



Norwegian University of  
Science and Technology

# Predicting the Distribution of the Atlantic Salmon Spot Price Using Quantile Regression

**Paal Olav Warbo Mjaugeto**  
**Kristian Sandaker**  
**Kjartan Berge Steinshamn**

Industrial Economics and Technology Management

Submission date: May 2017

Supervisor: Sjur Westgaard, IØT

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



## PROBLEM DESCRIPTION

---

During the past decades, the price of Atlantic salmon has been subject to increased fluctuations both in terms of frequency and magnitude, leading to uncertainty and costs imposed on the entire value chain for salmon farming. Alongside this rise in volatility, the industry has experienced significant growth. In particular, the global aquaculture production of Atlantic salmon has increased from a few thousand tonnes in 1980, to beyond two million tonnes in 2016. Today, production, processing, and marketing of salmon is a multi-billion dollar industry that has an increasing presence in the capital markets. It is correspondingly relevant to have a detailed outlook on the possible price paths of salmon. Operationally, a consideration such as harvest timing is sensitive to the spot price in the short term. Correspondingly, the production cycle spans up to three years, making a decision such as smolt release dependent on long-term prices. Financially, accurate price expectations can facilitate improved risk management, valuations, and investment decisions. The ability to make accurate predictions of the salmon spot price is thus of increasing importance; also, the capability to reason about the uncertainty in such predictions may prove beneficial.

In this study, we aim to assess and reason about the future uncertainty of the Atlantic salmon spot price through predicting the corresponding price distribution. To represent the price, we use the NASDAQ Salmon Index, denoted in USD. There are no past studies on predicting the distribution of the Atlantic salmon spot price. Our approach is twofold:

- i) Methodologically, we build a framework based on quantile regression to predict the distribution of the 1- to 12-month ahead log returns of the Atlantic salmon spot price. We use a genetic algorithm for variable selection, which in the context of quantile regression is a novel approach. Based on the chosen variables across these twelve months, we build corresponding quantile regression models to estimate the multi-step ahead conditional distribution.
- ii) Qualitatively, we provide industry hypotheses, corresponding results from the modelling, discussions, and scenario analyses to give insight into which exogenous factors affect the 1- to 12-month ahead distribution of the log returns of the spot price. According to the industry, the price is generally argued to be affected by factors such as harvest volumes, disease outbreaks, and prices of alternative proteins. Through the distribution model, we are able to give an increased understanding of these factors' importance and joint effect on the salmon spot price.



## PREFACE

---

This thesis concludes our Master of Science degree in Industrial Economics and Technology Management at the Norwegian University of Science and Technology (NTNU). It is original and independent work performed by Paal Olav Warbo Mjaugeto, Kristian Sandaker, and Kjartan Berge Steinshamn, written during the spring of 2017.

We would like to thank our supervisor, Professor Sjur Westgaard at the Department of Industrial Economics and Technology Management (NTNU), for helpful guidance and advice. His interest in our work has been truly valuable during the completion of our master's thesis. A thankful note is also directed to Professor Frank Asche, at the School of Forest Resources & Conservation at the University of Florida, for beneficial collaboration and feedback. Moreover, we have received guidance and valuable insights on the salmon market from an undisclosed company in the aquaculture industry. Our thesis and analyses have benefited considerably from their involvement, and we are grateful for their advice. Finally we would like to thank our fellow students at NTNU, friends and family for appreciated comments and support during our work with this thesis.



## ABSTRACT

---

The growth in salmon farming production has outperformed the average growth in aquaculture production during the last decades. Alongside this growth, the industry has strengthened its presence in the capital markets. However, the salmon price has become increasingly volatile, imposing uncertainty and additional costs across the entire value chain. This motivates the development of more accurate price distribution prediction models. Such models can support operational and financial considerations that are subject to price uncertainty, such as harvest timing, futures hedging, and investments.

This study aims to provide a framework for predicting the distribution of the Atlantic salmon spot price, and to identify the most important exogenous drivers in this respect. The spot price is represented by the NASDAQ Salmon Index (NQSALMON), denoted in USD. We build on a database of 25 carefully selected explanatory variables, and design a 1- to 12-month ahead quantile regression prediction model for the spot price. Each of the twelve months to be predicted are appointed a designated *submodel*, which is independent of the other submodels. Moreover, each submodel utilises eight explanatory variables that are selected from the database using a novel genetic algorithm-assisted variable selection approach.

There are several encouraging results from this study. First, the approach with a genetic algorithm for variable selection in quantile regression quickly finds submodels with favourable goodness-of-fit. Generally, the 2- to 7-month submodels explain slightly more variation in the NQSALMON distribution compared to the 8- to 12-month submodels. Regarding model specification, the regressions provide the correct unconditional coverage; however, the exceedances at quantiles are often clustered. This might be the result of unexplained, non-linear effects and higher-order statistical moments. However, the quantile regression equation specification error test (QRESET) suggests that linear quantile regression in most cases is a sound functional form for the given data. Finally, the results indicate that exogenous variables such as standing biomass, feed consumption, and prices of alternative proteins are important predictors of the future spot price of Atlantic salmon.





I løpet av de siste tiårene har produksjonsveksten innen lakseoppdrett overgått den gjennomsnittlige produksjonsveksten innen akvakultur for øvrig. Parallelt med denne veksten har industrien befestet sin tilstedeværelse i kapitalmarkedene. Samtidig har lakseprisen blitt stadig mer volatil, hvilket har resultert i økt risiko og økte kostnader i hele verdikjeden for lakseoppdrett. Dette motiverer utvikling av mer nøyaktige modeller for prediksjon av prisfordelingen til laks, som senere kan bli brukt til å bistå ved operasjonelle og finansielle beslutninger som er særlig sensitive til prisusikkerhet. Eksempler her er planlegging av utslakt, prissikring og investeringer.

Dette studiet forsøker å kartlegge de viktigste prisdriverne for atlantisk laks, samtidig som det foreskriver et rammeverk for prediksjon av den statistiske fordelingen til spotprisen. Spotprisen er representert ved NASDAQ Salmon Index (NQSALMON), angitt i USD. Vi tar utgangspunkt i en database bestående av 25 omstendelig utvalgte forklaringsvariabler, og vi predikerer spotprisen 1- til 12-måneder frem i tid ved bruk av kvantilregresjon. Hver og én av de tolv månedene som predikeres tilegnes én unik *delmodell*, som er uavhengig av de andre delmodellene. Videre er hver delmodell basert på åtte forklaringsvariabler, som velges fra den nevnte databasen ved bruk av en genetisk algoritme.

Vi oppnår en rekke lovende resultater i dette studiet. Anvendelsen av en genetisk algoritme for variabelseleksjon viser seg å være passende, ettersom algoritmen raskt klarer å identifisere delmodeller med tilfredsstillende forklaringskraft. Videre indikerer resultatene at delmodellene for 2- til 7-måneder presterer best i å forklare variasjonen i prisfordelingen til NQSALMON. Hva gjelder modellspesifikasjon, så oppnår regresjonene korrekt dekning av kvantilene. Samtidig observerer vi at kvantiloverskridelser ofte er grupperte, hvilket kan være en konsekvens av at ikke-lineære effekter og høyere ordens momenter ikke blir tilstrekkelig inkorporert i modellen. Videre antyder resultatene fra "quantile regression equation specification error test" (QRESET) at lineær kvantilregresjon er en passende funksjonell form for dataen modellen forsøker å beskrive. Avslutningsvis indikerer resultatene våre at eksogene variabler som stående biomasse, fôrforbruk og priser på alternative proteiner har stor prediktiv kraft på prisen for atlantisk laks.



# CONTENTS

---

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Literature Review</b>	<b>4</b>
2.1	Literature on Modelling Salmon Prices . . . . .	4
2.2	Literature On Predicting the Price of Other Commodities . . . . .	6
2.3	This Study in the Context of Existing Literature . . . . .	7
<b>3</b>	<b>Factors Affecting the Salmon Price</b>	<b>9</b>
3.1	Supply-Driving Factors . . . . .	9
3.2	Demand-Driving Factors . . . . .	12
3.3	Other Factors . . . . .	14
3.4	Hypothesis Overview . . . . .	14
<b>4</b>	<b>Data Analysis</b>	<b>16</b>
4.1	The NASDAQ Salmon Index . . . . .	16
4.2	Data Sources . . . . .	17
4.3	Data Pre-Processing . . . . .	17
4.4	Descriptive Statistics . . . . .	18
4.5	Choosing the Appropriate Time Window . . . . .	22
<b>5</b>	<b>Methodology</b>	<b>27</b>
5.1	Methodology Overview . . . . .	27
5.2	The Linear Quantile Regression Model . . . . .	29
5.3	Limiting the Number of Quantile Regression Predictors . . . . .	30
5.4	Variable Selection Through a Genetic Algorithm-Assisted Search . . . . .	31
5.5	Model Rationale and Challenges . . . . .	36
5.6	Specification Tests . . . . .	39
<b>6</b>	<b>Results and Discussion</b>	<b>42</b>
6.1	Genetic Algorithm Search Evaluation . . . . .	42
6.2	In-Sample Model Performance . . . . .	44
6.3	In-Sample Model Specification . . . . .	45
6.4	Key Explanatory Variables . . . . .	51
<b>7</b>	<b>Scenario analysis</b>	<b>56</b>
<b>8</b>	<b>Conclusion</b>	<b>59</b>
8.1	Motivation, Findings, and Remarks . . . . .	59
8.2	Further Research . . . . .	60
<b>A</b>	<b>Appendix</b>	<b>62</b>
A.1	Factors Affecting the Salmon Price . . . . .	62
A.2	Data Series Used in the Model . . . . .	63
A.3	In-sample Quantile Regression Results . . . . .	73
A.4	In-Sample Specification Test Results . . . . .	78
A.5	Cross-Correlation Tables for the Explanatory Variables . . . . .	79
A.6	Tests for Statistical Properties . . . . .	81
A.7	Model Performance Measure . . . . .	83
A.8	Test For Model Specification . . . . .	84
A.9	Data Sources . . . . .	85
A.10	Background: Important Aspects of the Salmon Farming Industry . . . . .	87
	<b>References</b>	<b>90</b>

## LIST OF FIGURES

---

Figure 1	Monthly index prices for salmon, meat and poultry . . . . .	12
Figure 2	Weekly index prices of the NQSALMON . . . . .	16
Figure 3	Descriptive figures of the monthly log returns of NQSALMON . . . . .	19
Figure 4	Cross-correlation surfaces . . . . .	24
Figure 5	Methodology overview . . . . .	29
Figure 6	Genetic algorithm (GA) iteration plots . . . . .	43
Figure 7	1- to 12-month in-sample prediction . . . . .	45
Figure 8	Distribution of $p$ -values for the submodels' quantiles . . . . .	46
Figure 9	In-sample exceedance plot . . . . .	49
Figure 10	Results from the scenario analysis . . . . .	57
Figure 11	Sensitivity of the NQSALMON to changes in harvest volumes . . . . .	62
Figure 12	Plot of the explanatory variable time series (p. 1). . . . .	65
Figure 13	Plot of the explanatory variable time series (p. 2). . . . .	66
Figure 14	Plot of the explanatory variable time series (p. 3). . . . .	67
Figure 15	Plot of the explanatory variable time series (p. 4). . . . .	68
Figure 16	Plot of the explanatory variable time series (p. 5). . . . .	69
Figure 17	1-month ahead Pearson correlation (p. 1). . . . .	70
Figure 18	1-month ahead Pearson correlation (p. 2). . . . .	71
Figure 19	1-month ahead Pearson correlation (p. 3). . . . .	72
Figure 20	Historical harvest volumes of Atlantic salmon. . . . .	87

## LIST OF TABLES

---

Table 1	Variable and hypothesis overview. . . . .	15
Table 2	Descriptive statistics of the log returns of the data series . . . . .	21
Table 3	Lags and time windows: Assumptions versus empirical evidence . . . . .	26
Table 4	Combinatorial optimisation program for variable selection . . . . .	33
Table 5	Genetic algorithm parametrisation . . . . .	36
Table 6	Pseudo-R-squared test results . . . . .	44
Table 7	Overview of coefficient significances for the model . . . . .	46
Table 8	Aggregated specification test results split by quantiles . . . . .	47
Table 9	Aggregated specification test results for the submodels . . . . .	47
Table 10	Cross-correlation between log returns for horizon $h = 1$ . . . . .	53
Table 11	Overview of the covariates employed for each submodel . . . . .	55
Table 12	Data used in the scenario analysis. . . . .	56
Table 13	Complete overview of the time series in the database . . . . .	63
Table 14	Descriptive statistics of the arithmetic returns of the data series . . . . .	64
Table 15	Coefficients overview of in-sample quantile regressions, $h \in \{1, 2, 3\}$ . . . . .	74
Table 16	Coefficients overview of in-sample quantile regressions, $h \in \{4, 5, 6\}$ . . . . .	75
Table 17	Coefficients overview of in-sample quantile regressions, $h \in \{7, 8, 9\}$ . . . . .	76
Table 18	Coefficients overview of in-sample quantile regressions, $h \in \{10, 11, 12\}$ . . . . .	77
Table 19	Extended specification test results . . . . .	78
Table 20	Cross-correlation between log returns, $h \in \{1, \dots, 6\}$ . . . . .	79
Table 21	Cross-correlation between log returns, $h \in \{7, \dots, 12\}$ . . . . .	80

## LIST OF ALGORITHMS

---

Algorithm 1	Genetic algorithm for variable selection . . . . .	34
-------------	--	----

## ACRONYMS

---

**ADF** augmented Dickey-Fuller

**ARIMA** autoregressive integrated moving average

**ARMA** autoregressive moving average

**CAD** classical additive decomposition

**DF** Dickey-Fuller

**FPI** Fish Pool Index<sup>TM</sup>

**GA** genetic algorithm

**GARCH** generalised autoregressive conditional heteroscedasticity

**HOG** head-on-gutted

**HW** Holt-Winters exponential smoothing

**i.i.d.** independent and identically distributed

**JB** Jarque-Bera

**LASSO** least absolute shrinkage and selection operator

**MoM** month-over-month

**NQSALMON** NASDAQ Salmon Index

**OLS** ordinary least squares

**PCA** principal component analysis

**PCR** principal component regression

**QRESET** quantile regression equation specification error test

**RESET** regression equation specification error test

**SCAD** smoothly clipped absolute deviation

**SVM** support vector machine

**VAR** vector autoregression

**VaR** value-at-risk

**YoY** year-over-year



## INTRODUCTION

---

Seafood prices are volatile, and that is also the case for the salmon price (Asche, Dahl, & Steen, 2015; Dahl & Oglend, 2014). During the past decades, the salmon price has been subject to increased fluctuations both in terms of frequency and magnitude (Oglend, 2013). This high volatility brings uncertainty, leading to additional, and possibly unnecessary, costs imposed on the entire value chain for salmon farming. In addition, the market for salmon is becoming more globalised, forcing salmon farming companies to stay highly competitive to maintain their market positions (Tveteras & Asche, 2008).

The price volatility and the increasingly competitive market motivate the creation of a model for predicting the salmon price. A reliable model could be of use for the entire value chain for salmon, ranging from farmers, and their suppliers, to exporters and retailers. As an example, a price model would provide vital decision support in determining when to harvest the fish,<sup>1</sup> or on which terms producers should engage in salmon forward contracts. Furthermore, alongside the extensive production growth within salmon farming during the last decades (Brækkan & Thyholdt, 2014), the industry has strengthened its presence in the capital markets. A number of companies have been listed on stock exchanges, there have been numerous corporate bond issuances, and the need for financial hedging instruments has become evident, leading to the establishment of Fish Pool.<sup>2</sup> Accordingly, investors and analysts pay increased attention to the salmon farming industry. From a financial standpoint, an accurate price model could facilitate better stock and bond valuations, investment decisions, risk management, and trading. Finally, for a more thorough assessment of future uncertainty, the ability to predict a range of possible future prices, rather than a price point, could prove useful. This enables relevant parties to plan according to the different outcomes from a predicted future price distribution (Chatfield, 2000).

This study has two key purposes: Firstly, to provide market participants with a tool for prediction of the salmon price distribution. Secondly, to achieve increased understanding of how fundamental explanatory variables, such as harvest volumes, smolt release, and sea lice occurrence, affect the price formation of salmon. Hence, we develop a probabilistic model based on linear quantile regression, as by Koenker and Bassett (1978), enabling us to obtain a complete picture of how *exogenous* variables affect the salmon price at various quantiles. We use the NASDAQ Salmon Index (NQSALMON), denoted in USD, to represent the spot price of salmon.<sup>3</sup>

Our model selects its explanatory variables from a database consisting of 25 time series, of which 24 are exogenous. Each of the 24 exogenous variables is carefully included in the database based on discussions with industry experts. Moreover, most of the variables are expected to have a lagged relationship with respect to the spot price, and industry experts have therefore helped us to develop sets of realistic lags for each variable. The literature provides no insight on which combinations of variables, at what lags, have strong predictive power for the spot price. Hence, we let a variable selection procedure search the space of such combinations. As this search space grows exponentially in the number of variables, an exhaustive search is computationally demanding. Therefore, we employ a genetic algo-

---

<sup>1</sup>A. Guttormsen (2008) provides an example of a harvesting model showing how harvest responds to changes in prices and other decision variables.

