log returns	0.002	0.064	-0.12	0.171	0.091	-0.074	0.996	-17.5***
Note: ***, **, * denote significance at 10%, 5%, and 1% level, respectively.								

Table 1: Descriptive statistics on salmon spot price and logarithmic return series, 2012 - 2019.

The logarithmic returns are centered at zero with standard deviation 6.4%. As confirmed by the Shapiro-Wilk test result of 0.996, the logarithmic returns closely align with a normal distribution. Although commodity prices, including salmon, often exhibit asymmetry and non-normal distributions, our data reveals different characteristics. The skewness of the returns is slightly above zero, hinting that large positive returns might be marginally more common than large negative ones. However, this near-zero value suggests a symmetric distribution. The balance between the largest negative return (12%) and the largest positive return (17.1%) corroborates this symmetry, thereby indicating normality in the returns distribution.

Moreover, the kurtosis indicator is less than zero, a sign of platykurtosis. This implies that the returns are less densely populated in the tails and more concentrated around the mean than one would expect from a normal distribution. Given its relative proximity to zero, we interpret this platykurtosis as insignificant, reinforcing our assumption of normality. In the context of the LSTM and ARIMA models, both of which are being applied to forecast the volatility of the salmon market in this study, the absence of significant asymmetry and non-normal distribution in our data offers a less complex environment for these models to capture the underlying patterns. As such, the characteristics of our data provide a more straightforward foundation for evaluating the effectiveness of LSTM and ARIMA in predicting salmon market volatility.

4 Measurement and Model Assessment

4.1 Volatility Measure

For a measure of volatility, which serves as the target value for the supervised learning process in the neural network (refer to Section 5.2), we utilize the sample standard deviation of logarithmic returns, computed over 4-week intervals using a rolling approach[2].

The variance calculation incorporates the mean of returns from the same rolling period. Consequently, the formula for calculating the variance proxy is presented as follows:

$$V_t = \frac{1}{T}\sum_{j=0}^{T-1}(r_{t-j} - \bar{r}_t)^2, \tag{7}$$

where r_j represents the logarithmic return at time j, and $\bar{r}_t$ signifies the average of the logarithmic returns from time t to $t+T$. This ensures that the mean return value utilized in the variance calculation is derived from the same period for which the variance is computed. The volatility is estimated using a rolling-window approach, reducing the length of the logarithmic returns to $(k-4)$, where k represents the length of the logarithmic return series.

Even though implied volatility would serve as an ideal forward-looking measure, the absence of a robust market for options contracts on spot prices at Fish Pool requires the employment of historical volatility measures. Despite this constraint, we are confident that the volatility proxy introduced in Eq.7 provides a reliable measure for our analysis.

4.2 Benchmark Model: ARMA

To provide a meaningful comparison for the performance of the neural network in forecasting the volatility measure presented in Eq. 7, it's crucial to employ a benchmark model. Based on insightful suggestions from a reviewer, we've chosen to incorporate the ARMA model as this benchmark.

The ARMA(p,q) process that generates the volatility series $\{V_t\}_{t=1}^{\alpha}$ is formulated as follows:

$$\mathrm{V}_t = \mu + \phi_1(V_{t-1}-\mu) + \phi_2(V_{t-2}-\mu) + ... + \phi_p(V_{t-p}-\mu) + \epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} - ... - \theta_q\epsilon_{t-q}, \tag{8}$$

[2]Take note that the data frequency is weekly.

In this formulation, V_t is the actual value at time t, and ϵ_t is the random error at the same time point. The parameter μ represents the intercept, and $\phi_i\,(i = 1, 2, ..., p)$ and $\theta_j\,(j = 1, 2, ..., q)$ are the model parameters. The quantity p denotes the number of autoregressive terms, and q denotes the number of random error terms, also known as moving average terms.

The selection of the lag order values p and q is done using the Akaike Information Criterion (AIC), a well-known method for model selection. Maximum allowable lag orders are specified as $(p_{max},\ q_{max})$, and the optimal p and q are those for which the AIC is minimized.

Estimation of the ARMA model parameters is performed via the method of maximum likelihood, which aims to find the parameters that make the observed data most probable.

This ARMA benchmark model will allow us to conduct a rigorous and fair comparison of the forecasting capabilities of the neural network, thereby highlighting the strengths and potential areas of improvement in our approach.

The ARMA model is chosen for its simplicity, interpretability, and the flexibility it offers in modelling various types of temporal dependencies, making it a robust choice for a benchmark model.

4.3 Model Assessment

To evaluate the accuracy of the models' forecasting performance, we employ four statistical error measures. These include the popular mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). These measures are commonly used in forecasting studies and provide a comprehensive evaluation of the models' performance. The error measures are defined as follows:

$$MSE = \frac{1}{N}\sum_{t=1}^{N}(x_t - V_t)^2, \tag{9}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(x_t - V_t)^2}, \tag{10}$$

$$MAE = \frac{1}{N}\sum_{t=1}^{N}\left|x_t - V_t\right|, \tag{11}$$

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|1 - \frac{x_t}{V_t}\right| \times 100\% \tag{12}$$

where V_t is the volatility value, and x_t is the predicted volatility value. These measures provide a holistic assessment of the prediction accuracy, taking into account both the magnitude of the errors (through the MSE and RMSE) and their relative size compared to the actual values (through the MAE and MAPE).

4.4 Diebold-Mariano Test

The performance of any pair of forecasts can be measured utilising the DM test (Diebold & Mariano, 2002). The null hypothesis of the test is that the expected loss due to forecast errors is equal for both forecast models, implying that both underlying models are equally accurate. The forecast errors from the two different forecasts are transformed into corresponding losses using selected loss functions. If the expected value of the loss function from each forecast is equal, the population mean of the loss differential series should be equal to zero.

For the DM test, we produce two series of volatility forecasts, $\hat{\sigma}_{i,N}, ..., \hat{\sigma}_{i,N}$ where $i = 1, 2$ from two different forecasting models. Next, we evaluate the accuracy of these forecasts against the series of volatility proxies denoted as $\sigma_1, ..., \sigma_N$ with loss function $L(\hat{\sigma}_{i,N}, \sigma_N) = (\hat{\sigma}_{i,N} - \sigma_N)^2$. Considering that Laurent and Violante (2012) showed that loss functions such as MAE are not suitable for comparing volatility models, we choose MSE and RMSE as the loss functions for the DM test. The null hypothesis of a DM test is denoted as: $E(d_t) = 0$, where $d_t = L(\hat{\sigma}_{1,t}, \sigma_t) - L(\hat{\sigma}_{2,t}, \sigma_t)$ is a loss differential sequence with a given loss function $L(\bullet)$. The DM statistic is denoted as follows:

$$DM = \frac{\frac{1}{N}\sum_{t=1}^{N} d_t}{\sqrt{2\pi \hat{f}_d(0)/N}}, \tag{13}$$

where N is the sample size, and $\hat{f}_d(0)$ is a consistent estimate of $f_d(0)$, which stands for the spectral density of the loss differential at frequency 0.

5 Proposed Models

As discussed in the introduction, the literature available for salmon price volatility mainly utilises autoregressive financial time-series models. Deep learning techniques have not been applied before in the context of salmon price volatility. A number of academic studies investigating financial market volatility have integrated feed-forward neural network techniques and found significant evidence that they strengthen volatility predictions (Kristjanpoller & Minutolo, 2016;

Roh, 2007; Tseng et al., 2008). We acknowledge these findings and we aim at exercising their validity for the salmon market. To do so, we assess the forecasting performance of each model individually and examine which one (if any) is able to accurately forecast salmon spot price volatility over a given forecast horizon.

5.1 Implementation of ARMA Model

Initially, we calculate a proxy for volatility using a 4-week measure of volatility, as defined in Eq. 7. Subsequently, we employ an ARMA model to forecast the series of this volatility proxy. Before setting up the ARMA model, we test whether the volatility proxy series contains a unit root using the ADF test. The results indicate that the series doesn't contain a unit root, suggesting it's likely stationary and therefore employing an ARMA Model is indeed feasible. The model parameters are determined by the Maximum Likelihood (ML) estimator, and their selection is guided by the Akaike Information Criterion (AIC), with the model yielding the lowest AIC being deemed the optimal fit (Akaike, 1969). Our model selection process identifies an Autoregressive model of order 5, or AR(5).

The selected AR(5) model is further validated by diagnosing the behavior of the residual series. The Ljung-Box test is used to confirm the linear independence of residuals across time, while the Shapiro-Wilk normality test establishes that the residuals follow a normal distribution.

In terms of forecasting, the volatility proxy series is divided into a training set, including 80% of the data, and a testing set containing the remaining 20%. The AR(5) model is fitted on the training data, and walk-forward validation over the testing set is used to produce multi-step ahead forecasts. These forecasts cover 1-step, 4-steps, 8-steps, and 12-steps ahead, as defined by our forecasting horizon.

Finally, to evaluate the performance of the AR(5) model, we calculate forecast error measures. These are then compared to the volatility forecasts generated by the LSTM, which serves as our benchmark model for forecasting.

5.2 LSTM Experiment

The initial step in our methodology involves establishing the LSTM network, where we employ the volatility proxy series as the input variable. The series' stationarity is confirmed through the application of the ADF test, resulting in evidence against the presence of a unit root.

The LSTM deep learning method employed in this study is characterized by *supervised learning*, an automatic search process for superior representations. We transform the input series into a supervised learning format via a lag transformation, whereby the value at time $(t-k)$ - where k signifies the number of lags - represents the input variable. The input variable from time $(t-k)$ to time t is then fed through an LSTM layer, pushed forward via one fully connected dense layer, and utilized to forecast volatility at $(t+n)$, where n represents the different forecasting horizons. The optimal number of lags was determined through the application of the Autocorrelation Function (ACF). Given that the volatility series is constructed based on a 4-week rolling standard deviation, it was not unexpected to observe autocorrelation up to lag 4 (see Appendix A.2 for more details). However, in order to fully leverage the long-term memory characteristic inherent in the LSTM model, we opted for a more extended lag length. Specifically, we set the timesteps[3] to 52, corresponding roughly to one year's worth of data. This decision allows the LSTM to capture and utilize long-term temporal dependencies, a feature that is central to its design and operation.

Prior to training the network, both the input and output variables are scaled to a range between -1 and 1 to enhance the training process. The data series is then split into training and testing samples, with the same 80%-20% split as the benchmark model.

We utilize the Random Search tuning technique for the optimization of the network's hyperparameters, with candidate hyperparameters presented in Table 2. The random search technique functions by sampling randomly from the specified pool of hyperparameters. Rather than undergoing exhaustive training and evaluation with each sampled set, the model is trained for a restricted number of iterations (epochs) based on these sampled hyperparameters. Optimal hyperparameters are then determined according to the results of these limited iterations. Random Search method is preferred due to its effectiveness in hyperparameter tuning, particularly in scenarios with limited computational resources or time.

The selection range for the hyperparameters has been carefully chosen based on their potential impact on the LSTM model. The units in the LSTM layer represent the dimensionality of the output space, and the selection range provides the model with enough flexibility to capture complex patterns in the data. The dropout layer helps in preventing overfitting by ignoring

[3]The context of an LSTM, "timesteps" refer to the sequence length, a parameter that dictates how many steps in time we allow the model to look back while learning. This is essentially the model's "memory". A longer sequence means that the model can learn from past patterns spanning a more extended period, thus improving its capability to understand long-term dependencies.

randomly selected neurons during training, and the activation function determines the output of a neuron given an input. The learning rate influences how much the model changes in response to the estimated error each time model weights are updated, and the epsilon parameter aids in maintaining numerical stability.

As a result of the Random Search technique, the neuron specification of the input LSTM layer is set to 30, the dropout value is adjusted to 0.4, and the activation function is determined to be the hyperbolic tan function. To further improve model convergence during training, we use kernel, recurrent, and bias regularizers. Regularization is a method used to avoid overfitting by adding a penalty term to the loss function. The penalty term corresponds to large weights in the model, forcing the model weights to be small, and therefore simpler. This enhances generalization and model performance on unseen data. The dense layer was specified with neurons equal to the forecasting horizon, which leads to a varying number of neurons based on the forecasting horizon $n = 1, 4, 8, 12$. Changing the number of neurons of the fully-connected layer can impact the convergence of the neural network (see Appendix. A.3).

Hyperparameters	Selection range
Units LSTM layer	$\{10,\ 20,\ 80,\ \dots\ ,\ 10i\}$ where $i = 12$
Dropout layer	$\{0,\ 0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5\}$
Activation function	{hyperbolic tan, ReLu, sigmoid, softmax}
Learning rate	$\{0.01,\ 0.001,\ 0.0001\}$
epsilon	$\{0.00001,\ 0.0001,\ 0.001,\ 0.01,\ 0.1,\ 1\}$

Table 2: Candidates for the Hyperparameters to be tuned.

We then define a loss function and an optimization algorithm. The “mean squared error” and “ADAM” are used as the loss function and the optimization algorithm, respectively. The “ADAM” optimization method is chosen due to its efficient performance as a stochastic optimization algorithm (Kingma & Ba, 2014). The hyperparameters for “ADAM” optimization algorithm are tuned with the help of the Random Search tuning technique. This technique sets the learning rate at 0.0001. The learning rate determines how much the model changes in response to the estimated error each time the model weight updates. The epsilon hyperparameter, which is used for numerical stability in the “ADAM” optimizer, is also tuned during this process. Numerical stability is crucial in preventing potential divisions by zero during the optimization process, hence tuning epsilon contributes to the robustness and stability of the learning process.

The model is trained using 200 epochs, where the training sample is passed through the single network to update the weights and to develop a more precise prediction model. During the training process, we employ an Early Stopping strategy, which monitors a designated metric on the validation data and halts the training procedure once the performance stops improving. Specifically, we observe the validation loss and set a patience level equal to 20 to control the number of epochs with no improvement after which training will be stopped. The necessary number of epochs for training the network can vary based on the characteristics and behavior of the underlying data series.

6 Empirical Results

The main focus of this study has been an investigation into the predictive capability of LSTM model regarding salmon market volatility. We have benchmarked these against the traditional ARMA model, employing various error measures to critically assess the performance of each. This section discusses the findings, emphasizing the utility of LSTM networks for volatility forecasting in the context of the salmon market.

Table 3 presents the out-of-sample multi-step ahead volatility forecasts, evaluated using MSE, MAE, RMSE, and MAPE. It is evident that the ARIMA model outperforms the LSTM model across all forecasting horizons in terms of the four metrics, showing lower errors for all. Specifically, when looking at the 1-step-ahead forecasts, it is noticeable that the ARIMA model performs better than the LSTM model in terms of all four metrics. Even though the errors have increased for both models when examining rhe 4-steps ahead forecasts, the ARIMA model remains superior to the LSTM. The results for the 8-steps-ahead and 12-steps-ahead forecasts also show a similar pattern. Despite an increase in errors for both models in these longer forecasting horizons, the ARIMA model consistently outperforms the LSTM. This indicates that ARIMA might be a better option for forecasting short-term salmon market volatility, potentially due to its simplicity and efficiency in situations with less complex data patterns.

To further examine these results, we analyzed the percentage changes in forecasting error measures when comparing the LSTM model against the benchmark ARMA model. These changes are presented in Figure 7.

Evidently, the most significant disparities between the two models are observed in 1-step

	1-step-ahead		4-steps-ahead		8-steps-ahead		12-steps-ahead	
Metric	ARIMA	LSTM	ARIMA	LSTM	ARIMA	LSTM	ARIMA	LSTM
MSE	0.00029	0.00049	0.00068	0.0008	0.00065	0.00072	0.00072	0.00073
MAE	0.01275	0.018377	0.02178	0.02396	0.02126	0.02243	0.02234	0.02249
RMSE	0.01699	0.02219	0.02613	0.02821	0.02551	0.02699	0.02683	0.02698
MAPE	37.5839	49.1988	55.0841	61.2407	53.1498	56.1876	55.1837	56.1174

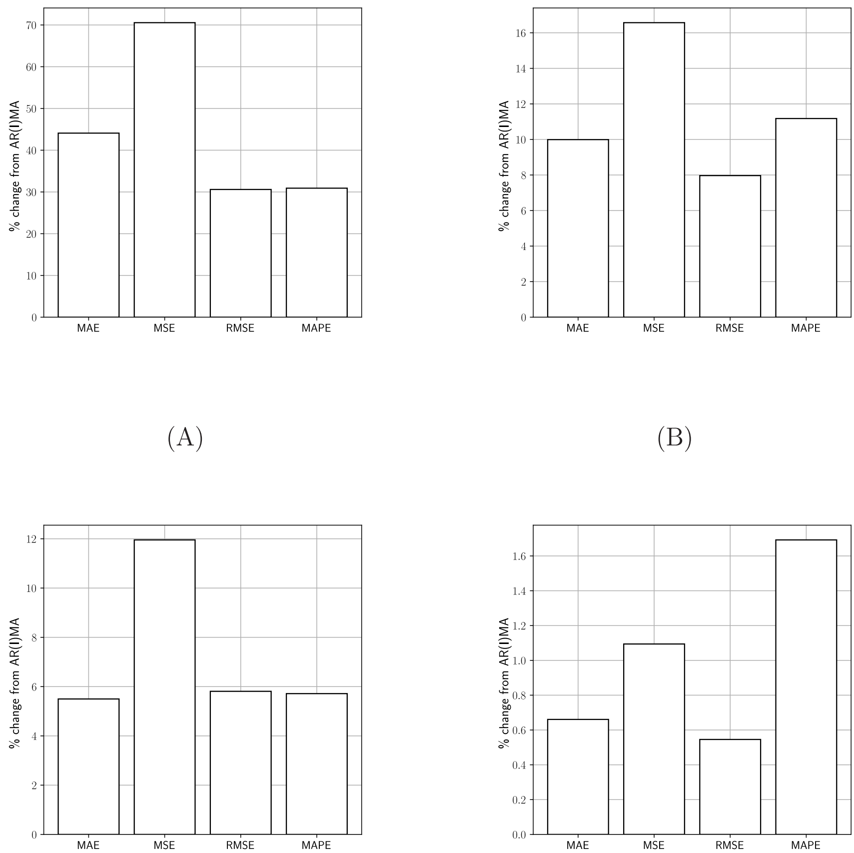
Table 3: Results of the out-of-sample multi-step ahead forecasts with the MSE, MAE, RMSE, and MAPE loss functions.

ahead forecasting, where the LSTM model reports a 70% larger MSE score compared to the ARMA model. Although the ARMA model consistently outperforms the LSTM across all forecast horizons, we notice a diminishing divergence between the two models as the forecast horizon increases. For instance, when forecasting 12-steps ahead, the percentage difference in error measures between the two models is minimal.

As the forecast horizon extends to 12 steps, the differences in MSE, MAE, and RMSE between the two models diminish, indicating that the two models start to converge in their predictive performance. However, the MAPE value shows the largest relative discrepancy between the models at this horizon. This behavior could be attributed to the MAPE metric's sensitivity to situations where the actual observations are close to zero. Given that we're predicting market volatility, which inherently involves predicting changes that can be close to zero, it's likely that longer-term forecasts—which are likely to involve greater uncertainty and more instances of small changes—would result in more pronounced relative differences in MAPE between the LSTM and ARMA models.

In Figure 8, we illustrate the development of the forecast error metrics across various forecasting horizons for each metric separately. It is evident that the overarching trends across all error metrics exhibit similarity. We observe a sharp surge in forecast errors when moving from a 1-step to 4-step ahead forecast for all metrics. However, all metrics reveal a subtle decrease in forecast errors when progressing from a 4-step to 8-step forecast horizon, only to rise again when forecasting 12 steps ahead.

The slight dip in forecast errors for the 8-step horizon, as compared to the 4-step and 12-step



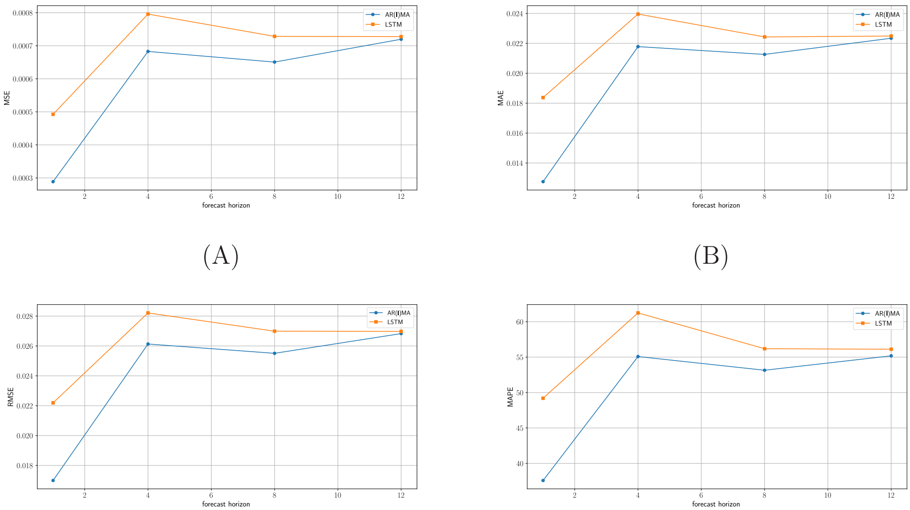
(A) (B)

(C) (D)

Figure 7: Percentage change in error metrics when forecasting with the LSTM model compared to the benchmark ARMA model. (A) One-step ahead forecasting; (B) Four-step ahead forecasting; (C) Eight-step ahead forecasting; (D) Twelve-step ahead forecasting.

horizons, could be attributed to certain temporal characteristics inherent in the salmon market volatility data. It is plausible that the dataset contains information patterns that resonate more with an 8-step ahead forecasting cycle, possibly due to underlying economic or seasonality cycles associated with the salmon market.

Moreover, the discrepancies between the error metrics of the ARMA and LSTM models diminish progressively as the forecast horizon expands. When forecasting 12 steps ahead, the differences become rather marginal. This suggests that as we forecast further into the future, the ability of the two models to predict salmon market volatility begins to converge. In other words, both the traditional ARMA and the more advanced LSTM methods prove to be comparable in their volatility forecasting performance for longer-term predictions.



(A) (B)

(C) (D)

Figure 8: Development of forecast error metrics across various forecasting horizons for each model. (A) Mean square error (MSE); (B) Mean absolute error (MAE); (C) Root mean square error (RMSE); (D) Mean absolute percentage error (MAPE).

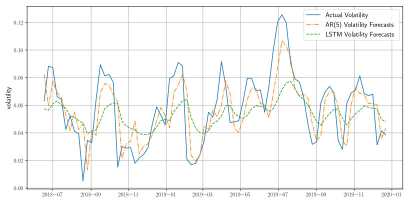
To further investigate the forecasting capabilities of each model, and to shed light on the observed drop in error measures from forecasting 4-steps ahead to 8-steps ahead, we opt to visualize the forecasts. Figure 9 showcases the predictive performance of each model at different forecasting horizons compared against the actual volatility measure (as represented by the volatility proxy measure in Eq. 7).

The ARMA model evidently performs better than the LSTM in capturing the spikes in the actual volatility series when forecasting 1-step ahead, which also corresponds to the forecasting horizon with the most marked differences between the two models. However, as we expand the forecasting horizon, the forecasting abilities of the two models begin to converge, and the forecasts tend to level off when forecasting 12-steps ahead, particularly for the LSTM model.

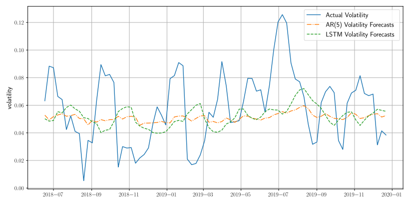
Interestingly, both models – ARMA and LSTM – report lower error metrics for 8-step ahead forecasts than for 4-step ahead ones (see Table 3). In Figure 9 it is evident that in the case of the ARMA model, the 4-step ahead forecast exhibits more fluctuations, but these often run counter to the direction of the actual volatility. However, its 8-step ahead forecast, while appearing flatter, more accurately mirrors the direction of the actual volatility.

For the LSTM model, the 8-step ahead forecast is even more flattened and seems to fluctuate minimally, mostly moving around the mean. Despite its apparently reduced dynamism, this relatively steady, mean-revolving forecast aligns better with the actual volatility series than its 4-step counterpart, which might explain the lower error metrics at this 8-step horizon.

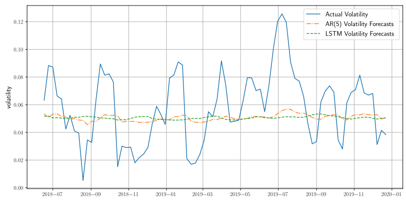
These visualizations suggest that, despite their differences, both models' forecasts better align with the actual volatility when forecasting 8-steps ahead, potentially due to inherent cycles in the salmon market data. While the ARMA model manages to capture the direction of volatility more accurately, the LSTM model's forecasts stay closer to the mean, providing a smoother, albeit less volatile, estimation that still manages to lower the error metrics compared to the 4-step ahead forecasts.



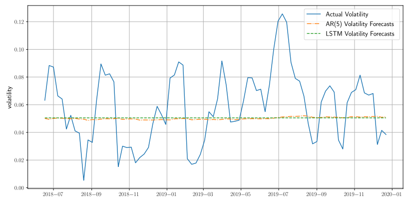
(A)



(B)



(C)



(D)

Figure 9: Forecasts of the ARMA and LSTM model against the actual volatility (A) One-step ahead forecasting; (B) Four-step ahead forecasting; (C) Eight-step ahead forecasting; (D) Twelve-step ahead forecasting.

To further establish the predictive ability of each model, we applied the Diebold-Mariano (DM) test—a statistical tool for comparing the predictive accuracy of two forecasting methods—over different horizons. The results are displayed in Table 4.

The DM test was conducted using two loss functions: the Mean Squared Error (MSE) and

Mean Absolute Error (MAE). The DM test statistic and corresponding p-values were calculated for each combination of loss function and horizon.

The results varied across horizons and loss functions. For a horizon of 1, neither MSE nor MAE showed a significant difference between the two forecast models, with p-values of 0.7848 and 0.6841 respectively. This trend continued for horizon 4 with the MSE loss function, with a p-value of 0.2167 indicating no significant difference.

However, at horizon 4 with the MAE loss function, the DM test statistic was highly significant, with a p-value of 0.0000. This suggests that there is a significant difference in the accuracy of the two models' forecasts at this horizon when assessed using MAE.

The DM test results were also significant for the MSE loss function at horizons 8 and 12, with p-values of 0.0000, again indicating a significant difference in the forecast accuracy of the two models at these horizons. In contrast, the MAE results at these horizons were not significant.

The results of the Diebold-Mariano test emphasize the substantial influence the selection of the loss function can have on the comparative evaluation of model predictions. The variance in predictive accuracy between the LSTM and ARMA models can notably fluctuate or remain consistent, depending on the chosen loss function. To ensure a more robust analysis, this study employed both Mean Squared Error (MSE) and Mean Absolute Error (MAE) as loss functions. These choices help incorporate the potential impacts of both squared and absolute errors on the evaluation of our forecasting models.

Horizon	Loss function	DM-Test	$p-$value
1	MSE	0.2739^{-}	0.7848
	MAE	0.4084^{-}	0.6841
4	MSE	1.2453^{-}	0.2167
	MAE	7.9843***	0.0000
8	MSE	5.2006***	0.0000
	MAE	0.2418^{-}	0.8095
12	MSE	5.7280***	0.0000
	MAE	0.4810^{-}	0.6318

Notes: *** significant at 1%; $-$ indicates that there is no significant difference between the forecasts generated by the two models.