<sup>2</sup>Fish Pool is an international marketplace for price hedging of seafood products. See Appendix A.10 for more information.

<sup>3</sup>All variants in the text for expressing the salmon price will from here on concern the NQSALMON, i.e., the export price received by Norwegian producers of fresh *Atlantic* salmon, unless otherwise stated. See Section 4.1 for a complete definition of the NQSALMON. Note that all salmon farmed in Norwegian waters belong to the *Atlantic salmon* species (*Salmo salar*).

rithm (genetic algorithm (GA)), which quickly is able to find good candidate solutions in this setting (Oreski & Oreski, 2014). We utilise these selected variables (at some particular lags) as covariates in a quantile regression framework for modelling the 1- to 12-month ahead distributions of the NQSALMON. To obtain each distribution, we develop a *sub-model* with corresponding regressions for five quantiles. Each submodel has a fixed set of eight covariates selected from the aforementioned, GA-assisted search. On a final note, we predict the density of the *log returns* of the salmon spot price in this framework; however, this density may be converted to a corresponding *prediction* of the price distribution.<sup>4</sup>

Furthermore, for illustrative purposes we implement a scenario analysis based on the above framework. Here, the value of a single covariate is varied to create a conditional distribution of the salmon price. Although not covered in detail by this study, our framework can be used to further examine how a particular covariate impacts various parts of the price distribution, and whether it can be considered a significant driver of risk.

The existing literature on both point and density prediction of the salmon spot price is scarce. To our knowledge, this study is the first in applying exogenous variables to predict the distribution of the salmon price, irrespective of the salmon species under consideration. Also, we have not encountered any recent work that model the 1- to 12-month ahead salmon price. Hence, our study represents a significant contribution to the literature. It should provide risk managers in the industry, traders, and analysts with an important tool for handling price risk and developing hedging strategies. Moreover, the use of genetic algorithms for variable selection is, to our knowledge, not represented in the literature for quantile regression.

From a methodological standpoint, there are several encouraging results from this study. Firstly, the novel approach with a genetic algorithm for variable selection in quantile regression shows promising properties. It relatively quickly finds subsets of the database that yield favourable goodness-of-fit for the respective regressions, and it is able to identify covariates whose coefficients jointly have high significances – particularly at more central quantiles. Generally, the submodels are able to explain a somewhat higher degree of the variation in the 2- to 7-month ahead log returns' respective quantiles, to the disadvantage of slightly longer horizons. Regarding model specification, the regressions provide the correct unconditional coverage; however, the quantile exceedances are often clustered. Clustered exceedances at quantiles might, however, be explained by non-linear effects and higher-order statistical moments. Nonetheless, functional specification tests indicate that, in most cases, linear quantile regression is a sound model for the given data. In sum, we have been able to fit quantile regression submodels with often highly significant coefficients and relatively favourable goodness-of-fit; however, there are some misbehaviour that can be explored in further research.

Next, from a qualitative point of view, the results indicate that exogenous variables such as standing biomass, feed consumption, prices of alternative animal proteins, and sea temperatures are important predictors for the Atlantic salmon spot price. Over the twelve submodels, these variables tend to have the largest coefficients and greatest significance. However, as the regressions appear to be subject to multicollinearity effects in their variables, we are unable to give a full interpretation of the submodels' respective regression coefficients.

The remainder of the study is structured as follows: **Chapter 2** discusses previous research related to modelling different aspects of the salmon market, particularly the salmon price and returns, and corresponding volatility. We also elaborate on past applications of

---

<sup>4</sup>We write that we develop a model for *prediction* of the salmon price. As we do not evaluate the model out-of-sample, we are reluctant to define this study as one that conducts *forecasts*, as this is often synonymously with out-of-sample evaluations in the literature. However, when the model is specified in-sample, it predicts the salmon price 1- to 12-months ahead, based on the information available at the start of that period. Hence, the model can easily be applied to forecast the price, both out-of-sample and in actual forward-looking situations.



quantile regression for commodity modelling. **Chapter 3** provides a thorough discussion of explanatory variables and associated lags, whereas **Chapter 4** describes the processing of the collected data, presents descriptive statistics for the NQSALMON and the explanatory variables in the database, and elaborates on the data transformations that have been carried through in order to generate the respective log returns. **Chapter 5** outlines the methodology, and puts particular emphasis on the concepts of quantile regression and GA-aided variable subset selection. **Chapter 6** then proceeds with the results, presenting the performance of the variable selection procedure, the different model formulations, in-sample model performance, and key explanatory variables. Next, **Chapter 7** provides an example of applying the model for scenario analysis. Finally, in **Chapter 8**, we present our conclusions and suggestions for further research.

## LITERATURE REVIEW

---

This study may be placed in the context of two research areas: (i) to reveal the fundamental factors affecting the price of salmon, and (ii) to model its price distribution. As follows from the consecutive parts of this chapter, it is clear that understanding the formation of the salmon price, and modelling of its price distribution, are novel topics in the existing literature. Thus, we believe this study contributes substantially to the field of salmon price prediction, and it could relatively easily be extended to forecasting.

### 2.1 LITERATURE ON MODELLING SALMON PRICES

The existing literature on modelling salmon prices is rather twofold. On the one hand, there are studies modelling salmon prices *directly* and pointwise, which are scarce and dated. By *directly*, we refer to research examining explicit modelling and prediction of the salmon price. To our knowledge, the last published study modelling the salmon price directly is carried out by [A. G. Guttormsen \(1999\)](#). On the other hand, studies modelling the salmon price *indirectly* are somewhat more frequently represented. By *indirectly*, we refer to studies that discuss other aspects related to the price, such as volatility and salmon futures. However, the existing literature on price modelling and prediction has not dealt specifically with the NQSALMON. Moreover, general price distribution forecasts and scenario analyses are scarce, irrespective of commodity investigated. In the following sections, contributions to modelling and predicting salmon prices directly and indirectly are considered.

#### *Research Papers on Modelling and Predicting Salmon Prices Directly*

Although the existing literature on modelling and predicting the salmon price directly is limited and dated, a variety of econometric methods have been employed. In all instances, *pointwise*, not distributional, predictions have been generated.

[Vukina and Anderson \(1994\)](#) and [Gu and Anderson \(1995\)](#) both use a state-space approach to model and forecast salmon prices directly and pointwise. [Vukina and Anderson \(1994\)](#) model and forecast the price of five different salmon products on the Tokyo wholesale market, using four state-space models for modelling vector-valued nonstationary time series. The price series evidently exhibit cyclical behaviour, and the out-of-sample forecasts generated by the models are satisfactorily accurate, encouraging future research in the area.

[Gu and Anderson \(1995\)](#) develop deseasonalised state-space time series models to provide short-term price forecasts for the U.S. salmon market. The time series used include the U.S. fresh Atlantic salmon wholesale price index, fresh salmon (that is; Atlantic, coho and Chinook) monthly U.S. import quantities and prices, and U.S. chum and sockeye salmon monthly export prices. [Gu and Anderson \(1995\)](#) embed seasonal factors in the modelling, by combining an ordinary least squares (OLS)-approach with a state-space time series model. Out-of-sample forecasts 3-, 6-, and 12-months ahead are implemented to validate the models' performance. The study demonstrates that deseasonalisation improves the overall performance of the state-space models, and models that apply deseasonalisation produce substantial out-of-sample predictive power. The paper provides meaningful insight, but the methodology may be considered intricate for market participants with limited knowledge of statistics and econometrics, and thus troublesome to apply for practical use. However, the fact that certain salmon prices exhibit seasonal movements, as emphasised by [Vukina and Anderson \(1994\)](#) and [Gu and Anderson \(1995\)](#), might be a key result impacting the formulation of our model. Certainly, a well-performing forecasting model needs to conform to this property.

A somewhat broader approach, and probably more easily applicable, for forecasting salmon prices directly is employed by [A. G. Guttormsen \(1999\)](#). The author applies in total six different methods to forecast weekly producer prices for the Norwegian salmon market: classical additive decomposition (CAD), Holt-Winters exponential smoothing (HW), autoregressive moving average (ARMA), vector autoregression (VAR), and two naïve techniques. Moreover, 4-week, 8-week, and 12-week ahead forecasts are calculated. The models might be considered simpler and more intuitive to understand and apply for market participants, compared to the procedure carried out by [Gu and Anderson \(1995\)](#). However, the forecasting horizon in this study might be regarded short for some applications; for example, salmon production planning may require a prediction horizon several months ahead due to long farming cycles. For a thorough discussion on the salmon farming industry and its fundamentals, see [Appendix A.10](#).

The out-of-sample forecast results of [A. G. Guttormsen \(1999\)](#) were generally promising. Of the six models, he found that the CAD model forecast the *direction* of price movements best. Moreover, concerning accuracy measures for out-of-sample forecasting, the VAR model performs best. However, [A. G. Guttormsen \(1999\)](#) finds no evidence for a general, superior model for point forecasts of the price of Norwegian salmon. This motivates further research on forecasting salmon prices, particularly for the Norwegian market.

#### *Research Papers on Modelling and Forecasting Salmon Price Volatility*

As previously discussed, [Vukina and Anderson \(1994\)](#), [Gu and Anderson \(1995\)](#) and [A. G. Guttormsen \(1999\)](#) all model a selection of different salmon prices *directly* and *pointwise*. However, the price forecasting techniques they apply somewhat rely on accurate knowledge of the *noise* generating part of the prices. That is, the precision of the forecasts is highly linked to the volatility term in the salmon price process.

Some more recent studies focus on modelling the salmon price volatility, such as [Oglend and Sikveland \(2008b\)](#), [Solibakke \(2012\)](#), and [Oglend \(2013\)](#). The studies present several interesting findings with practical implications. First, they indicate that the salmon price volatility has increased over the past decade, and that the volatility is higher in periods of elevated prices. For industry actors, this implies that greater expected profits usually follow larger price risks.

In particular, [Oglend and Sikveland \(2008b\)](#) employ a generalised autoregressive conditional heteroscedasticity (GARCH)-model to, among other things, test for volatility clustering in Norwegian salmon prices. They find that the previous week's volatility provides predictive power for the next week's volatility. However, they do not find any predictive patterns in the volatility over horizons of one month and above. This implies that the inclusion of volatility parameters might be of limited use for our model, as we aim to model the NQSALMON 12-months ahead using monthly granularity.

[Oglend \(2013\)](#) also models volatility, using weekly salmon prices received by Norwegian producers. He applies GARCH to parametrically model conditional variance with salmon price returns as input, and discovers that the volatility of the salmon price has increased since the start of the 2000s. Oglend argues that the increasing volatility trend is driven by a wide range of factors, but he highlights tighter conditions for supply and demand as an important cause. Moreover, he suggests that prices of *alternative animal proteins* are important drivers of the salmon price, and that they account for a major part of the volatility in the salmon price.

#### *Research Papers on Salmon Futures*

Another topic discussed in the literature is the use of salmon futures for predicting the price of salmon. [Asche, Misund, and Oglend \(2016b\)](#) and [Ankamah-Yeboah, Nielsen, and](#)

Nielsen (2016) study the spot-forward relationship on the Fish Pool Index<sup>TM</sup> (FPI), using different empirical approaches. Here, Asche et al. (2016b) conclude that the futures market is immature; hence, salmon futures lack a price discovery role and the spot price leads the futures prices. However, this contradicts the results of Ankamah-Yeboah et al. (2016), who find that the futures market is mature and closed. Another study on salmon futures, carried out by Asche, Misund, and Oglend (2016a), analyses the relationship between changes in the futures-spot basis and the spot price. They find that the basis does not appear as a robust determinant for changes in the salmon spot price. This is in line with the findings of Asche et al. (2016b), and the study validates their suspicion of limited price discovery abilities in the salmon futures market.

## 2.2 LITERATURE ON PREDICTING THE PRICE OF OTHER COMMODITIES

As previously stated, past research on modelling salmon prices directly is scarce, and, specifically, studies on predicting the NQSALMON are non-existent. However, other commodity markets, such as markets for oil, cattle, and gold are widely discussed in the literature, and there are numerous publications attempting to predict and explain the price of these commodities. Nevertheless, past studies on price distribution forecasts are generally limited irrespective of commodity. In the following sections, a selection of contributions on modelling and forecasting the price of commodities such as crude oil, electricity, and meats are considered.

### *A Selection of Research Papers on Modelling and Forecasting Commodities*

Coleman (2012) models crude oil prices using exogenous variables and linear regression, obtaining satisfactory explanatory power. Furthermore, Ye, Zyren, and Shore (2005) forecast crude oil prices using petroleum inventory levels, achieving superior out-of-sample forecast performance compared to that of a naïve approach. Moreover, Ates, Lusk, and Brorsen (2016) use consumer price expectations to predict a selection of meat prices, concluding that survey data is a suitable indicator for projecting price changes. In other words, a selection of past studies on forecasting commodity prices have proven quite successful. This indicates that it should be possible to develop a model using exogenous variables for forecasting salmon prices as well, including the NQSALMON.

### *Studies on Price Distribution Forecasting*

The literature on *density* and *interval forecasts* is somewhat limited compared to that of *point forecasts* (Diebold, Gunther, & Tay, 1997; Tay & Wallis, 1999).<sup>5</sup> According to Tay and Wallis (1999), the evaluation and use of density forecasts is at an elementary stage. Nevertheless, recent studies have provided promising results with respect to accurate density and interval forecasts. In this section, a selection of papers applying such methods for various commodities are discussed.

Morana (2001) employs GARCH properties in oil price changes to forecast oil price distributions over short-term horizons. The out-of-sample results suggest that the forecast approach might be used to obtain a performance measure for the forward price and estimate interval forecasts for the oil price. Moreover, Zhou, Yan, Ni, Li, and Nie (2006) employ an extended autoregressive integrated moving average (ARIMA) approach to develop hourly spot price forecasts for electricity markets. They incorporate confidence interval estimations in the forecast procedure, and obtain acceptable accuracy even in market conditions with significant volatility. Furthermore, Zhao, Dong, Xu, and Wong (2008) apply a novel

---

<sup>5</sup>*Density forecasts*: Refer to estimation of a probability distribution for the future value of a random variable. They provide a description of the underlying uncertainty of the forecast. *Interval forecasts*: The specification of intervals of future values, based on e.g. quantiles.

data mining-based approach to accurately estimate the prediction interval of electricity price series. They find this method strikingly efficient compared to existing methods such as GARCH models. Hence, overall, past research suggests that density or interval forecasts applied on commodity prices may result in decent accuracy.

### 2.3 THIS STUDY IN THE CONTEXT OF EXISTING LITERATURE

In this study, we predict the 1- to 12-month ahead density of the log returns of the salmon spot price on a monthly resolution. We also seek to identify exogenous variables that, conceivably, have significant explanatory power with respect to the NQSALMON. Thus, the modelling approaches applied in the literature for direct and indirect forecasting of the salmon price are of limited value to our study. This is due to:

- i) The econometric methods applied in the literature are different from the one used in this study. There are few, if any, that use a regression-based approach to predict the price of salmon.
- ii) Our study applies a multi-step ahead procedure to predict the distribution of the NQSALMON log returns, while previous studies have predicted prices pointwise, and mostly over the short term.
- iii) The models specified in the literature for salmon price forecasting do not utilise any exogenous variables – in general, these are purely endogenous models. Hence, they are of no use in determining fundamental price drivers of the salmon price.

These observations also underpin the value of this study. As discussed in [Chapter 1](#), salmon market participants are in need of greater knowledge of which drivers that affect the price of salmon, and a model for predicting its distribution over medium to long term horizons. These two aspects are both covered by our study.

Although past research on the salmon market is of limited value for this study, we have identified a few key results from the literature. These should be accounted for in the model formulation:

- i) Several studies, including [Vukina and Anderson \(1994\)](#) and [Gu and Anderson \(1995\)](#), state that salmon prices exhibit seasonal movements.
- ii) [Oglend and Sikveland \(2008b\)](#), [Solibakke \(2012\)](#), and [Oglend \(2013\)](#) suggest that salmon price volatility has increased in recent years and that raised volatility follows increased salmon prices.
- iii) [Oglend \(2013\)](#) provides empirical evidence suggesting that protein substitutes for salmon might explain a significant part of the salmon price volatility. Also, he highlights tight supply and demand dynamics as an important driver of volatility.
- iv) [Asche and Misund \(2016\)](#) and [Asche et al. \(2016a\)](#) find that salmon futures do not provide any predictive power for the salmon price, suggesting that futures prices should not be included as explanatory variables.

As salmon prices exhibit seasonal movements, an accurate price forecasting model needs to conform to this property. A reasonable strategy, as employed in this study, is to use explanatory variables causally related to the NQSALMON, which also exhibit similar seasonal properties. Moreover, the recent years' increased salmon price volatility suggests that the incorporation of volatility dynamics is essential to create accurate predictions of the salmon price distribution. Thus, our modelling approach needs to accommodate these

dynamics.<sup>6</sup> Furthermore, [Oglend \(2013\)](#) finds that food substitutes for salmon may explain parts of the salmon price volatility dynamics; hence, we have included several explanatory variables for the price of alternative animal proteins in our database. Moreover, Oglend points to the balance between supply and demand as an important trigger of volatility. As both supply and demand are inevitable price determinants for most commodity markets with perishable products, it is essential to capture their dynamics when implementing distribution forecasts. In our modelling procedure, we have attempted to capture their dynamics by identifying their most important determinants, as discussed in the next chapter. Finally, due to the aforementioned findings of [Asche and Misund \(2016\)](#) and [Asche et al. \(2016a\)](#), we choose not to include time series for futures prices in our database.

---

<sup>6</sup>Although volatility is not included in our database as a parameter to be chosen in the variable selection, we construct a time-varying volatility estimate that is added on top of the other explanatory variables used by the model. This is discussed more closely in the results, [Section 6.3](#).

## FACTORS AFFECTING THE SALMON PRICE

---

Our model selects its explanatory variables from a database of 25 time series. The database contains time series highlighted by an undisclosed industry participant and [Marine Harvest \(2016\)](#),<sup>7</sup> a handbook for the salmon farming industry. It provides a discussion on indicators most frequently used by the industry in order to make assumptions on where the salmon price is heading in the future. The rest of this chapter presents the explanatory variables highlighted by the industry participant and [Marine Harvest \(2016\)](#), and suggests the appropriate *lag structure* and *time window* for each variable.<sup>8</sup> We divide the potential explanatory variables into two groups, based on whether they are assumed to affect the supply or demand for salmon.

### 3.1 SUPPLY-DRIVING FACTORS

The NQSALMON has been closely related to the supply of salmon in the past, as depicted by [Figure 11](#) in [Appendix A.1](#). The figure shows that changes in the NQSALMON, *ceteris paribus*, are negatively correlated with changes in the supply of salmon. This relationship is the basis for the discussion that follows.

#### *Smolt Release*

A smolt is defined as a young salmon at the stage when it is transferred from fresh water to seawater cages. Smolt release is an indicator of future supply, and changes in smolt release are expected to be negatively correlated with changes in the NQSALMON.

Smolts typically spend about 16 months in seawater before harvest, implying that figures for smolt release are suitable indicators for predicting medium-term supply ([Marine Harvest, 2016](#)). Appropriate number of lags is assumed to be 15-17 months.

The release of smolt is cyclical, as seen in [Figure 12](#) in [Appendix A.2](#), and it typically peaks in May and September for Norwegian producers ([Directorate of Fisheries, 2017a](#)). This implies that year-over-year (YoY) time windows should be used in the modelling. As figures for smolt release are proprietary for most salmon producing countries, only data for Norway is employed in the modelling.

#### *Feed Consumption*

Feed consumption, defined as the total feed consumption from the fish in a given area over a specified period, is an indicator of future supply. The larger consumption of feed, the larger is the expected future supply of salmon. Hence, changes in feed consumption are expected to be negatively correlated with changes in the NQSALMON.

The amount of feed consumed is a function of the size of the fish. The larger fish, the higher is the feed consumption. Accordingly, an increase in total feed consumption for a given area is mostly expected to be attributed to the consumption from larger fish. This implies that changes in feed consumption are expected to induce changes in supply after a relatively short period of time, such as 2-4 months.

Feed consumption is seasonal, as shown in [Figure 12](#) in [Appendix A.2](#), and it typically

---

<sup>7</sup>The industry participant is a significant player in the global industry for farmed salmon.

<sup>8</sup>*Lag structure*: The variables employed in this study are expected to have a lagged relationship with the salmon price. The lag structure for a particular variable refers to the range of lags that are deemed appropriate.

*Time window*: The time window used for any variable refers to whether changes between consecutive observations for a particular time series are used (that is, changes month-over-month, as all variables are obtained monthly), or whether we rather apply changes year-over-year. The year-over-year time window is likely to be most appropriate for many of the variables, as they often exhibit seasonal patterns.



peaks shortly before most volumes are harvested.<sup>9</sup> Hence, YoY time windows will be used in the modelling. Time series are obtained for the consumption of feed in Norway, as retrievable time series for other producing nations, such as Chile, United Kingdom, and the Faroe Islands, are too short (starting Jan. 2010).

### *Standing Biomass*

Standing biomass is defined as the total salmon biomass at a given area at the present time. It is typically measured in tonnes or number of individuals. Standing biomass is an important parameter for predicting future supply, as it indicates the volumes expected to be harvested within the foreseeable future (*Marine Harvest, 2016*). Hence, we expect changes in standing biomass to be negatively correlated with changes in NQSALMON.

Ideally, time series for standing biomass should be split according to *generation*,<sup>10</sup> yielding greater visibility of the size distribution of the fish in the cages. In turn, this enables more accurate estimates for when the fish, or equivalently, the standing biomass, is expected to be harvested. However, we have not been able to retrieve time series split by generation, implying great uncertainty attached to when changes in standing biomass will materialise in altered harvest volumes. Hence, we apply a relaxed lag structure for this variable, allowing lags of 3, 6, and 9 months. Here, a discrete range is chosen to achieve more consistency for the lag chosen by the GA across the models. Ideally, the majority of the submodels will utilise either 3, 6, or 9 months lag.

The time series for standing biomass exhibits strong seasonality, as shown in *Figure 12* in *Appendix A.2*. Therefore, YoY time windows will be used in the modelling. Yet again, time series are only obtained for the production in Norway, as time series for other producers are too short (starting Jan. 2010).

### *Harvest Volume*

Harvest volume is the most reliable measure of the supply of Atlantic salmon in the very short term, and the variables above are merely indicators of future harvest volumes. Changes in harvest volume are expected to be negatively correlated with changes in NQSALMON.

As Atlantic salmon primarily is marketed as a fresh product, the majority of the harvested volumes is sold immediately after harvest. Accordingly, there should not be a lagged relationship between harvest volume and the NQSALMON. However, we include time series for Norwegian harvest volumes at 1 month lag, in case the variable is able to provide any predictive power in the very short term. We also include average harvest weights for Norway, as the size of the fish has a rather large impact on the price of salmon on a per kg basis.<sup>11</sup> Here, we also allow 1 month lag. Moreover, we include time series for global harvest volumes, with Norwegian harvest volumes stripped out. We assign a longer lag structure to these volumes, as we assume that the volumes use time to find new distribution channels and flow to important markets for Norwegian producers, such as Europe. For instance, we assume that a sudden increase in Chilean production will not be consumed by Chile's usual markets (US), and that volumes to a larger extent will find new markets, such as Europe, and thereby affect the price of NQSALMON directly. We allow lags of 3 and 6 months.

Again, by referring to *Figure 13* in *Appendix A.2*, it is clear that harvest volumes are seasonal, suggesting that YoY time windows is appropriate.

---

<sup>9</sup>For Norway, most volumes are harvested during the autumn, with a peak in October (*Directorate of Fisheries, 2017a*).

<sup>10</sup>The *generation* refers to the time of smolt release. In Norway, smolts are mainly released into seawater twice a year. Hence, the generation of a particular amount of biomass is assigned the year of release, and whether it was released in the first or second half of the given year.

<sup>11</sup>Larger fish typically trade at a premium to smaller fish. See *Figure 2*.



### *Sea Lice*

Sea lice are naturally occurring seawater parasites, and can cause lesions, act as catalysts for infections, and slow down the growth rate of the fish (Marine Harvest, 2016). High concentration of sea lice may also lead to premature harvest, due to legal obligations constraining the maximum allowed occurrence of sea lice. Also, the prevalence of sea lice requires treatment of the fish, leading to elevated mortality rates due to rough treatment methods. In other words, a bloom in sea lice populations leads to reduced supply of salmon in terms of total volumes supplied over the medium term. However, in the short term, increased sea lice levels may lead to increased supply as a consequence of premature slaughtering. We choose to focus on both the short and medium term. Hence, we allow 12 month lag to meet the effect from lower total volumes supplied over the medium term, and 3 month lag to meet the effect from premature harvest and increased volumes over the short term. We expect changes in sea lice levels to be positively correlated with changes in the NQSALMON over the medium term, and negatively correlated over the short term.

The occurrence of sea lice exhibits seasonal patterns, as seen in Figure 13 in Appendix A.2. Accordingly, we apply YoY time windows in the modelling. Time series for sea lice levels are obtained for Norway.

### *Sea Temperatures*

The sea temperature affects the length of the production cycle, implying that large variations in sea temperatures may cause changes in supply due to changes in productivity. At times of higher sea temperatures, the production cycle is generally expected to be shorter, leading to increased harvest rates (Marine Harvest, 2016). Therefore, a negative relationship between sea temperatures and salmon prices is expected.

Changes in temperatures affect all the fish in the cages, irrespective of their sizes. Hence, an increase in the sea temperature is likely to trigger a consistent wave of higher harvest rates. We expect the effect from elevated temperatures to materialise rather promptly, and allow lags of 3 and 6 months.

Sea temperatures are highly seasonal, implying that YoY time windows should be applied in the modelling. Time series are obtained for Norway.

### 3.2 DEMAND-DRIVING FACTORS

The global demand for Atlantic salmon is not directly observable, but as all the fish is marketed and sold, it is reasonable to assume that consumption equals supply (Marine Harvest, 2016). Below we discuss four potential demand-driving factors.

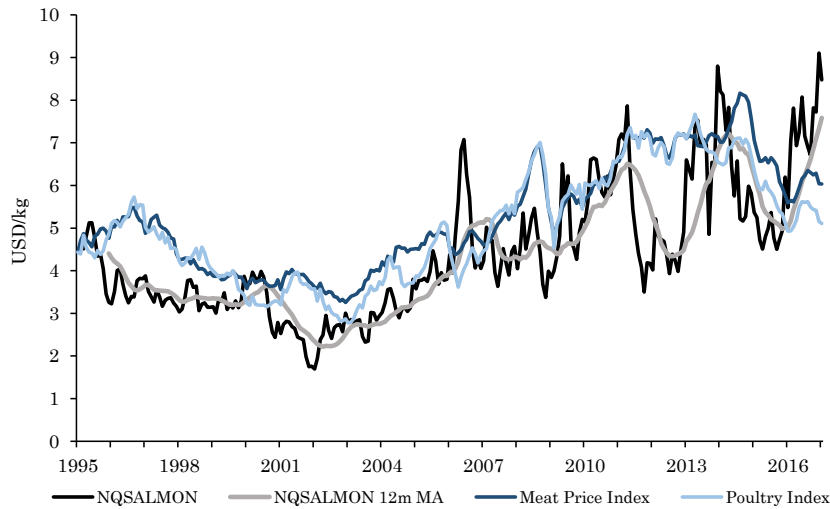


Figure 1: Monthly index prices for salmon, meat and poultry. Rebased to NQSALMON. 12-months moving-average is included for NQSALMON.

#### *Alternative Animal Proteins (Excluding Trout)*

Due to salmon being a source of protein, the demand for salmon is likely to be affected by the price of alternative animal protein sources. Among the substitutes, there are land-based animal proteins such as beef and chicken, and more direct substitutes such as cod and shrimp. Higher prices on these protein sources, *ceteris paribus*, should result in higher consumption of salmon. In turn, this is likely to give support to higher salmon prices. Hence, a positive relationship between changes in the price of alternative animal proteins and changes in the NQSALMON is expected. This relationship is confirmed by Figure 1, illustrating the NQSALMON versus price indices for poultry and meat. Here, the overall trend is that price changes are positively correlated among the different types of proteins, although with a few severe discrepancies.<sup>12</sup>

The appropriate lag structure is not easily determined, but our hypothesis is that there is no instant effect on the NQSALMON from changes in the prices of alternative animal proteins. Retailers need time to adjust their prices of alternative proteins, and consumers need time to adjust their consumption patterns. Also, retailers often negotiate long-term contracts with their suppliers of both salmon and other proteins, which imposes additional uncertainty to when price changes of alternative proteins will materialise in altered salmon prices (Ripegutu, 2017). Hence, we apply relaxed lag structures for the price of the different alternative proteins, allowing lags of 3, 6, 9, and 12 months.

Prices of alternative animal proteins do not exhibit seasonality, as shown by the time series for the meat price index, poultry index, beef, and shrimp in Figure 14 and Figure 15 in Appendix A.2. Hence, we use changes between consecutive price observations for these time series, i.e. month-over-month (MoM) time windows.

<sup>12</sup>Most of the discrepancies were due to extraordinary events. The discrepancy in 2006-07 was largely driven by a fear of bird flu combined with an E. Coli outburst in Norway, leading to strong demand for salmon. The deviation in 2011-13 was driven by a strong recovery of Chilean production, after a disease outbreak in Chile in 2008. The strong salmon prices seen in 2016 were triggered by an algae bloom in Chile, and to some extent high sea lice levels in Norway.

### *Trout*

Trout is a direct substitute to salmon, with global production volumes of 0.81 million tonnes as of 2014, or 35% of the global production of Atlantic salmon (FAO, 2017a). We include data for Norwegian trout production, represented by time series for standing biomass, harvest volumes, and the export price.

We choose to consider trout as an alternative protein, and base the lag structure on the preceding discussion on the prices of other alternative proteins. Also, we build on the discussions of the lag structures for standing biomass and harvest volumes of salmon. This leads to the following lag structures:

- i) *Standing biomass of trout.* We argued that standing biomass of salmon is related to the NQSALMON at lags of either 3, 6, or 9 months. As the production cycle for trout resembles that of salmon (FAO, 2017a), we assume the same lag structure between changes in standing biomass of trout and the price of trout. Moreover, we argued that changes in prices of other alternative proteins affect the NQSALMON after either 3, 6, 9, or 12 months. Drawing on the two lag structures, we get that changes in standing biomass for trout should be related to changes in NQSALMON at lags of 6-21 months. We choose to focus on the lower end of the implied lag structure, and allow lags of 6, 9, and 12 months. Also, we assume negative correlation between changes in standing biomass of trout and the price of trout, and positive correlation between changes in the price of trout and the NQSALMON. This implies a negative correlation between changes in standing biomass of trout and the NQSALMON.
- ii) *Harvest volume of trout.* We argued that changes in harvest volumes of salmon should affect the NQSALMON instantly, as most salmon is sold fresh. This is also the case for trout (FAO, 2017a), implying that changes in harvest volumes of trout should affect the price of trout instantly. Hence, we use the same line of argument as for changes in the price of other alternative proteins. Lags of either 3, 6, 9, or 12 months should be appropriate between changes in the harvest volume of trout and the NQSALMON.
- iii) *Price of trout.* Same line of argument as for other alternative proteins. Changes in the price of trout should be correlated with the NQSALMON at lags of either 3, 6, 9 or 12 months.

Time series for standing biomass and harvest volume of trout exhibit seasonality. Hence, we apply YoY time windows in the modelling. Moreover, we use MoM time windows for the price of trout, as it is not seasonal. See Figure 13 and Figure 15 in Appendix A.2 for plots of the time series.

### *Exchange Rates*

We expect correlation between the currency pair USD/EUR and the NQSALMON. The relationship is rather complex, and it is expressed as follows: First, we expect the price of alternative proteins to have an impact on the price of salmon, as elaborated on above. Secondly, the prices of proteins in Europe, adjusted for agricultural subsidies, are affected by the prices of commodities (Harri, Nalley, & Hudson, 2009; Saghaian, 2010), which again are partly driven by the U.S. Dollar (Akram, 2009; Kowalski, 2016). Noting that nearly one half of the Atlantic salmon volumes are consumed in Europe (Marine Harvest, 2016), and that the price is quoted in Euro, it is clear that the Euro is an important driver for the NQSALMON. Ultimately, when the U.S. Dollar strengthens against the Euro, we expect

negative price pressure on commodities from European consumers (Akram, 2009), which again results in lower prices of proteins (Harri et al., 2009; Saghaian, 2010). Subsequently, this will have a negative impact on the price of Atlantic salmon. Hence, changes in the USD/EUR are expected to be negatively correlated with changes in the NQSALMON.