Table 4: Results of the Diebold-Mariano (DM) test comparing the forecasting accuracy of ARIMA and LSTM models for salmon spot prices using MSE and MAE loss functions.

To conclude our analysis, we examine whether a significant difference exists between the volatility predicted by the ARMA model and the LSTM model, compared to the actual volatility values. A method similar to that used by (Fritz & Berger, 2015) is adopted here, involving the use of a paired sample t-test. However, prior to implementing this test, we must ensure that the forecasts produced by each model across all forecasting horizons are normally distributed. To verify this, we utilise both the Shapiro-Wilk test and QQ-plots (see Appendix A.4 and A.5).

Our investigations indicate that the forecasts predominantly follow a normal distribution, with the exception of the 1-step ahead forecast generated by the ARMA model. Consequently, for this particular case, we resort to employing the Wilcoxon signed-rank test—a non-parametric test that doesn't necessitate the forecasts to be normally distributed. This test proves suitable for the ARMA 1-step ahead forecasts as it caters specifically to non-normally distributed data.

The outcomes of these statistical tests are illustrated in Table 5. A clear observation from these results is that both the ARMA and LSTM models exhibit a statistically significant difference from the actual volatility values across all forecast horizons, as evidenced by the p-values below the standard thresholds of 0.1, 0.05, and 0.01.

However, the extent of the difference varies with the forecast horizon and between the two models. For instance, the ARMA model shows a relatively larger t-statistic for the 1-step ahead forecast, suggesting a greater discrepancy between its predictions and the actual values. On the contrary, this difference narrows down as we extend the forecasting horizon, with the LSTM model showing slightly better alignment with the actual values, especially at the 12-steps ahead forecast.

The results underscore the complex nature of volatility forecasting and the nuanced performances of the ARMA and LSTM models across different forecast horizons.

7 Concluding remarks

An accurate multi-step ahead salmon market volatility forecasting model holds considerable value for various participants in the salmon market. Despite the existence of hedging opportunities, these can often be limited due to the lack of speculative traders. Consequently, the implementation of a predictive model for salmon market volatility, which can reliably anticipate market fluctuations, could be a significant asset for all participants in the market.

	ARIMA		LSTM	
Horizon	t-statistic	p-value	t-statistic	p-value
1	1242*^{a}	0.0698^a	1.9398*	0.056
4	2.5251**	0.0135	2.0131***	0.0475
8	2.7583***	0.0072	2.6522***	0.0097
12	2.8415***	0.0057	2.6847***	0.0088

Notes: ***, **, and * denote significance at 1%, 5%, and 10% significance level, respectively; The null hypothesis for the tests is that there is no significant difference between the forecasts produced by the underlying model and the actual volatility values; [a] The Wilcoxon signed-rank test is applied for the forecasts generated by the ARMA model when forecasting 1-step ahead as they do not follow a normal distribution.

Table 5: Paired two-tailed t-test and Wilcoxon signed-rank test results for ARIMA and LSTM models at different forecasting horizons.

To date, there has been no exploration of the use of deep-learning techniques for predicting salmon market volatility. Recognizing this gap, this study seeks to examine the application of such advanced methodologies in this domain, hoping to bring about improved forecasting performance and valuable insights for stakeholders. Therefore, we explored and compared the forecasting capabilities of two time-series prediction models, ARMA and LSTM, with respect to predicting salmon market volatility. The analysis presented a clear, albeit complex, picture of their comparative effectiveness across different forecast horizons.

Our results indicate that the ARMA model has a slight edge over the LSTM model in terms of capturing the spikes in volatility at a 1-step ahead forecasting horizon. However, as the forecast horizon expands, the performance differences between the two models begin to narrow, and the forecasts from both models tend to level off when forecasting 12-steps ahead.

The results indicated a reduction in error measures when transitioning from a 4-step to an 8-step ahead forecast for both models. This behavior, further illustrated by visualizing the forecast results, suggested a potential alignment between the 8-step forecasting cycle and inherent patterns present in the salmon market volatility data. It's plausible that this may be attributed to the seasonal patterns reflected within the out-of-sample series over a mid-term future period. Although our analysis has demonstrated that seasonally adjusting the salmon spot price series is not required for the purpose of this study, these findings emphasize the potential significance of seasonality when conducting mid-term forecasting studies.

To establish whether the forecasts generated by the two models were significantly diffifenet we employed the Diebold-Mariano (DM) test. Both Mean Squared Error (MSE) and Mean

Absolute Error (MAE) loss functions were employed to present a comprehensive understanding of the predictive accuracy of both models. The results indicated discrepancies between the two forecasts for all horizons except the 1-step ahead forecast. As these results varied based on the underlying error metrics, we argue that the choice of loss function plays a significant role in the comparative evaluation of model predictions

Last, the paired $t-$test and Wilcoxon signed-rank test results emphasized the nuanced performances of the models across different forecast horizons. It was found that both models exhibited a statistically significant difference from the actual volatility values across all forecast horizons, indicating the intricate nature of volatility forecasting.

This study highlights the importance of using a range of statistical methods and taking a comprehensive approach in analyzing and comparing forecast models. While both ARMA and LSTM models demonstrate their unique strengths, their efficacy is significantly influenced by the characteristics of the data and the chosen forecast horizon. Moreover, despite the ability of the LSTM to model complex nonlinear relationships, the ARMA model proved superior in predicting salmon market volatility. Thus, we suggest that the salmon market may be more linear than expected, with negligible or even no non-linear volatility patterns for an LSTM model to exploit.

Future research could build on this study by exploring additional factors that might impact the accuracy of these models, such as the influence of different market dynamics or even the impact of macroeconomic variables. Further studies could also consider the implementation of hybrid models, combining the strengths of ARMA and LSTM, to improve forecasting accuracy. Moreover, given that our study period concludes before the COVID-19 pandemic, it would be interesting to test this framework with adequate post-COVID data.

Our findings contribute to the ongoing dialogue around the best practices in volatility forecasting, providing valuable insights for stakeholders in the salmon market. Furthermore, the framework of this study could be adapted for use in other commodity markets where accurate volatility forecasting is also critical.

Disclosure Statement

The author(s) report there are no competing interests to declare.

Funding Details

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

References

Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the institute of Statistical Mathematics*, *21*(1), 243–247.

Andersen, B. P., & de Lange, P. E. (2021). Efficiency in the atlantic salmon futures market. *Journal of Futures Markets*, *41*(6), 949–984.

Ankamah-Yeboah, I., Nielsen, M., & Nielsen, R. (2017). Price formation of the salmon aquaculture futures market. *Aquaculture Economics & Management*, *21*(3), 376–399.

Asche, F. (2008). Farming the sea. *Marine Resource Economics*, *23*(4), 527–547.

Asche, F., Dahl, R. E., & Steen, M. (2015). Price volatility in seafood markets: Farmed vs. wild fish. *Aquaculture Economics & Management*, *19*(3), 316–335.

Asche, F., Misund, B., & Oglend, A. (2016). The spot-forward relationship in the atlantic salmon market. *Aquaculture Economics & Management*, *20*(2), 222–234.

Asche, F., Misund, B., & Oglend, A. (2019). The case and cause of salmon price volatility. *Marine Resource Economics*, *34*(1), 23–38.

Asche, F., & Oglend, A. (2016). The relationship between input-factor and output prices in commodity industries: The case of norwegian salmon aquaculture. *Journal of Commodity Markets*, *1*(1), 35–47.

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, *5*(2), 157–166.

Bergfjord, O. J. (2007). Is there a future for salmon futures? an analysis of the prospects of a potential futures market for salmon. *Aquaculture Economics & Management*, *11*(2), 113–132.

Bloznelis, D. (2016). Salmon price volatility: A weight-class-specific multivariate approach. *Aquaculture Economics & Management*, *20*(1), 24–53.

Bloznelis, D. (2018a). Hedging salmon price risk. *Aquaculture Economics & Management*, *22*(2), 168–191.

Bloznelis, D. (2018b). Short-term salmon price forecasting. *Journal of Forecasting*, *37*(2), 151–169.

Dahl, R. E., & Oglend, A. (2014). Fish price volatility. *Marine Resource Economics*, *29*(4), 305–322.

Dahl, R. E., & Jonsson, E. (2018). Volatility spillover in seafood markets. *Journal of Commodity Markets*, *12*, 44–59.

Dahl, R. E., Oglend, A., & Yahya, M. (2021). Salmon stock market prices revealing salmon price information. *Marine Resource Economics*, *36*(2), 173–190.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, *74*(366a), 427–431.

Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *20*(1), 134–144.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211.

Ewald, C.-O., Haugom, E., Kanthan, L., Lien, G., Salehi, P., & Størdal, S. (2022). Salmon futures and the fish pool market in the context of the capm and a three-factor model. *Aquaculture Economics & Management*, *26*(2), 171–191.

Forsberg, O. I., & Guttormsen, A. G. (2006). The value of information in salmon farming. harvesting the right fish at the right time. *Aquaculture Economics & Management*, *10*(3), 183–200.

Fritz, M., & Berger, P. D. (2015). *Improving the user experience through practical data analytics: Gain meaningful insight and increase your bottom line*. Morgan Kaufmann.

Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with lstm. *Neural Computation*, *12*(10), 2451–2471.

Graves, A., Mohamed, A.-r., & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. *2013 IEEE international conference on acoustics, speech and signal processing*, 6645–6649.

Guttormsen, A. G. (1999). Forecasting weekly salmon prices: Risk management in fish farming. *Aquaculture Economics & Management*, *3*(2), 159–166.

Guttormsen, A. G. (2008). Faustmann in the sea: Optimal rotation in aquaculture. *Marine Resource Economics*, *23*(4), 401–410.

Hamid, S. A., & Iqbal, Z. (2004). Using neural networks for forecasting volatility of s&p 500 index futures prices. *Journal of Business Research*, *57*(10), 1116–1125.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Hyndman, R., J. (2014). Forecasting weekly data.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice*. OTexts.

Kim, H. Y., & Won, C. H. (2018). Forecasting the volatility of stock price index: A hybrid model integrating lstm with multiple garch-type models. *Expert Systems with Applications*, *103*, 25–37.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kristjanpoller, W., & Minutolo, M. C. (2016). Forecasting volatility of oil price using an artificial neural network-garch model. *Expert Systems with Applications*, *65*, 233–241.

Kvaløy, O., & Tveterås, R. (2008). Cost structure and vertical integration between farming and processing. *Journal of Agricultural Economics*, *59*(2), 296–311.

Laurent, S., & Violante, F. (2012). Volatility forecasts evaluation and comparison. *Wiley Interdisciplinary Reviews: Computational Statistics*, *4*(1), 1–12.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PloS one*, *13*(3), e0194889.

Manogna, R., & Mishra, A. K. (2021). Forecasting spot prices of agricultural commodities in india: Application of deep-learning models. *Intelligent Systems in Accounting, Finance and Management*, *28*(1), 72–83.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347–370.

Oglend, A. (2013). Recent trends in salmon price volatility. *Aquaculture Economics & Management*, *17*(3), 281–299.

Oglend, A., & Sikveland, M. (2008). The behaviour of salmon price volatility. *Marine Resource Economics*, *23*(4), 507–526.

Oglend, A., & Straume, H.-M. (2019). Pricing efficiency across destination markets for norwegian salmon exports. *Aquaculture Economics & Management*, *23*(2), 188–203.

Roh, T. H. (2007). Forecasting the volatility of stock price index. *Expert Systems with Applications*, *33*(4), 916–922.

Selvin, S., Vinayakumar, R., Gopalakrishnan, E., Menon, V. K., & Soman, K. (2017). Stock price prediction using lstm, rnn and cnn-sliding window model. *2017 international conference on advances in computing, communications and informatics (icacci)*, 1643–1647.

Solibakke, P. B. (2012). Scientific stochastic volatility models for the salmon forward market: Forecasting (un-) conditional moments. *Aquaculture Economics & Management*, *16*(3), 222–249.

Tseng, C.-H., Cheng, S.-T., Wang, Y.-H., & Peng, J.-T. (2008). Artificial neural network model of the hybrid egarch volatility of the taiwan stock index option prices. *Physica A: Statistical Mechanics and its Applications*, *387*(13), 3192–3200.

Verma, S. (2021). Forecasting volatility of crude oil futures using a garch–rnn hybrid approach. *Intelligent Systems in Accounting, Finance and Management*, *28*(2), 130–142.

Xu, X., & Zhang, Y. (2021). Corn cash price forecasting with neural networks. *Computers and Electronics in Agriculture*, *184*, 106120.

Xu, X., & Zhang, Y. (2022). Commodity price forecasting via neural networks for coffee, corn, cotton, oats, soybeans, soybean oil, sugar, and wheat. *Intelligent Systems in Accounting, Finance and Management.*

A Appendix

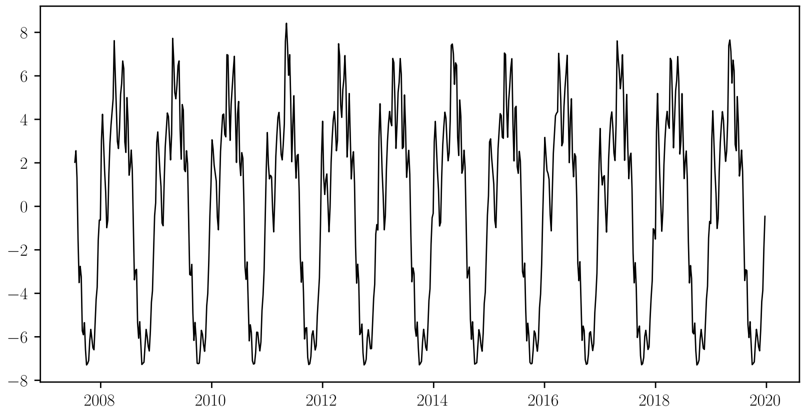


Figure A.1: Seasonal Component of the Salmon Spot Price Series.

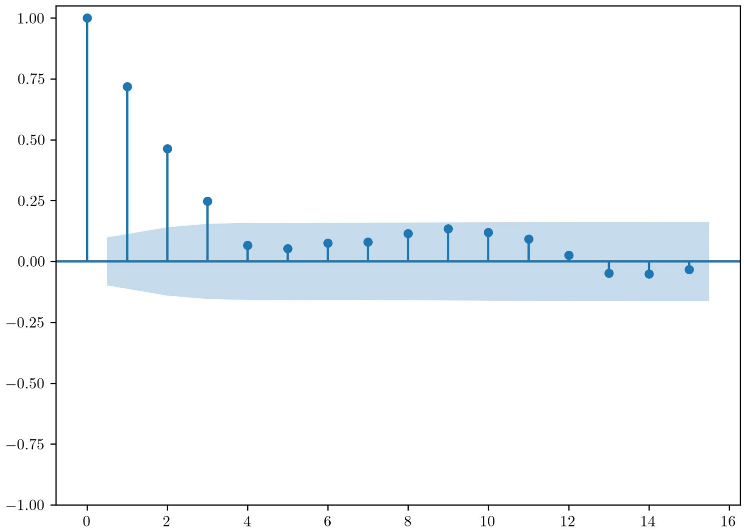
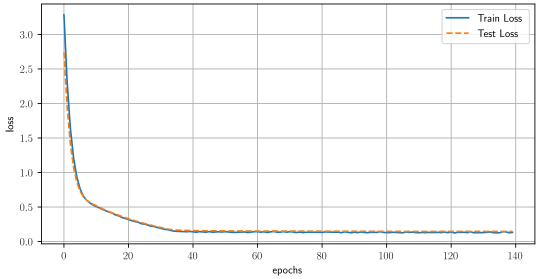
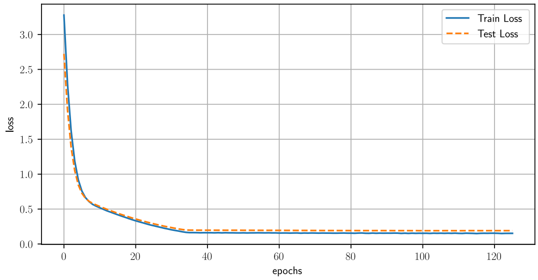


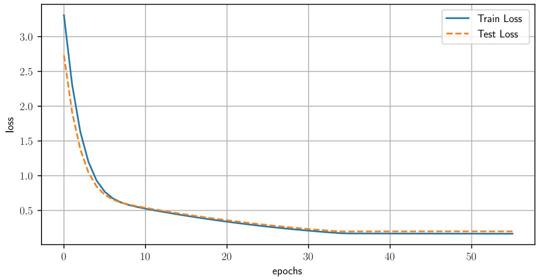
Figure A.2: Autocorrelation Function (ACF) for the volatility proxy series from 2007:W27 to 2019:W51.



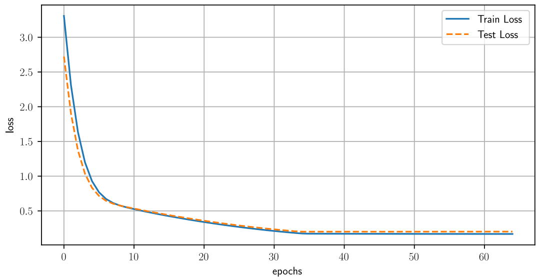
(A)



(B)

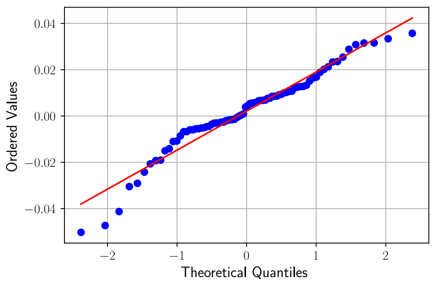


(C)

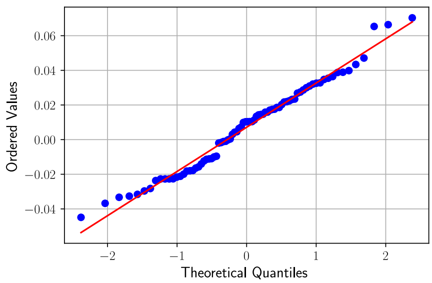


(D)

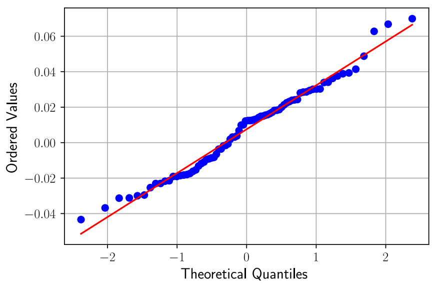
Figure A.3: Curves of the Mean Squared Error (MSE) loss function during LSTM training: (A) Loss curves for forecasting 1-step-ahead; (B) Loss curves for forecasting 4-steps-ahead; (C) Loss curves for forecasting 8-steps-ahead; (D) Loss curves for forecasting 12-steps-ahead.



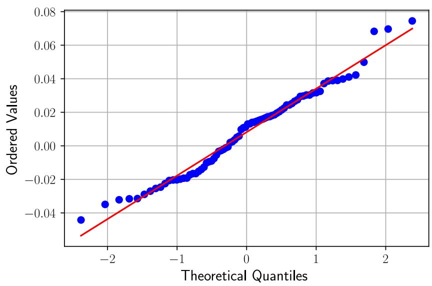
(A)



(B)

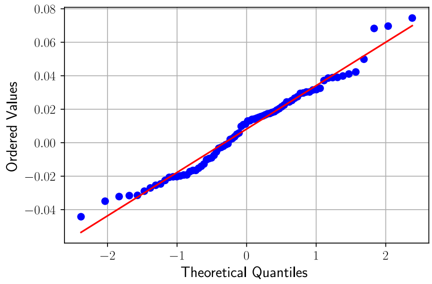


(C)

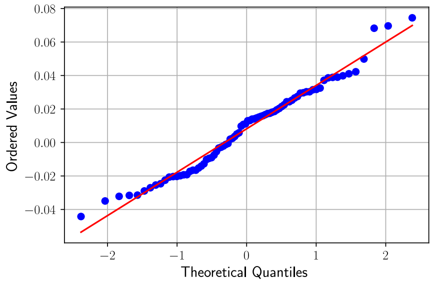


(D)

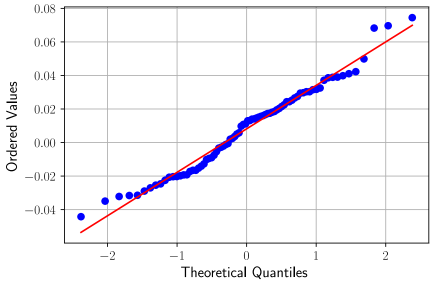
Figure A.4: QQ-Plots for normality test of ARMA model forecasts. (A) 1-step-ahead; (B) 4-steps-ahead; (C) 8-steps-ahead; (D) 12-steps-ahead.



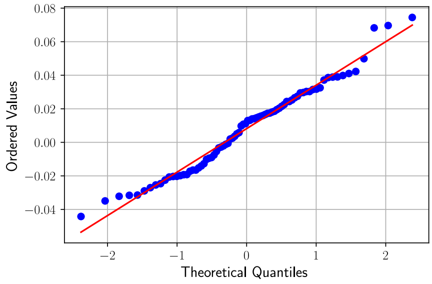
(A)



(B)



(C)



(D)

Figure A.5: QQ-Plots for normality test of LSTM model forecasts. (A) 1-step-ahead; (B) 4-steps-ahead; (C) 8-steps-ahead; (D) 12-steps-ahead.

Research Article III

The Influence of Market Adaptability on Forecasting Salmon Spot Prices: A Comparison of Hybrid Deep Learning and Traditional Models

Mikaella Zitti[1]

[1]Department of Economics and Resource Management, Norwegian University of Life Sciences, Ås, Norway.

Abstract

This research conducts a comparative analysis between the traditional Vector Autoregression (VAR) model and a hybrid VAR-Long Short-Term Memory (VAR-LSTM) model for salmon spot price prediction. The study employs LASSO for dimensionality reduction and performs residual diagnostics to ensure model robustness. VAR and hybrid models are estimated based on different futures contract maturities and evaluated using various window methods. Predictive performance is assessed through error metrics. Contrary to expectations, the VAR and VAR-LSTM models show near-identical forecasting capabilities, suggesting a lack of exploitable non-linear patterns in the salmon spot price series. This indicates the efficiency of the salmon market, with price trends quickly reflected by investors. Further comparison of the VAR model with a random walk with a drift model confirms the VAR model's superior performance across different forecasting horizons.

1 Introduction

Salmon price prediction is important to various market participants, such as farmers, processors, traders, and hedgers. Specifically, farmers control salmon production, aiming to harvest when profitability is at its peak (Asche, 2008). A high salmon spot price encourages farmers to harvest to maximize their profits. Yet, the market's volatility makes it challenging to plan the production process early enough to avoid price fluctuations. Beyond short-term profitability, farmers have a vested interest in the technological development of the salmon industry, as it bolsters industry growth. Consequently, they often set price quotas to maximize sales performance and raise funds for technological innovation and investment.

The salmon market is highly volatile (Bloznelis, 2016; Guttormsen, 1999; Oglend, 2013; Oglend & Sikveland, 2008). This volatility affects not only farmers but also other market participants. For instance, processors plan their operations based on the expected input cost level, indicated by the salmon spot prices. Without control over production, their operations are directly influenced by fluctuations in salmon spot prices. Traders aim to minimize their buying costs while maintaining profitability, but high market volatility can significantly harm their operations. Other participants, such as hedgers, also depend on salmon price predictions, especially since hedging opportunities in the salmon market are thin due to a lack of speculative traders (Andersen & de Lange, 2021; Asche et al., 2016b; Ewald et al., 2022).

Considering the challenges posed by the market's high volatility, a major concern for aquaculture economists, a number of studies provide insights into its causes and behavior (Asche et al., 2015; Asche et al., 2019; Bloznelis, 2016; Dahl & Oglend, 2014; Dahl & Jonsson, 2018; Oglend, 2013; Oglend & Sikveland, 2008). These studies shed light on how farmers' harvesting and investment decisions can affect salmon market volatility which in turn can impact the operations of other market participants. Specifically, Bloznelis (2018b) and Oglend (2013), found that salmon price volatility has doubled since the launch of Fish Pool, a futures and options exchange of farmed salmon, and since then it remains persistently high. Many studies analyze the efficacy and role of the salmon futures market (Ankamah-Yeboah et al., 2017; Ewald et al., 2017; Larsen & Asche, 2011; Misund & Asche, 2016; Solibakke, 2012) and specifically its relationship with the spot market (Asche et al., 2016a, 2016b; Chen & Scholtens, 2019). Many raise the concern that this market suffers from low liquidity and infrequent trades (Andersen & de Lange, 2021; Bloznelis, 2018a; Dahl et al., 2021; Ewald et al., 2022), but a study showed that the majority of the salmon price variation is common, justifying the representation of a salmon price index (Oglend & Straume, 2019).

Given the volatility in the salmon market presents significant challenges to various participants, salmon futures contracts unfortunately cannot serve as a reliable hedging mechanism to reduce their risk exposure. However, so far there has been a limited number of studies attempting to predict salmon prices. Specifically, Guttormsen (1999) argued that if prices are predicted within reasonable confidence bounds, salmon volatility would be reduced. In a more recent study, Bloznelis (2018b) found that using a simple trading strategy based on price forecasts can increase the net profit of a salmon farmer by approximately 7%. To come to these conclusions, Guttormsen (1999) used standard econometric models for forecasting salmon prices and found that the Vector Autoregressive (VAR) model performed best amongst its peers based on forecasting accuracy measures. Bloznelis (2018b) on the other hand, used sixteen alternative forecasting methods and found that $k-$nearest neighbours and vector error correction model are the best performing ones.

While these models have brought useful insights, there is a gap in research in the integration of traditional statistical methods with advanced deep learning techniques for salmon price prediction. To address this gap, this study develops a hybrid architecture model that combines the forecasting capabilities of a traditional multivariate forecasting model, the Vector Autoregressive (VAR), with a deep learning technique known as Long Short-Term Memory (LSTM). The idea behind incorporating a recurrent neural network (RNN) such as the LSTM, is that traditional statistical methods such as time series forecasting models (e.g., VAR) are not able to detect any potential complexities and non-linearity in financial data series. A number of studies investigated the forecasting performance of neural networks in forecasting commodity prices (Ramyar & Kianfar, 2019;

RL & Mishra, 2021; Xu & Zhang, 2022), with some using hybrid methods (Kristjanpoller & Minutolo, 2015, 2016; Parot et al., 2019) and found that they have significant forecasting accuracy. Therefore, we decided to further explore their performance in predicting salmon spot prices for various forecasting horizons ranging from 1-step ahead to 52-steps ahead (1 year).

Employing a hybrid architecture technique can be beneficial when forecasting longer horizons as VAR models are found to perform better at short-term forecasting horizons (Baumeister & Kilian, 2012, 2015). Deep learning techniques such as LSTM are able to deal with long temporal dependencies (Hochreiter & Schmidhuber, 1997) and therefore can recognize long term patterns that emerged in the past. Moreover, neural networks use *supervised learning* which is an input-to-output mechanism that takes a set of input variables and tries to match them to a given output. In other words, the network knows the output and after a number of training iterations it is trying to match this output as close as possible. Therefore, employing the LSTM comes with a number of advantages as it is able to recognize potential nonlinear and complex patterns in the data and can also be used for longer-term forecasting.