We assume that changes in the U.S. Dollar will have a rather instant impact on prices of commodities, such as crude oil (Kowalski, 2016). According to Harri et al. (2009), the price of crude oil is correlated with prices of agricultural products at lags of 2 to 4 months. Thus, we assume that changes in prices of commodities will impact the prices of the most widely used sources of animal proteins after 3 months. Moreover, changes in prices of alternative proteins will impact the price of salmon after 3, 6, 9, or 12 months, as concluded above. Hence, implied lag structure between changes in USD/EUR and the NQSALMON is 6-15 months. We apply lags of 6, 9, 12, and 15 months.

The USD/EUR does not exhibit seasonality, implying that we use MoM time windows.

#### *Consumption of Salmon Split by Country*

Although the demand for Atlantic salmon is not directly observable, it is possible to obtain total import volumes split by country. Our hypothesis is that the NQSALMON is sensitive to changes in consumption from a particular country or region, and *vice versa*. When the NQSALMON increases, we expect that the consumption of salmon decreases, implying weaker salmon prices in the following period. We have obtained import figures, or equivalently, consumption figures, for the EU, the U.S., Japan, Russia, and other emerging markets (excluding Russia). In total, the amount of salmon consumed by these markets constituted 2.2 million tonnes as of 2016, or close to the entire global production of salmon. Ideally, we should have obtained import figures for Norwegian exports, but these are proprietary for observations prior to 2015. Of the five markets above, Europe is by far the most important market for Norwegian producers, with 64% of their volumes flowing to Europe as of 2015 (Marine Harvest, 2016). Hence, the NQSALMON is expected to be most sensitive to changes in European consumption.

We expect changes in consumption from the aforementioned markets to have an instant impact on the NQSALMON, and allow lags of 1 to 3 months. Time series for consumption are seasonal, following from harvest volumes being seasonal and the fact that most salmon is consumed immediately after harvest. Hence, we use YoY time windows in the modelling.

### 3.3 OTHER FACTORS

#### *NQSALMON*

A study carried out by Oglend and Sikveland (2008b) reveals short-term predictability in the price received by Norwegian salmon exporters. They find that the log return in any week is positively correlated with the log return of the preceding week and five weeks ago. Hence, we choose to include the NQSALMON itself at lags of one to two months. Moreover, the NQSALMON has exhibited signs of seasonality in the past, with prices being somewhat higher from the fourth to the first quarter of any given year. See Section 4.1 in the next chapter for a plot of the salmon price. However, we do not consider the pattern consistent enough to model on a YoY-basis. Thus, we employ MoM time windows in the modelling.

### 3.4 HYPOTHESIS OVERVIEW

Table 1 provides a brief overview of the variables previously discussed, and the expected impact on the NQSALMON log return from a change in the log return of one of the variables. The impact is illustrated by the expected coefficient sign in the regression model. Note that some of the table entries below represent several time series, such as the entry

*Price of alternative proteins.* For a complete list of all variables included in the database, and details regarding lag structures and time windows, see [Table 13](#) in [Appendix A.2](#).

Table 1: Variable and hypothesis overview.

CATEGORY	VARIABLE	UNIT	GEOGRAPHY	IMPACT <sup>1</sup>
Supply-driving				
	Smolt release	#Individuals	Norway	-
	Feed consumption	Tonnes	Norway	-
	Standing biomass	Tonnes, #Individuals	Norway	-
	Harvest volume	Tonnes	Norway, Global	-
	Sea lice occurrence	#Lice/fish	Norway	+/-
	Sea temperature	Degrees celsius	Norway	-
Demand-driving				
	Price of alternative proteins	Index, USD	-	+
	Supply of trout	Tonnes, #Individuals	Norway	-
	Exchange rates	USD/EUR	-	-
	Consumption split by country	Tonnes	EU, US, Russia, Japan, Emerging markets	+
Other				
	NQSALMON	USD/kg	Norway	+

1) The impact is illustrated by the expected coefficient sign in the regression model.

## DATA ANALYSIS

### 4.1 THE NASDAQ SALMON INDEX

In this study, we model the log return distribution of the spot price of salmon, which is represented by the NASDAQ Salmon Index (NQSALMON) denoted in USD. The index is the volume-weighted average of weekly reported sales prices of fresh Atlantic Superior Salmon, head-on-gutted (HOG). Similar to the Fish Pool Index<sup>TM</sup>, it is reported on a weekly basis – being published on Tuesdays (Fish Pool ASA, 2017; The Nasdaq Group Inc., 2016). The index contributors are Norwegian salmon exporters and producers with export license, who report prices for salmon in nine weight classes. There are eight weight classes for the 1 kg-subintervals 1-2 kg, 2-3 kg, and up to 8-9 kg, and one weight class for larger fish that belong to the 9 kg+ class. The particular rules and details concerning the requirements for index contributors, reporting of spot reference transactions, and adjustments for costs such as transportation are described in the rulebook provided by The Nasdaq Group Inc. (2016).

The NASDAQ OMX Group (NASDAQ) provides all relevant data for the calculation of the weighted price; that is, the total reported volume of Atlantic salmon that has been exported (thousand metric tonnes), the distribution of volume in the aforementioned weight classes (in percentage terms), and the average unit (kg) price in each weight class.

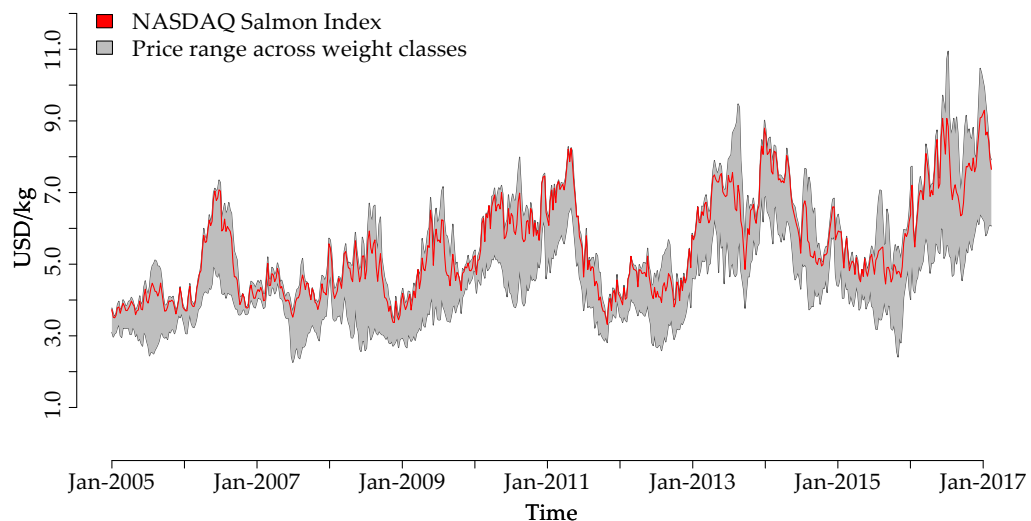


Figure 2: Weekly index prices of the NQSALMON, from Jan. 2005 to Feb. 2017, with the grey area representing the overall price range across the different weight classes (1 to 7+ kg).<sup>13</sup>

Regarding price history, NASDAQ provides weekly quotes in NOK dating back to 1995. As a consequence of the historical reporting in NOK, we have converted unit prices to USD with the average of the Monday through Thursday in the *spot reference week* weighted 60% plus the Friday *prior* to the spot reference week weighted 40%. This is similar to how fresh Atlantic Superior Salmon (HOG) transactions are converted in terms of foreign currencies when published by Nasdaq Clearing (The Nasdaq Group Inc., 2016).<sup>14</sup> Note that the reporting week is the week *after* the spot reference week in question. Thus, the price that is reported on Tuesdays before 14:00 represent *spot reference transactions* where prices and

<sup>13</sup>Generally, there is a price premium on the upper weight classes; larger fish tend to have higher unit prices relative to smaller fish.

<sup>14</sup>We employ daily foreign exchange rates, in this case the NOK/USD pair, which have been retrieved through Oanda.

corresponding volumes, belonging to the week of invoicing (i.e., the *spot reference week*), or the week prior to invoicing, are agreed upon (The Nasdaq Group Inc., 2016).

Figure 2 displays the weekly prices of the NQSALMON. When utilised in modelling of the log returns, we will effectively work on returns starting in Aug. 2007 – this due to various transformations of the data that primarily will be discussed in Chapter 5.

Now, as our sample data of the NQSALMON stretches back approximately one decade, one might expect structural breaks and shocks due to e.g. political and economic events. However, we note that the index did stay at low levels before and throughout the financial crisis of 2007 to 2008 – not taking any extraordinary hit compared to the time period as a whole. That being said, it plateaued at a high level from 2010 to 2012. During this period, Norwegian fish farming facilities experienced a severe situation of high sea lice levels,<sup>15</sup> with a corresponding long-term negative impact on the produced volume. Furthermore, the salmon spot price might have historical indications of being somewhat sensitive to political events. For instance, the index significantly declined during the autumn of 2014. This might to some extent be attributed to Russia’s imposed sanctions on the import of Norwegian salmon in Aug. 2014. Analysts expected that the prices for the fish could fall by 10% or 12% in Europe (Reklaitis, 2014), which, *ex-post*, seem to be relatively good predictions. Equivalently, the decline in Norway’s market share of salmon exports to China from 2010 to 2013 might have had a political cause (Milne, 2013), which again may have influenced the salmon price.

## 4.2 DATA SOURCES

In addition to the index data provided by NASDAQ, we use various other data sources to gather in total 24 exogenous variables. These will be employed in determining candidate predictors for the log returns of the salmon spot price. An overview of the variables and units, along with corresponding description, pre-determined lag range, respective data source, and retrieved time resolution (e.g. weekly or monthly data points) is given in Table 13. Furthermore, as the level of quality and accuracy of the data provided by various sources can never be assured, we provide a brief description of the data sources that have been used in Appendix A.9.

## 4.3 DATA PRE-PROCESSING

The NQSALMON and all the exogenous variables are assessed on a monthly time resolution. Even though the salmon spot price index is quoted every week, the majority of the exogenous variables are reported on a monthly basis, as is evident in Table 13, Appendix A.2. Furthermore, there are certain irregularities in the data sets, such as missing data points due to non-trading weekdays. This motivates the usage of a pre-processing and cleaning procedure prior to the modelling. A selection of robust approaches are discussed in Hellerstein (2008); we build on these concepts and employ the procedure as outlined below.

### PROCEDURE 1: General data pre-processing

- 1) To account for irregularities in the time series, replace N/As, zeros, and other erroneous values with the previous value in each fundamental time series.
- 2) For the time series with monthly observations, select the reported value. Generally, these correspond to the *end-of-the-month*-value. For the time

---

<sup>15</sup>Specifically, female sea lice levels in Norwegian waters peaked at historically high levels in the years from 2009 to 2011, confirmed by data provided by Sjømat Norge (2016).



series with weekly and daily observations, keep the value constituting the latest observation in the respective month, and discard the others.

- 3) Perform a stationarity transformation by calculating the associated log returns of the observations. This transform should apply a *time window* being either month-over-month (MoM) or year-over-year (YoY), in correspondence with Eq. 2 in the upcoming Section 4.5.
- 4) Calculate the descriptive statistics for the log return series, and create corresponding scatter plots (versus the NQSALMON log return), histograms (see Figures 12-16 in Appendix A.2), and normal Q-Q plots to get an idea of the series' distribution and potential outliers. Replace visual outliers that significantly skew the data's distribution with the previous observed log return.

#### 4.4 DESCRIPTIVE STATISTICS

Table 2 displays a selection of descriptive statistics for the monthly log returns of the NQSALMON and the other 24 exogenous variables. Furthermore, a normality test, two stationarity tests, and an autocorrelation test are included.<sup>16</sup> Arithmetic returns have *not* been used in the price distribution prediction. Nevertheless, for reference, they are displayed in Table 14 in Appendix A.2.

##### *Descriptive Statistics of the NQSALMON*

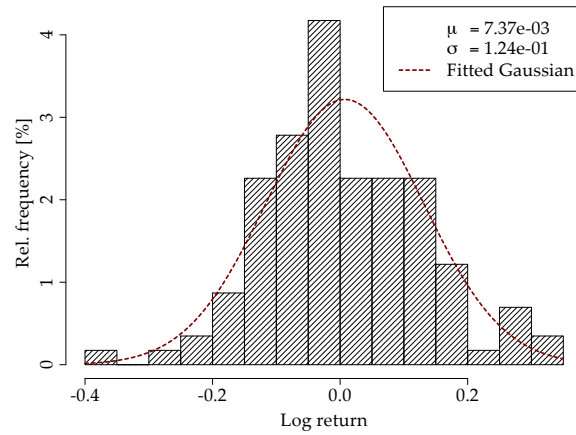
With reference to Table 2, we note that the log returns of the NQSALMON yield an insignificant Jarque-Bera (JB) statistic, indicating the presence of normality. In particular, the skewness is quite low; thus, the empirical distribution is more or less symmetric. Furthermore, the excess kurtosis is almost negligible. In other words, skewness and kurtosis seem to fit well with a normal distribution. The same conclusions can be drawn from the descriptive plots in Figure 3, in particular by the histogram in Figure 3a. That being said, by visual inspection, the log returns have slightly more mass in the tails rather than at the peak, principally with respect to the excess observations at the upper quantiles. The extra mass in the tails is also present in the shape of the normal Q-Q plot in Figure 3d – upper and lower empirical quantiles deviate somewhat from the normal distribution. Furthermore, by the log return time series plot, some months exhibit large price jumps. In conclusion, even though the JB statistic is insignificant, there are indications of some non-normal behaviour. Consequently, more data points could certainly be useful in determining distributional properties of the NQSALMON log returns.

Moreover, the (augmented) Dickey-Fuller (ADF) stationarity statistics are rejected at high significance levels, which indicates stationary log returns. The Ljung-Box Q test yield a sufficiently low test statistic, corresponding to a somewhat higher  $p$ -value than the 10%-level, suggesting no presence of significant autocorrelation in the NQSALMON at lag  $\rho = 6$  months. However, this is slightly at odds with the plot of the autocorrelation function in Figure 3e, which suggests a modest sinusoid memory. In particular, there seems to be an annual pattern with negative autocorrelation at every  $6 + 12k$  lags,  $k \in \mathbb{Z}_0^+$ , as well as an annual pattern with positive autocorrelation at every  $12k$  lags.

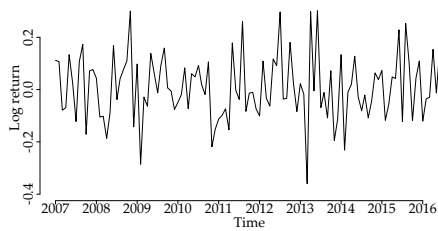
Figure 3b and Figure 3c suggest time-varying volatility, i.e. heteroscedasticity, in the monthly log returns, as there seems to be some clustering of "spikes" in the squared log returns. This conforms to previous studies; for instance do Oglend and Sikveland (2008a) find evidence for persistence and reversion in the salmon price volatility.

<sup>16</sup>For a description of the tests, see Appendix A.6.

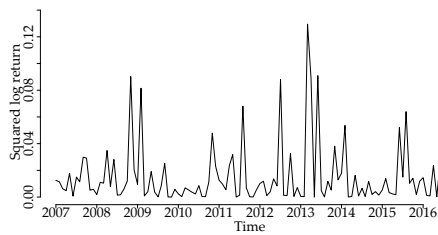




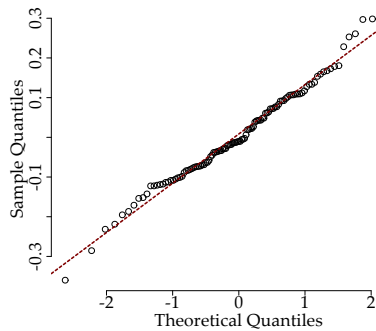
(a) Histogram of monthly log returns



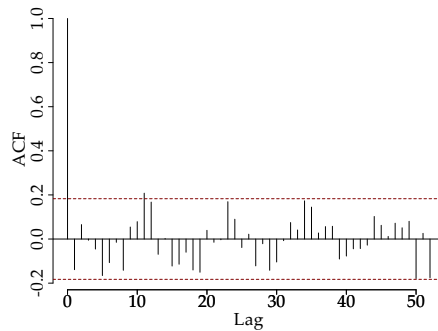
(b) Monthly log returns



(c) Squared monthly log returns



(d) Normal Q-Q plot of monthly log returns



(e) Autocorrelation of monthly log returns

Figure 3: Descriptive figures of the monthly log returns of the NQSALMON.<sup>17</sup>

#### Descriptive Statistics for the Exogenous Variables

By Table 2, we observe that the monthly log returns for  $\sim 40\%$  of the exogenous variables exhibit non-normal behaviour, characterised by respective JB statistics significant at the 1% level. A prominent category of such variables is prices of alternative proteins; log returns of meat, poultry, beef, and shrimp reject normality at the 1% level according to this test. Also, salmon biomass-related variables exhibit positive skew and are non-normal at the 5% significance level. Variables with the high positive excess kurtosis, i.e., leptokurtosis, are *Average harvest weight, kg (Norway)* and *Shrimp price, US cents/pound*, which might help explain the excess probability mass in the tails of the monthly log return of the NQSALMON. The category of variables having the most significant skewness is that representing consumption figures of Atlantic salmon across various geographies (e.g. *Consumption of Atl. salmon*,

<sup>17</sup>Regarding Figure 3a: The *fitted Gaussian* curve represents the corresponding normal distribution of the log return of the NQSALMON, parameterised by the sample mean and sample variance as given in the figure.

1000 tonnes (Russia)), in which the skewness generally is negative. That is, the presence of excess mass in the left tails of these figures might help explain some of the tail behaviour of the NQSALMON log return as well. However, most variables exhibit only slight skewness – the kurtosis seems to be the main driver for rejection of normality by the JB statistic.

Some variables have significant outliers in their log returns, such as *Smolt release, #individuals (Norway)* and *Sea lice treatments, % of fish being treated (Norway)*. This is observed from their minimum and maximum values. However, on a monthly basis, this might be explained by the binary nature of these variables. For instance, smolts are usually released at particular times at the year, particularly in May and September for Norwegian producers, as discussed previously in [Section 3.1](#). High fluctuations of release levels between months might again yield correspondingly extreme log returns.

Next, we consider various time series test results of the exogenous variables. Firstly, from the statistics of the Dickey-Fuller (DF) and ADF tests, it is evident that all the exogenous variables' monthly log returns exhibit stationarity significant at the 1% level. Furthermore, the Ljung-Box Q test at  $\rho = 6$  lags reveals significant autocorrelation at the 1% level for approximately half of the exogenous variables. This is particularly present in seasonal variables, such as the biomass levels, sea temperature, and smolt release.

Table 2: Descriptive statistics of the log returns of the data series, time period: Jul. 2007 – Jan. 2017.<sup>18</sup>

GENERAL INFORMATION		DESCRIPTIVE STATISTICS							TESTS			
Series name	N #	Mean	Std. dev.	Min	Max	Skew	Ex. kurt.	NQSALMON corr.	JB	DF	ADF	Ljung-Box $\rho = 6$
$\Delta \ln[\text{NQSALMON, USD/kg}]$	115	$7.37 \times 10^{-3}$	$1.24 \times 10^{-1}$	$-3.60 \times 10^{-1}$	$3.02 \times 10^{-1}$	$1.34 \times 10^{-1}$	$2.59 \times 10^{-1}$	1.00	0.67	-12.21	-7.63	7.81
$\Delta \ln[\text{Standing biomass, \#Individuals (Norway)}]$	115	$2.87 \times 10^{-3}$	$5.58 \times 10^{-2}$	$-7.20 \times 10^{-2}$	$1.84 \times 10^{-1}$	1.02	$7.99 \times 10^{-1}$	-0.20	22.81	-6.89	-9.41	44.94
$\Delta \ln[\text{Standing biomass, Tonnes (Norway)}]$	115	$5.56 \times 10^{-3}$	$4.77 \times 10^{-2}$	$-7.52 \times 10^{-2}$	$1.05 \times 10^{-1}$	$4.34 \times 10^{-1}$	$-9.50 \times 10^{-1}$	-0.22	7.94	-4.00	-6.29	218.25
$\Delta \ln[\text{Feed consumption, Tonnes (Norway)}]$	115	$3.63 \times 10^{-3}$	$1.94 \times 10^{-1}$	$-3.43 \times 10^{-1}$	$3.82 \times 10^{-1}$	$2.79 \times 10^{-1}$	-1.09	-0.22	7.22	-4.16	-4.85	269.50
$\Delta \ln[\text{Smolt release, \#Individuals (Norway)}]$	115	$-1.82 \times 10^{-2}$	1.40	-5.77	6.45	$2.26 \times 10^{-1}$	4.87	-0.08	114.77	-9.57	-7.90	34.12
$\Delta \ln[\text{Harvest volume, Tonnes (Norway)}]$	115	$2.63 \times 10^{-3}$	$1.30 \times 10^{-1}$	$-4.17 \times 10^{-1}$	$2.89 \times 10^{-1}$	$-1.02 \times 10^{-1}$	$4.95 \times 10^{-2}$	-0.23	0.21	-11.77	-9.20	21.01
$\Delta \ln[\text{Standing biomass of trout, \#Individuals (Norway)}]$	115	$-4.68 \times 10^{-3}$	$5.14 \times 10^{-2}$	$-1.23 \times 10^{-1}$	$1.48 \times 10^{-1}$	$3.75 \times 10^{-1}$	$-1.07 \times 10^{-1}$	-0.12	2.75	-7.71	-7.40	35.46
$\Delta \ln[\text{Standing biomass of trout, Tonnes (Norway)}]$	115	$-4.68 \times 10^{-3}$	$5.93 \times 10^{-2}$	$-1.50 \times 10^{-1}$	$1.22 \times 10^{-1}$	$2.05 \times 10^{-1}$	$-3.73 \times 10^{-1}$	-0.19	1.47	-6.36	-5.48	68.54
$\Delta \ln[\text{Harvest volume of trout, Tonnes (Norway)}]$	115	$-4.38 \times 10^{-3}$	$2.57 \times 10^{-1}$	$-8.20 \times 10^{-1}$	$5.58 \times 10^{-1}$	$-3.61 \times 10^{-1}$	$1.49 \times 10^{-2}$	-0.16	2.50	-12.62	-9.63	16.35
$\Delta \ln[\text{Sea lice occurrence, \#Lice/fish (Norway)}]$	115	$-8.62 \times 10^{-4}$	$3.98 \times 10^{-1}$	-1.05	$9.48 \times 10^{-1}$	$-2.73 \times 10^{-2}$	$-4.51 \times 10^{-4}$	-0.24	0.01	-7.22	-6.05	56.22
$\Delta \ln[\text{Sea lice treatments, \% of fish being treated (Norway)}]$	115	$-6.10 \times 10^{-3}$	$4.91 \times 10^{-1}$	-2.39	2.18	$-3.28 \times 10^{-1}$	6.92	-0.14	231.68	-15.89	-9.62	23.59
$\Delta \ln[\text{Sea temperature, Degrees celsius (Norway)}]$	115	$-4.83 \times 10^{-3}$	$1.99 \times 10^{-1}$	$-3.26 \times 10^{-1}$	$4.83 \times 10^{-1}$	$4.02 \times 10^{-1}$	-1.09	-0.17	8.80	-3.40	-7.96	276.80
$\Delta \ln[\text{Meat price index, Index}]$	115	$1.47 \times 10^{-3}$	$2.89 \times 10^{-2}$	$-9.08 \times 10^{-2}$	$5.97 \times 10^{-2}$	$-8.12 \times 10^{-1}$	$7.60 \times 10^{-1}$	0.08	15.42	-6.00	-4.21	75.22
$\Delta \ln[\text{Poultry index, Index}]$	115	$-4.01 \times 10^{-4}$	$3.96 \times 10^{-2}$	$-1.49 \times 10^{-1}$	$1.14 \times 10^{-1}$	$-4.39 \times 10^{-1}$	2.52	0.02	34.10	-9.83	-5.40	13.82
$\Delta \ln[\text{Beef price, US cents/pound}]$	115	$3.48 \times 10^{-3}$	$4.82 \times 10^{-2}$	$-1.80 \times 10^{-1}$	$1.80 \times 10^{-1}$	$-4.48 \times 10^{-2}$	2.70	0.15	34.88	-7.93	-7.42	18.63
$\Delta \ln[\text{Shrimp price, US cents/pound}]$	115	$1.81 \times 10^{-3}$	$5.02 \times 10^{-2}$	$-2.43 \times 10^{-1}$	$1.84 \times 10^{-1}$	$-3.31 \times 10^{-2}$	8.16	-0.04	319.30	-7.59	-6.17	14.63
$\Delta \ln[\text{Currency pair, USD/EUR}]$	115	$2.03 \times 10^{-3}$	$2.47 \times 10^{-2}$	$-6.30 \times 10^{-2}$	$7.82 \times 10^{-2}$	$2.71 \times 10^{-1}$	$5.77 \times 10^{-1}$	-0.24	3.00	-8.11	-6.66	10.03
$\Delta \ln[\text{Trout price, NOK/kg (Norway)}]$	115	$8.91 \times 10^{-3}$	$8.89 \times 10^{-2}$	$-2.76 \times 10^{-1}$	$2.39 \times 10^{-1}$	$-1.13 \times 10^{-1}$	$4.37 \times 10^{-1}$	0.32	1.16	-9.61	-8.57	11.43
$\Delta \ln[\text{Average harvest weight, kg (Norway)}]$	115	$-1.18 \times 10^{-3}$	$5.21 \times 10^{-2}$	$-2.41 \times 10^{-1}$	$2.28 \times 10^{-1}$	$-1.30 \times 10^{-1}$	5.35	-0.03	137.40	-11.87	-8.70	10.98
$\Delta \ln[\text{Consumption of Atl. Salmon, 1000 Tonnes (EU)}]$	115	$3.51 \times 10^{-3}$	$1.14 \times 10^{-1}$	$-3.58 \times 10^{-1}$	$2.62 \times 10^{-1}$	$-4.72 \times 10^{-1}$	$7.12 \times 10^{-1}$	-0.13	6.70	-12.02	-9.79	34.42
$\Delta \ln[\text{Consumption of Atl. Salmon, 1000 Tonnes (US)}]$	115	$1.59 \times 10^{-3}$	$9.42 \times 10^{-2}$	$-2.21 \times 10^{-1}$	$2.79 \times 10^{-1}$	$3.53 \times 10^{-1}$	$1.02 \times 10^{-1}$	-0.16	2.45	-19.87	-11.61	53.49
$\Delta \ln[\text{Consumption of Atl. Salmon, 1000 Tonnes (Russia)}]$	115	$4.19 \times 10^{-3}$	$2.25 \times 10^{-1}$	$-8.77 \times 10^{-1}$	$4.50 \times 10^{-1}$	$-9.38 \times 10^{-1}$	1.46	-0.01	27.10	-11.43	-9.46	8.04
$\Delta \ln[\text{Consumption of Atl. Salmon, 1000 Tonnes (Japan)}]$	115	$7.11 \times 10^{-3}$	$1.89 \times 10^{-1}$	$-5.66 \times 10^{-1}$	$3.92 \times 10^{-1}$	$-7.06 \times 10^{-1}$	$8.34 \times 10^{-1}$	0.09	12.89	-14.81	-10.24	14.85
$\Delta \ln[\text{Consumption of Atl. Salmon, 1000 Tonnes (Emerging markets)}]$	115	$8.94 \times 10^{-3}$	$9.01 \times 10^{-2}$	$-2.67 \times 10^{-1}$	$2.04 \times 10^{-1}$	$-3.48 \times 10^{-1}$	$7.24 \times 10^{-2}$	-0.07	2.34	-13.20	-9.63	13.04
$\Delta \ln[\text{Harvest volume, 1000 Tonnes (Global, excl. Norway)}]$	115	$4.89 \times 10^{-3}$	$6.81 \times 10^{-2}$	$-1.60 \times 10^{-1}$	$2.08 \times 10^{-1}$	$-1.56 \times 10^{-1}$	$6.37 \times 10^{-1}$	-0.21	2.41	-11.74	-7.61	3.25

Critical values:

JB:  $\chi^2_{2,A=10\%} > 4.61$ ;  $\chi^2_{2,A=5\%} > 5.99$ ;  $\chi^2_{2,A=1\%} > 9.21$   
 DF, ADF:  $\tau_{\alpha=10\%} < -1.62$ ;  $\tau_{\alpha=5\%} < -1.95$ ;  $\tau_{\alpha=1\%} < -2.58$   
 Ljung-Box:  $\chi^2_{6,A=10\%} > 10.64$ ;  $\chi^2_{6,A=5\%} > 12.59$ ;  $\chi^2_{6,A=1\%} > 16.81$

<sup>18</sup>Note the choice of the time period, namely Jul. 2007 to Jan. 2017, corresponding to  $N = 115$  observations present in the table. We, however, note that the *modelling* data set is of 103 observations of 1- to 12-step ahead NQSALMON log returns. For the case of shorter horizons, say horizon  $h = 1$ , the data set of log returns will stretch from Jul. 2007 to Jan. 2016. Similarly, for longer horizons, e.g. horizon  $h = 12$ , the data set comprises observations from Jul. 2008 to Jan. 2017, as it is pushed forward with  $h$  months.

#### 4.5 CHOOSING THE APPROPRIATE TIME WINDOW

In the literature, we find that forecasting a stationary time series as a function of stationary explanatory variables is the most common approach (Nau, 2017). Predicting a stationary series involves assessing whether or not its statistical properties will be the same in the future as they have been in the past. Furthermore, sample statistics such as mean, standard deviation, and higher-order moments are only useful as descriptors for future behaviour if the series is stationary. The same rationale motivates to stationarise the explanatory variables; joint distribution metrics such as correlation will only be meaningful when the underlying series are stationary.

In our case, the stationarity transform of some explanatory variable series  $i \in \mathcal{N}$ , where  $\mathcal{N}$  is the set of explanatory variables, is the natural logarithm of the relative change between time increments. This transform will be called the *log return*. We will refer to a price or observation at some time  $t \in T$  of some explanatory variable  $i$  by  $z_{t,i}$ , whereas the log return or change will be denoted  $x_{t,i}$ .<sup>19</sup> With the increment in time, often called *time resolution*, being one month in this study, the log returns of some series  $i$  are given by

$$x_i \equiv \{x_{t,i}\}_{t=1}^T \quad : \quad x_{t,i} = \ln \left( \frac{z_{t,i}}{z_{t-1,i}} \right) = \ln(z_{t,i}) - \ln(z_{t-1,i}). \quad (1)$$

When predicting the log returns of some commodity price, conditional on the observation of the log returns of some explanatory variables, the notion of *change* is of importance. Some explanatory variables might have cyclical behaviour; in our case, sea temperatures and lice levels are appropriate examples – as visualised in [Appendix A.2](#). Furthermore, all cyclical variables that are used in this study have annual cycles – i.e., they can be considered *seasonal*. In such cases, as discussed in the previous chapter, one might argue for the consideration of YoY (logarithmic) returns. Usually, such variables are also transformed to YoY percentage changes in e.g. industry reports and exhibits (Marine Harvest, 2016). Variables without cyclical patterns, such as foreign exchange rates, alternative protein prices, and feed prices, might be more interesting on a shorter time window.

There are a few studies that have considered transforming time series through various time windows. Although not used here, stock indices have tended to have positive auto-correlations, significantly different from zero, on daily, weekly, and monthly time windows (Campbell, Lo, & MacKinlay, 1996, p. 68), and one might possibly discover such patterns with respect to e.g. currencies on both monthly and multi-monthly returns. Although intraday returns are not in the scope of this study, Politi, Millot, and Chakraborti (2012) use time windows to investigate near-extreme densities of higher-frequency log returns.

We will proceed on to consider the predictive accuracy of various time windows. In line with Shen, Jiang, and Zhang (2012), varying the time window, denoted  $\delta$ , we can generate new, auxiliary time series  $\Delta_\delta z_i \forall i \in \mathcal{N}$  and add them to the set of *candidate predictors* in the following way:

$$\Delta_\delta z_{t,i} = \ln \left( \frac{z_{t,i}}{z_{t-\delta,i}} \right) = \ln(z_{t,i}) - \ln(z_{t-\delta,i}). \quad (2)$$

The set of candidate predictors will be those that are available in the variable subset selection search procedure for the quantile regressions. Now, in the case of variables exhibiting seasonality, we would in a YoY transform consider the case  $\delta = 12$ ; likewise, the MoM transform, equivalent to [Eq. 1](#), consider the case  $\delta = 1$ . Intuitively, a larger time window  $\delta$  corresponds to a change representing a longer trend of the underlying, explanatory variable.