This research study aims to fill a significant gap in the literature by investigating the potential superiority of a hybrid forecasting approach, integrating traditional econometric models with deep-learning techniques, for predicting salmon prices. Given the controversy regarding the efficacy of salmon futures contracts, we use the salmon spot price series as the predictive variable. The forecasting ability of the VAR-LSTM is assessed against the traditional VAR model. The forecasting performance of each model is assessed under different forecasting horizons. Once the forecasts are generated, their predicting accuracy is evaluated using different forecast error measures. The model with the lowest forecast error measures is deemed as the best performing one. To assess the robustness of the results, we compare the expected value of the forecast losses generated by each model using the Diebold-Mariano (DM) test. In essence, this study aims to determine whether the integration of traditional forecasting models with deep-learning methods can significantly improve the prediction of salmon spot prices for various market participants.

The remaining sections are organized as follows: Section 2 discusses the specifications of the underlying predictive and predictor variables, their seasonality adjustments, and summary statistics. Section 3 describes the forecasting techniques used in this study, as well as the forecast error metrics and tests employed to evaluate their forecasting ability. Section 4 presents the results. Finally, Section 5 discusses the implications of the findings.

2 Data and Statistics

2.1 Data

This study utilizes weekly salmon spot prices as the predictive variable. These prices are procured from the NASDAQ Salmon Index[1], covering the period from the first week of 2013 up to the fifty-third week of 2020 (2013:W1 – 2020:W53).

Additionally, we incorporate data on salmon futures contract prices, sourced from the Fish Pool website[2]. Fish Pool offers daily price quotes on salmon forward prices with monthly maturities extending up to 60 months. To align these prices with the spot price series, we converted them into a weekly format by taking the final observation each week. We focused on contracts maturing after intervals of one week, one month, six months, and twelve months ahead, reflecting our interest in predicting prices at these specific horizons.

In the case of front-month futures, we employ the contract whose price settlement month aligns with the month of issue. However, if a contract is issued during the last week of its maturity month, we shift our focus to the contract set to expire in the following month. For instance, for a forward price issued on the 19$^{\text{th}}$ of June 2015, we would consider the price that expires in June of 2015. On the other hand, if a futures contract is issued on the 26$^{\text{th}}$ of June 2015, we choose the contract that is due to expire in July 2015. This methodology extends to the six-month and twelve-month contracts as well. The reasoning behind this approach is because the price for the contract that expires in the same month as the issue date does not carry meaningful predictive information during the last week of that month. By selecting the next month's contract, we can ensure the inclusion of

[1] https://salmonprice.nasdaqomxtrader.com

[2] www.fishpool.eu/price-information/forward-prices-3/

relevant and informative futures contract data in our price forecast analysis for salmon.

Although futures prices are often preferred when forecasting commodity prices (Wang & Li, 2018; Xu & Zhang, 2022), their utility in the salmon market is controversial (Andersen & de Lange, 2021; Ewald et al., 2022). Previous studies suggest that the salmon futures market remains immature and has not yet reached the stage where forward prices can predict spot prices (Asche et al., 2016b). Additionally, Dahl et al. (2021) found that stock prices from salmon farming firms reflect salmon price information earlier than the Fish Pool Index (FPI) due to bias in the salmon futures pricing structure, which relies solely on the FPI. Despite these challenges, futures prices are still included as they offer a market perspective on expected price movements, adding a valuable dimension to our forecast. Figure 1 graphically represents the evolution of prices for the spot price, front-month, six-month, and twelve-month futures contract series.

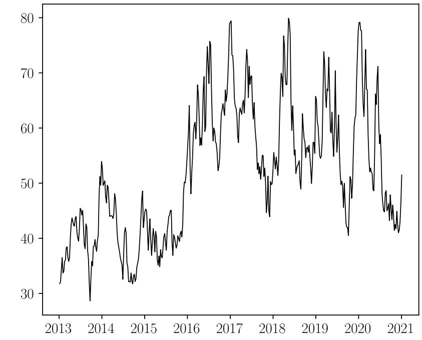
In addition to futures contract prices, there are other factors that contain predictive information for the development of the salmon spot price, which are also important to consider in this study. Weekly data on the export volume is available from Statistics Norway (SSB)[3]. Despite the availability of data regarding both fresh and frozen salmon volumes on the SSB website, we focused solely on fresh salmon volumes, given their substantially larger quantity compared to frozen salmon volumes. Figure 2 presents the development of the salmon export volume.

Another important factor to incorporate is soybean prices. Soy has been used to replace fish meal and fish oil for feeding farmed salmon (Egerton et al., 2020). Therefore, weekly soybean prices were obtained from Refinitiv Eikon[4] and transformed into NOK. The development of the soybean price series is presented in Figure 3. Incorporating soybean prices in our analysis is particularly relevant for this analysis as the use of soy in salmon feed has become more prevalent in recent years, and any changes in soy prices could affect the cost of production for salmon farmers.

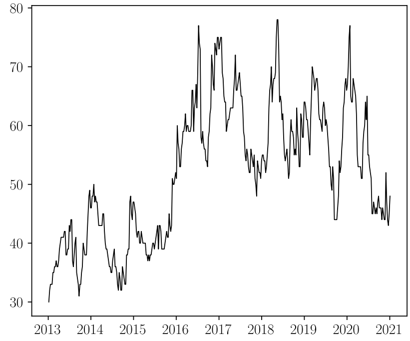
Straume (2014) found that the currency exchange rate can have an impact on the supply and demand relationship of the salmon spot market. As Europe is the primary market

[3]www.ssb.no/en/statbank/

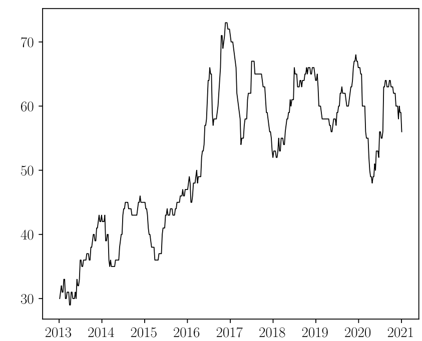
[4]Chicago Board of Trade (CBoT) Soybeans Composite Commodity Future is obtained from Refinitiv.



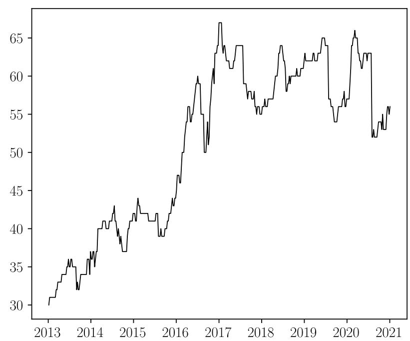
(A)



(B)



(C)



(D)

Figure 1: (A): Development in salmon spot price; (B): Front-month futures contracts price; (C): Six-month futures contracts price; (D): Twelve-month futures contracts price. The figure suggests that all the underlying series are non-stationary in levels.

for Norwegian salmon, information regarding the EUR/NOK exchange rate is considered an important predictive factor of the salmon spot price. We obtained the EUR/NOK exchange rate data series from Norges Bank (Norwegian Central Bank) website[5] for the same period from 2013:W1 until 2020:W53. However, the exchange rates were not available on a weekly basis. Therefore, we transformed the series by taking the last observation of each week. Figure 4 illustrates the development of the EUR/NOK exchange rate series. The inclusion of the EUR/NOK exchange rate in our analysis allows us to identify how changes in the exchange rate affect the competitiveness of Norwegian salmon in the

[5] www.norges-bank.no/en/topics/Statistics/exchange_rates/

international market.

Lastly, the analysis includes the stock prices of Norwegian salmon farming companies. Dahl et al. (2021) found that stock prices of these companies reflect information on salmon prices earlier than the Fish Pool Index. Therefore, we collected the weekly closing prices of the Norwegian salmon farming companies with the highest market capitalization (Marine Harvest, 2020): Marine Harvest (MOWI), Grieg Seafood (GSFG), Lerøy Seafood (LSG), and SalMar (SALM) from Refinitiv. To incorporate the information from these stock prices into our analysis, we created a Share Price Index (SPI) that corresponds to an equally weighted portfolio of the four shares. We then normalized each price $SPI_t; (t = 0, 1, \ldots, T)$ by dividing it with the first price of the series SPI_0 to show the growth rate of these shared prices. The choice to create a share price index is because it is less sensitive to company-specific news and can reflect the development of the salmon spot price better than the individual share prices. The SPI series is presented in Figure 5. By including the stock prices of salmon farming companies in our analysis, we can determine how changes in these companies' performance affect salmon prices.

2.2 Seasonal Adjustment

As a result of factors related to supply and demand a key characteristic of salmon production is seasonality. Seasonality in supply does not match the seasonality in demand and that generates seasonal patterns in salmon price and production volume. Modelling seasonality when having weekly time series is complicated. In the existing literature, the most common technique is the Fourier series; that is, sums of trigonometric functions (Bloznelis, 2016, 2018b; Oglend, 2013). Here, we follow a technique introduced by Hyndman and Athanasopoulos (2018), and also applied by Bloznelis (2018b), that uses a regression with autoregressive moving average (ARMA) errors, having Fourier terms as regressors. The number of Fourier terms could be up to 26 pairs for weekly data. However, the number of Fourier terms for the fitted model was selected by minimising the Akaike information criterion (AIC) and choosing between none to 26 pairs, while the same applies for selecting the order of the ARIMA model (Hyndman & Athanasopoulos,

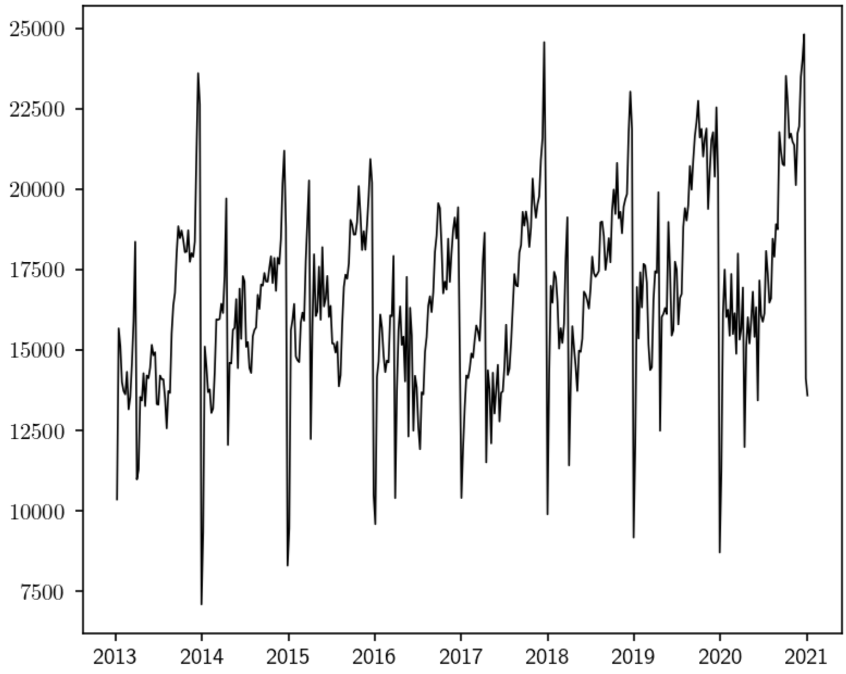


Figure 2: Salmon export volume in metric tons. The figure depicts strong seasonality patterns and suggests stationarity.

2018)[6].

The salmon spot price and the export volume display large movements around Christmas and Easter. To incorporate the possibility of deterministic seasonality occurring, we add four dummy variables to specify the weeks before and after Christmas (including Christmas week) and four more to specify the same weeks before and after Easter (including Easter week), as it was suggested by Bloznelis (2018b). The deterministic seasonality is expressed by the means of these eight seasonal dummy variables. The combination of the Fourier terms and these eight seasonal dummy variables exhibit the seasonal component, which is subtracted from the data before proceeding. The eight seasonal dummy

[6]The fitted model has 23 pairs of Fourier terms and can be written as: $\sum_{j=1}^{23}[\alpha_j \sin(\frac{2\pi jt}{r}) + \beta_j \cos(\frac{2\pi jt}{r})] + \eta_t$ where $r = \frac{365.25}{7} = 52.18$ seasonal cycle, η_t is an ARIMA(p, d, q) process (this depends on the underlying data series). There are 48 parameters to capture the seasonality according to the AIC selection.

Figure 3: Soybeans (CBoT) futures price series in NOK.

variables are not incorporated when seasonally adjusting the front-month, six-month, and twelve-month futures prices variables because the underlying of a future contract is the average FPI recorded over a month and not a week.

Seasonality in soybeans is observed from May until August when the soybeans are ready for harvesting. It is evident from Fig.3 that the prices increase around May, and they decline around July-August. The same technique as for the spot price and the export volume is also employed for obtaining the soybeans seasonally adjusted series. To incorporate seasonality, 17 dummy variables are constructed to specify the weeks from May until August[7]. A combination of the Fourier terms and the 17 seasonal dummy variables form the seasonal component. This is then deducted from the original series to construct the seasonally-adjusted version.

[7]If August had 5 observations, the last one was deleted.

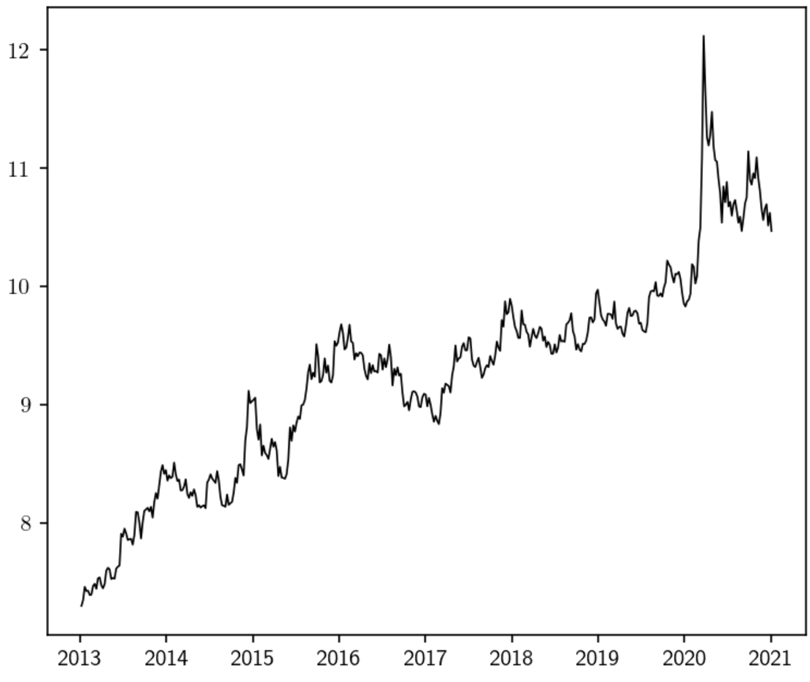


Figure 4: EUR/NOK currency exchange rate.

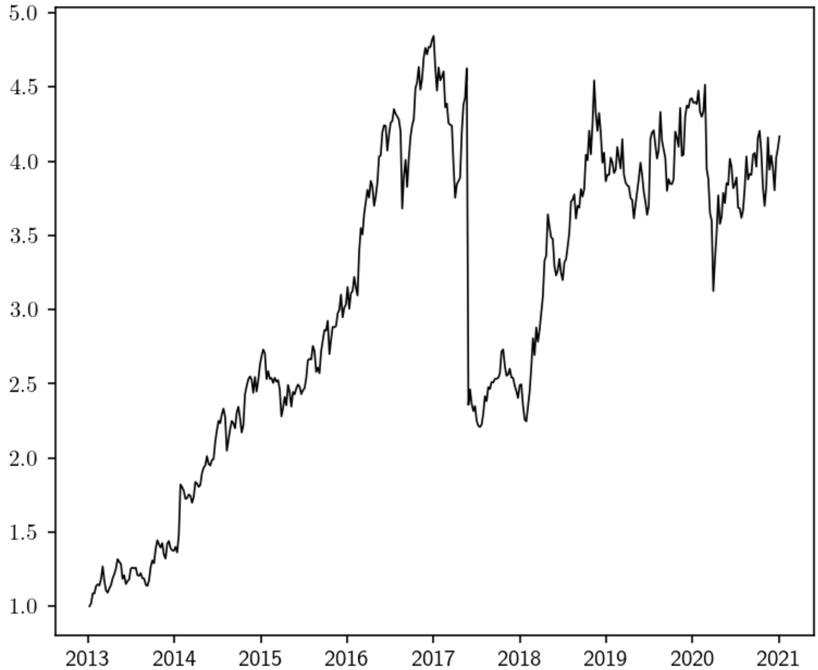
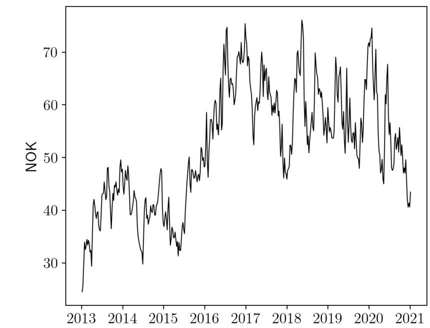
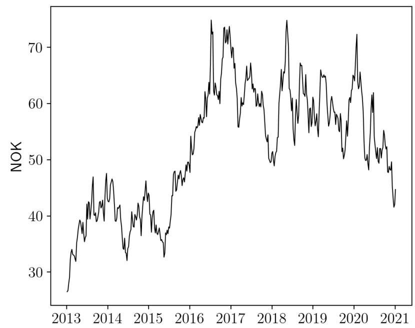


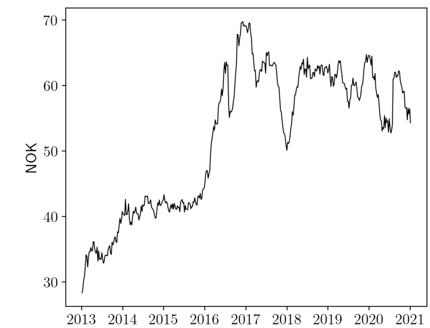
Figure 5: Share Price Index (SPI). The figure presents the shared prices of four salmon producing companies publicly trading on the Oslo stock exchange (OSE). The prices are normalized and the figure presents the exponential of the logarithmic series.



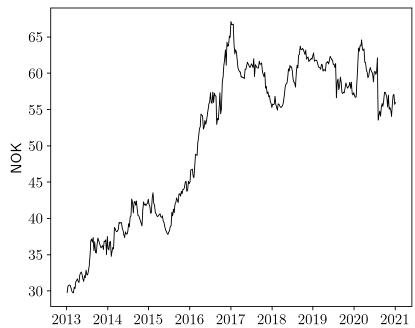
(A)



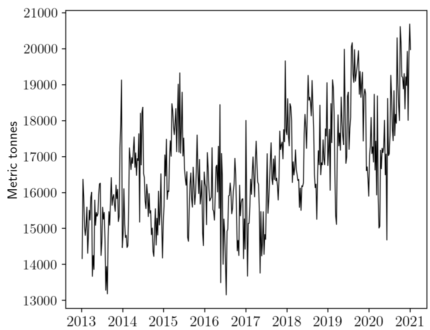
(B)



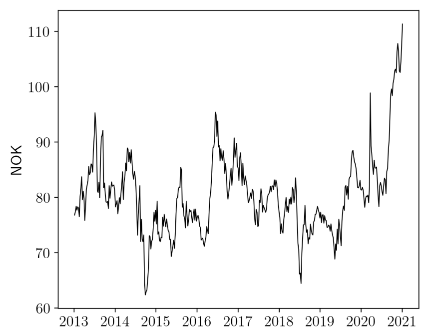
(C)



(D)



(E)



(F)

Figure 8: Seasonally adjusted versions for: spot price (A); front-month futures contracts price (B); six-month futures contracts price (C); twelve-month futures contracts price (D); export volume (E); soybeans (F). The figure suggests that all the underlying series but export volume are non-stationary. The export volume seasonally adjusted series is depicted as a random walk.

Therefore, seasonally-adjusted versions of the spot price, front-month, six-month, and twelve-month futures price, export volume, and soybeans series are employed in the remainder of this study instead of the original series. Figure 8 presents the seasonally adjusted version of these series.

2.3 Descriptive Statistics

Statistics such as mean, standard deviation, skewness, and kurtosis can be found in Table 1. Results from the Shapiro-Wilk normality test and the augmented Dickey-Fuller (ADF) test are also presents in Table 1. The results indicate that all the series are non-normally distributed. The skewness and excess kurtosis indicators show that the distributions of the spot price, futures contracts price (front-month, six-month, and twelve-month) series, and SPI series have thinner tails than a normal distribution with prices close to the minimum price appearing more often than those close to the maximum price. The soybeans price series is the furthest from a normal distribution indicating that higher prices appear more often than lower prices. The ADF test is applied to examine the presence of a single unit root in each of the series (Dickey & Fuller, 1979). All but the seasonally adjusted export volume series indicate the presence of a unit root. In other words, except from the export volume series, the remaining ones are likely non-stationary and therefore should be transformed into stationary. A logarithmic transformation is applied in all the series including the export volume. Even though the seasonally adjusted export volume does not contain a unit root, for applying the VAR, two versions of the variable are considered; the logarithmic transformation and a scaled transformation where each observation is divided by 100000.

The logarithmic returns of the predictive variable are centered at zero with standard deviation 6%. The logarithmic transformation is aimed so that the skewness and the kurtosis indicators come closer to zero and any potential multiplicative seasonality and shocks turn into additive ones. The Shapiro-Wilk normality test suggests that the seasonally adjusted logarithmic series of the spot price is normally distributed. The skewness indicator is above zero implying that large positive returns are likely to be more common

	Mean	Min	Max	SD	Skewness	Kurtosis	SW test	ADF
Spot price*	52.06	24.52	76.08	11.31	-0.017	2.109	0.982***	-2.84
Front-month futures*	52.03	26.5	74.85	11.05	-0.101	1.996	0.970***	-2.31
Six-month futures*	52.24	28.38	69.78	10.88	-0.33	1.67	0.902***	-1.94
Twelve-month futures*	51.02	29.75	67.13	10.7	-0.47	1.7	0.878***	-1.86
EV[a]*	16671	13155	20684	1491	0.273	2.654	0.99***	-3.07**
EUR/NOK	9.243	7.3	12.12	0.89	0.046	2.932	0.98***	-1.91
Soybean price*	80.50	63.4	111.35	7.33	1.113	5.6	0.93***	-2.29
SPI[b]	3.068	1	4.84	1.06	-0.339	1.918	0.934***	-1.93

Note: SD represents the standard deviation, Kurtosis the excess kurtosis, and SW provides the t-statistics from the Sahpiro-Wilk test of normality. Notations *** and ** denote statistical significance at a 1% and 5% significance level, respectively.

[a]Export volume; [b]Share Price Index.

*Statistics are applied on the seasonally adjusted version.

Table 1: Descriptive statistics for the predictor and predictive variables.

than large negative returns, or similar. The largest negative return is 18% and the largest positive return is 19%, indicating no significant difference between the two. However, the excess kurtosis indicator of the spot logarithmic returns is larger than zero showing signs of leptokurtosis, in other words that returns are more densely concentrated in a narrow interval around the mean and more spread out in tails as compared to the normal distribution. Since it is relatively close to zero, the leptokurtosis of the returns distribution is not significant and therefore the normality assumption is valid. The ADF test indicates that none of the underlying data series contains a unit root and hence all factors are likely to be stationary.

	Mean	Min	Max	SD	Skewness	Kurtosis	SW test	ADF
Spot price	0.001	-0.18	0.19	0.06	0.117	3.389	0.994	-18.42***
Front-month futures price	0.001	-0.16	0.14	0.043	-0.129	3.965	0.99***	-9.27***
Six-month futures price	0.002	-0.1	0.12	0.02	0.181	6.355	0.969***	-8.48***
Twelve-month futures price	0.002	-0.15	0.08	0.02	-0.839	11.239	0.909***	-23.87***
EV[a]	0.0008	-0.31	0.24	0.067	-0.124	5.119	0.977***	-14.95***
EV[b]	0.167	0.13	0.21	0.015	0.275	2.660	0.99***	-3.07**
EUR/NOK	0.0009	-0.04	0.09	0.01	1.26	10.7	0.926***	-20.46***
Soybean price	0.0009	-0.13	0.19	0.031	-0.02	7.547	0.951***	-22.23***
SPI	0.003	-0.67	0.21	0.05	-5.963	83.514	0.689***	-21.34***

Note: SD represents the standard deviation, Kurtosis the excess kurtosis, and SW provides the t-statistics from the Shapiro-Wilk test of normality. Notations *** and ** denote statistical significance at a 1% and 5% significance level, respectively.

[a]Export volume based on the logarithmic transformation; [b]Export volume based on the re-scaling transformation.

Table 2: Descriptive statistics for the logarithmic transformed series.

2.4 Variable Selection Using LASSO

To make the forecasts more accurate and reduce dimensionality we use a variable selection method, the Least Absolute Shrinkage and Selection Operator (LASSO) regression (Tibshirani, 1996). LASSO is a shrinkage and variable selection method for linear regression models, which minimizes the usual sum of squared errors with a bound on the sum of the absolute values of the coefficients (Tibshirani, 1996).

The objective function of LASSO can be formally defined as:

$$\min_{\beta} \left\{ \frac{1}{2n} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right\} \tag{1}$$

where $\|y - X\beta\|_2^2$ is the residual sum of squares, $\|\beta\|_1$ is the $L1$ norm of the parameter vector, and $\lambda \geq 0$ is a tuning parameter controlling the amount of shrinkage.

The distinctiveness of LASSO lies in its ability to shrink some of the coefficients to zero, thereby performing variable selection. This characteristic is especially beneficial in our context, as our objective is to generate precise forecasts, and only a subset of the predictors contributes significantly to the prediction outcome (Tibshirani, 1996).

To implement LASSO, we initially conduct a Grid Search technique for hyperparameter tuning to determine the most suitable value for the regularization parameter, λ, via cross-validation. This search is performed over a logarithmically spaced grid of values ranging from 10^{-5} to 10^{1}. Once the optimal λ value is identified, we then fit a LASSO model using this λ value and select the features for which the coefficients are not shrunk to zero. These selected features are then used as input for the subsequent VAR and LSTM models. Furthermore, a crucial step before feeding these selected variables into the models is their normalization. This process ensures all variable values are transformed to lie within a 0 to 1 range, providing consistency and preventing any influence from variable magnitudes on the models' learning process.

In our research, the LASSO method is employed before the VAR estimation. Through this process, we ensure that our model only includes the most relevant predictors, reducing dimensionality, improving interpretability, and potentially enhancing the the robustness of VAR predictions.

3 Forecasting Methods and Accuracy

3.1 Vector Autoregressive (VAR)

Vector Autoregressive (VAR) model is a popular multivariate time series model. It models linear dependencies among multiple features that evolve in time. A reduced form of the VAR model of order p and Gaussian error, is denoted as follows:

$$y_t = \nu + A_1 y_{t-1} + ... + A_p y_{t-p} + \epsilon_t,\ t = 1, .., T \tag{2}$$

where y_t is a $n \times 1$ vector of estimated endogenous variables; ν is also a $n \times 1$ vector of intercepts; $A_i\ \ (i = 1, ..., p)$ are $n \times n$ matrices denoting the model's coefficients of each endogenous variable; ϵ_t is a $n \times 1$ column vector that represents the error terms of each endogenous variable; and T is the length of the time series; p determines the number of lags of the model, i.e., the degree on which the data at time t is dependent on the data at time $t-p$. The distribution of the error term ϵ_t series is Gaussian, without autocorrelation, zero mean and constant variance.