The performance of some predictor of a dependent variable  $\mathbf{y}$ , the NQSALMON log

<sup>19</sup>The variable  $\mathbf{y} \equiv \{y_t\}_{t=1}^T$  is reserved for the NQSALMON log return at time  $t$ .

return in this case, will, not considering the effects of other candidate predictors, depend significantly on their cross-correlation. In the context of multi-step ahead prediction, we label  $y_t^{(h)}$  as the  $h$ -step ahead log return of the dependent variable. Correspondingly, in the same way as Eq. 2, a particular  $h$ -step ahead log return can be written as  $y_t^{(h)} = \ln(p_{t+h,i}) - \ln(p_{t,i})$ , where  $p_{t,i}$  is the *price* of the dependent variable, i.e. the value of the NQSALMON, at time  $t$ . Shen et al. (2012) investigate temporal autocorrelation and cross-correlation among global stock markets and different financial products, by exploring combinations of lags  $l$  and time windows  $\delta$ , to select predictors with the highest temporal correlation. The results yield a high one-day ahead prediction accuracy for indices such as NASDAQ, S&P500, and DJIA. They also find that temporal correlation between various markets increase with the time window  $\delta$ . In addition, the authors suggest that an increased time window  $\delta$  might remove noise and result in a clearer correlation between various stock market indices. Thus, such an approach might prove successful in building prediction models for other indices – in our case the NQSALMON.

#### *A Preliminary Candidate Predictor Analysis*

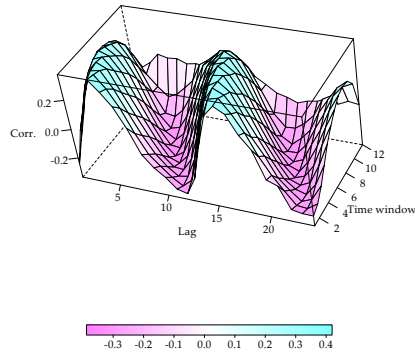
As discussed in Chapter 3, there might be some structure in the lagged values and the time windows of the explanatory variables. Therefore, in a preliminary candidate predictor analysis, we initially combine lags with time windows to create auxiliary time series. Next, the cross-correlation relationships between the auxiliary time series and the  $h$ -step ahead log return of the salmon spot price are assessed to investigate candidate predictors. As we will consider forecasts of up to one year at monthly time increments, we have that  $h \in \{1, \dots, 12\} \equiv \mathcal{H}$ . We refer to  $\mathcal{H}$  as the index set of steps ahead, and interchangeably refer to  $h$  as the *horizon* or the steps ahead for the NQSALMON log return. Now, based on the resulting cross-correlation metrics, we will discuss whether or not the presumed structure as discussed in the variable analysis section has empirical evidence, and which predictors to include in the modelling of the NQSALMON.

Formally, we consider a space of candidate predictors generated from the explanatory variables  $i \in \mathcal{N}$ . The candidate predictors are lagged with respect to the  $h$ -month ahead log return of the NQSALMON at lags  $l \in \mathcal{L}$  with time windows  $\delta \in \{1, \dots, 12\}$ . That is, we consider up to 12-month trends. The lags  $l \in \mathcal{L}$  for each explanatory variable  $i \in \mathcal{N}$  are determined according to the discussion in Chapter 3. However, in the preparatory analysis, we set  $\mathcal{L} = \{1, \dots, 24\} \forall i \in \mathcal{N}$ , that is, we consider explanatory variables lagged up to two years. Now, each candidate predictor generated by combining lags and time windows with respect to some explanatory variable  $z_i$  can be denoted  $\Delta_{\delta,l}z_i$ , where each observation at time  $t \in T$  is given by:

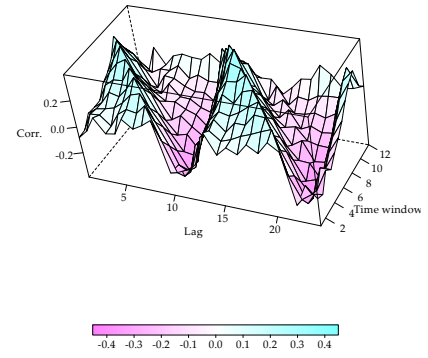
$$\Delta_{\delta,l}z_{t,i} = \ln \left( \frac{z_{t-l,i}}{z_{t-l-\delta,i}} \right) = \ln(z_{t-l,i}) - \ln(z_{t-l-\delta,i}), \quad (3)$$

$$\forall i \in \mathcal{N}, l \in \mathcal{L} = \{1, \dots, 24\}, \delta \in \{1, \dots, 12\}.$$

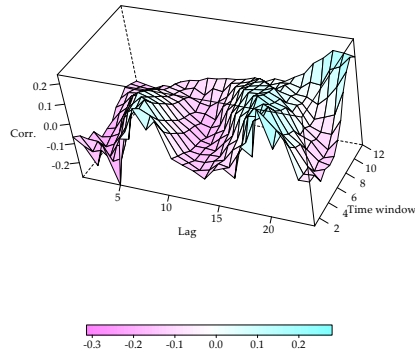
Based on this preparatory candidate predictor analysis, we are at first able to check whether or not the particular time windows discussed in Chapter 3 are sound – e.g., whether to consider MoM returns ( $\delta = 1$ ), YoY returns ( $\delta = 12$ ), or some other time window ( $1 \leq \delta \leq 12$ ). That is, we use the analysis to validate how the industry considers these parameters, as summed up in Table 13 in Appendix A.2. Secondly, we are able to visually identify which combinations of lags  $l$  and time windows  $\delta$  that empirically have strong correlation (in absolute terms) with the  $h$ -month ahead price and therefore, conceivably, strong predictive power.



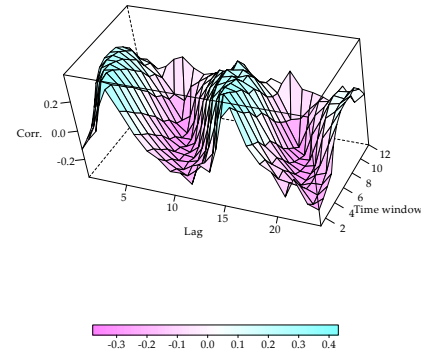
(a) Cross-correlation between NQSALMON 1-step ahead log returns and the log returns of *Standing biomass, Tonnes (Norway)*



(b) Cross-correlation between NQSALMON 1-step ahead log returns and the log returns of *Harvest volume, Tonnes (Norway)*



(c) Cross-correlation between NQSALMON 1-step ahead log returns and the log returns of *Meat price index, Index*



(d) Cross-correlation between NQSALMON 1-step ahead log returns and the log returns of *Sea lice occurrence, #Lice/fish (Norway)*

Figure 4: Examples of cross-correlation surfaces generated by the 1-step ahead NQSALMON log returns and four arbitrarily chosen time series' log returns across combinations of lags  $l \in \{1, \dots, 24\}$  and time windows  $\delta \in \{1, \dots, 12\}$ .

### Revision of Lag Structure and Time Window

Although the following analysis may result in conclusions regarding lags and time windows in conflict with the discussion in [Chapter 3](#), we choose to not implement changes that are considerably in conflict with the relationships suggested by [Marine Harvest \(2016\)](#). Hence, this approach can not be criticised as data mining. Moreover, changes that can not be supported by fundamental relationships will not be implemented.

Each plot in [Figure 4](#) displays a surface in which the lattice points represent respective cross-correlations between the 1-month ahead NQSALMON log return and the given explanatory variable  $i$  lagged at some lag  $l \in \{1, \dots, 24\}$  relative to the 1-month ahead index return at a certain time window  $\delta \in \{1, \dots, 12\}$ . Formally, each point is the correlation between  $y^{(1)}$  versus the candidate predictor  $\Delta_{\delta,l}z_i$ . Correlation surfaces for all the explanatory variables with respect to the 1-step ahead NQSALMON log return are visualised in [Figures 17-19](#) in [Appendix A.2](#). Note that most of the cross-correlation plots in [Figures 17-19](#) propose both positive and negative correlation between the NQSALMON and the variables being examined. This is due to seasonality effects in the explanatory variables, and we account for this by holding on to the directional impacts concluded on in [Chapter 3](#). That is, for e.g. *Standing biomass, tonnes (Norway)*, we only consider areas of the surface where the correlation with the NQSALMON is negative. Furthermore, with respect to seasonal variables, we note that correlation effectively "levels out" for larger time windows – particularly in the YoY case with  $\delta = 12$ . This confirms the previous discussion regard-

ing the usage of large time windows, as they will remove cyclicity in the corresponding candidate predictor  $\Delta_{\delta,l}z_i$ .

In [Table 3](#), we present the lag structures and time windows suggested by the cross-correlation plots, and benchmark the results against the lag structures and time windows concluded upon in [Chapter 3](#) and [Table 13](#) in [Appendix A.2](#). Note that the relationships in [Table 3](#) are obtained from visual inspection of the cross-correlation plots, and consequently they are not entirely accurate and represent qualitative judgement. However, the relationships should provide enough evidence to consider whether or not empirical data are in line with industry assumptions.

By examining [Table 3](#), it is clear that the industry assumptions, regarding time windows, deviate substantially from the results of the candidate predictor analysis. Firstly, the preparatory analysis is very reluctant to propose YoY time windows as a means of increasing cross-correlation with the salmon spot price. For all variables, the correlation surfaces essentially propose the usage of MoM time windows, as the corresponding correlation will be the strongest (in absolute terms). We account for these results by introducing the following changes to the application of time windows:

- i) *Variables that previously were allowed YoY time windows.* We allow the use of both YoY and MoM time windows in the modelling.
- ii) *Variables that previously were allowed MoM time windows.* No alterations in the use of time windows.

The lag structures proposed by this candidate predictor analysis also deviate substantially from the industry assumptions. Of the 25 potential explanatory variables in the database, the analysis indicates that the lag structure suggested by two approaches are more or less aligned for 16 of the variables. Hence, for nine of the variables the industry assumptions do not comply with this preliminary analysis. In particular, it is worth highlighting the lag structure proposed for *Sea lice occurrence, #lice/fish (Norway)* and *Sea lice treatments, % of fish being treated (Norway)*. We argued in [Chapter 3](#) that changes in sea lice levels and sea lice treatments should be correlated with changes in NQSALMON with varying signs, depending on the lag structure used. Over the short term, we expected changes in sea lice levels and sea lice treatments to be negatively correlated with the NQSALMON, and positively correlated over the long term. Interestingly, the results from the candidate predictor analysis suggest the opposite. However, we consider the reasoning from [Chapter 3](#) as sufficiently robust, and choose to not alter the lag structure for these two variables. Moreover, we decide to base the lag structure for all variables only on industry assumptions, as we are not able to support potential changes with fundamental relationships. Hence, we do not make any changes to the lag structure.

Overall, we find the results from the candidate predictor analysis surprising, and although we choose to put limited emphasis on the analysis, it provides useful insight for industry participants, as well as for further modelling of salmon prices.



Table 3: Lag structures and time windows: Assumptions versus empirical evidence from cross-correlation plots, 1-month-ahead.

TIME SERIES, UNIT	INDUSTRY ASSUMPTIONS		CANDIDATE PREDICTOR ANALYSIS	
	Lag(s) <sup>1</sup>	Time window	Lag(s) <sup>1</sup>	Time window
NQSALMON, USD/kg	1-2	MoM	8-11	MoM
Standing biomass, #Individuals (Norway)	3,6,9	YoY	10-12	MoM
Standing biomass, Tonnes (Norway)	3,6,9	YoY	11-13	MoM
Feed consumption, Tonnes (Norway)	2-4	YoY	11-13	MoM
Smolt release, #Individuals (Norway)	15-17	YoY	15-17	MoM
Harvest volume, Tonnes (Norway)	1	YoY	10-12	MoM
Standing biomass of trout, #Individuals (Norway)	6,9,12	YoY	11-13	MoM
Standing biomass of trout, Tonnes (Norway)	6,9,12	YoY	12-14	MoM
Harvest volume of trout, Tonnes (Norway)	3,6,9,12	YoY	10-12	MoM
Sea lice occurrence, #Lice/fish (Norway)	3,12	YoY	10-12, 4-6 <sup>2</sup>	MoM
Sea lice treatments, % of fish being treated (Norway)	3,12	YoY	10-12, 4-6 <sup>2</sup>	MoM
Sea temperature, Degrees celsius (Norway)	3,6	YoY	10-12	MoM
Meat price index, Index	3,6,9,12	MoM	7-10	MoM
Poultry index, Index	3,6,9,12	MoM	6-8	MoM
Beef price, US cents/pound	3,6,9,12	MoM	6-8	MoM
Shrimp price, US cents/pound	3,6,9,12	MoM	20-24	MoM
Currency pair, USD/EUR	6,9,12,15	MoM	9-11	MoM
Trout price, NOK/kg (Norway)	3,6,9,12	MoM	9-11	MoM
Average harvest weight, kg (Norway)	1	YoY	12-14	MoM
Consumption of Atl. Salmon, 1000 Tonnes (EU)	1-3	YoY	1-3	MoM
Consumption of Atl. Salmon, 1000 Tonnes (US)	1-3	YoY	1-3	MoM
Consumption of Atl. Salmon, 1000 Tonnes (Russia)	1-3	YoY	1-3	MoM
Consumption of Atl. Salmon, 1000 Tonnes (Japan)	1-3	YoY	1-3	MoM
Consumption of Atl. Salmon, 1000 Tonnes (Emerging markets)	1-3	YoY	1-3	MoM
Harvest volume, 1000 Tonnes (Global excl. Norway)	3,6	YoY	10-12	MoM

1) All lags are denoted in months.

2) Recalling the discussion from [Chapter 3](#), we assumed two regimes with different direction of price impacts from changes in sea lice levels.



## METHODOLOGY

---

### 5.1 METHODOLOGY OVERVIEW

#### *Formalisation of the Prediction Problem*

We consider the empirical modelling of the  $h$ -step ahead, conditional distribution function of the monthly log returns of the NQSALMON. That is, for each timestep  $t$ , construct a model to predict the conditional distribution of returns at  $t + h$ , where

$$h \in \{1, 2, \dots, 12\} \equiv \mathcal{H}.$$

This corresponds to a modelling *horizon* of up to 12 months. Formally, the problem is to make a prediction model for the  $h$ -step ahead *distribution* of the log return

$$y_t^{(h)} | \mathcal{F}_t, \quad \text{where } y_t^{(h)} = \ln \left( \frac{p_{t+h}}{p_t} \right), \quad \forall h \in \mathcal{H}, \quad (4)$$

where  $p_t$  and  $p_{t+h}$  are the prices (in USD terms) of the NQSALMON at times  $t$  and  $t + h$ , and  $\mathcal{F}_t$  is a suitable *filtration* representing the information available at time  $t$  (Mikosch, 1998, p. 77). In our case,  $\mathcal{F}_t$  typically contains the data available in a (subset) of covariates up to  $t$ ,<sup>20</sup> i.e.  $\{x_1, x_2, \dots, x_t\} \subset \mathcal{F}_t$ . As the key problem here is to address the distribution of the  $y_t^{(h)}$ , we approximate the  $h$ -step ahead distribution, conditional on  $\mathcal{F}_t$  at various quantiles

$$\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\} \equiv \mathcal{T},$$

this through a procedure called *quantile regression*, which will be described in Section 5.2. That is, we create estimates

$$\hat{y}_{t,\tau}^{(h)} | \mathcal{F}_t, \quad \forall h \in \mathcal{H} \quad \forall \tau \in \mathcal{T}, \quad (5)$$

where  $\hat{y}_{t,\tau}^{(h)} | \mathcal{F}_t$  is an approximation of the  $\tau^{\text{th}}$  quantile of the true conditional distribution  $y_t^{(h)} | \mathcal{F}_t$ , which, in practice, is unobservable at time  $t$  given  $\mathcal{F}_t$ .

#### *The Application of Variable Selection*

A key issue in this analysis is to fit an in-sample model with a sufficiently small number of degrees of freedom in order to avoid problems such as overfitting, as the variable search space consists of 357 candidate predictors and as the number of observations is small (approximately 100 monthly returns). That is, a significant part of the methodology will explain how we perform variable selection in our quantile regression framework through a genetic algorithm-assisted combinatorial optimisation search (in shorthand referred to as the GA). To our knowledge, this is a novel approach to variable selection in the context of quantile regression. Furthermore, as candidate predictors represent explanatory variables being lagged and transformed at up to two different time window sizes, the variable search space is also somewhat constrained, which also has to be accounted for in the optimisation problem. Thus, we expand existing methodology in a twofold way: Firstly, through exploring a new approach to variable selection in quantile regression (the GA); secondly, in applying constraints in this optimisation problem to be able to make prediction models for quantiles  $\tau \in \mathcal{T}$  for *several*  $h$ -step ahead conditional distributions,  $h \in \mathcal{H}$ .

---

<sup>20</sup>We refer to the covariates as the subset of candidate predictors included in the model.

## Model Overview

In essence, the modelling procedure involves the creation of candidate predictors, followed by the selection of a subset of these for each horizon  $h$  in the overall model, and finally the design of corresponding quantile regressions. It can be summarised in the following steps:

### PROCEDURE 2: The $h$ -step ahead log return conditional distribution prediction model

- 1) 25 explanatory variables, indexed  $i \in \mathcal{N}$ , are selected based on the discussion in previous chapters.  $\mathcal{N}$  contains 24 exogenous explanatory variables as well as the NQSALMON itself (as lagged versions of the index might have predictive power).
- 2) The explanatory variables are lagged and transformed through time windows outlined in the previous chapter. These comprise a set of 357 candidate predictors. In total, all these candidate predictors will comprise the relevant information available in  $\mathcal{F}_t$ .
- 3) A quantile regression framework is applied to make estimates of the  $h$ -step ahead conditional log return distribution of the NQSALMON. To determine the  $U = 8$  most predictive covariates in this framework, we consider a variable selection problem across the joint set of  $h$ -step ahead quantiles  $\tau \in \mathcal{T}$ . That is, for each horizon  $h \in H$ :
  - i) A genetic algorithm (GA) is employed to determine the combination of candidate predictors that yields the lowest sum of quantile regression loss functions of the respective quantiles  $\tau \in \mathcal{T}$ . That is, the predictors employed across the quantiles are held constant, differing only in the regression coefficients.
  - ii) The result from the algorithm's search is a set of covariates that will be applied in multivariate quantile regressions for horizon  $h$ . I.e.,  $|\mathcal{T}| = 5$  linear quantile regression models are created for each particular  $h$ .
- 4) Finally,  $|\mathcal{H}| \cdot |\mathcal{T}| = 12 \cdot 5 = 60$  quantile regression models are obtained, providing estimates of the  $h$ -step ahead conditional distribution  $\hat{y}_{i,\tau}^{(h)} | \mathcal{F}_t$ . This set of models can then be used to perform forecasts of the NQSALMON  $h$ -step ahead (conditional) log return distribution.

An overview of the procedure is depicted in [Figure 5](#) on the next page.

### Terminology and Other Practical Considerations

Due to the many aspects of the model as summarised previously, we find it convenient to introduce some modelling-related terminology that will be used throughout the rest of the study, as well as some other practical considerations. At first, by the term *model*, we refer to the overall set of quantile regression models across all the twelve horizons, in each of which five quantiles are considered. The variable selection search will be performed in such a way that the set of quantile regression models for *one particular horizon*  $h \in \mathcal{H}$  will contain the same set of covariates, although with varying coefficients (and corresponding significances) across the five quantiles. Now, with the term *submodel*, we refer to the set of quantile regressions at a particular horizon  $h \in \mathcal{H}$ . As there are  $|\mathcal{T}| = 5$  quantiles under consideration, a submodel will therefore contain five quantile regressions. Correspondingly, the model will contain a set of  $|\mathcal{H}| = 12$  submodels. Next, as the overall application of variable selection and quantile regression *only* has been done in an in-sample setting,

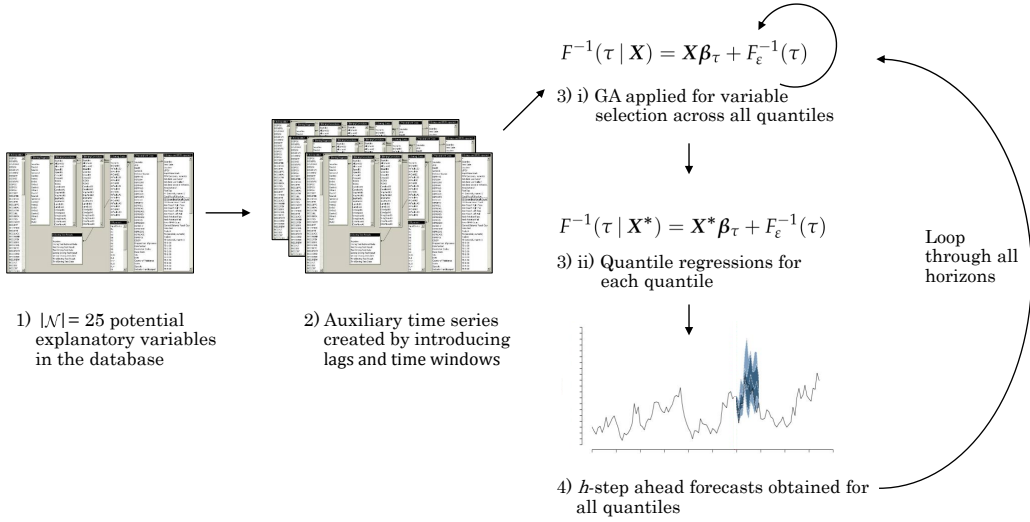


Figure 5: Overview of the  $h$ -step ahead log return conditional distribution prediction model.

the notions of in-sample and sample will be used interchangeably. Finally, we note that the size of the modelling data set is of 103 observations of 1- to 12-step ahead NQSALMON log returns. For the case of shorter horizons, say horizon  $h = 1$ , the data set of log returns will stretch from Jul. 2007 to Jan. 2016. Similarly, for longer horizons, e.g. horizon  $h = 12$ , the data set comprises observations from Jul. 2008 to Jan. 2017, as it is pushed forward with  $h$  months. Also, note that the data set is somewhat shrunk due to several data transformations, which will be discussed later on.

## 5.2 THE LINEAR QUANTILE REGRESSION MODEL

Quantile regression applies concepts from standard linear regression to parametrically estimate the quantiles of the conditional distribution of the dependent variable (Koenker & Bassett, 1978). Whereas standard linear regression provides a prediction of the mean and variance of the dependent variable  $\mathbf{y} = (y_t)_{t=1}^T$ , conditional on some given value of  $n$  covariates  $\mathbf{X} = (x_{t,i})_{t=1,i=1}^{T,n}$ , quantile regression enables the computation of the set of empirical quantiles  $\tau \in (0, 1)$  of the whole conditional distribution. This helps building a more thorough picture of the conditional distribution, denoted  $\mathbf{y} | \mathbf{X}$ . We will work with the *linear* quantile regression model (Alexander, 2008a, p. 305), which considers a linear relationship in the empirical quantile of the dependent variable with respect to its covariates. This particular framework will be described in the next paragraphs, and will thus be referred to as *the* quantile regression model.

### Theoretical Aspects of Quantile Regression

In our case,  $\mathbf{y} = (y_t)_{t=1}^T$  is the log return vector of the NQSALMON at times  $t$ , whereas  $x_{t,i}$  is the value of the covariate  $i$  at time  $t$  (also some log return). In the quantile regression model setup, we let  $\mathbf{X}$  denote the *design matrix* that incorporates the covariates. Also, as per the notation of Härdle and Simar (2014, p. 253), we let the leftmost column in  $\mathbf{X}$  correspond to a column of ones  $\mathbf{1}_T$  – thus, the overall design matrix has dimension  $T \times (n + 1)$ . Now, to define the quantile regression model, we start by recalling the formulation of the linear regression model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2 \mathbf{I}), \quad (6)$$

where  $\mathbf{y}$  and  $\mathbf{X}$  are as above,  $\boldsymbol{\beta} = (\beta_i)_{i=0}^n$  is the  $(n + 1) \times 1$  column vector of true model parameters where  $\beta_0$  is defined as the constant intercept term, and  $\boldsymbol{\varepsilon}$  is the  $T \times 1$  col-

umn vector of the independent and identically distributed (i.i.d.) normal error process terms. Now, we define our quantile regression model in similar terms as Alexander (2008a, pp. 305-307). Consider the random variable  $\mathcal{Y}$  with a cumulative distribution function  $F_{\mathcal{Y}}^{-1}(y) = P(\mathcal{Y} \leq y)$ . Let the  $\tau^{\text{th}}$  quantile of  $\mathcal{Y}$ ,  $\tau \in (0, 1)$ , be given by

$$Q_{\mathcal{Y}}(\tau) = F_{\mathcal{Y}}^{-1}(\tau) = \inf(y : F_{\mathcal{Y}}(y) \geq \tau). \quad (7)$$

Now, we have a set of  $T$  samples  $\mathbf{y}$  of  $\mathcal{Y}$ , and corresponding  $T$  observations of  $n$  covariates  $\mathbf{X}$ . We define a specific cumulative error distribution function,  $F_{\varepsilon}$ , and denote the  $\tau$  quantile of the error by  $F_{\varepsilon}^{-1}(\tau)$ . Taking conditional  $\tau$  quantiles of Eq. 6 gives:

$$F^{-1}(\tau | \mathbf{X}) = \mathbf{X}\boldsymbol{\beta}_{\tau} + F_{\varepsilon}^{-1}(\tau), \quad (8)$$

This equation is often referred to as the (simple) linear quantile regression model (Alexander, 2008a, p. 305). We would like to obtain sample estimates  $\hat{F}_{\tau}^{-1}(\mathbf{y} | \mathbf{X})$  of the conditional quantiles  $\tau$  under consideration, which require the computation of coefficient estimates  $\hat{\boldsymbol{\beta}}_{\tau}$  for the true model parameter vector  $\boldsymbol{\beta}_{\tau}$ . Koenker and Bassett (1978) show that this corresponds to the following optimisation problem:

$$\hat{\boldsymbol{\beta}}_{\tau} = \arg \min_{\boldsymbol{\beta}_{\tau}} \sum_{t=1}^T (\underbrace{\tau - \mathbf{1}_{y_t \leq \mathbf{x}_t \boldsymbol{\beta}_{\tau}}}_{\rho_{\tau}}) (\underbrace{y_t - \mathbf{x}_t \boldsymbol{\beta}_{\tau}}_{e_t}) \quad : \quad \mathbf{1}_{y_t \leq \mathbf{x}_t \boldsymbol{\beta}_{\tau}} = \begin{cases} 1, & \text{if } y_t \leq \mathbf{x}_t \boldsymbol{\beta}_{\tau} \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

Here,  $\mathbf{x}_t$  denotes the  $t$ -th row of  $\mathbf{X}$ , i.e. the  $1 \times (n + 1)$  row vector of intercept and covariate realisations at time  $t$ . Now, large residuals  $e_t$  will be penalised according to the loss function  $\rho_{\tau}$ . Thus, for small quantiles  $\tau$  the penalisation will be larger for data points *below* the estimated regression line  $\mathbf{X}\hat{\boldsymbol{\beta}}_{\tau}$ ,<sup>21</sup> moving the regression line towards smaller observations. *Mutatis mutandis*, for large quantiles  $\tau$ , data points *above* the regression line will be significantly penalised, effectively moving the line towards these observations.

### 5.3 LIMITING THE NUMBER OF QUANTILE REGRESSION PREDICTORS

In the case of forecasting the distribution of log returns of the salmon price, a large number of candidate predictor variables  $p$  are available. In our case, the space of candidate predictor variables is defined by the number of explanatory variables  $|\mathcal{N}| = n = 25$ , the allowed lags for each explanatory variable  $l_i \subseteq \{1, \dots, 18\} \forall i \in \mathcal{N}$  (as variables are at most lagged 18 months), and the *time window* per explanatory variable  $\delta_i \subseteq \{1, 12\} \forall i \in \mathcal{N}$  – being it month-over-month (MoM) or year-over-year (YoY). In the completely unrestricted case, and, for simplicity, disregarding the fact of multiple horizons, there would be  $p = 25 \cdot 18 \cdot 2 = 900$  candidate predictors. However, the data set of observations, i.e. the number of monthly log returns, effectively has size  $T = 103$ , thus in-sample overfitting is a significant concern. When a fitted model is too complex, that is, it has too many free parameters to estimate for the amount of information in the data, the in-sample *worth* of the model will be overestimated (Harrell, 2015, p. 72). In our case, the worth in the context of quantile regression is represented by the goodness-of-fit metric *pseudo-R-squared* of quantile regression, denoted  $R^1(\tau)$ . This metric is described in Appendix A.7. In the out-of-sample-case (not modelled in this study), many covariates in the regression model can yield the potential consequence of *overfitting*. That is, future observed values may not necessarily agree with respect to the quantiles predicted by the regression model. As previously discussed, we only consider in-sample modelling of the conditional distribution of the NQSALMON log returns. Consequently, this imposes discipline in the number of pre-

<sup>21</sup>For simplicity, we refer to this as the regression *line*; however, in the case of multivariate quantile regression, this is essentially a regression *plane*.

dictors to include as covariates in a quantile regression model. This motivates the process of *variable selection*.

The goal of variable selection, often called variable *subset* selection, is to identify and select the most important and non-redundant variables from a large pool of candidate variables – often called predictors in this context (Mao, 2004). Variable selection is often considered the hardest part of model building (Ratner, 2010). To begin with, each variable will to a larger extent add noise to the estimation of the dependent variable. Secondly, too many variables might increase the chance of multicollinearity (Ruczinski, 2005). This is associated with information overlap in the covariates (Kutner, Nachtsheim, Neter, & Li, 2005), and might in turn yield erratical behaviour of the coefficient estimates. Multiple studies have showed that in many situations, a fitted *standard linear* regression model is likely to be reliable when the number of predictors  $p$  is less than  $T/10$  or  $T/20$  (Harrell, 2015, p. 72), where  $T$  is the number of observations. Taking the average of  $T/15$ , this would in our case imply a subset size of  $U \leq 103/15 \simeq 7$ , as there are 103 observations available to the quantile regressions. However, in *low signal-to-noise environments* such as short- to medium-term financial time series, one can relax the  $T/15$  rule to some extent. The intuition is that such domains usually are complex in the sense that they represent the interaction between a large set of variables. That being said, being on the safe side and taking the phenomenon of overfitting into consideration, we have decided to set the subset size, i.e. the number of covariates in each quantile regression model, to  $U = 8$ .

For standard linear regression, there are a plethora of frameworks for the selection of  $U \leq p$  variables from a pool of candidate predictors. Given some fixed model size  $U$ , the simplest simultaneous selection method is the exhaustive-subset search. However, due to the combinatorial nature of this search problem, in a minimally-restricted case, described above, one would have to search  $\binom{900}{8} > 10^{19}$  combinations. Clearly, such combinatorial search problems quickly become computationally infeasible. Heuristic approaches such as adding or removing variables in a model to effectively yield the highest increase or the lowest decline, respectively, in the model worth often prove useful. Common methods following such approaches are stepwise regression, forward, and backward selection (Hastie, Tibshirani, & Friedman, 2009, pp. 57-61). However, some of them are greedy and effectively omit large parts of the search space, i.e. combinations of candidate predictors. Regularisation through various penalties is another approach.<sup>22</sup> Recently, some of these approaches have been adapted for quantile regression; for example, Wu and Liu (2009) apply norm-based penalties such as the adaptive-LASSO and smoothly clipped absolute deviation (SCAD) in this context. In this study, we apply a genetic algorithm, which is a stochastic optimisation technique, to search the pool of candidate predictors.

#### 5.4 VARIABLE SELECTION THROUGH A GENETIC ALGORITHM-ASSISTED SEARCH

The variable subset selection problem for quantile regression, and in our case with multiple  $h$ -step ahead submodels across multiple quantiles  $\tau$ , can be considered a constrained non-linear combinatorial optimisation problem. As discussed, exhaustive testing of all combinations of variables will find the optimal solution according to some objective function – in our case, the models' respective *worth* metrics – but is a computationally infeasible approach. The branch-and-bound algorithm and genetic algorithms (GAs), the latter being a particular approach of stochastic optimisation, can provide optimal variable selection without exhaustive search in the context of standard linear regression (Siedlecki & Sklansky, 1989). We utilise a tailored genetic algorithm for our particular problem, which, to our knowledge, in the context of quantile regression is a novel approach to variable subset selection. GAs have been used successfully in the context of variable selection for e.g.

<sup>22</sup>See e.g. the seminal introduction of the  $L_1$ -penalty in this context, called the least absolute shrinkage and selection operator (LASSO), by (Tibshirani, 1996).

standard linear regression (i.e., OLS) problems (Acosta-González & Fernández-Rodríguez, 2007; Broadhurst, Goodacre, Jones, Rowland, & Douglas, 1997). However, we have not found examples in the literature for the application of GAs for this selection problem in connection with (linear) quantile regression. Thus, we expand the limited work on variable selection for this class of statistical models.

#### *The Quantile Regression Variable Selection Optimisation Problem*

For each  $h$ -timestep ahead, where  $h \in \{1, \dots, 12\} \equiv \mathcal{H}$ , we seek to find a subset of candidate predictors that will be fixed across the corresponding  $h$ -step ahead regressions for quantiles  $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\} \equiv \mathcal{T}$  (differing only in regression coefficients). A main reason for this design choice is interpretability; we are able to assess how each predictor influences the  $h$ -step ahead distribution at various quantiles. Each timestep is 1 month, thus the overall prediction model will span up to one year ahead. We will refer to each step ahead as a quantile regression model's respective *horizon*. Correspondingly, for each horizon we model five quantiles, resulting in  $12 \cdot 5 = 60$  quantile regressions. Together, these comprise an approximation of the 1- to 12-step ahead conditional distributions of the log returns of the NQSALMON index.