Lag order selection is key when employing the VAR model. Selecting the number of lag orders is not straightforward for the VAR model. The total number of parameters in a VAR(p) model is $n + n^2p + n(n+1)/2$; where n is the intercepts vector, n^2p are the coefficients of the lagged independent variables, and $n(n+1)/2$ are the variances and covariances of the errors. In other words, the total number of parameters p in a VAR increases with the square of the VAR order. Therefore, zero restrictions on the parameter matrices may be desirable. We use the model selection criteria method to choose the optimal VAR order, and specifically the Akaike's information criterion (Akaike, 1974). The optimal VAR order is selected such that the AIC criterion is minimized over the possible orders $m = 0, \ldots, p_{max}$. We choose $p_{max} = 15$ so that the optimal parameter is in the set of possibilities.

Given a sample size T, $y_1, \ldots, y_T$, and p presample vectors $y_{-p+1}, \ldots, y_0$, the parameters of the VAR can be estimated efficiently using the least squares (LS) estimator. If no restrictions are imposed on the parameters, then the OLS estimator is identical to the generalized least squares (GLS) estimator. For a normally distributed process y_t with $\epsilon_t \sim \mathcal{N}(0, \sum_{\epsilon})$, the LS estimator is also identical to the maximum likelihood (ML) estimator. Since we assume normality of the error terms and we also find that the input salmon price series is normally distributed (see Section 2.3), we employ the OLS to estimate each equation of the VAR separately.

Once we select the optimal parameter and estimate the VAR model we test for residual autocorrelation and normality. Depending on the optimal VAR parameter either the Breusch-Godfrey-LM test or the Portmanteau test is employed. The Breusch-Godfrey-LM test is more suitable for low order autocorrelation and it examines whether the coefficient matrices in a VAR are equal to zero. The Portmanteau test is applied primarily to test for high order autocorrelation and it tests whether all residual autocovariances are zero, $E\left(\epsilon_t \epsilon'_{t-1}\right) = 0 \;\; (i = 1, 2, ...)$. For the purpose of this study it is assumed that the Portmanteau test is best applied on a VAR with order 5 and higher. Next, to test for normality of the residual series the Shapiro-Wilk test is employed. Considering that we are interested in examining whether a hybrid architecture that combines a standard VAR model with a neural network generates better forecasting results than a standard VAR alone, we do not test for residuals conditional heteroscedasticity. We assume that any such features will be captured by the neural network.

Once we estimate the VAR($\hat{p}$) and perform residuals diagnostics we use the y_t series generated by the process to forecast long-term salmon prices. The conditional expectation of y_{T+h} given y_t, $t \leq T$, is:

$$y_{T+h|T} = E\left(y_{T+h}|y_T, y_{T-1}, \dots\right) = \nu + A_1 y_{T+h-1|T} + \dots + A_p y_{T+h-p|T}, \tag{3}$$

where $y_{T+j|T} = y_{T+j}$ for $j \leq 0$. Considering that y_t with $\epsilon_t \sim \mathcal{N}\left(0, \sum_\epsilon\right)$, the forecast errors are also normal, $y_{T+h} - y_{T+h|T} \sim \mathcal{N}\left(0, \sum_y(h)\right)$.

3.2 Long Short-term Memory (LSTM)

LSTM is an advanced type of Recurrent Neural Network (RNN) able to carry out information for longer periods (Hochreiter & Schmidhuber, 1997). In combination with an appropriate gradient-based algorithm, the LSTM uses a memory cell and gates to store information for longer periods. Gers et al. (2000) specify that the cell state (C_t) represents the 'memory' feature of the network. The gates are mechanisms that filter the information passing through the cell state using activation functions. There are three gates: the forget gate (f_t), the input gate (i_t), and the output gate (o_t). The three gates (f_t, i_t, o_t), the cell state (c_t), and the hidden state (h_t) of the LSTM layer are described using the following equations:

$$f_t = \sigma\big[W_f h_{t-1} + U_f x_t + b_f\big], \tag{4}$$

$$i_t = \sigma\big[W_i h_{t-1} + U_i x_t + b_i\big], \tag{5}$$

$$\tilde{C}_t = \tanh\big[W_c h_{t-1} + U_c x_t + b_c\big], \tag{6}$$

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t, \tag{7}$$

$$o_t = \sigma\big[W_o h_{t-1} + U_o x_t + b_o\big], \tag{8}$$

$$h_t = o_t \tanh[C_t], \tag{9}$$

where h_{t-1} is the hidden state from period $t-1$, x_t the input from period t, and $\tilde{C}_t$ is a vector of values that determines how much 'new' information the cell state should receive. $\sigma(\cdot)$ and $\tanh(\cdot)$ denote the sigmoid and hyperbolic tan activation functions, respectively. W is a weight matrix associated with the hidden state, U is a weight matrix associated with the current input, and b is the bias term.

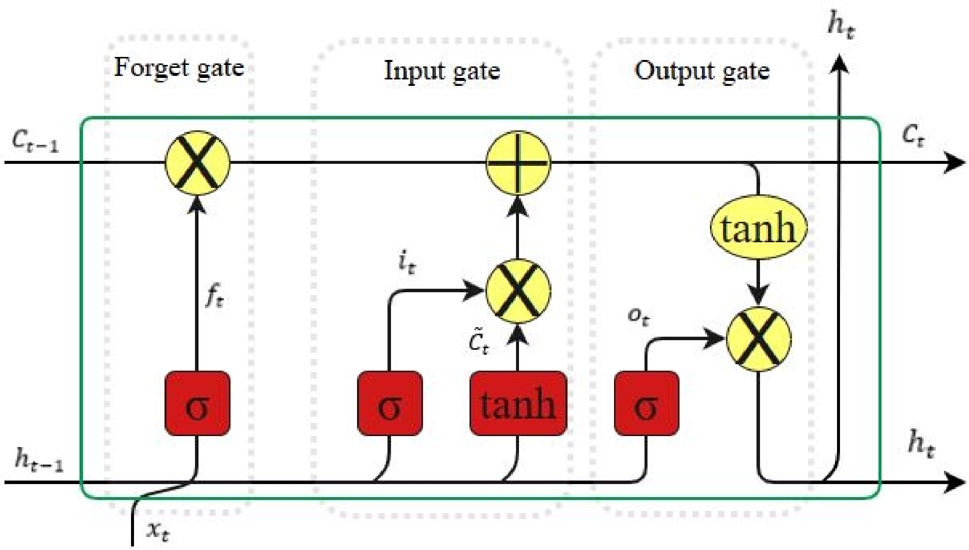


Figure 9: Long Short-term Memory (LSTM) neuron inspired by Gers and Schmidhuber (2000)

Fig.9 illustrates the structure of a hidden LSTM layer. First, the forget gate (f_t), takes as inputs the hidden state from the previous period (h_{t-1}), and the input from the current period (x_t), passes them via a sigmoid function that returns values between 0 and 1, where 0 means that "nothing passes through" and 1 means that "everything passes through" to the cell state. The output from the forget gate is described by Eq.4. The input gate (i_t), has two parts. The first part is the same as for the forget gate, and the output is denoted in Eq.5. The second part of the input gate passes the same inputs via a hyperbolic tan function which returns values between -1 and 1, and creates a vector of 'new' information ($\tilde{C}_t$), as in Eq.6. These two outputs from Eq.5 and 6 are added on the product of the forget gate's output (f_t) with the information carried from the previous cell state (C_{t-1}) to calculate the current period's cell state (C_t) as shown in Eq.7. The scalar multiplication is denoted as $\times$ and the addition is denoted as $+$ (see Fig9). Last, the output gate (o_t) takes the same inputs as the forget and the input gate and outputs Eq.8. In the meantime, the cell state (C_t) passes via a hyperbolic tan function which then is multiplied with the output from Eq.8 to separate necessary from unnecessary information and create the 'new' hidden state (h_t) which is carried on to the next period.

3.3 Hybrid Model: VAR-LSTM

VAR models are known to perform well on linear time-series data. Even though financial time-series data is rarely, perhaps never linear, VAR models are still widely used in financial time-series forecasting. Specifically, in the salmon price forecasting field, VAR models were employed by Guttormsen (1999) and Bloznelis (2018b). While Guttormsen (1999) found VAR to outperform other traditional forecasting models based on error measures, Bloznelis (2018b) found that the VAR performed poorly compared to a number of different models.

As discussed in Section 3.1, the residuals of the VAR process are linearly independent. Even though residual autocorrelation tests such as the Breusch-Godfrey-LM test are able to capture linear relationships, there can still be non-linear dependencies present. However, there are no general diagnostic tests for non-linear autocorrelation relationships. Therefore, to incorporate any non-linearity features of the residuals, we employ an integrated model architecture that combines a VAR model with a deep learning method, the Long Short-term Memory (LSTM). Modelling the residuals of a linear model using a neural network increases the possibility of capturing non-linear relationships in the residuals, improve forecasting performance, and produce more reliable results.

We consider that the spot logarithmic return series constitutes of two components, a linear and a non-linear one as follows:

$$y_{1,t} = L_{1,t} + N_{1,t}, \tag{10}$$

where $L_{1,t}$ denotes the linear component and $N_{1,t}$ denotes the non-linear component. The linear component is estimated using the VAR model. We retrieve the residuals from the VAR estimation and test for linear dependencies. Based on our tests, we assume that any remaining structure in the residuals can be attributed to non-linear relationships Let $\hat{\epsilon}_{1,t}$ denote the residual at time t from the VAR model:

$$\hat{\epsilon}_{1,t} = y_{1,t} - \hat{L}_{1,t}, \tag{11}$$

where $\hat{L}_{1,t}$ is the forecast value for time t generated from a VAR process.

Next, we model the residuals using an LSTM model, with k input nodes, the LSTM model configuration will be:

$$\hat{N}_{1,t} = f\left(\hat{\epsilon}_{1,t}, \hat{\epsilon}_{1,t-1}, \hat{\epsilon}_{1,t-2}, \ldots, \hat{\epsilon}_{1,t-k}\right) + u_{1,t}, \tag{12}$$

where f is a non-linear function determined by the LSTM and $u_{1,t}$ is the random error from employing the LSTM model. It is important to consider that if the non-linear function f is not a suitable one, the error term $u_{1,t}$ might not be random and hence identifying the model correctly is crucial. The forecast obtained from the hybrid architecture is:

$$\hat{y}_{1,t} = \hat{L}_{1,t} + \hat{N}_{1,t}, \tag{13}$$

$$\hat{L}_{1,t} = g(y_{1,t-1}, y_{1,t-2}, \ldots, y_{1,t-p}) + N_{1,t}, \tag{14}$$

$$\hat{N}_{1,t} = f(\hat{N}_{1,t-1}, \hat{N}_{1,t-2}, \ldots, \hat{N}_{1,t-k}) + u_{1,t}, \tag{15}$$

where g is the multivariate function defined by the VAR model, $y_{1,t}$ is the vector of predictor variables at time t, and p is the lag order of the VAR model, f is the multivariate function defined by the LSTM model, $\hat{N}_{1,t}$ is the vector of residuals at time t from the VAR model, and k is the number of time steps in the LSTM model.

A hybrid model utilizes the strength of each model, VAR and LSTM, in determining different data patterns. Therefore, we acknowledge the advantages of modelling linear and non-linear patterns separately by using two different models and then combine their predictions to improve the forecasting performance. This methodology is built upon the approach developed by Zhang (2003).

3.4 Proposed Model

The predictive variable in our VAR model is the salmon spot series, while the remaining series are the predictor variables. To ensure the data meet the assumption of stationarity, a staple in financial studies, we calculated the logarithmic returns for all variables. Subsequently, we split the data into training and testing sets following an 80-20 split, which allowed for validation of the models' predictive power.

Before employing the VAR, we used LASSO to reduce the dimensionality of the predictor variables and a grid search to tune the regularization hyperparameter, λ, across a range of values from 10^{-5} to 10^{1} (see Eq. 1). The selection of an optimal λ is crucial for the effectiveness of the LASSO and was achieved via cross-validation, targeting a value that minimized the out-of-sample error.

Following the determination of the best λ, we applied the LASSO method and derived a subset of explanatory variables with non-zero coefficients. The selected predictors—futures contracts prices, export volume, soybean prices, and SPI prices—were consistent across all VAR model versions. Thus, our application of the LASSO technique helps shape a parsimonious, yet robust model for predicting salmon spot prices.

Next, we estimate three different VAR models, each based on different futures contract maturities: one model uses the front-month futures contract, another the six-month futures contract, and the last the twelve-month futures contract. When estimating the VAR using the front-month futures contracts price series, the AIC criterion indicates that the optimal lag parameter is $p = 2$. The same optimal parameter selection method is applied on the different VAR models. The optimal lag parameter is $p = 1$ for all remaining models (using the six-month and twelve-month futures contracts). Upon estimating the VAR models, we retrieve their residual series.

Next, residual diagnostics are performed to test their adequacy. The Breusch-Godfrey-LM test and Autocorrelation Functions (ACF) were utilized to determine if the residuals demonstrate linear independence over time. Our results pointed to evidence of autocorrelation in the residuals across all three VAR models, according to the Breusch-Godfrey LM test. Given that the optimal parameter for all VAR models is less than 5 (see Section 3.1), we chose the Breusch-Godfrey-LM test over the Portmanteau test. We further substantiated the presence of autocorrelation in the residuals using the ACF (see Appendix A.1). Next, we applied the Shapiro-Wilk normality test to the residual series for each model. When using the front-month and twelve-month futures contracts, the Shapiro-Wilk test suggested normality in both residual series at a 5% significance level. However, when incorporating the six-month futures contracts, the residual series did not demonstrate normality at the 5% significance level.

We use estimated VAR models for the purpose of forecasting individual series of salmon spot prices. Three distinctive VAR models are utilized in this analysis. The first VAR(2) model projects short-term forecast horizons of $h = 1$ and $h = 4$, while the subsequent VAR(1) models are configured with longer-term horizons of $h = 26$ and $h = 52$. The forecasting horizon in each model is purposefully aligned with the maturity period of the corresponding futures contracts price series, ensuring coherence between the forecast and the financial instrument in question.

To confirm the efficacy of the forecasting strategy, we implement three varied window methods: a fixed window, an expanding window, and a rolling window approach. Each estimation window is set to 52, closely mirroring the full life cycle of salmon production. Among these, the rolling window approach yields the lowest error metrics and is thus selected to generate the forecasts. This technique advances each estimation window $h-$steps forward, generating $h-$step ahead forecasts until all data in the testing subset has been utilized. The predictive capabilities of each VAR model are then evaluated using forecast error measures.

However, this methodology is not without its limitations. The futures contracts price generally encapsulates an anticipated spot price for an entire month, not a specific week, which may decrease the precision of forecasts predicting weekly salmon spot prices. Despite this shortcoming, the futures contracts price series remain a valuable forecasting instrument, contributing significant insights into the potential trajectory of the spot price series.

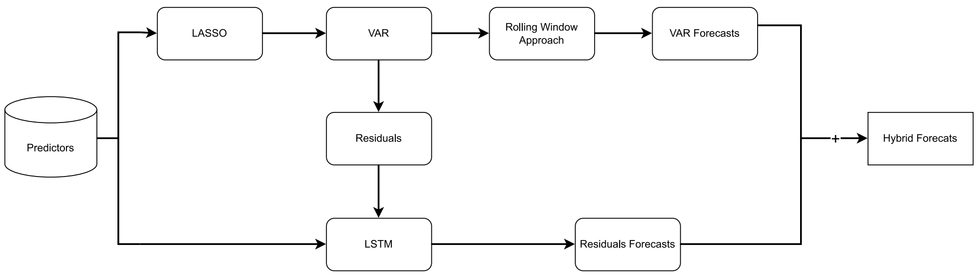


Figure 10: A flow chart of the proposed VAR-LSTM hybrid architecture.

Out of the n sub-equations estimated in each VAR model, our primary interest lies in forecasting $y_{1,t}$ for $n = 1$ (see Eq.2). As such, we utilize the residual series from the $y_{1,t}$ equation as input for the LSTM model. It's important to note that these residual series for each VAR model contain fewer observations than the logarithmic returns series, due to the influence of degrees of freedom.

Deep learning models, such as LSTM, utilize a method known as 'supervised learning'. In supervised learning, the model is trained on a labelled dataset. In the case of time-series data, a lag transformation is employed to create these labels. To convert time-series data into a supervised learning format, a lag transformation is employed. Here, the value at time $(t-k)$ (where k is the number of lags) serves as the input variable, while the value at time $(t+k)$ acts as the predictive or target variable.

Hyperparameters	**Selection range**
LSTM layer	
Units	$\{10, 20, 3i, \dots, 8i\}$ where $i = 10$
Activation function	{tanh, ReLu, sigmoid, softmax}
Dropout	{0, 0.1, 0.2, 0.3, 0.4, 0.5}
"ADAM" optimization	
Learning rate	{0.01, 0.001, 0.0001}
epsilon	{0.0001, 0.00001, 0.01, 0.001, 1.0}

Table 3: Configuration Grid for Neural Networks' Hyperparameter selection range.

The LSTM model processes the input variables from time $(t-52)$ to time t. This data is then passed through an LSTM layer, which has a varying number of LSTM neurons, referred to as 'Units' in Table 3. After processing through the LSTM layer, the data is forwarded via a fully-connected dense layer. This dense layer plays a critical role in translating the high-level features learned by the LSTM into predictions, eventually producing forecasts $\hat{\epsilon}_{1,t+h}$ (or $\hat{N}_{1,t+h}$ for consistency with Eq. 14) at $(t+h)$, where h represents the different forecasting horizons. Regularization methods, such as L1 and L2, are also implemented in the LSTM and dense layers to prevent overfitting by adding a penalty term to the loss function. The specific hyperparameters for these regularization techniques were determined during the tuning process. An illustrative flow chart of the VAR-LSTM model predicting salmon spot prices is presented in Figure 10.

For consistency, the size of the test sample is equal to the one used for the VAR models. Therefore, we must ensure that the test sample consists of the last 83 observations. The train data is scaled between -1 and 1 to accelerate the training process.

Hyperparameter tuning is essential in machine learning model training to achieve the best performance by selecting the optimal set of hyperparameters. For this task, we employed the Random Search tuning technique. This method involves randomly sampling

(a) Hybrid Model: VAR(2)-LSTM

	Layer		
	Neurons	Activation	Dropout
LSTM	10	tanh	0.5
Dense	1	tanh	
	Optimizer		
Learning rate		0.0001	
Epsilon		0.00001	

(b) Hybrid Model: VAR(2)-LSTM

	Layer		
	Neurons	Activation	Dropout
LSTM	10	tanh	0.5
Dense	4	tanh	
	Optimizer		
Learning rate		0.0001	
Epsilon		0.00001	

(c) Hybrid Model: VAR(1)-LSTM

	Layer		
	Neurons	Activation	Dropout
LSTM	10	tanh	0.5
Dense	26	tanh	
	Optimizer		
Learning rate		0.0001	
Epsilon		0.00001	

(d) Hybrid Model: VAR(1)-LSTM

	Layer		
	Neurons	Activation	Dropout
LSTM	10	tanh	0.5
Dense	52	tanh	
	Optimizer		
Learning rate		0.0001	
Epsilon		0.00001	

Table 4: Optimal hyperparameters based on the RandomSearch tuning technique for each neural network based on window size = 52 and the corresponding forecast horizon h. (a) based the input residuals series from the VAR(2) model and $h = 1$; (b) based the input residuals series from the VAR(2) model and $h = 4$; (c) based on the input residuals series from the VAR(1) model and $h = 26$; (d) based on the input residuals series from the VAR(1) model and $h = 52$.

combinations of hyperparameters from a predefined search space, and is utilized with the training data. The potential hyperparameters are presented in Table 3. This approach, described by Bergstra and Bengio (2012), offers an efficient and less computationally-intensive alternative to exhaustive searches, such as grid search.

We repeated this procedure three times, each time altering the input data to the predictor variables and the residual series from the corresponding VAR model. The optimal hyperparameters identified for each model are presented in Table 4.

To compile the network, we utilize the "mean squared error" as the loss function and "ADAM" as the optimization algorithm (Kingma & Ba, 2014). The learning rate hyperparameter is set to 0.0001 for all three models, as indicated by the selection process. The model was trained for 200 epochs with a stopping condition, known as 'early stopping,' which is defined as no improvement in validation loss after 20 epochs. This condition was evaluated on the validation data, preventing the model from overfitting by stopping it from learning noise in the training data and saving computational resources.

To ensure the convergence of the training process, we examine the loss function for both the training and test samples (refer to Appendix A.2). Following the training process, we obtain the residual forecasts and add them to the VAR forecasts to form the forecasts of the hybrid models. The forecasting performance of these hybrid models is then compared to that of the VAR models using various forecast error measures.

3.5 Loss Functions

To assess the predictive accuracy of the models, we use loss functions. The mean squared error (MSE), the mean absolute error (MAE), the root mean squared error (RMSE), and

Theil's $U-$statistic. These are defined as follows:

$$MSE = \frac{1}{T}\sum_{t=1}^{T}(y_{1,t} - \hat{y}_{1,t})^2, \tag{16}$$

$$MAE = \frac{1}{T}\sum_{t=1}^{T}\left|y_{1,t} - \hat{y}_{1,t}\right|, \tag{17}$$

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_{1,t} - \hat{y}_{1,t})^2}, \tag{18}$$

$$U = \sqrt{\frac{\sum_{t=1}^{T}(y_{1,t} - \hat{y}_{1,t})^2}{\sum_{t=1}^{T} y_{1,t}^2}}. \tag{19}$$

These metrics compute the total error between the forecast $\hat{y}_{1,t}$ and the actual $y_{1,t}$ for $n = 1$, which corresponds to the salmon spot price (see Eq. 2 and 14). After calculating these forecasting error metrics (see Eq.17-19) for both the hybrid model and the VAR, we can compare their predictive capabilities. The best model is indicated by lower values of MSE, MAE, and RMSE, and a Theil's U statistic value closer to 0.

3.6 The Diebold–Mariano test

The comparison of the performance of two distinct forecasts can be achieved via the DM test (Diebold & Mariano, 2002). The test's null hypothesis posits that both forecast models have an identical expected loss due to forecast errors, suggesting that the models have comparable accuracy. Forecast errors from the two distinct models are converted into associated losses via chosen loss functions. If the expected value of the loss function for each forecast is equal, then the mean of the population in the loss differential series should be zero.

For the DM test, we generate two series of volatility forecasts, $\hat{\sigma}_{i,N}, ..., \hat{\sigma}_{i,N}$ where $i = 1, 2$, derived from two distinct forecasting models. The next step involves measuring the accuracy of these forecasts against the series of volatility proxies denoted as $\sigma_1, ..., \sigma_N$. We employ two different loss functions: Mean Squared Error (MSE), $L(\hat{\sigma}_{i,N}, \sigma_N) = (\hat{\sigma}_{i,N} - \sigma_N)^2$, and Mean Absolute Error (MAE), $L(\hat{\sigma}_{i,N}, \sigma_N) = |\hat{\sigma}_{i,N} - \sigma_N|$. The DM test's null hypothesis, $E(d_t) = 0$, involves a loss differential sequence $d_t = L(\hat{\sigma}_{1,t}, \sigma_t) - L(\hat{\sigma}_{2,t}, \sigma_t)$

with the chosen loss function $L(\bullet)$. The DM statistic is expressed as:

$$DM = \frac{\frac{1}{N}\sum_{t=1}^{N} d_t}{\sqrt{2\pi \hat{f}_d(0)/N}}, \tag{20}$$

where N represents the sample size and $\hat{f}_d(0)$ is a reliable estimate of $f_d(0)$, which corresponds to the spectral density of the loss differential at frequency 0.

4 Empirical Results

We explored the potential of two distinct models, the VAR and the VAR-LSTM hybrid, for forecasting $h-$steps ahead. Our methodology incorporated creating three unique VAR models, adjusting the predictor variable on futures contracts prices based on their respective time to maturity (e.g., front-month, six-month, and twelve-month). The selected futures contract for prediction changes in accordance with the forecast horizon. For instance, for short-term forecasts such as $h = 1$ and 4, we employ the front-month futures series. On the other hand, when making forecasts for a medium-term horizon, specifically for $h = 26$ and 52, we employ futures series that cover a period of six and twelve months, respectively. Implementing this strategy, we successfully developed three diverse hybrid models.

Table 5 presents a comparison of the out-of-sample forecast accuracy between the VAR and the VAR-LSTM model, at different forecast horizons. The metrics used for evaluation are Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Theil's U statistic. Interestingly, the VAR and VAR-LSTM models exhibit near identical performance across all horizons and measures, indicating that the inclusion of LSTM in a hybrid framework did not significantly improve the forecasting accuracy over the single VAR model.

In terms of MSE, both models exhibit a slight increase in error as the forecast horizon expands, while MAE and RMSE values show minimal differences between the two models across all forecasting horizons. Moreover, Theil's U statistic, a relative measure of forecast accuracy, implies similar predictive ability for both models at each forecasting horizon.

Measure	Horizon	VAR	VAR-LSTM
MSE	1	**0.00390**	**0.00390**
	4	**0.00435**	**0.00435**
	26	**0.00486**	**0.00486**
	52	**0.00487**	**0.00486**
MAE	1	0.04975	0.04973*
	4	0.05364	0.05365*
	26	0.05636	0.05635*
	52	0.05638	0.05637*
RMSE	1	**0.06249**	**0.06249**
	4	**0.06593**	**0.06593**
	26	**0.06974**	**0.06974**
	52	**0.06976**	**0.06976**
Theil's U	1	0.77160	0.77154*
	4	0.92434	0.92433*
	26	1.00326	1.00318*
	52	1.00350	1.00343*

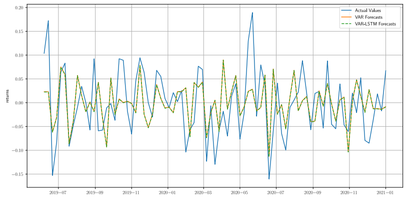
Note: Forecasting error metrics in bold indicate the same forecasting performance and those with an asterisk indicate a better forecasting performance than its peer.

Table 5: Out-of-sample forecast accuracy of the VAR and the hybrid VAR-LSTM under different forecast horizons.

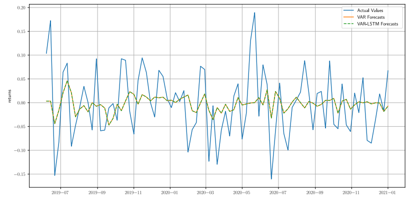
Notably, the Theil's U values are below 1 at shorter horizons (1 and 4), implying a satisfactory forecasting performance of both models. However, the Theil's U statistic slightly surpasses 1 at the longer horizons of 26 and 52, suggesting a less than perfect fit to the actual data at these forecasting horizons.

These findings indicate that both the VAR and VAR-LSTM models offer comparable performance across multiple forecasting horizons and forecasting error metrics, with minor differences that do not unequivocally favor one model over the other. The results prompt a need for further investigation to establish whether the hybrid approach's added complexity translates to a tangible advantage in this forecasting context.

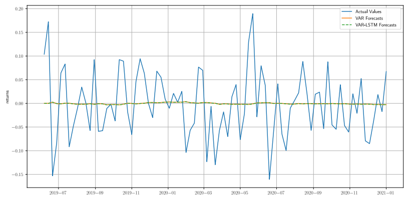
Figure 11 depicts the actual and forecasted salmon spot prices at different forecasting horizons using the VAR and VAR-LSTM models. It is evident from these figures that the forecast trajectories of both the VAR and VAR-LSTM models closely coincide across all



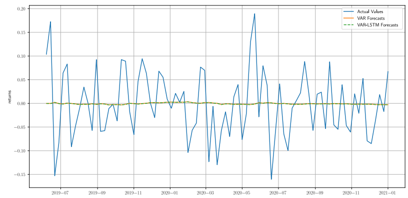
(A)



(B)



(C)



(D)

Figure 11: Actual and forecasted salmon spot prices at different forecasting horizons using VAR, and VAR-LSTM. (A); forecasting 1-step ahead (B); forecasting 4-steps ahead (C); forecasting 26-steps ahead (D); forecasting 52-weeks ahead.

forecasting horizons. This denotes a striking resemblance in their performance, suggesting that the incorporation of LSTM in a hybrid framework does not yield an apparent improvement in the forecast outcomes over the single VAR model. The striking similarity in the behavior of these models suggests a lack of significant non-linearity in the salmon spot price series. Thus, the VAR-LSTM model may not have any complex patterns to exploit for improved forecasting. Even though LSTMs are renowned for detecting and learning non-linear patterns, the comparable performance of the two models suggests that the salmon spot price series likely depicts linear patters, adequately captured by the VAR model, with no substantial non-linear patterns to exploit. Consequently, it appears that the added complexity of the VAR-LSTM model does not offer a significant advantage in this specific forecasting context. These observations highlight the importance of understanding the underlying dynamics of the time series in question before opting for more complex models, as simpler linear models might provide equally effective forecasts in certain scenarios.

Our findings, which show an absence of exploitable non-linear patterns in the salmon spot price series, validate a key characteristic of financial markets: the adaptiveness of investors. This trait is widely documented in behavioral finance literature, reflecting a 'learning by doing' process (Gervais & Odean, 2001; Shiller, 2003). Given the dynamic nature of markets, rapidly identified trends can swiftly lose their relevance or even disappear, posing challenges to their effectiveness in future price predictions. Our results align well with this theory.

This behavior might be an essential driver of market dynamics, possibly resulting in the linear pattern we observed in the salmon spot price series. The complex interplay between investor behavior and market patterns is emphasized in these findings. In financial markets where changes are continual, the predictive advantage of more complex models like the VAR-LSTM may be constrained.This limitation arises as these models depend on past data for future predictions. However, in adaptive markets, past patterns may not reliably forecast future trends due to investors' continuous strategic adjustments.

Horizon	Loss function	DM-Test	$p-$value
1	MSE	-1.1023	0.999991*
	MAE	-0.00147	0.99883*
4	MSE	0.0000	0.9999998*
	MAE	0.0000	0.999999*
26	MSE	0.0000	0.999994*
	MAE	-0.00014	0.99989*
52	MSE	0.0000	0.999994*
	MAE	0.00014	0.999889*

Note: The * indicates that there is no sufficient evidence to reject the Diebold-Mariano's hypothesis and therefore there is no significant difference between the forecasts generated by the two models.

Table 6: Comparing VAR and VAR-LSTM forecasts using the Diebold-Mariano (DM) test with MSE and MAE loss functions.

From the results presented in Table 6, we used the Diebold-Mariano (DM) test to evaluate the forecasts generated by the VAR and VAR-LSTM models across different forecasting horizons. The Mean Squared Error (MSE) and Mean Absolute Error (MAE) were chosen as loss functions to give a comprehensive picture of the predictive accuracy, with MSE emphasizing larger errors and MAE representing the average magnitude of the errors. For each of the forecast horizons (1, 4, 26, and 52), both MSE and MAE loss functions resulted in DM test statistics close to zero and very high p-values indicating no statistically significant differences between the VAR and VAR-LSTM forecasts. This lack of variance between the results of the two models further highlights the comparative robustness of the two models across varying forecasting periods in forecasting the salmon spot prices.

These results also verify that while the VAR-LSTM hybrid model introduces an additional level of complexity, it does not significantly outperform the traditional VAR model in terms of forecasting accuracy for salmon spot prices. This may suggest that the data does not contain additional complex dynamics that the LSTM component could exploit. These findings contribute valuable insights for future research in hybrid models, suggesting that increased complexity does not necessarily lead to improved forecasting accuracy.

Measure	Horizon	VAR	Random-walk (with drift)
MSE	1	0.00390*	0.00877
	4	0.00435*	0.00840
	26	0.00486*	0.00941
	52	0.00487*	0.00557
MAE	1	0.04975*	0.07392
	4	0.05364*	0.07415
	26	0.05636*	0.08311
	52	0.05638*	0.05920
Note: Forecasts with an asterisk have a better forecasting performance than its peer.			

Table 7: Out-of-sample forecast accuracy of the VAR and the hybrid VAR-LSTM under different forecast horizons.

Since our results indicate that there are no non-linear patterns in the salmon spot price data to be exploited by incorporating an LSTM, we decide to further enhance the robustness of the VAR model by comparing its forecasting ability with that of a random walk with a drift. Table 7 presents the results of applying the MSE and MAE forecasting error measures on various forecasting horizons to compare the forecasting ability of the VAR with that of a random walk with a drift.

The results reveal that the VAR model consistently outperforms the random walk model with a drift, at all considered forecast horizons. This is demonstrated by lower MSE and MAE values, indicating smaller forecast errors for the VAR model. This is a significant result as a random walk with drift is often a challenging benchmark to beat in financial forecasting due to the random nature and efficient market hypothesis. These results therefore provide a strong support for the effectiveness of the VAR model in forecasting salmon spot prices and underline the importance of this model as a powerful tool for such predictions. It also suggests that while salmon spot prices may possess random walk properties, they are not purely random and do exhibit predictable dynamics that can be captured by VAR.

5 Concluding remarks

The objective of this study was to determine whether a hybrid VAR-LSTM model could improve the accuracy of salmon spot price forecasts and compare its performance with that of a traditional VAR model. We adopted a comprehensive and robust approach involving a detailed series of steps, which included dimensionality reduction using the LASSO technique, VAR model estimation, residuals diagnostics, forecast generation, and, finally, the introduction of deep learning through LSTM. Despite the expectation that integrating VAR and LSTM methods into a hybrid model would enhance forecasting performance, our results showed no statistically significant difference in predictive accuracy between the single VAR and the hybrid VAR-LSTM model.

Interestingly, the forecast trajectories of both the VAR and VAR-LSTM models closely coincided across all forecasting horizons, suggesting the absence of significant non-linearity in the salmon spot price series. This implies that the VAR-LSTM model found no complex patterns to exploit for improved forecasting. It seems that the added complexity of the VAR-LSTM model does not offer a significant advantage in this specific forecasting context. These observations emphasize the importance of understanding the underlying dynamics of the time series before opting for more complex models, as simpler linear models might provide equally effective forecasts in certain scenarios.

Our findings highlight that investors in financial markets often adapt their strategies, which in turn impact the development of salmon spot prices. We observed that investor behavior, which involves continual strategic adjustments, might lead to a linear pattern in the salmon spot price series. This behavior suggests a limitation to the predictive advantage of complex models like VAR-LSTM in markets where past patterns may not reliably forecast future trends due to the adaptive nature of these markets.

The robustness checks confirmed the comparable performances of the VAR and VAR-LSTM models across varying forecasting horizons, emphasizing that the increased complexity of LSTM did not enhance the forecast accuracy beyond the traditional VAR model's capabilities. This study serves as a valuable reference for future research into hybrid models, suggesting that increased complexity does not automatically correlate with improved forecasting accuracy.

Lastly, we benchmarked the VAR model against a random walk with a drift model. We found that the VAR model consistently outperformed the random walk model across all considered forecast horizons. This result provides strong support for the VAR model's effectiveness in forecasting salmon spot prices, indicating its importance as a powerful tool for such predictions. It also suggests that salmon spot prices, although potentially exhibiting random walk properties, are not purely random and do contain predictable dynamics that can be captured effectively by VAR.

In conclusion, the results suggest that while hybrid models can theoretically combine the strengths of individual methodologies, they do not necessarily lead to improved forecast accuracy. Sometimes, simpler models, such as VAR, can perform just as effectively, if not better, than their more complex counterparts. Future research could explore other hybrid models or leverage emerging deep learning techniques, such as Transformers, to predict salmon spot prices. Exploring how these methods perform across different markets could also be insightful, potentially leading to more accurate and reliable salmon price forecasts in the future.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716–723.

Andersen, B. P., & de Lange, P. E. (2021). Efficiency in the atlantic salmon futures market. *Journal of Futures Markets*, *41*(6), 949–984.

Ankamah-Yeboah, I., Nielsen, M., & Nielsen, R. (2017). Price formation of the salmon aquaculture futures market. *Aquaculture Economics & Management*, *21*(3), 376–399.

Asche, F. (2008). Farming the sea. *Marine Resource Economics*, *23*(4), 527–547.

Asche, F., Dahl, R. E., & Steen, M. (2015). Price volatility in seafood markets: Farmed vs. wild fish. *Aquaculture Economics & Management*, *19*(3), 316–335.

Asche, F., Misund, B., & Oglend, A. (2016a). Determinants of the atlantic salmon futures risk premium. *Journal of Commodity Markets*, *2*(1), 6–17.

Asche, F., Misund, B., & Oglend, A. (2016b). The spot-forward relationship in the atlantic salmon market. *Aquaculture Economics & Management*, *20*(2), 222–234.

Asche, F., Misund, B., & Oglend, A. (2019). The case and cause of salmon price volatility. *Marine Resource Economics*, *34*(1), 23–38.

Baumeister, C., & Kilian, L. (2012). Real-time forecasts of the real price of oil. *Journal of business & economic statistics*, *30*(2), 326–336.

Baumeister, C., & Kilian, L. (2015). Forecasting the real price of oil in a changing world: A forecast combination approach. *Journal of Business & Economic Statistics*, *33*(3), 338–351.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of machine learning research*, *13*(2).

Bloznelis, D. (2016). Salmon price volatility: A weight-class-specific multivariate approach. *Aquaculture economics & management*, *20*(1), 24–53.

Bloznelis, D. (2018a). Hedging salmon price risk. *Aquaculture Economics & Management*, *22*(2), 168–191.

Bloznelis, D. (2018b). Short-term salmon price forecasting. *Journal of Forecasting*, *37*(2), 151–169.

Chen, X., & Scholtens, B. (2019). The spot-forward relationship in the atlantic salmon market. *Reviews in Fisheries Science & Aquaculture*, *27*(2), 142–151.

Dahl, R. E., & Oglend, A. (2014). Fish price volatility. *Marine Resource Economics*, *29*(4), 305–322.

Dahl, R. E., & Jonsson, E. (2018). Volatility spillover in seafood markets. *Journal of Commodity Markets*, *12*, 44–59.

Dahl, R. E., Oglend, A., & Yahya, M. (2021). Salmon stock market prices revealing salmon price information. *Marine Resource Economics*, *36*(2), 173–190.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, *74*(366a), 427–431.

Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & economic statistics*, *20*(1), 134–144.

Egerton, S., Wan, A., Murphy, K., Collins, F., Ahern, G., Sugrue, I., Busca, K., Egan, F., Muller, N., Whooley, J., et al. (2020). Replacing fishmeal with plant protein in atlantic salmon (salmo salar) diets by supplementation with fish protein hydrolysate. *Scientific reports*, *10*(1), 1–16.

Ewald, C.-O., Haugom, E., Kanthan, L., Lien, G., Salehi, P., & Størdal, S. (2022). Salmon futures and the fish pool market in the context of the capm and a three-factor model. *Aquaculture Economics & Management*, *26*(2), 171–191.

Ewald, C.-O., Ouyang, R., & Siu, T. K. (2017). On the market-consistent valuation of fish farms: Using the real option approach and salmon futures. *American Journal of Agricultural Economics*, *99*(1), 207–224.

Gers, F. A., & Schmidhuber, J. (2000). Recurrent nets that time and count. *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, *3*, 189–194.

Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with lstm. *Neural computation*, *12*(10), 2451–2471.

Gervais, S., & Odean, T. (2001). Learning to be overconfident. *The review of financial studies*, *14*(1), 1–27.

Guttormsen, A. G. (1999). Forecasting weekly salmon prices: Risk management in fish farming. *Aquaculture Economics & Management*, *3*(2), 159–166.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice*. OTexts.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kristjanpoller, W., & Minutolo, M. C. (2015). Gold price volatility: A forecasting approach using the artificial neural network–garch model. *Expert systems with applications*, *42*(20), 7245–7251.

Kristjanpoller, W., & Minutolo, M. C. (2016). Forecasting volatility of oil price using an artificial neural network-garch model. *Expert Systems with Applications*, *65*, 233–241.

Larsen, T. A., & Asche, F. (2011). Contracts in the salmon aquaculture industry: An analysis of norwegian salmon exports. *Marine Resource Economics*, *26*(2), 141–150.

Marine Harvest. (2020). *Salmon farming industry handbook*.

Misund, B., & Asche, F. (2016). Hedging efficiency of atlantic salmon futures. *Aquaculture Economics & Management*, *20*(4), 368–381.

Oglend, A. (2013). Recent trends in salmon price volatility. *Aquaculture Economics & Management*, *17*(3), 281–299.

Oglend, A., & Sikveland, M. (2008). The behaviour of salmon price volatility. *Marine Resource Economics*, *23*(4), 507–526.

Oglend, A., & Straume, H.-M. (2019). Pricing efficiency across destination markets for norwegian salmon exports. *Aquaculture Economics & Management*, *23*(2), 188–203.

Parot, A., Michell, K., & Kristjanpoller, W. D. (2019). Using artificial neural networks to forecast exchange rate, including var-vecm residual analysis and prediction linear combination. *Intelligent Systems in Accounting, Finance and Management*, *26*(1), 3–15.

Ramyar, S., & Kianfar, F. (2019). Forecasting crude oil prices: A comparison between artificial neural networks and vector autoregressive models. *Computational Economics*, *53*(2), 743–761.

RL, M., & Mishra, A. K. (2021). Forecasting spot prices of agricultural commodities in india: Application of deep-learning models. *Intelligent Systems in Accounting, Finance and Management*, *28*(1), 72–83.

Shiller, R. J. (2003). From efficient markets theory to behavioral finance. *Journal of economic perspectives*, *17*(1), 83–104.

Solibakke, P. B. (2012). Scientific stochastic volatility models for the salmon forward market: Forecasting (un-) conditional moments. *Aquaculture Economics & Management*, *16*(3), 222–249.

Straume, H.-M. (2014). Currency invoicing in norwegian salmon export. *Marine Resource Economics*, *29*(4), 391–409.

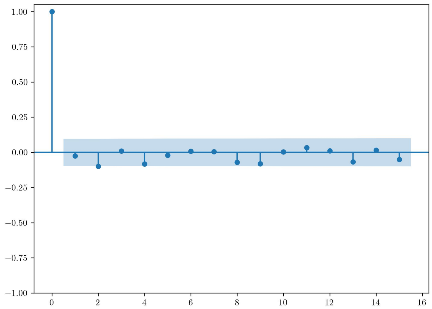
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Wang, J., & Li, X. (2018). A combined neural network model for commodity price forecasting with ssa. *Soft Computing*, *22*(16), 5323–5333.

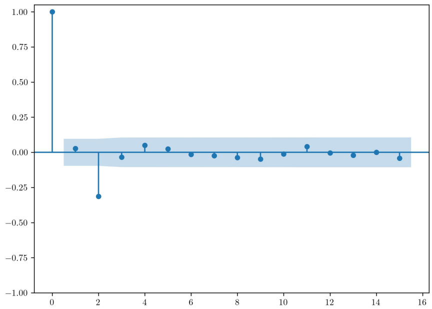
Xu, X., & Zhang, Y. (2022). Commodity price forecasting via neural networks for coffee, corn, cotton, oats, soybeans, soybean oil, sugar, and wheat. *Intelligent Systems in Accounting, Finance and Management*, *29*(3), 169–181.

Zhang, G. P. (2003). Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, *50*, 159–175.

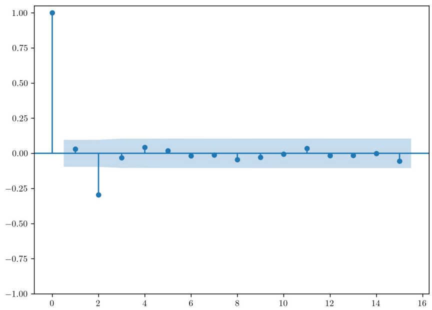
A Appendix



(A)

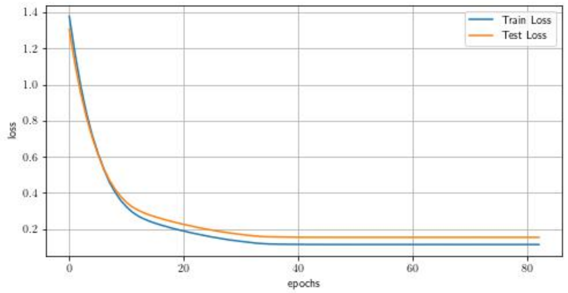


(B)

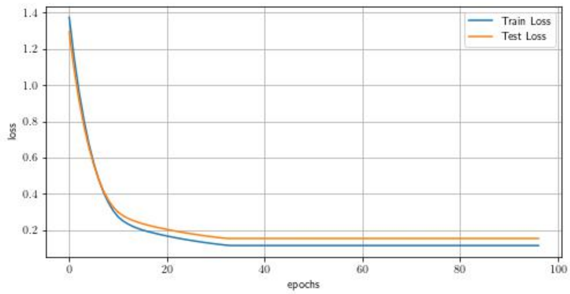


(C)

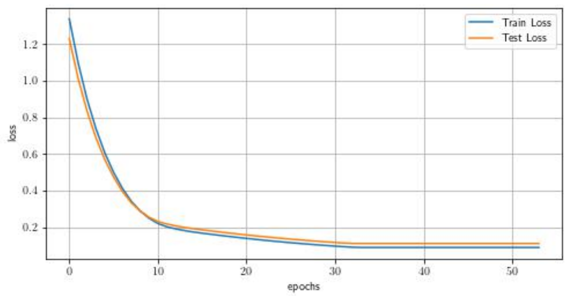
Figure A.1: Autocorrelation functions (ACF) up to 15 lags for the residuals series of the three versions of the VAR models: (A) VAR(2) model implementing front-month futures contracts prices; (B) VAR(1) model implementing six-month futures contracts; (C) VAR(1) implementing twelve-month futures contracts.



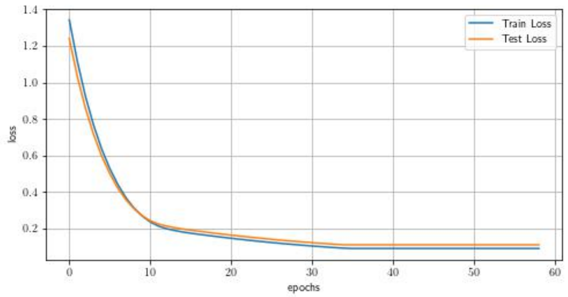
(A)



(B)



(C)



(D)

Figure A.2: Train and Test sample loss functions for the hybrid LSTM forecasting: (A) $h = 1$; (B) $h = 4$; (C) $h = 26$; (D) $h = 52$.

Research Article IV

Salmon stock returns around market news

Clemens Knoppe[1], Mariia Okuneva[2], and Mikaella Zitti[3]

[1]Kiel University, Institute of Economics, Kiel, Germany, knoppe@economics.uni-kiel.de

[2]Kiel University, Institute of Statistics and Econometrics, Kiel, Germany, mokuneva@stat-econ.uni-kiel.de

[3]Department of Economics and Resource Management, Norwegian University of Life Sciences, Ås, Norway, mikaella.zitti@nmbu.no

Abstract

We examine the relationship between salmon-related news and trading behaviour in the salmon market. For this, we create a share price index (SPI) based on five salmon aquaculture companies trading on the Oslo Stock Exchange (OSE) with the highest market cap. We utilise the Latent Dirichlet Allocation (LDA) machine learning algorithm to derive topics from the news data set. Further, we adopt a lexicon-based sentiment analysis, assigning a sentiment score to each article based on the number of negative and positive words it contains. These methods impose structure on the otherwise unstructured text data, enabling the application of standard econometric analyses to identify effects of news on stock returns. To explore the impact of topics on market volatility, we extract principal components based on all the topics. We find that the components related to COVID-19 and corporate news and stocks exert the most pronounced effect on the market. A surprise increase in sentiment series, based on the Loughran-McDonald lexicon, resulted in a marginally significant effect on logarithmic returns with an unexpected sign, highlighting the limitations of a lexicon exclusively focused on financial-related words for industry-specific studies. However, we can overcome this issue by adding relevant words to the dictionary. Taking into account the competitive market structure of salmon markets, we invert sentiment in articles that pertain to competitor firms. By making these adjustments, we establish a positive correlation between extended sentiment and returns. Moreover, upon extracting components based on topics and extended sentiment, we discover that the components related to business, algal blooms and Chilean aquaculture, and the salmon industry play a pivotal role in driving the market. Importantly, through our out-of-sample forecasting experiment, we provide compelling evidence that the incorporation of news data can significantly improve the performance of models predicting stock market price movements.

Keywords— Salmon Market; Text Mining; Topics Modelling; LDA; Sentiment Analysis

1 Introduction

The aquaculture industry has experienced rapid growth as a food production sector (FAO, 2018; Garlock et al., 2020), with global production surging from approximately 14.9 million tonnes in 1995 to 82.1 million tonnes in 2018 (FAO, 2020). Originating in the late 1960s, the Norwegian aquaculture industry has developed into one of the country's most vital export industries alongside oil and gas within a span of 50 years. Salmon has emerged as one of the most successful species in the industry (Asche, 2008), leading Norway to become the world's largest salmon producer, accounting for over half of the total production (Hersoug, 2021). The industry's expansion has facilitated its consolidation, simultaneously generating increased demand, elevated salmon prices, and augmented volatility (Asche et al., 2019; Bloznelis, 2016; Oglend, 2013).

Guttormsen (1999) showed that salmon is a volatile commodity, with its volatility more than doubling and now surpassing that of most comparable commodities (Dahl & Oglend, 2014). Previous research has revealed that volatility exhibits clustering (Dahl & Yahya, 2019), and that it is partially explained by trends in other food commodities (Oglend, 2013). Building upon this literature, our study investigates the types of news that influence the salmon market. Given that investor behavior is driven by news surrounding the companies they are invested in, we analyze the impact of salmon production-related news articles on stock prices, which have been found to be more reactive than commodity prices (Dahl et al., 2021).

Text mining methods have yielded promising results in macroeconomics- and finance-related studies. Hansen et al. (2018) employed the Latent Dirichlet Allocation (LDA), a topic modeling algorithm, to gauge the effect of transparency on monetary policymakers' deliberation. Similarly, Larsen (2021) used LDA to classify various news types and identify the type of uncertainty that significantly affects the economy. Further in financial literature, topics extracted with LDA have been used to predict stock movements (Nguyen et al., 2015). Given its ability to provide an overview of the primary topics discussed in a news dataset, we also incorporate LDA in our analysis. Additionally, we examine news sentiment using a lexicon-based approach, as it has been widely applied in stock market analyses (Karalevicius et al., 2018; Khedr, Yaseen, et al., 2017; Li et al., 2020; Li et al., 2014). Our focus is on the lexicon constructed by Loughran and McDonald (2011, henceforth: LM) as it is specifically tailored for financial applications. Thus, by utilizing text mining approaches such as LDA and lexicon-based sentiment analysis, our goal is to uncover how news topics and sentiment influence investor behavior in the salmon market.

Complementing text mining techniques, we utilize a Vector Autoregressive (VAR) model to investigate the relationship between investors' trading behavior and salmon market-related financial news. To capture investors' trading behavior, we examine stock price data from the top five salmon market producer companies in terms of market capitalization, listed on the Oslo Stock Exchange (OSE): Marine Harvest (MOWI), SalMar (SALM), Grieg Seafood (GSF), Lerøy Seafood Group (LSG), and Bakkafrost (BAKKA). Although futures contracts trading is available in the salmon market, its low liquidity and sporadic trades limit its usefulness (Andersen & de Lange, 2021; Bloznelis, 2018; Dahl et al., 2021; Ewald et al., 2022). Building upon Dahl et al. (2021)'s finding that stock prices assimilate salmon price information more swiftly than the salmon futures exchange market, we focus on stock prices. For our textual analysis, we employ news articles retrieved from Intrafish[1], spanning from January 2010 to July 2022, and filtered using the keywords "salmon", "finance", and "prices". This approach enables us to achieve a comprehensive understanding of the factors that drive investors' behavior within the salmon market, extending beyond the interplay between supply and demand.

Applying principal component analysis (PCA) to the topics estimated using the LDA algorithm enables us to identify the most dominant themes shaping market news. The inclusion of sentiment analysis allows us to consider the directional effects as well. One of our key contributions in this regard is the improvement of the typical sentiment calculation method by considering the competitive market structure. Moreover, we augment the sentiment dictionary with industry-specific terminology, which enables us to capture information that may not be effectively detected using the financially-oriented LM dictionary alone.

The remainder of this paper is organized as follows. Section 2 elaborates on the data sets incorporated and the pre-processing methods employed. Section 3 outlines the methodologies and techniques utilized to estimate the relationship between salmon market investors' behavior and related financial news. Section 4 presents the results, and Section 5 discusses the primary conclusions.

[1]https://www.intrafish.com/

2 Data & Pre-processing

2.1 Financial Data

We obtain daily stock prices from January 2016 until July 2022 from Refinitiv on five salmon producing companies with the highest market cap trading on the Oslo Stock Exchange (OSL). These are the Marine Harvest (MOWI), SalMar (SALM), Grieg Seafood (GSF), Lerøy Seafood Group (LSG), and Bakkafrost (BAKKA). We choose the salmon market stock prices over the salmon spot or futures contracts because Dahl et al. (2021) found that stock prices reflect salmon price information earlier than the Fish Pool Index, the primary price index of farmed salmon. We construct a Share Price Index (SPI) corresponding to an equally weighted portfolio of the five stocks. We normalize the price index by dividing each price P_t, where $t = 1, 2, 3 \ldots, T$, with the first price of the series P_0, to demonstrate the growth rate of the shared prices. We constructed the SPI instead of using individual share prices because an index should be less sensitive to company-specific information and better reflect on general news related to the salmon market. The index is depicted in Figure 1.

Figure 1: Share Price Index (SPI) for daily price data.

To measure investors' behavior we determine a reaction measure. The reaction is defined as the return from day to day, denoted as $Y_t = P_t/P_{t-1}$, where P_t is the SPI price at time t and P_{t-1} is the SPI price at time $t-1$. To account for proportional changes in the returns, we

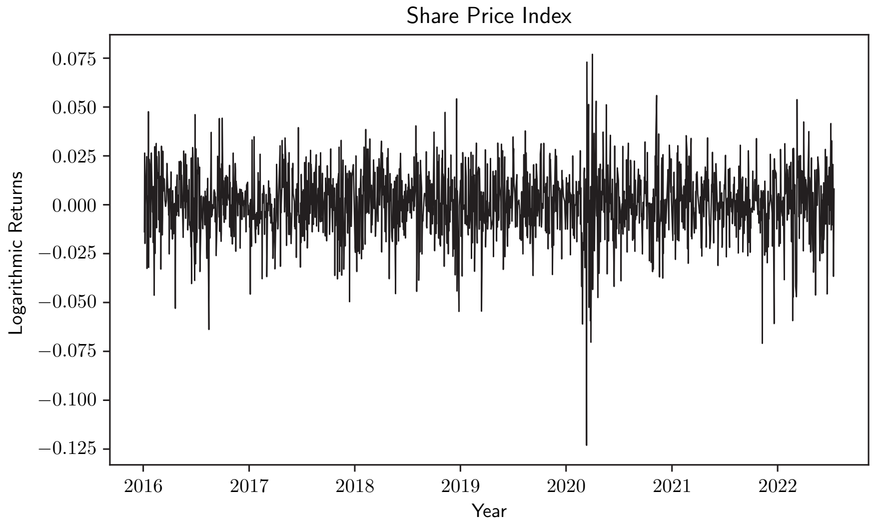


Figure 2: Investors' Reaction or SPI logarithmic returns.

calculate the logarithmic returns as follows::

$$R_t = ln(P_t/P_{t-1}), \tag{1}$$

where R_t denotes the investors' reaction or logarithmic returns of SPI. The development of R_t is illustrated in Figure 2 where the impact of COVID-19 pandemic is evident.

2.2 Text Data

We collected news articles that were published on IntraFish[2] between November 2012 and July 2022. We performed a keyword-based search, filtering articles containing at least one of the following terms either in the text or in the metadata: "salmon", "prices", or "finance". Our scope encompassed all articles relevant to the salmon market, as well as financial articles relating to salmon and other fish or seafood commodities, which are prominent topics on the website. Due to a low number of articles published per month between 2012 and 2016, we limited the scope to articles published from 1 January 2016 to 8 July 2022 (see Figure 3).

In our next step, we filtered the dataset to focus only on articles that provide meaningful insights. To do this, we removed articles that contained many short entries, which were difficult to

[2]https://www.intrafish.com/

analyze using our chosen text mining methods. Specifically, articles with titles containing strings such as "IntraFish Price Tracker", "Top Headlines", "Top Stories", "LIVE Updates", "Reports", and "Conference Updates" were discarded.

Further, we removed articles that held no relevance to our research question, such as those carrying promotional content indicated by phrases like "IntraFish Podcast", "IntraFish App", etc., in their titles. Articles with little text accompanying videos and photos, denoted by "VIDEO:" or "PHOTOS:" in the title, along with some other short articles that seemed irrelevant, were also excluded from the analysis.

Moreover, we disaggregated blog posts into individual entries based on their posting times, recognizing that each small article in a blog conveys unique sentiment and discusses a distinct topic. By employing these filtering and transforming strategies, we enhanced the quality of our dataset.

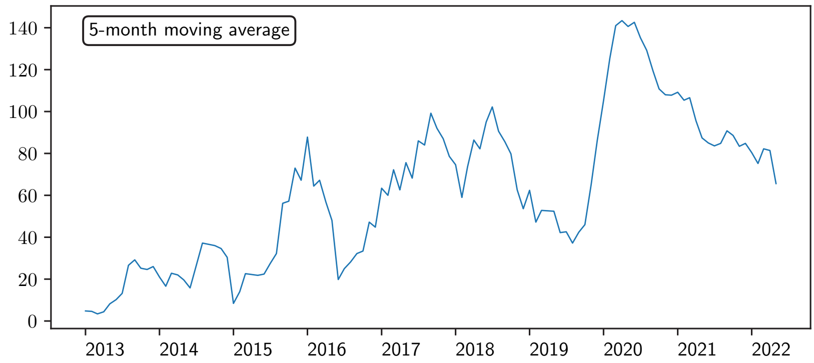


Figure 3: Number of articles per month, 5-month moving average.

In the next step, we pre-processed the remaining articles. This involved resolving encoding issues and removing certain strings and text fragments that could potentially introduce noise to the data rather than contributing meaningful narratives. This included the removal of HTML tags, web addresses, and links to videos, photos, tweets, and ads. Text segments supporting graphs, photos, or tables within the articles were also discarded. We further omitted information that was quantitative in nature, such as the price per salmon weight class, export volume, and export earnings, as this information did not aid in either sentiment or topic identification. Frequently recurring text segments in articles that didn't offer any new insights, such as the methodology for calculating the Nasdaq Salmon Index, were also eliminated. In the interest of brevity, we won't list all the strings and text parts that were removed, but the guiding principle was to focus on text that is pertinent, provides new information to the market, and is compatible with our chosen text mining algorithms. Following this thorough cleaning process, articles with 20 or less words were deleted, as short text can hinder topic modelling effectiveness.

Before finalizing the text data pre-processing, we ensured consistency by converting all the articles to a uniform time zone - the Greenwich Mean Time (GMT). The final version of our news data, which comprises 6082 articles, includes the article's title, the time it was posted, and the text content.

For alignment with the daily log returns data, we assigned dates to articles based on their potential impact on stock returns. Since the Oslo Stock Exchange (OSE) is open from 9:00 am to 4:20 pm Central European Summer Time (GMT+02:00), we perceived articles published after 2:20 pm GMT as potentially impacting the next trading day's stock market. Additionally, articles published post 2:20 pm GMT on Fridays and over the weekends were attributed to influence the forthcoming trading day's stock return. Likewise, articles released on public holidays - in accordance with the Oslo Stock Exchange's operational hours on such days - were understood to bear an impact on the stock market on the subsequent trading day.

Following the completion of standard pre-processing and organization of the textual data, we move on to more specialized pre-processing steps, tailored for topic modelling. Firstly, we transform collocations into single terms. Collocations are identified using part-of-speech patterns as proposed by Justeson and Katz (1995). These are combinations of two or three words that together hold a specific meaning, such as "salmon farming", "United States", or "earnings before interest". In line with Hansen et al. (2018), we retain 194 two-word collocations with a frequency of at least 100 and 85 three-word collocations with a frequency of at least 50.

Next, we transform all upper-case letters to lower-case and split contractions into their constituent words (e.g., 'aren't' becomes 'are not'). Subsequently, we carry out tokenization, where each token represents a sequence of characters that are being treated as a group. We then remove non-alphabetic characters such as numbers, punctuation, currency symbols etc., as well as stop words. These stop words[3] are commonly used words that carry little standalone meaning (examples include 'but', 'I', 'at'). Next, we remove IntraFish journalist names from each article as their high frequency within our corpus effectively categorizes them as stop words.

[3] http://snowball.tartarus.org/algorithms/english/stop.txt

A crucial step in our pre-processing is stemming. We utilize the Porter Stemmer, a popular algorithm for the English language, to reduce words of varying grammatical forms but with a common root to their base form or stem (for example, 'consist' and 'consisted' would be stemmed to 'consist').

To manage dimensionality, we employ the Term Frequency-Inverse Document Frequency (TF-IDF) method as in Blei and Lafferty (2009) and Hansen et al. (2018). The TF-IDF score for each token v is calculated as:

$$log\,(1 + N_v) \times log\left(\frac{D}{D_v}\right) \tag{2}$$

where N_v is the count of the token v in the corpus, D_v is the number of articles that contain the term v, and D is the total number of articles in the corpus. Tokens appearing either very rarely or very frequently in all the articles have a lower TF-IDF score, so we discard the tokens with the lowest scores. At the end of this process, our corpus includes 11,666 unique stems, which are then used for the subsequent topic modelling.

3 Methodologies

3.1 Topic Modelling

3.1.1 Latent Dirichlet allocation (LDA)

The Latent Dirichlet Allocation (LDA), an unsupervised generative probabilistic model, was initially developed by Blei et al. (2003). The central premise of LDA is that documents, in our case, news articles, are represented as random mixtures of latent topics. Each of these topics is defined as a particular distribution of words.

In a corpus of D articles, we can identify V unique words. An article d ($d \in 1, \ldots, D$) is characterized by a collection of topics K. Each topic k ($k \in 1, \ldots, K$) is a distribution $\beta_k \in \Delta^{V-1}$ over V unique words in the vocabulary. Notably, these notations align with those provided by Hansen et al. (2018). The probability distributions of the K topics allow the same term to occur in different topics, potentially with varying weights. Thus, one can think of a topic as a list of words, each weighted to reflect their relevance to the topic. LDA is a mixed-membership model where each article is associated with multiple topics. Therefore, each article d is described by its

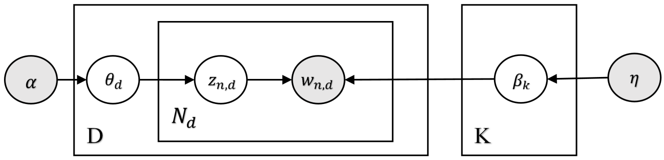


Figure 4: Schematic of LDA algorithm. Shaded circles are observed ($w_{n,d}$) or hyperparameters of the Dirichlet priors (α, η), while white cells are latent, unobserved random variables. The underlying assumption is that a document d is a mixture over topics (θ_d), leading to topic assignments ($z_{n,d}$) of the word n in document d. Meanwhile, topics (β_k) are probability vectors over words that are observed in the corpus. The same words can have non-zero weights in several topics, but a word within a specific document is associated with one topic only.

own distribution $\theta_d \in \Delta^{K-1}$ over the K topics.

The Latent Dirichlet Allocation (LDA) model's inner workings can be best comprehended by visualizing the generative process, which details how documents are created according to the model's perspective. Given a document d, let N_d denote the number of words it contains. We also consider two single-value prior hyperparameters, α and η. The data generating process according to LDA is as follows:

1. For each topic $k = 1, \ldots, K$, draw a distribution over words $\beta_k \sim \text{Dirichlet}(\eta)$, independently.

2. For each document $d = 1, \ldots, D$, draw a distribution over topics $\theta_d \sim \text{Dirichlet}(\alpha)$, independently.

3. For each document $d = 1, \ldots, D$, and for each word $n = 1, \ldots, N_d$ within the document:

   (a) Draw a topic assignment $z_{n,d} \sim \text{Multinomial}(\theta_d)$, where $z_{n,d} \in \{1, \ldots, K\}$.

   (b) Draw a word $w_{n,d} \sim \text{Multinomial}(\beta_{z_{n,d}})$, where $w_{n,d} \in \{1, \ldots, V\}$.

This conceptual generative process can be conveniently visualized as a directed acyclic graph (see Figure 4). Notations used in Figure 4 are shown in Table 1.

Within the LDA model, both θ_d and β_k are treated as random variables. The necessity for a Bayesian formulation of the model is primarily motivated by the extensive number of parameters to be estimated, namely $K \times V$ for β_k and $D \times K$ for θ_d. Dirichlet priors are an optimal choice due to their conjugacy with the multinomial distribution. The hyperparameters η and α govern the sparsity of β_k and θ_d, respectively. When α decreases, fewer topics exhibit a high probability, while the remaining topics maintain a small but positive probability.

Given the presumption that all hidden variables (topic-specific vocabulary distributions β_k, document-specific topic proportions θ_d, and per-word topic assignments $z_{n,d}$) and observed variables (words $w_{n,d}$, hyperparameters α and η) are known, the joint distribution of all these variables can be denoted as:

$$\Pr(B, T, z, w|\alpha, \eta) = \prod_{k=1}^{K} \Pr[\beta_k|\eta] \prod_{d=1}^{D} \Pr[\theta_d|\alpha] \prod_{n=1}^{N_d} \Pr[w_{n,d}|\beta_{z_{n,d}}]\Pr[z_{n,d}|\theta_d]. \tag{3}$$

Here, we have defined $B = (\beta_1, ..., \beta_K)$, $T = (\theta_1, ..., \theta_D)$, $z_d = (z_{1,d}, ..., z_{N_d,d})$, $z = (z_1, ..., z_D)$, and $w = (w_1, ..., w_D)$.

However, the actual challenge lies in inferring the latent variables β_k, θ_d, and $z_{n,d}$ from the observed documents. Hence, the posterior of interest is $\Pr(B, T, z|w, \alpha, \eta)$.

Notation	Definition
α, η	Hyperparameters of the Dirichlet prior distributions
β_k	The distribution over words
K	The number of topics
θ_d	The article-specific topic distribution
$z_{n,d}$	The assignment of a word w_n ($n \in 1, \ldots, N_d$) in an article d to a given topic
$w_{n,d}$	The observed word w_n in an article d
N_d	The number of words in an article d
D	The number of articles in the data set

Table 1: Notations in LDA

3.1.2 Estimation

To estimate the parameters of the LDA model, we apply collapsed Gibbs sampling, an approach introduced by Griffiths and Steyvers (2004). The essence of this method lies in the conjugacy of the Dirichlet prior to the multinomial distribution, enabling us to analytically marginalize the parameters B and T out of the joint distribution $Pr(B, T, z, w|\alpha, \eta)$. Consequently, we can express the probability of the observed and latent variables as $Pr(z, w|\alpha, \eta)$.

In practical terms, our goal is to determine the posterior $Pr(z|w, \alpha, \eta)$, as the topic assignments z are not observed. Using conditional-marginal factorization of the joint probability, this posterior can be expressed as:

$$Pr(z|w,\alpha,\eta) = Pr(z_{n,d} = k|z_{(-n,d)}, w, \alpha, \eta)Pr(z_{(-n,d)}|w, \alpha, \eta). \tag{4}$$

Consequently, the computational task is simplified to sampling the topic assignments $z_{n,d}$ for each word in the corpus, given all the words w and other topic assignments $z_{(-n,d)}$. The advantage of this procedure is that it eliminates the need to sample topic proportions θ_d and topic-specific vocabulary distributions β_k.

A complete derivation of the conditional distribution $Pr(z_{n,d} = k|z_{(-n,d)}, w, \alpha, \eta)$ can be found in the technical appendix of Hansen et al. (2018). The resulting distribution is expressed as:

$$Pr(z_{n,d} = k|z_{(-n,d)}, w, \alpha, \eta) \propto \frac{m^k_{v,-(n,d)} + \eta}{\sum_v m^k_{v,-(n,d)} + V\eta} \times (m^d_{k,-n} + \alpha), \tag{5}$$

where $m^d_{k,-n}$ represents the count of words in document d assigned to topic k, excluding the current assignment $z_{n,d}$, and $m^k_{v,-(n,d)}$ is the number of occurrences of word $w_{n,d}$ assigned to topic k throughout the corpus, excluding the current assignment $z_{n,d}$.

Intuitively, the probability of assigning the current word $w_{n,d}$ to topic k increases if many other words in document d are also assigned to topic k and if the word $w_{n,d}$ has a high probability under topic k.

With the conditional distribution at hand, we can now proceed to detail the collapsed Gibbs sampling algorithm. Firstly, we initialize the topic assignment variables z to the values in $\{1, \ldots, K\}$ by randomly drawing $z_{n,d}$ from a uniform distribution. For each document $d = 1, \ldots, D$ and each word $n = 1, \ldots, N_d$, we sequentially draw a new topic assignment $z_{n,d}$ through multinomial sampling using Eq. 5, based on all the updated topic assignments $z_{(-n,d)}$. We then repeat this procedure for iterations 2 to 4000 as part of a burn-in phase and again for iterations 4001 to 8000, keeping samples with a thinning interval of 50 to ensure that the autocorrelation between samples is low. To be precise, we retain 80 samples corresponding to iterations $\{4050, 4100, \ldots, 8000\}$.

We need to choose three parameters to estimate the model: hyperparameters α and η, and the number of topics K. The values for hyperparameters are derived from Griffiths and Steyvers (2004) and are $\alpha = 50/K$, and $\eta = 200/V$. The number of topics is set to 100.

The collapsed Gibbs sampling procedure yields a set of samples with estimated topic assignments z. Yet, we do not gain direct insights about the primary parameters of interest β_k and θ_d. For each stored sample, we can estimate topic-specific vocabulary distributions and document-specific topic proportions using predictive distributions over new topics and new words. The probability that a new $(N_d + 1)$-th word in a document d is assigned to topic k is given by

$$\hat{\theta}_d^k = Pr(z_{(N_d+1),d} = k|z_d) = \frac{m_k^d + \alpha}{\sum_{k=1}^{K}(m_k^d + \alpha)}, \tag{6}$$

where m_k^d is a count of words in document d assigned to topic k.

In a similar fashion, the predictive distribution over new words is expressed as

$$\hat{\beta}_k^v = Pr(w_{(N_d+1),d} = v|w, z) = \frac{m_v^k + \eta}{\sum_{v=1}^{V}(m_v^k + \eta)}, \tag{7}$$

where m_v^k is the count of times word $w_{n,d}$ is assigned to topic k in the entire corpus. We estimate β_k and θ_d for each iteration in $\{4050, 4100, \ldots, 8000\}$.

We use a measure called perplexity to determine if the chain has converged. The formula for perplexity is given by

$$\exp\left[-\frac{\sum_{d=1}^{D}\sum_{v=1}^{V} n_{d,v}\log\left(\sum_{k=1}^{K}\hat{\theta}_d^k\hat{\beta}_k^v\right)}{\sum_{d=1}^{D} N_d}\right], \tag{8}$$

where $n_{d,v}$ is a count of word v in document d and $\hat{\theta}_d^k$ and $\hat{\beta}_k^v$ are introduced above.

This is a measure of how well the LDA model fits the data. Perplexity, often used in topic modeling, is monotonically decreasing in the log-likelihood of the unobserved documents. Therefore, a model that predicts the data well has a low perplexity. The first 4000 replications of the chain are characterized by rapidly decreasing perplexity values (see Figure 5 and note that we do not save perplexity for the first 500 iterations) and hence are discarded.

To understand the content of the estimated topics, we save the most probable stems under predictive topic-specific vocabulary distributions $\hat{\beta}_k$ obtained at the 8000th iteration. The predictive document-specific topic distribution $\hat{\theta}_d$ is the average over 80 samples.

Finally, to estimate daily topic frequencies, we collapse all the articles for one specific day into one document. Following Hansen et al. (2018), we re-sample topic assignments $z_{n,d}$ for day-specific articles. The topics in this re-sampling step are kept fixed at the values of their

Figure 5: Perplexity values along the chain drawn for the 100-topic model, corresponding to samples ranging from 1 to 150. These sample indices align with perplexity calculations performed across specific iterations: $\{550, 600, \ldots, 8000\}$. We have retained the last 80 samples, representing the point at which the chain has converged.

predictive distributions. Therefore, we use only 20 iterations of the Gibbs sampler in this step. After re-sampling, we obtain the predictive day-specific topic distribution.

3.1.3 Cross-Validation

To choose the optimal number of topics K we employ 10-fold cross-validation. In this process, we randomly split the article data set into 10 folds. Next, we estimate the LDA model on the first 9 folds that represent the training set and keep the last 80 samples where we have observed convergence of the chain (see Figure 5). Thereafter, we estimate perplexity on the 10^{th} fold that represents the test set. For this, we calculate the perplexity using Eq. 8 for each of the 80 samples. When implementing Eq. 8 for cross-validation, $\hat{\beta}_k^v$ corresponds to the topic distributions from the LDA estimation on the training set. Moreover, we re-sample topic assignments $z_{n,d}$ in the test set for each of the 80 samples using 20 iterations of Gibbs sampling and obtain the corresponding $\hat{\theta}_d^k$. The final perplexity for the 10^{th} fold is calculated by averaging over the 80 samples. This procedure is repeated 9 more times, each time changing the fold that serves as the test set. The final perplexity is calculated as the average across the 10 folds. We choose the optimal number of topics K based on Figure 6, where the horizontal axis corresponds to the number of topics and the vertical axis to the perplexity values. Even though the perplexity measure continues to decrease for numbers of topics $K > 100$, we choose it as the cut-off, since the marginal improvements due to additional topics become relatively small, and the number of topics should remain small relative to the number of documents ($D = 6082$). Also, as the number of topics becomes too large, they tend to be overly detailed and less interpretable, reducing their utility.

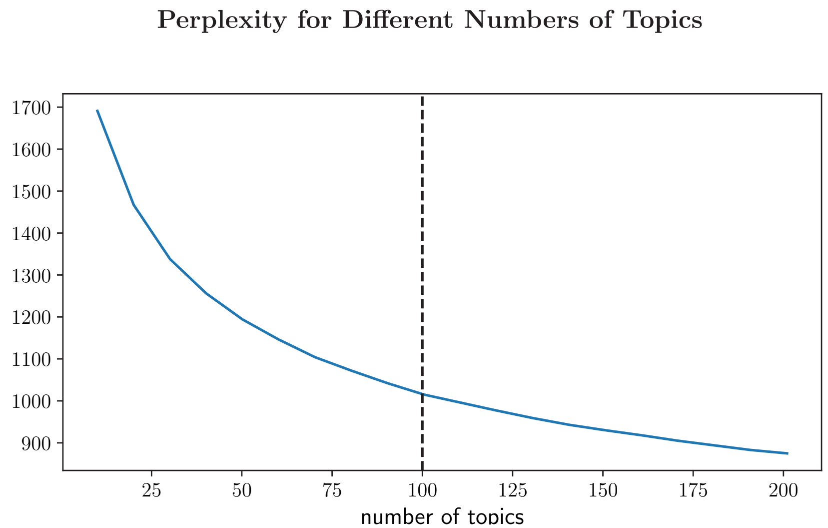


Figure 6: The average perplexity of test data for different topics, calculated according to the formula presented in Eq. 8. These data show that as the number of topics increases, the goodness-of-fit of the model improves. Given the relatively small size of our data set, we choose 100 topics, as the improvements thereafter are non-substantial.

3.2 Sentiment Approach

We believe that both topics and sentiment might be important for explaining stock returns because they provide different types of information. Topics refer to the content of the news, or what is being discussed, while sentiment refers to the emotional tone or valence of the news, or whether it is perceived as positive, negative, or neutral.

To quantify sentiment, we have chosen to use a lexicon-based approach. A lexicon-based approach involves using a dictionary of words or phrases and their corresponding sentiment scores to determine the overall sentiment of a text. The advantages of this method are that it is relatively simple, efficient, and does not require an annotated training set. Additionally, it is more interpretable than machine learning methods, as it allows researchers to examine the specific words and phrases that contribute to the overall sentiment score.

The literature on sentiment analysis differentiates between general and domain-specific lexicons. General lexicons are designed to be applicable across a range of different contexts. In contrast, domain-specific lexicons are tailored to a particular topic, market, or industry. For instance, in the finance literature, Tetlock (2007) examined the relationship between sentiment expressed in the news and stock market movements by using the general Harvard IV dictionary. However, Loughran and McDonald (2011) argued that a domain-specific lexicon is better suited for capturing the tone in financial texts because certain words that may have a negative sentiment in a general lexicon (such as “liability”, “cost”, or “capital”) are rather neutral in the finance domain.

Our approach to sentiment computation is similar to Shapiro et al. (2022). The authors highlight that combining lexicons can improve sentiment analysis performance, particularly when using domain-specific lexicons. In our study, we combine the Loughran-McDonald (LM) dictionary, which is specific to the finance industry, with our own domain-specific dictionary tailored to the salmon market.

We develop a daily sentiment index by following a systematic approach. Firstly, we create a list of seed sentiment words and phrases based on our industry expertise and initial analysis of the news articles. For instance, we consider “algal bloom” as a negative phrase since the rapid growth of algae in the water can have detrimental effects on the health and survival of salmon, while “recapture” is a positive word since it implies that the salmon that have escaped from farms have been successfully brought back. We include all grammatical forms of seed words to expand our list. Secondly, we estimate a word2vec model on our corpus to identify data-driven synonyms for the seed words. Thirdly, we manually select relevant synonyms and their grammatical forms to extend our dictionary. Fourthly, we combine our dictionary with the LM dictionary.

We then use this combined dictionary to estimate sentiment scores for each news article, where sentiment is defined as the number of positive words minus the number of negative words divided by the total number of words in the text. Numbers are not included in the total word count. To account for negation, we multiply the sentiment scores of words by -1 if the word is preceded by a negation term (such as, but not exclusively “neither”, “never”, “not”, or “no”) within a three-word window. Once we have the sentiment score for each article, we compute the daily sentiment index by taking the average of the sentiment scores across all articles on that day.

4 Empirical results

4.1 The impact of industry-specific topics

To investigate the drivers of behavior in financial markets, we conducted Principal Component Analysis (PCA) on a set of estimated topics, in line with (Larsen, 2021). By extracting principal components, we effectively reduced the data's dimensionality, which enabled us to concentrate on major "themes", i.e. uncorrelated linear combinations of original topics, that might impact returns. For the insights from this analysis to hold value, it is essential that these themes are highly correlated with the topics that are related to each other in a meaningful way. In order to understand and appropriately label the themes, we pinpointed the five topics exhibiting the highest absolute correlation coefficients with each component.

The inclusion of absolute returns in this stage of analysis is necessary because topics can be framed positively or negatively. While some topics may be considered inherently "good" or "bad" news, influencing the markets correspondingly, other topics may be directional in nature, impacting markets positively or negatively based on their portrayal. To isolate the impact of specific themes on returns, we account for these effects by combining the topics with sentiment.

However, at this point, our focus is primarily on the topics themselves, and less so on the sentiment surrounding them. We are therefore concentrating on the absolute value of returns rather than simply returns. Our aim is to uncover which topics hold relevance to the market, as this would manifest through market volatility. We use standard bivariate vector autoregressions (VAR) to identify the dynamic responses of absolute log returns to surprise increases in the identified principal components. In these VAR models, our text measures are ordered first, and the number of lags is determined based on the Akaike Information Criterion (AIC). Furthermore, we generate cumulated Impulse Response Functions (IRFs) for the subsequent 20 working days to examine the continued impact of these surprise increases.

Our analysis reveals several components that exhibit a significant impact on absolute returns, indicating a pronounced response from investors. The accompanying VAR results of a few chosen ones are displayed in Figure 7. Moreover, other components are interesting due to the particular combination of topics that these components are most highly correlated with. Table A1 in the appendix provides more details on some of these components, including their respective shares of total variation in topics. It also lists the five topics that exhibit the highest absolute correlation with each component.

The first component, which accounts for approximately 4% of total variation in topic values, is labelled "Business figures". This label has been assigned due to the component's high positive correlation with topics related to results (topics 53 and 41), as well as changes in numbers (topic 67), and its negative correlation with topics 19 (public and social) and 12 (future challenges). The names of the topics listed in Table A1 are assigned based on the most probable words under each topic according to the estimated LDA model. For example, topic 53 is labelled "quarterly results" because the words with the highest probability under this topic are 'quarter', 'earnings', 'revenue', 'tax', 'ebitda', and the like. The fact that this component explains the highest percentage of variance in topic values shows that news on business aspects is covered substantially. This matters for our purposes of analysing the effect of these articles on returns of the companies and is hence a reassuring result.

However, the effect of this component on absolute returns is small, and only significant one day after a surprise increase in the component. While one could argue that the reported figures are results from past business activity, which may be partially priced in already, and that the future prospects may be more relevant, direct news about the state of the business would still be expected to matter for expectations on the next rounds of dividends, i.e. cash flow to stock holders. Moreover, it indicates whether a company is on a profitable course. As we will explain in more detail below, we rather interpret this as a need to provide more structure to our analysis, as business results could be reported for any company in the salmon industry, not only those reflected in our index. Especially the competitive relationships between salmon producers are shown to matter in the latter course of this analysis. Hence, all articles cannot be treated equally.

Given the time frame of our data set, it is not surprising that the Covid pandemic had significant effects on stock markets. Components five and six are specifically related to the pandemic, with the former exhibiting a positive correlation with topic 36 (Covid news) and topic 83 (Corona infections in production facilities such as farms and processors). As seen from Figure 7, following a surprise increase in component five, which we label as 'Covid and production', absolute log returns increase by 0.0013 (or 0.13 percentage points, given that log returns are expressed in percentages) four days post-shock, maintain that level for another 6 days, and then revert to the original level 10 days subsequent to the shock. An explanation for the link between Covid and volatility may be attributed to the uncertainty surrounding the pandemic's duration and severity, as well as its potential impact on the global economy.

Additionally, component five displays negative correlations with topics 37, 28, and 10, which pertain to kilo prices for Norwegian salmon, salmon harvesting results, and outlooks on prices, respectively. This negative correlation implies that these topics have a calming effect on the markets. Although it may seem counter-intuitive that articles about prices and production results are related to lower volatility, it can be explained by the fact that such articles appear regularly and occupy significant shares of daily news only during periods of relative market stability. In times of turmoil such as the beginning of the Covid pandemic, these topics are overshadowed and occupy relatively little space in salmon-related reporting.

Components five and six both pertain to Covid, but they differ in their specific focus. While component five combines news about Covid and production, the sixth component contains information on the role of Covid in salmon markets. This is indicated by the high and positive correlation (36%, see Table A1) of this component with topic 13, which is frequently featured in articles discussing business challenges, risk, and uncertainty in the markets. Given this focus, it is not surprising that this component has a stronger effect on absolute returns, as seen in Figure 7. Specifically, absolute returns increase immediately one day after a surprise increase in component six (labelled "Covid and markets"), and this rise continues for the next 20 days (excluding days 8 and 9), peaking at an increase of about 0.0025. Overall, this response is more persistent and prolonged. The negative correlation of the sixth component with topics on salmon commodity and wholesale prices (topics 59 and 39, respectively), as well as mergers and acquisitions, can be attributed to these topics being less prominent in discussions during turbulent times.

Shifting from the Covid-related components, we now turn our focus to another key driver of market volatility: Component 17, labelled "Corporate & Stocks". This component exhibits a slowly building and prolonged influence on absolute returns, as depicted in the bottom-right panel of Figure 7. The peak positive response of absolute log returns occurs 9 days after the shock, and there is no evidence of a rebound effect. Being highly and positively correlated with topics 41 (business results), 48 (management and boards, i.e., high-level personnel of producers), and 69 (stock market news), this component covers various aspects related to business, which are expected to significantly influence investors' decisions.

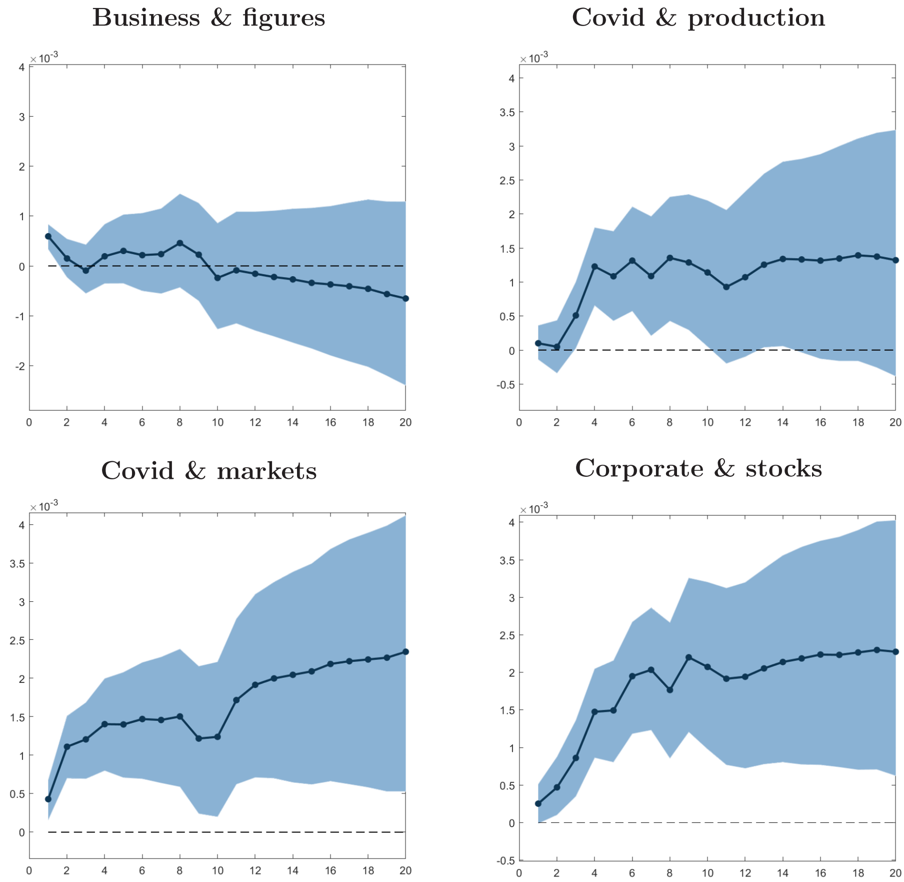


Figure 7: Cumulated impulse responses of SPI absolute returns to one standard deviation innovations in selected components. Shaded regions are 68 percent confidence bands, computed with bootstrap standard errors, using 1000 replications. The horizontal axis represents the number of lagged days in the Impulse Response Functions (IRFs).

Another component of interest is component nine, which pertains to salmon production in Chile. This is most notably indicated by its high and positive correlations with topics 76 and 52 (see Table A1), relating to Chilean producers and salmon farming in Chile, respectively. Interestingly, topic 61, the topic on algal blooms, correlates positively with the component, too. This suggests that algal blooms, deadly to salmon, are a concern that is more relevant (albeit not exclusively) to production in Chile. This does not come as a surprise as at the beginning of 2016, high and persistent harmful algal blooms (HABs) took place in the marine ecosystems of southern Chile. A major mortality event of about 27 million salmon i.e. 39,000 tonnes) was caused by blooms in the Los Lagos Region (León-Muñoz et al., 2018; Montes et al., 2018). Indeed, the articles in our data set with the highest values for shares of topic 61, algal blooms, cover events taking place in Chile almost exclusively. Topic 33, which discusses SalMar's offshore farming - a firm in our index - correlates negatively with component nine, indicating once again that this component focuses more on the competitors. In response to this component the absolute returns of the firms within our index are seen to experience a brief, significant decline for three days, before a swift recovery (IRF omitted). Hence, when news reporting is more focused, in relative terms, on issues in Chile, stocks of the firms in our index exhibit relatively less volatility.

On the other hand, Component 12 is most highly and positively correlated with topics that are relevant to our index: SalMar offshore farming, NTS acquiring Norwegian Royal Salmon (NRS), and project licenses in Norway. This component has a positive and short-lived effect on market volatility (IRF omitted). Again, it is worthwhile to investigate the topics that correlate negatively as well: here, the ones worth mentioning are topics 85 and 3, which pertain to Atlantic Sapphire land-based farming, and Aquabounty land-based and genetically-modified (GM) farming (see Table A1). These two topics cover not only competitor companies, but also competitive technologies. The Norwegian and Faeroese (i.e. Bakkafrost) companies in our index produce mostly in pens in the sea. Their competitive advantage lies in access to high-quality locations in natural waters. However, these locations are typically remote and far removed from consumer markets. Hence, land-based farming, which in the news reporting here is opposed to Norwegian sea-based farming, may become a significant threat to the companies in our index, in case it can be scaled to produce substantial amounts of high-quality salmon.

4.2 Incorporating lexicon-based sentiment

In the following, we multiply the daily topic values with the daily sentiment, specifically calculated using the LM dictionary often used in financial literature, see e.g., (Li et al., 2020). Our goal is to differentiate between positive and negative news concerning topics that inherently may not possess any sentiment. Given this distinction, we can proceed and analyse the effect of news on returns, including their direction. Prior to this step, human judgement was only involved in the choice of methods, as the procedures used were mathematical and algorithmic and did not require external guidance. However, the use of a dictionary approach for sentiment extraction introduces an element of subjectivity, as the choice of words is based on the expertise of the creators of the dictionary.

The PCA results are strikingly different to the analysis above. The first component now explains 26.5% of total variation in the values of topics multiplied with sentiment, compared to 4% in the analysis without sentiment. Furthermore, the topics that most highly correlate with extracted components, when multiplied with sentiment, have changed, as well as the impact of these components on returns. The correlations between key components and the sentiment-weighted topics are provided in Table A2 in the appendix.

As shown in Figure 8, the first component, labelled "business expectations", due to its co-movement with topics 43 (plans & strategy), 35 (fear, harm, negative outlook), 68 (business data), 12 (future challenges), and 47 (contracts & agreements), has a marginally significant, small, and prolonged effect on logarithmic returns. The effect is negative, which, given that the topic values are multiplied with sentiment, stands in contrast to the expected positive effect. With sentiment included in the analysis, one would anticipate that a surprise increase in sentiment about business expectations would lead to an increase in logarithmic returns. However, we observe the opposite, i.e. a negative relationship. As this component does not differentiate between different producers, or clusters of producers, the negative correlation is mostly driven by the fact that there are only five firms in our index, and many more competitors, and hence there are also more news articles about other firms, which hence dominate the effect.

Components 3 and 4, and their respective effects on log returns, reveal one reason why sentiment does not function as expected. We find two components, one labelled "Investments Norway", and the other one labelled "Investments rest of the world (ROW)" (investments including M&A). As shown in Figure 9, a surprise increase in sentiment about investments in Norway is followed by an increase in returns. Conversely, an unanticipated increase in sentiment regarding investments in the ROW has the opposite effect. This difference in market reaction hints to the prevailing effects of competition amongst firms in salmon markets. Increased sentiment regarding investments into the productive capacities of firms reflected in our index raises the expectations of future cash flows, and hence returns increase, and vice versa for a decrease in sentiment about investment projects. Investments in the rest of the world, however, increase competitive pressures on the companies in our index. Hence, a rise in sentiment about those projects relates to declines in returns of our index, and vice versa.

The observation that some sentiment-weighted topics lack correlation with the components

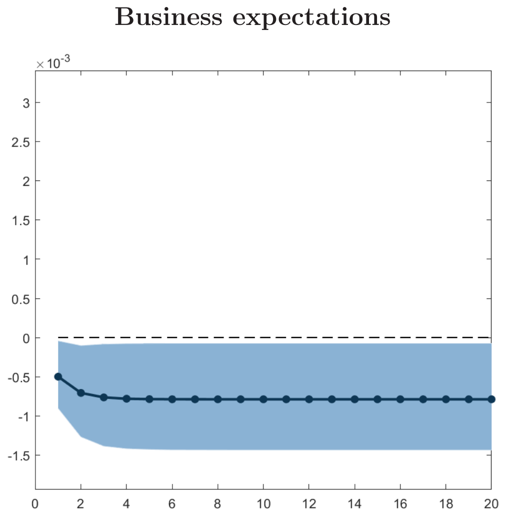


Figure 8: Cumulated impulse response of logarithmic stock returns to one standard deviation innovation in component 1 labelled “Business expectations”. Shaded regions are 68 percent confidence bands, computed with bootstrap standard errors, using 1000 replications. The horizontal axis represents the number of lagged days in the Impulse Response Functions (IRFs).

driving the market, as estimated in this section, is important. Specifically, topics related to Covid, which had shown to be particularly relevant in the analysis without sentiment, and biological aspects of salmon farming, such as algal blooms, do not seem to strongly correlate with the first, third, and fourth components when their values are multiplied with the LM sentiment. While it is anticipated that business and finance-related news would greatly influence returns, the lack of prominent correlation of these other topics with the components driving the market can be largely attributed to the nature of the sentiment dictionary used. The dictionary was initially designed to identify sentiment in business reports from a wide range of companies, and thus it effectively identifies sentiment in a business and finance-related context. Articles in which other topics are discussed are consequently evaluated to have neutral sentiment. As a result, variation in sentiment-weighted topics is driven by the topics that the sentiment dictionary is able to recognize, rather than topics that may be relevant to salmon producers specifically, such as those related to production processes.

The weaknesses of LM dictionary are clearly shown in Figure 10, which portrays the effect of LM sentiment (left panel), as well as of topic 36 (Covid) multiplied with LM sentiment (right panel) on logarithmic returns. It is generally anticipated that an increase in sentiment would

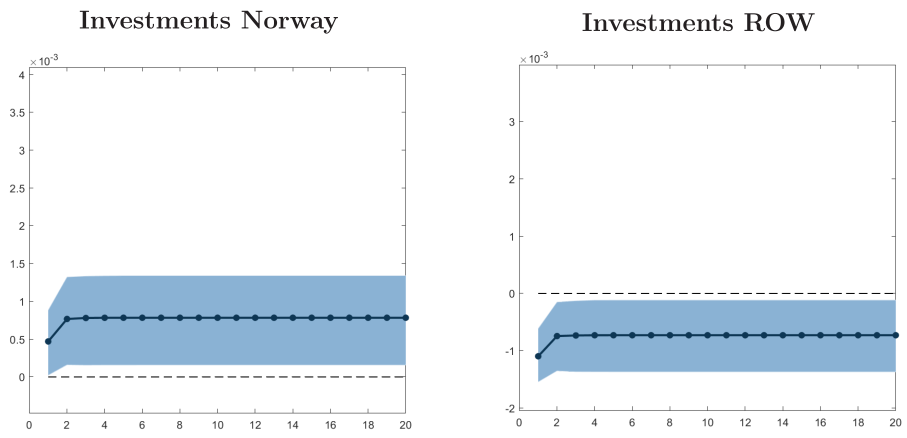


Figure 9: Cumulated impulse responses of logarithmic stock returns to one standard deviation innovations in components 3, labelled “Investments Norway”, and 4, labelled “Investments rest of the world (ROW)”. Shaded regions are 68 percent confidence bands, computed with bootstrap standard errors, using 1000 replications. The horizontal axis represents the number of lagged days in the Impulse Response Functions (IRFs).

result in a corresponding rise in logarithmic stock returns. However, an unexpected increase in LM sentiment is observed to precede a marginally significant and prolonged decrease in these returns (Figure 10, left panel). Moreover, an unexpected surge in sentiment about topic 36, Covid, does not have a statistically significant influence on logarithmic returns (Figure 10, right panel). The unusual negative correlation between returns and LM sentiment could potentially be explained by the dictionary’s inability to account for competition dynamics in the salmon market. As discussed earlier, sentiment is calculated based on articles not only about the firms in our index but also their numerous competitors. At the same time, the insignificant effect observed from the sentiment-weighted Covid topic could suggest the LM dictionary’s limitations in accurately capturing the sentiment associated with non-financial news.

Moreover, we observe a striking similarity between the impulse response of stock returns to LM sentiment and to the first component. This is likely due to every daily topic value being scaled by the LM sentiment value for that same day, resulting in the variation in sentiment-weighted topics being dominated by the variation in LM sentiment. Consequently, the topics that correlate strongly with the first component serve as a reliable indicator of the type of news that the sentiment index captures - primarily, financial and business-related news in this case.

Another example highlighting the importance of considering competition in the salmon mar-

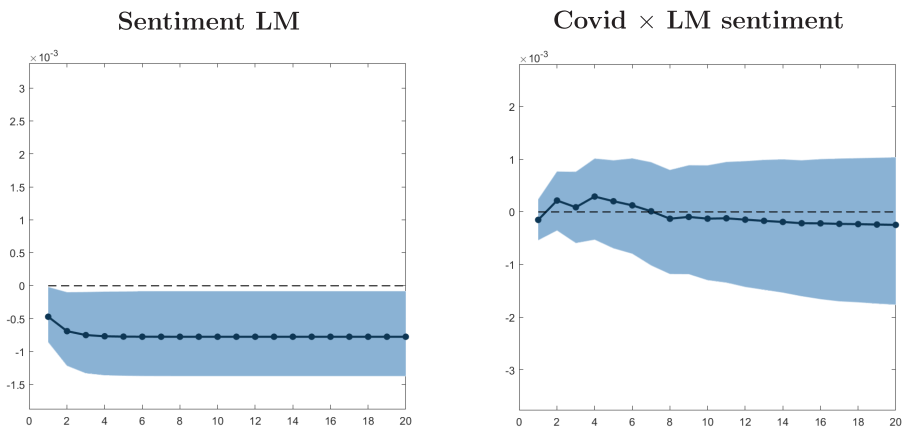


Figure 10: Left: cumulated impulse response of logarithmic stock returns to one standard deviation innovation in LM sentiment. Right: cumulated impulse response of logarithmic stock returns to one standard deviation innovation in topic 36 (Covid) multiplied with LM sentiment. Shaded regions are 68 percent confidence bands, computed with bootstrap standard errors, using 1000 replications. The horizontal axis represents the number of lagged days in the Impulse Response Functions (IRFs).

ket when computing sentiment is topic 61, algal blooms. This topic correlates strongly and positively with component 9, which pertains to salmon production in Chile and is extracted based on topics only. This strong correlation suggests that during the period our data set covers, algal blooms were a more prominent issue in that part of the world. This is confirmed by examining the articles that exhibit a high proportion of the topic about algal blooms. These articles are almost exclusively related to aquaculture in Chile. The positive correlation of this topic (not multiplied with sentiment) with the logarithmic returns of our index (top left panel in Figure 11) reveals the competition effect in the data. Since news about algal blooms is reported more frequently in relation to Chilean farms, we observe a positive correlation with log returns of our index. Some algal blooms that had been reported in this period led to devastating losses in Chilean farms, hence creating substantial contractions in total supply. For firms that did not suffer from these losses, the supply shock could only be experienced as increased prices on global salmon markets, thus increasing profits. Hence, this example vividly demonstrates the competition effect: the logarithmic return of our index reacts positively to news about deadly algal blooms, but the sign of the correlation changes once we account for sentiment (top right panel in Figure 11). This means that a surprise increase in sentiment about algal blooms was followed by a decrease in log returns of our index. It is worth noting that the LM dictionary effectively captures this topic, given the presence of words such as “loss”, “incidence”, and “suffer” that are denoted as negative in the dictionary and also bear high probabilities in the distribution of the topic on algal blooms.

Technological competition and competitive products are two specific kinds of competition effects that might partially explain the negative correlation between the logarithmic returns of our companies and the sentiment index calculated with the LM dictionary. For instance, topic 25 focuses on Nordic Aquafarms’ land-based salmon farming project in Maine. When multiplied with LM sentiment, it exhibits a negative correlation with the stock returns of our index (see the bottom right panel of Figure 11). Land-based production is a new and innovative technology that could potentially threaten the competitive advantage of traditional salmon producers in remote parts of the world with good access to high-quality water. If land-based salmon farming can compete at relevant scales, remote production sites could potentially become a liability for traditional producers, as production could then move closer to consumption markets. Thus, successful investments by competitors in this technology can be seen as bad news for the companies represented in our index. Similarly, the sentiment-weighted topic 42 on substitute products, such as plant-based alternatives and meat, also demonstrates a negative correlation with the logarithmic returns of salmon producers (see the bottom left panel of Figure 11). Positive news about products that consumers may replace their salmon consumption with can be bad news for salmon producers as well.

It is common practice to use sentiment dictionaries to identify the effects of news on stock prices. However, we find that using a sentiment dictionary that is not tailored to the specific industry can result in inaccurate analyses. The Loughran-McDonald dictionary, for example, was not designed for the salmon aquafarming industry and thus misses important market structure and industry-specific vocabulary. As a result, the version of the LM dictionary that we applied does not capture the effects of natural disasters, such as algal blooms, storms or diseases that impact salmon production. Additionally, the LM dictionary does not account for the impact of the Covid-19 pandemic, which has been a major driver of market volatility in recent times. To address these limitations, we suggest modifying the dictionary by adding domain-specific vocabulary and differentiating between news about competitors and the firms in our index. Such modifications could improve the accuracy of sentiment analysis and help investors make more informed decisions based on news about the salmon aquafarming industry.

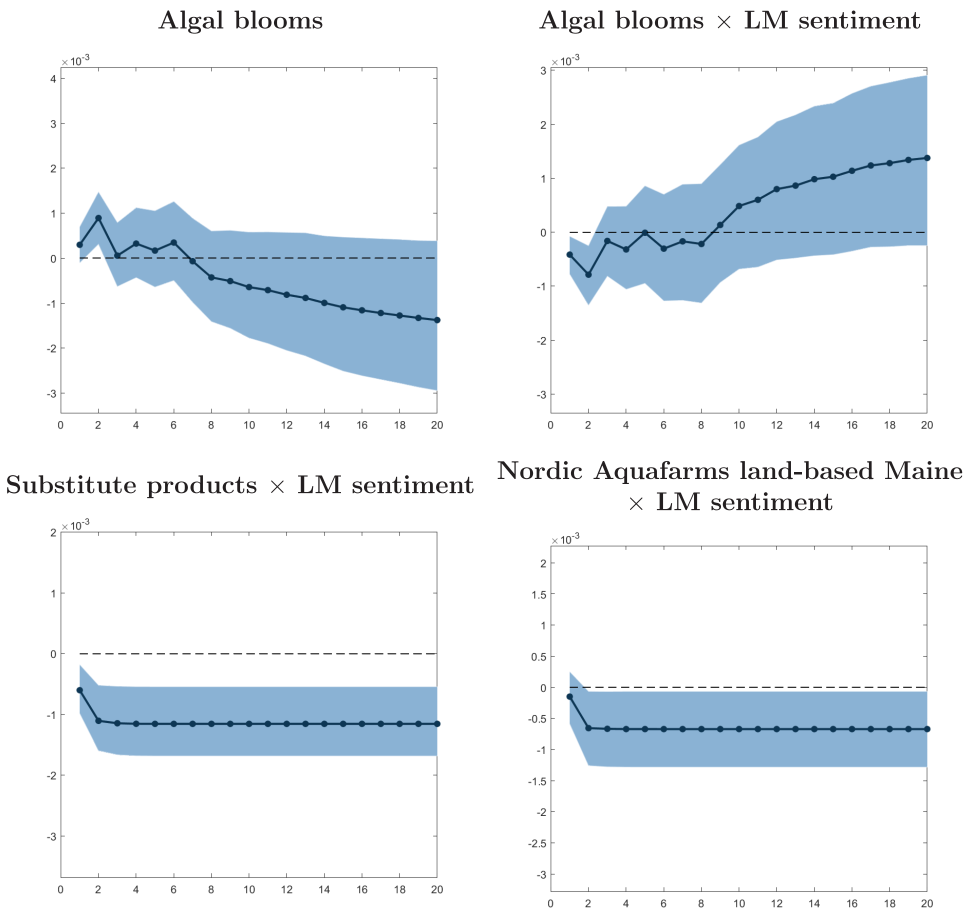


Figure 11: Cumulated impulse responses of logarithmic stock returns to one standard deviation innovations in topic 61 (algal blooms), topic 61 multiplied with LM sentiment, topic 42 (substitute products, such as meat and plant-based alternatives) multiplied with LM sentiment, and topic 25 (Nordic Aquafarms' land-based project in Maine, US) multiplied with LM sentiment. Shaded regions are 68 percent confidence bands, computed with bootstrap standard errors, using 1000 replications. The horizontal axis represents the number of lagged days in the Impulse Response Functions (IRFs).

4.3 Resolving limitations of the Loughran-McDonald dictionary

As highlighted earlier, the direct application of the Loughran-McDonald dictionary to our specific data suffers from two limitations: firstly, the presence of a competitive market structure, where news about our index firms elicits an opposite reaction to news on their competitors, renders the approach inappropriate. To overcome this issue, we modify the sentiment of articles mentioning at least one competitor but none of our index firms by multiplying the sentiment score by

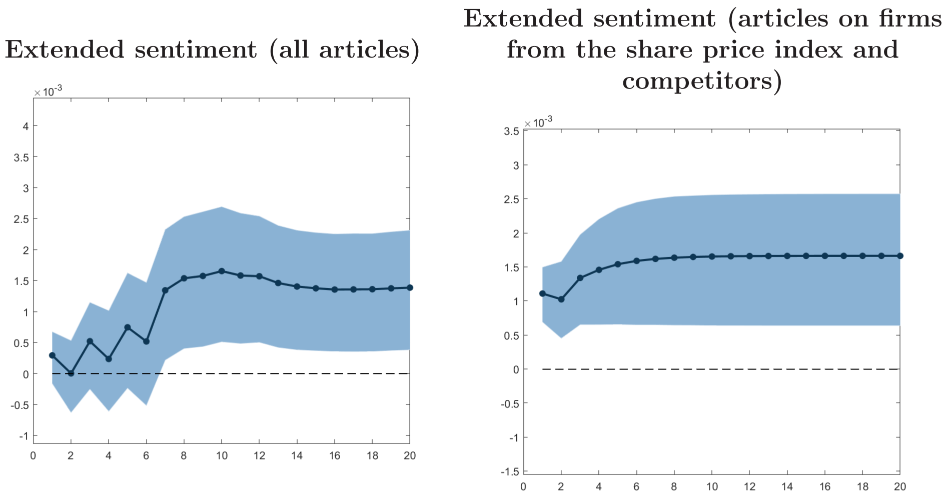


Figure 12: Cumulated impulse responses of logarithmic stock returns to one standard deviation innovations in sentiment, as measured with the extended dictionary. Left panel: all articles; right panel: only articles that discuss firms from the share price index or competitors. Shaded regions are 68 percent confidence bands, computed with bootstrap standard errors, using 1000 replications. The horizontal axis represents the number of lagged days in the Impulse Response Functions (IRFs).

−1. Secondly, the dictionary's lack of sensitivity to industry-specific language necessitates its expansion to include terms relevant to salmon production. These terms encompass technology, diseases and natural disasters, market-specific expressions, and general sentiment-related words that we deemed relevant but were not included in LM dictionary. The modified dictionary contains 187 additional positive words and 336 negative words, thereby supplementing LM's 347 positive and 2345 negative words. We deliberately excluded words that pertain exclusively to the Covid crisis, such as "Corona" or "pandemic", since their relevance was only apparent ex post and would have been unpredictable ex ante. However, we included "virus" since viral infections pose a significant challenge to salmon production. Nonetheless, as we shall demonstrate below, the inclusion of "virus" is insufficient to elicit substantial variation in articles related to Covid.

In Figure 12, we present the impulse response of logarithmic returns following a shock to the extended sentiment index. Remarkably, this small increment of words (amounting to only 16% of the words in the extended dictionary), in combination with competition effects, leads to a significant positive (0.0015, or 0.15 percentage points, given that log returns are expressed in percentages) effect of sentiment on returns seven days after the shock. Notably, our individual modifications to the dictionary could not yield such a substantial outcome, as evidenced by our detailed results (omitted for brevity). Indeed, the combination of both extensions is crucial to obtain the desired results. In the right panel, we display the impulse response for articles that solely mention either the companies in our index or their competitors, indicating the potential of filtering the data for pertinent articles. However, this unsophisticated analysis only provides a preliminary indication of the scope of data filtering, since it potentially discards relevant articles that do not mention firms directly, but yet carry important information for salmon markets. Developing more efficient and refined methods of data filtering remains an open research question, outside the scope of this paper.

Having shown that our solutions of extending the dictionary to include domain-specific vocabulary, and imposing market structure on the data analysis has the desired effect such that sentiment has a positive impact on markets, we can now proceed to present the results of a few chosen topics combined with the extended sentiment index (Figure 13).

Algal blooms topic (topic 61, top left panel) multiplied with extended sentiment now has the desired positive correlation with returns for three days after the shock. This is likely to be driven mainly by the explicit introduction of a competitive market structure to our analysis, since algal blooms are more likely to be associated with Chilean aquaculture in our data set, and therefore are more likely to affect competitors. Word inclusions related to this well-known problem in aquafarming may however also have had an improving effect on the results.

Covid (topic 36, top right panel) on the other hand is still not adequately accounted for. We consciously decided against the inclusion of words that are straightforwardly linked to the pandemic, as it could not have been foreseen before its outbreak. Hence, exogenous shocks from unexpected directions are still not possible to analyse with this approach. However, this is an expected results, as some shocks may indeed be unforeseeable and cannot be planned for.

The inclusion of domain-specific vocabulary has clear benefits to this approach as well, as can be seen in the bottom panels of figure 13: topic 23 (bottom left), which pertains to diseases in production facilities, multiplied with extended sentiment now has a highly significant effect on returns, with the largest magnitude (0.002) observed. Naturally, the LM dictionary could not pick up news on such topics, and overall there may be good or bad news about diseases (outbreaks, or successful mitigation, for instance). By including words that relate to typical farmed salmon diseases, we could achieve this positive correlation between a sentiment-weighted

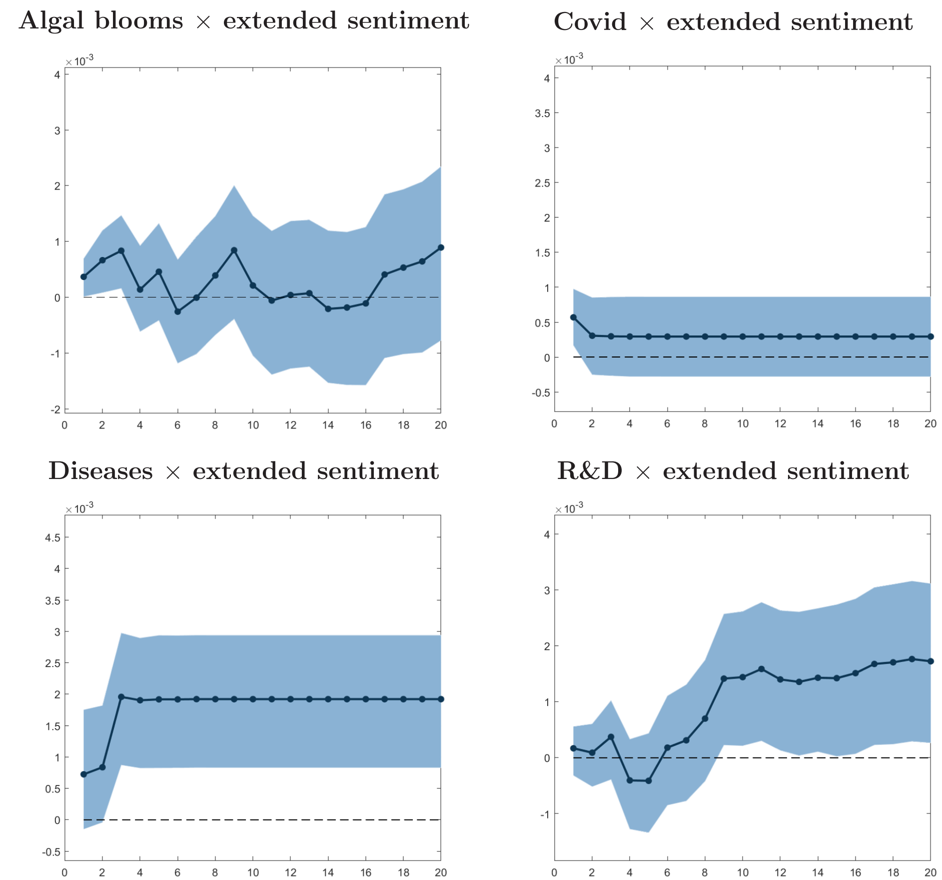


Figure 13: Cumulated impulse responses of logarithmic stock returns to one standard deviation innovations in topics 61 (algal blooms), 36 (Covid), 23 (diseases), and 71 (research & development), all multiplied with the extended sentiment index. Shaded regions are 68 percent confidence bands, computed with bootstrap standard errors, using 1000 replications. The horizontal axis represents the number of lagged days in the Impulse Response Functions (IRFs).

topic that is highly specific, and the respective returns on company shares.

A similar effect can be observed for topic 71, which pertains to news on research and development (R&D) in salmon farming. These articles discuss technological and biological solutions to issues or difficulties in salmon production, which are also highly field-specific. With the extended sentiment, we find a positive correlation between the sentiment-weighted topic and returns in the long-run (9 days after the shock). Other examples (omitted) are the effects of topics 24 (escapes), 29 (Scottish salmon, Brexit), 57 (risk measures), 73 (construction of land-based facilities), 79 (biological performance), 87 (technology), 95 (project licenses Norway), 97 (Mowi business & management), all multiplied with extended sentiment.

Our analysis is supported by the VAR findings derived from components based on topics multiplied by the extended sentiment. As illustrated in Figure 14, after a short lag period, the impulse response of logarithmic stock returns to one standard deviation innovation in the first component turns significantly positive. Notably, the response to this innovation in the first component is nearly indistinguishable from the impulse response to an innovation in sentiment alone (see Figure 12), as seen in the case of topics multiplied with LM-sentiment (Figures 8, and 10, left panel). This suggests that a substantial portion of sentiment is effectively captured by the first component. This is underlined by first component's large share in explaining total variation of topic values combined with sentiment of 28.9%.

Furthermore, the sentiment-weighted topics that exhibit strong correlation with this component (see Table A2) mirror the sentiment-weighted topics that have substantial correlation with the first component in the LM-sentiment analysis (see Table A3 in the appendix). The top three sentiment-adjusted topics that share the strongest correlation are the same, albeit in a different order. The fourth and fifth sentiment-adjusted topics, which demonstrate a notable correlation with the first component - namely, "People, Projects & Perspectives" and "Green Bonds" - integrate seamlessly within the overarching theme of the component. This similarity can be ascribed to the relatively small addition of words to the sentiment dictionary. It is promising that such a minor adjustment to the financially-oriented LM dictionary can nudge the response towards the expected direction and improve the results substantially.

We omit the second component as it primarily pertains to business-related news, similarly to the first component. However, since components are orthogonal by construction (being eigenvectors of the covariance matrix), it appears to capture rather spurious aspects in the behavior of topics. Instead, we draw attention to two additional components that, when subjected to a one standard deviation shock, produce statistically significant responses in log returns (Figure 15).

Component 4, labelled "Chile & algal blooms", demonstrates a positive association with topic 61, concerning algal blooms, as well as topics 52 and 76, both of which address salmon aquaculture in Chile (see Table A3). More precisely, Component 4 suggests that, considering the increased frequency of algal blooms in Chile as opposed to Norway or the Faroe Islands, calamities occurring

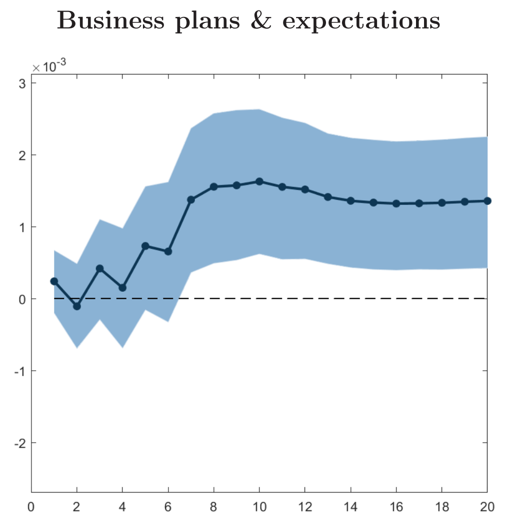


Figure 14: Cumulated impulse response of logarithmic stock returns to one standard deviation innovation in component 1 based on topics multiplied with extended sentiment. Shaded regions are 68 percent confidence bands, computed with bootstrap standard errors, using 1000 replications. The horizontal axis represents the number of lagged days in the Impulse Response Functions (IRFs).

within the Chilean aquaculture industry may contribute to a rise in our constructed index. This verifies that the salmon market is subjected to competition. Furthermore, Component 7, broadly labelled "Salmon Industry", relates to various topics specifically addressing aspects that are specific to salmon markets and production, such as topic 37 "Norwegian Salmon Prices", topic 11, which covers articles about a collaboration between Salmon Evolution and Dongwon to establish a land-based production facility in Korea, and topic 28, "Salmon Harvesting Results".

Shocks to either component generate positive impulse responses, as anticipated due to the multiplication of topics with sentiment values. Most importantly, however, our findings indicate that the added terms in the dictionary enable us to uncover salmon industry-specific topics that are prominent in the data, as evidenced by their strong correlation with components based solely on topics. Thus, the addition of domain-specific vocabulary not only improves impulse responses to sentiment and its combination with topics but also enhances our ability to analyze a specific market with its unique characteristics.

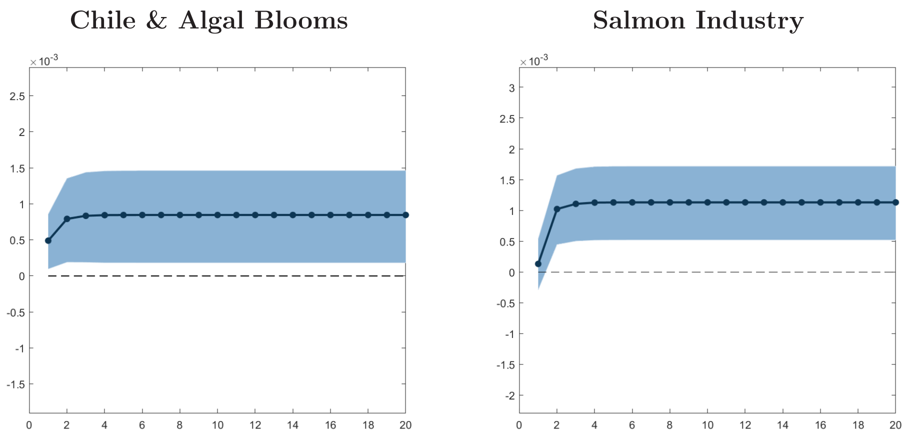


Figure 15: Cumulated impulse responses of logarithmic stock returns to one standard deviation innovations in components 4 ("Chile & algal blooms") and 7 ("Salmon Industry") based on topics multiplied with extended sentiment. Shaded regions are 68 percent confidence bands, computed with bootstrap standard errors, using 1000 replications. The horizontal axis represents the number of lagged days in the Impulse Response Functions (IRFs).

4.4 Out of Sample Forecasting Exercise

While the main purpose of the study is to understand what type of information drives returns of salmon producer stocks, we validate the findings through an out-of-sample forecasting exercise. This will provide a financial significance check, determining whether these results could underpin profitable trading strategies.

Following the methodology established by (Li et al., 2020), we attempt to predict stock price movements. Specifically, we sort the close-to-close stock price returns in ascending order to label each stock. We use the 25th and 75th percentiles of the close-to-close stock price returns as thresholds for this label determination. Hence, if the return is in the bottom 25%, the label is defined as "fall" (class 0); if it is in the top 25%, the label is "rise" (class 2); for the middle 50%, it is labeled as "horizontal" (class 1).

Our dataset is partitioned into a training set (spanning 13th January 2016 to 11th November 2021) and a test set, constituting 10% of the data (12th November 2021 to 11th July 2022).

The Support Vector Machine (SVM) model serves as our fundamental analytical tool, often used as the benchmark in literature predicting stock market movements. Despite the potential for better performance from models like Long Short-Term Memory (LSTM), our focus is not on finding the most sophisticated model, but rather understanding if our proposed analysis could enhance out-of-sample forecasting.

Model parameters are fine-tuned using 10-fold cross-validation, where we select penalty C and polynomial degree d. We consider a grid of values for C $(10^0, 10^{0.25}, ..., 10^3)$ and for d $(2, 3, 4, 5)$, aligned with (Li et al., 2020). Given the class imbalance in our data (40, 74, and 49 observations in classes 0, 1, and 2, respectively), the weighted average F1-score serves as our cross-validation metric.

Our SVM uses a polynomial kernel and adopts a "one versus one" strategy for this multi-classification task. To examine if components based on topics multiplied with extended sentiment can improve market movement prediction beyond price data alone, we estimate SVM models with varying feature sets.

The baseline model includes only the close price $Close_{t-1}$ and volume $Volume_{t-1}$, with a lag of one period, following the most popular practices (Li et al., 2020). Both variables are equally weighted for the five companies in our index. We experimented with using returns instead of close prices, but the performance on the test set was comparable, so we chose to retain the close prices.

Our second model incorporates sentiment estimated using the LM dictionary in addition to the price features, again with a lag of one period. The third model specification brings into play prices along with components derived from topics multiplied by the LM sentiment. In this instance, we factor in the first lag of the first component owing to its substantial role in explaining the variation. We also incorporate five other variables that represent lags ranging from 1 to 10 of the ten components most strongly correlated with log returns in the training dataset.

Our fourth model combines prices and extended sentiment from the period $t-1$. Lastly, our fifth model encompasses prices along with components based on topics multiplied by extended sentiment. Just as in the third model, the first lag of the first principal component is included due to the high percent of variance it explains. Additionally, five other variables are incorporated, which represent lags ranging from 1 to 10 of the ten components that are most highly correlated with log returns in the training data. Initially, we attempted to include the principal components at period t, however, the resulting weighted F1 metric was lower, and hence, in the final model, we only considered the lags of the components. Our feature selection process has been carefully designed to include the most meaningful variables for forecasting.

Model performance on the test set is evaluated using the weighted average F1-score. For models that incorporate text features, we compute the performance improvement attributable to the news information using the following metric:

$$\Delta_{\text{news}} = \frac{F1_{p,text} - F1_p}{F1_p} \tag{9}$$

Here, $F1_{p,text}$ represents the weighted average F1 score when utilizing both information sources—text and prices—while $F1_p$ stands for the F1 score with only price information.

Table 2 presents the F1-scores and Δ_{news} values for each model.

Model	Weighted Average F1	Δ_{news}
Prices	0.28	-
Prices + LM Sentiment	0.28	0
Prices + LM Components	0.31	0.10
Prices + Extended Sentiment	0.28	0
Prices + Extended Components	0.38	0.32

Table 2: Forecasting improvements due to the inclusion of sentiment and topic components (news).

These results highlight that incorporating only sentiment (either LM or extended) does not enhance performance relative to the baseline prices-only model. However, including components based on topics multiplied with sentiment (LM or extended) significantly increases the F1 score. Specifically, the increase is by 10% with LM sentiment and 32% with extended sentiment.

In conclusion, this study demonstrates the potential of our text information extraction approach in enhancing out-of-sample forecasting. While it remains a limited experiment using a single model, it serves its purpose: to verify that text matters.

5 Discussion

In this study, we conducted a comprehensive examination of over 6000 news articles covering salmon production and markets, intending to assess the influence of news on the stock returns of the largest salmon-producing companies listed on the Oslo Stock Exchange. To derive meaning from this unstructured data, we employed Latent Dirichlet Allocation (LDA) to generate topics, as well as a dictionary approach to analyze sentiment.

Initially, we explored the impact of topics on markets by consolidating them into components and utilizing Vector Autoregression (VAR) analyses on these components and logarithmic stock returns. Given that news coverage can encompass both positive and negative aspects, the effect's direction remains undefined, thus constraining our analysis to absolute returns. Owing to the specific time frame examined, news concerning the Covid-19 pandemic dominated the topics and their market repercussions. Nevertheless, we identified significant market reactions driven by the component related to corporate news and stocks, too.

Upon evaluating topics combined with sentiment, we found that sentiment dictionaries are not sufficiently adaptable to domains other than those they were originally designed for. Specifically, a surprise increase in sentiment constructed using the Loughran-McDonald dictionary, which was tailored for financial data, resulted in a marginally significant effect on logarithmic returns with an incorrect sign. One rationale for this outcome is that the dictionary was designed to detect sentimental expressions in the financial reporting of companies, neglecting industry-specific news.

Moreover, our stock index solely comprised the largest companies on the Oslo Stock Exchange, influencing the results due to the competitive market structure of salmon markets. Although concentrating only on news directly and specifically related to the underlying firms could solve this issue, it would simultaneously disregard vital general news about the salmon industry, such as research not conducted by the producers, or news about competitors that could impact other firms, either directly or indirectly. Alternatively, we addressed this issue by inverting sentiment in articles concerning competitor firms.

To tackle these challenges, we augmented the LM sentiment dictionary by incorporating industry-specific terms and considering the market structure, thereby constructing a sentiment index that has the desired impact on stock returns. This methodological contribution holds general relevance, since language is highly dependent on context. Employing the extended dictionary and explicitly considering competition between producers, we achieved the anticipated positive correlation between sentiment and returns. Additionally, we observed the effects of industry-specific topics on returns, including algal blooms, diseases, and R&D.

However, we discovered that the extended dictionary and explicit competition effects could not account for the Covid topic, representing an archetypal unanticipated exogenous shock to the market that could not have been represented in the dictionary ex ante. Although space constraints in this paper did not permit an in-depth discussion of numerous additional topics, we contend that the enhanced sentiment index we devised will prove beneficial for future investigations of news effects on financial markets.

In our out-of-sample experiment, we aimed to demonstrate the value of incorporating sentiment and topics into financial forecasting models. We found that integrating components based on topics multiplied with sentiment, particularly extended sentiment, significantly improved the performance of the SVM model in predicting stock market price movements, as evidenced by the higher weighted average F1 score. This improvement underlines that news information can provide substantial predictive power beyond what is captured by price data alone.

One constraint of our study was the limited number of news articles and the relatively brief time series under examination. Future studies could consider broadening the time series horizon and incorporating articles from additional news sources to address these limitations. We believe that our extended dictionary holds potential for application in similar studies within aquaculture economics beyond the salmon industry. Moreover, future research focusing exclusively on salmon markets could contemplate further expanding the dictionary to encompass competitive seafood markets, such as shrimp, tuna, and others, thereby enhancing the scope and applicability of the sentiment analysis. Finally, the exploration of enhanced data filtering techniques that preserve general market news and competitor information, while still removing unrelated noise, could be considered for future research. In particular, our study emphasized the importance of accounting for competition and market structure, but a trade-off emerged between focusing on articles directly related to the firms in question (thereby increasing significance), and retaining crucial news about the overall market.

References

Andersen, B. P., & de Lange, P. E. (2021). Efficiency in the atlantic salmon futures market. *Journal of Futures Markets*, *41*(6), 949–984.

Asche, F. (2008). Farming the sea. *Marine Resource Economics*, *23*(4), 527–547.

Asche, F., Misund, B., & Oglend, A. (2019). The case and cause of salmon price volatility. *Marine Resource Economics*, *34*(1), 23–38.

Blei, D. M., & Lafferty, J. D. (2009). Topic models. *Text mining: classification, clustering, and applications*, *10*(71), 71–89.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993–1022.

Bloznelis, D. (2016). Salmon price volatility: A weight-class-specific multivariate approach. *Aquaculture economics & management*, *20*(1), 24–53.

Bloznelis, D. (2018). Hedging salmon price risk. *Aquaculture Economics & Management*, *22*(2), 168–191.

Dahl, R. E., & Oglend, A. (2014). Fish price volatility. *Marine Resource Economics*, *29*(4), 305–322.

Dahl, R. E., Oglend, A., & Yahya, M. (2021). Salmon stock market prices revealing salmon price information. *Marine Resource Economics*, *36*(2), 173–190.

Dahl, R. E., & Yahya, M. (2019). Price volatility dynamics in aquaculture fish markets. *Aquaculture Economics & Management*, *23*(3), 321–340.

Ewald, C.-O., Haugom, E., Kanthan, L., Lien, G., Salehi, P., & Størdal, S. (2022). Salmon futures and the fish pool market in the context of the capm and a three-factor model. *Aquaculture Economics & Management*, *26*(2), 171–191.

FAO. (2018). *The state of world fisheries and aquaculture 2018*.

FAO. (2020). *The state of world fisheries and aquaculture*.

Garlock, T., Asche, F., Anderson, J., Bjørndal, T., Kumar, G., Lorenzen, K., Ropicki, A., Smith, M. D., & Tveterås, R. (2020). A global blue revolution: Aquaculture growth across regions, species, and countries. *Reviews in Fisheries Science & Aquaculture*, *28*(1), 107–116.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National academy of Sciences*, *101*(suppl_1), 5228–5235.

Guttormsen, A. G. (1999). Forecasting weekly salmon prices: Risk management in fish farming. *Aquaculture Economics & Management*, *3*(2), 159–166.

Hansen, S., McMahon, M., & Prat, A. (2018). Transparency and deliberation within the fomc: A computational linguistics approach. *The Quarterly Journal of Economics*, *133*(2), 801–870.

Hersoug, B. (2021). Why and how to regulate norwegian salmon production?–the history of maximum allowable biomass (mab). *Aquaculture*, *545*, 737144.

Justeson, J. S., & Katz, S. M. (1995). Technical terminology: Some linguistic properties and an algorithm for identification in text. *Natural language engineering*, *1*(1), 9–27.

Karalevicius, V., Degrande, N., & De Weerdt, J. (2018). Using sentiment analysis to predict interday bitcoin price movements. *The Journal of Risk Finance*, *19*(1), 56–75.

Khedr, A. E., Yaseen, N. et al. (2017). Predicting stock market behavior using data mining technique and news sentiment analysis. *International Journal of Intelligent Systems and Applications*, *9*(7), 22.

Larsen, V. H. (2021). Components of uncertainty. *International Economic Review*, *62*(2), 769–788.

León-Muñoz, J., Urbina, M. A., Garreaud, R., & Iriarte, J. L. (2018). Hydroclimatic conditions trigger record harmful algal bloom in western patagonia (summer 2016). *Scientific reports*, *8*(1), 1330.

Li, X., Wu, P., & Wang, W. (2020). Incorporating stock prices and news sentiments for stock market prediction: A case of hong kong. *Information Processing & Management*, *57*(5), 102212.

Li, X., Xie, H., Chen, L., Wang, J., & Deng, X. (2014). News impact on stock price return via sentiment analysis. *Knowledge-Based Systems*, *69*, 14–23.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-ks. *The Journal of finance*, *66*(1), 35–65.

Montes, R. M., Rojas, X., Artacho, P., Tello, A., & Quiñones, R. A. (2018). Quantifying harmful algal bloom thresholds for farmed salmon in southern chile. *Harmful Algae*, *77*, 55–65.

Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications*, *42*(24), 9603–9611.

Oglend, A. (2013). Recent trends in salmon price volatility. *Aquaculture Economics & Management*, *17*(3), 281–299.

Shapiro, A. H., Sudhof, M., & Wilson, D. (2022). Measuring news sentiment. *Journal of Econometrics*, *228*, 221–243.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, *62*, 1139–1168.

Appendix: component tables

Component number	Component label	variation explained	Component's Top 5 Topics	Correlation
1	Business Figures	4%	T53 Quarterly Results	57%
			T19 Public & Social	-49%
			T67 Drop, Fall, Decline	48%
			T12 Future Challenges	-47%
			T41 Business Results	44%
5	Covid & Production	2.2%	T37 Norwegian Salmon Prices (kg)	-50%
			T28 Salmon harvesting results	-48%
			T36 Covid Pandemic	37%
			T83 Covid Cases in Production Facilities	30%
			T10 Outlook and Opinions on Prices	-29%
6	Covid & Markets	2.1%	T59 Salmon Commodity Prices	-41%
			T36 Covid Pandemic	37%
			T39 Salmon Wholesale Prices	-36%
			T13 Challenges, Market risk & Uncertainty	36%
			T82 Mergers & Acquisitions	-32%
9	Chile	1.7%	T76 Chilean Producers	52%
			T52 Chilean Salmon Farming	48%
			T47 Contracts & Agreements	32%
			T61 Algal Blooms	27%
			T33 SalMar Offshore Farming	-25%
12	Norwegian Farming	1.5%	T33 SalMar Offshore Farming	36%
			T96 NTS Acquisition of NRS	33%
			T85 Atlantic Sapphire Land-based	-30%
			T95 Project Licenses Norway	24%
			T3 Aquabounty Land-based & GM	-24%
17	Corporate & Stocks	1.4%	T41 Business Results	32%
			T98 Global Salmon Markets	-29%
			T48 Management & Boards	29%
			T69 Stock Markets	29%
			T51 Salmon Market Analysis	-29%

Table A1: Topic components and the five topics that exhibit the highest absolute correlation with each component.

Component number	Component label	variation explained	Component's Top 5 Topics	Correlation
1	Business expectations	26.5%	T43 Plans & Strategy	82%
			T35 Fear, Harm, Negative Outlook	81%
			T68 Business Data	74%
			T12 Future Challenges	74%
			T47 Contracts & Agreements	73%
3	Investments Norway	2.1%	T33 SalMar Offshore Farming	39%
			T96 NTS Acquires NRS	38%
			T10 Opinion & Judgement on Future Prices	-33%
			T82 Mergers& Acquisition	30%
			T70 Smolt Production Facilities Investment	31%
4	Investments ROW	2.1%	T14 Processor Acquisitions (UK)	38%
			T78 US Fund Raising	36%
			T97 Mowi business & Management	-36%
			T63 Retail	36%
			T55 Strategy, Investments, Opportunities	28%

Table A2: LM sentiment-weighted topic components and the five LM sentiment-weighted topics that exhibit the highest absolute correlation with each component.

Component number	Component label	variation explained	Component's Top 5 Topics	Correlation
1	Business Plans & Expectations	28.9%	T35 Fear, Harm, Negative Outlook	85%
			T43 Plans & Strategy	83%
			T68 Business Data	78%
			T60 People, Projects & Perspectives	76%
			T66 Green Bonds	75%
4	Chile & Algal Blooms	2.0%	T61 Algal blooms	45%
			T76 Chilean Producers	38%
			T10 Opinion & Judgement on Future Prices	-37%
			T65 Seafood Markets & Consumption	-35%
			T52 Chilean Salmon Farming	33%
7	Salmon Industry	1.6%	T37 Norwegian Salmon Prices	36%
			T82 Mergers & Acquisitions	-36%
			T56 Bakkafrost Acquires SSC	-28%
			T11 Land-based Facility by Salmon Evolution & Dongwon	26%
			T28 Salmon Harvesting Results	26%

Table A3: Components of topics multiplied with the extended sentiment and the five topics multiplied with the extended sentiment that exhibit the highest absolute correlation with each component.

Mikaella Zitti

School of Economics and Business
Norwegian University of
Life Sciences (NMBU)
P.O Box 5003
N-1432 Ås, Norway

E-mail:
mikzitti@gmail.com

Mikaella Zitti was born in Limassol, Cyprus in 1993. She holds a BSc. Degree in Mathematics from the National and Kapodistrian University of Athens (2016) and a MSc. Degree in Financial Risk Management from the University of Glasgow (2017). The thesis consists of an introduction and four independent papers. The papers investigate the importance of climate-related financial disclosures in the salmon industry, the forecasting ability of advanced models such as Long Short-term Memory (LSTM), and factors that impact the salmon market based on news articles.

Paper I discusses the importance of climate-related financial disclosures in the salmon industry. Results suggest a trend of increased transparency, encouraged by organizations such as the TCFD and CDP, leading to better practices in handling climate-related risks.

Paper II evaluates the predictive power of LSTM in forecasting salmon market volatility. Results reveal that the LSTM model did not outperform the benchmark ARMA model in terms of forecasting accuracy, suggesting that salmon market volatility may not display complex temporal patterns.

Paper III examines a hybrid model combining a Vector Autoregressive (VAR) approach with LSTM against a benchmark model for predicting salmon prices. Results suggest the hybrid model doesn't enhance salmon price forecasting, indicating salmon prices might not follow non-linear patterns. This implies the salmon market's efficiency, where new information is quickly recognized and used by investors.

Paper IV investigates factors influencing the salmon commodity market using salmon-news-article data. Topic modelling algorithms identify key factors impacting the salmon market, complemented by a lexicon-based analysis. This study emphasizes the competitive nature of the global salmon market and identifies factors that contribute to its price fluctuations, with Covid-19 recognized as a significant contributor. An out-of-sample robustness test further validates the efficacy of the extended dictionary.

Thesis number: 2023:72
ISSN: 1894-6402
ISBN: 978-82-575-2103-5

Supervisor: Professor Atle G. Guttormsen.

ISBN: 978-82-575-2103-5
ISSN: 1894-6402

Norwegian University
of Life Sciences

Postboks 5003
NO-1432 Ås, Norway
+47 67 23 00 00
www.nmbu.no

118789 / ANDVORDGRAFISK.NO