The non-linear combinatorial optimisation program to determine the optimal subset of variables to be included is described in Table 4. Note that we consider the set of models for each horizon  $h \in \mathcal{H}$  as separate optimisation problems, quantified by the  $\forall h \in \mathcal{H}$  in the objective function in the table.

#### *The Concept of Genetic Algorithms*

Genetic algorithms (GAs) are stochastic optimisation techniques based on principles of natural evolution, initially developed by Holland (1992). Generally, they try to overcome problems of traditional optimisation algorithms, such as absence of continuity or differentiability of the loss function, or many local optima (Acosta-González & Fernández-Rodríguez, 2007). The original genetic algorithm works in binary search spaces, that is, search spaces consisting of binary strings (Siedlecki & Sklansky, 1989).<sup>23</sup>

Unlike hill climbing and other stochastic search approaches, a genetic algorithm does not evaluate and try to improve a single solution; rather, it does this simultaneously for a *population*, i.e. a set, of candidate solutions. To begin with, this set consists of randomly initialised solutions. Each member of the population is an *individual*, with a corresponding binary string representation that is called a *chromosome*. In each population *generation* (i.e., algorithm iteration), the algorithm will manipulate this set of solutions in a way similar to that of natural selection. Initially, each individual's *fitness* will be assessed. In general, better fitness is associated with a better approximation to the optimal solution. "Fit" individuals, i.e., promising solutions, are selected and recombined ("bred") through an operation called *crossover*. The child chromosome resulting from the crossover operation will represent a solution being a synthesis of its parents. To avoid local optima, various randomisation procedures might be introduced. The most common one is the *mutation* operation, in which a chromosome's binary representation, at random, is slightly altered. Now, the population size is fixed, and only a predefined number of the best individuals will "survive" to the next generation. Usually, more "fit" individuals in the pool of parents and children will proceed in this last selection procedure of each iteration. In the end, the objective of the GA is to create new population generations whose individuals gradually become better fit, and thus provide better approximations to the solution of the optimisation problem. For a thorough introduction to genetic algorithms, we refer to the seminal work of Goldberg (1989).

<sup>23</sup>Through various encoding schemes, continuous search spaces can be considered as well.



Table 4: Combinatorial optimisation program to determine the variable selection for quantile regression model across  $h$ -step ahead horizons and quantiles  $\tau \in \mathcal{T}$ .

SETS AND INDICES		
$i \in \mathcal{N}$		Explanatory variable $i$ . There are 25 such variables, as by Table 2
$\tau \in \mathcal{T}$		Quantile $\tau$
$h \in \mathcal{H}$		Horizon $h$ to be modelled; i.e., which log returns $y^{(h)}   \mathcal{F}_t$ to model
$l \in \mathcal{L}$		Lag $l$
$\delta \in \mathcal{D}$		Time window $\delta$
$t \in T$		Observation $t$
PARAMETERS		
$U \in \mathbb{N}$		The upper bound on the number of variables to be included in a model, set to 8
$\psi_{i,l}^{(h)} \in \{0,1\}$	$\forall (i,l,h) \in \mathcal{N} \times \mathcal{L} \times \mathcal{H}$	Indicator function with value 1 if lag $l$ of explanatory variable $i$ is allowed for models belonging to horizon $h$ , 0 otherwise
$\theta_{i,\delta} \in \{0,1\}$	$\forall (i,\delta) \in \mathcal{N} \times \mathcal{D}$	Indicator function with value 1 if time window $\delta$ of explanatory variable $i$ is allowed, 0 otherwise
VARIABLES		
$\phi_{i,l,\delta}^{(h)} \in \{0,1\}$	$\forall (i,l,\delta,h) \in \mathcal{N} \times \mathcal{L} \times \mathcal{D} \times \mathcal{H}$	Indicator with value 1 if variable $i$ at lag $l$ and time window $\delta$ is included in <i>all</i> quantile regression models at horizon $h$ , 0 otherwise. <b>Note:</b> If 1, it reserves one column vector slot $x_i$ in the design matrix $\mathbf{X}$ .
$\beta_\tau^{(h)} \in \mathbb{R}^{U+1}$	$\forall (\tau,h) \in \mathcal{T} \times \mathcal{H}$	The $U + 1$ -dimensional coefficient vector for the $h$ -step ahead $\tau^{\text{th}}$ quantile regression model
OBJECTIVE FUNCTION		
minimise $\sum_{\tau \in \mathcal{T}} \sum_{t \in T} \left( \tau - \mathbf{1}_{y_t^{(h)} \leq x_t \beta_\tau^{(h)}} \right) \left( y_t^{(h)} - x_t^{(h)} \beta_\tau^{(h)} \right) \quad \forall h \in \mathcal{H}$		
CONSTRAINTS		
$\sum_{i \in \mathcal{N}} \sum_{l \in \mathcal{L}} \sum_{\delta \in \mathcal{D}} \phi_{i,l,\delta}^{(h)} \leq U$	$\forall h \in \mathcal{H}$	Maximum model size
$\phi_{i,l,\delta}^{(h)} \leq \psi_{i,l}^{(h)}$	$\forall (i,l,\delta,h) \in \mathcal{N} \times \mathcal{L} \times \mathcal{D} \times \mathcal{H}$	Allowed lag at horizon
$\phi_{i,l,\delta}^{(h)} \leq \theta_{i,\delta}$	$\forall (i,l,\delta,h) \in \mathcal{N} \times \mathcal{L} \times \mathcal{D} \times \mathcal{H}$	Allowed time window

### Applying a Genetic Algorithm to Variable Selection

A high-level overview of the genetic algorithm that have been applied to solve the problem in [Table 4](#) is described in pseudocode in [Algorithm 1](#). Many of the functional aspects of the algorithm have been omitted in the pseudocode, and will be described in the upcoming paragraphs. Also, we will provide an overview of the parametrisation of the algorithm, e.g. values for the crossover rate  $P_c$  and the mutation  $P_m$ , which is applied in the variable selection search.

In the case of variable subset selection, each chromosome consists of bits  $\phi_{i,l,\delta}^{(h)}$ , and can therefore be represented as a binary vector (or, equivalently, a binary string). Each bit will then represent the presence (1) or absence (0) of the predictor, which again will be used to generate a corresponding quantile regression model formulation. For instance, one possible chromosome, denoted  $a_i$ , could be  $a_i = 00010011101$ . This can be translated such that predictors 4, 7, 8, 9, and 11 are to be used in the modelling process, whereas candidate predictors 1, 2, 3, 5, 6, and 10 are to be omitted.

---

#### Algorithm 1: Genetic algorithm for variable selection

---

**Input:** Horizon  $h$  for which the variable selection is to be solved  
**Output:** Set of covariates  $\{x_1, x_2, \dots, x_8\}$

- 1 Construct an initial population set  $\Pi \leftarrow \{a_i\}_{i=1}^{\text{pop\_size}}$
- 2 **for**  $a_i \in \Pi$  **do**
- 3      $\lfloor$  eval\_fitness( $a_i$ )
- 4 Initialise the elite set  $\mathcal{E} \leftarrow \emptyset$
- 5 **for**  $i \leftarrow 1$  to  $n\_generations$  **do**
- 6     Keep the elitism% of the most fit individuals in  $\Pi$  in  $\mathcal{E}$
- 7     Initialise mating set  $M \leftarrow \emptyset$ , offspring  $O \leftarrow \emptyset$
- 8     **for**  $j \leftarrow 1$  to  $pop\_size$  **do**
- 9          $\lfloor$   $M \leftarrow M \cup \text{run\_tournament}(\Pi)$
- 10     **for**  $j \leftarrow 1$  to  $pop\_size$  **do**
- 11         Select a pair  $a_j, a_k$  from  $M$
- 12         **if**  $\text{rand}() < P_c$  **then**
- 13              $\lfloor$   $O \leftarrow O \cup \text{do\_crossover}(a_j, a_k)$
- 14         **else**
- 15              $\lfloor$   $O \leftarrow O \cup a_j$
- 16     **for**  $a_j \in O$  **do**
- 17         **if**  $\text{rand}() < P_m$  **then**
- 18             Randomly select and flip one of the 1-bits in  $a_j$  to 0
- 19             Randomly select and flip one the 0-bits in  $a_j$  that belongs to  
            allowable\_indices( $h$ ) to 1
- 20     **for**  $a_j \in O$  **do**
- 21          $\lfloor$  eval\_fitness( $a_j$ )
- 22      $\lfloor$   $\Pi \leftarrow \text{combine}(\Pi, O, \mathcal{E})$
- 23 **return** The set of covariates  $\{x_1, \dots, x_8\}$  corresponding to the chromosome of the best-fit individual in the final population  $\Pi$

---

The first important function in [Algorithm 1](#) is the fitness evaluation. Denoted functionally as  $\text{eval\_fitness}(a_i)$ , it involves assigning an individual, with its chromosome  $a_i$ , a numeric value representing its fitness, the lower the better. In particular, solution  $a_i$  is translated into a design matrix  $X$  which incorporates an intercept column  $\mathbf{1}_T$  and up to  $U = 8$  candidate predictors, at some lag and time window (MoM or YoY). Recall that each index  $d \in a_i$  corresponds to a candidate predictor with a particular combination of lag and time window. Now, this design matrix is used in the objective function described in [Table](#)



4 to find coefficients  $\beta_{\tau}^{(h)}$ , while at the same time yielding a value for the corresponding objective function across quantiles  $\tau \in \mathcal{T}$  for the given horizon  $h \in \mathcal{H}$ . As mentioned, a lower fitness indicates that the corresponding set of quantile regressions in the submodel of a horizon  $h$  to a better extent fit the data.

The fitness function is, however, slightly different from the aforementioned objective function due to the fact that it also has to penalise infeasible solutions. In particular, one outstanding problem associated with the mapping from chromosome (bit string) to a set of covariates is that the latter may be infeasible solutions as specified by the optimisation program. This problem may become particularly severe for constrained and/or combinatorial optimisation problems (Gen & Cheng, 1999, p. 4), and the optimum may typically occur at the boundary between feasible and infeasible regions of the solution space. Therefore, instead of straight out discarding infeasible solutions, the fitness function penalises the number of constraint breaches and corresponding "severities". This will force the genetic search to approach potential optima from both sides of the feasible and infeasible regions. In our variable subset selection case, solution strings that have more than  $U = 8$  bits of value 1 are penalised, in addition to penalties regarding violations of the other constraints in Table 4. Intuitively, such constraint violations incur positive-valued penalties, as we seek to minimise the overall fitness value. The result is that one observes a large number of infeasible combinations of candidate predictors in the early generations; however, this share of individuals is generally close to zero in later generations (and discarded as final solutions).

Next, the crossover operation, defined in `do_crossover`, merges two candidate solutions, called parents, to create a new child chromosome. Before running crossover, a *mating set*  $M$  is initialised through the `run_tournament`( $\Pi$ ) function. In accordance with Broadhurst et al. (1997), in this tournament function, a weighted, random selection is applied to the original population  $\Pi$  where the probability of a particular individual (representing a particular subset) being selected is inversely proportional with its fitness value. Thus, "fit" individuals (low fitness values), with correspondingly "good" subsets, have a higher chance of entering the mating set and subsequently become parents. After having initialised  $M$ , iterations are run in which two parents  $(a_j, a_k)$  at random are selected from  $M$ . With a given probability  $P_c$ , the crossover is carried through and a child is added to the offspring set  $O$ , otherwise, the first parent  $a_j$  is added to  $O$ . In our implementation, the `do_crossover`( $a_j, a_k$ ) function saves the two parent chromosomes' respective indices of ones (recall that the chromosome length is fixed), and uniformly samples at most  $U$  indices from this joint set. For instance, given two bit strings 00010011101 and 10001010110, with  $U = 5$ , the joint index set of ones is  $\{4, 7, 8, 9, 11\} \cup \{1, 5, 7, 9, 10\} \equiv \{1, 4, 5, 7, 8, 9, 10, 11\}$ . Thus, a child chromosome might be represented by indices  $\{4, 5, 7, 9, 11\}$ , i.e. 000101010101.

The mutation is done with probability  $P_m$ , and works essentially as described in the pseudocode in Algorithm 1. Specifically, it is a variant of the *swap mutation* (Gen & Cheng, 1999, p. 88). In our case, one randomly selected bit of the  $U = 8$  1-bits is set to 0, and another randomly selected, "feasible" bit is set to 1. A feasible bit is present at indices found by the `allowable_indices`( $h$ ) function. For instance, a particular index of the chromosomes may represent a lag that is allowed at some horizons  $h$ , but not allowed at other horizons – as specified by the constraint  $\phi_{i,l}^{(h)} \leq \psi_{i,l}^{(h)}$  in the optimisation program in Table 4.

Finally, there are some principles and procedures regarding the management of the population from generation to generation. An important principle that is employed is *elitism*. Specifically, we retain the `elitism_rate` share of the most "fit" individuals as unaltered copies at the beginning of each generation, placing them *elite set*  $\mathcal{E}$ . At the end of each generation, the elite set is merged in together with the offspring set  $O$  and the current population  $\Pi$  to create a new population through the  $\Pi \leftarrow \text{combine}(\Pi, O, \mathcal{E})$  operation. The `combine` operation selects the `pop_size` most fit individuals from the union of these three

sets. In other words, even though the elite individuals are unaltered throughout the generation, they still will have to "compete" with the other individuals at the end. However, elitism is clearly an exploitation measure, ensuring that the best available solution has to monotonically improve for every generation.

#### *Parametrisation and Implementation of the Genetic Algorithm*

The most significant GA parameters are summarised in Table 5. The selection of parameters is governed by known, useful default values provided by Holland (1992), as well as by some preparatory experimentation. Also, the available processing power is an important factor: To achieve reasonable total running times for the selection procedure across *all* horizons, the number of generations is set to the relatively modest number of 200. Also, due to architecture of the multiprocessing system which is to be employed, the population size is set to 230 – a multiple of the available processing cores. The GA is implemented in **R**, supported by the quantile regression package `quantreg`, described by Koenker (2017). Due to the parallel nature of the operations performed in every generation of the GA, e.g. crossover and mutation, the code is designed to work in a multicore setting for appropriate speedup.

Table 5: Parameters for the genetic algorithm that is used in the variable selection search.

PARAMETER	VALUE	DESCRIPTION
Population size	<code>pop_size = 230</code>	The population size, i.e. the number of chromosomes that is worked with in parallel per generation. Essentially controls the search depth. This is set to a multiple of the number of cores on the computer that is used in the search.
Generations	<code>n_generations = 200</code>	The number of generations, i.e. iterations, of which to run the genetic algorithm. Essentially controls the search depth. The given number proves sufficient; in particular, the algorithm usually converges to a minimum fitness in less than half this number.
Elitism	<code>elitism_rate = 0.10</code>	The percentage of chromosomes in the current population (before crossover and mutation) to retain unaltered to the next generation. Important exploitation parameter.
Crossover rate	$P_c = 0.60$	The probability of carrying through a crossover for a particular pair of covariates. Important exploitation parameter (as it preferably merges "fit" chromosomes).
Mutation rate	$P_m = 0.80$	The probability of carrying through a mutation for a particular (child) chromosome. Important exploration parameter (as it may introduce entirely new variables).

## 5.5 MODEL RATIONALE AND CHALLENGES

The overall approach, building specific models based on a general pool of candidate predictors through some process of variable selection, draws resemblance to the *general-to-specific* paradigm of Campos, Ericsson, and Hendry (2005). That is, the modeller simplifies a general model, which satisfactorily fits the evidence, to a more specific theoretical framework. In our case, we start out with a slightly restricted pool of candidate predictors, as outlined

initially in this chapter – as well as [Chapter 3](#). Thus, one might argue that the overall approach is not entirely *general* in nature, as we have utilised industry expertise in order to restrict the problem. Equivalently, the argument of data mining, which frequently is used with respect to the general-to-specific framework ([Hoover & Perez, 2000](#)), does not carry as large weight here as in the "unrestricted search" case. However, we will discuss the choices done with respect to the distribution model (here being linear quantile regression) and overall problem parametrisation (e.g. the choices of quantiles  $\tau \in \mathcal{T}$  to be modelled, as well as horizons).

#### *On the Choice of Distribution Model: Linear Quantile Regression*

There are several advantages of quantile regression relative to simpler approaches, such as the standard linear regression framework. The main idea is that a dense set of quantiles essentially describes any conditional distribution function ([Alexander, 2008a](#)). This is useful in the case of predicting the NQSALMON (log) returns, as one might in a detailed and structured way assess how explanatory variables and information available at some time  $t$  might impact the  $h$ -step ahead returns. Now, as the justification for making prediction models for the conditional distribution of the log returns of the NQSALMON, we will reason for why quantile regression in particular is a suitable framework in this context.

Quantile regression provides a simple and intuitive framework for scenario analysis and assessing covariate effects ([Fahrmeir, Kneib, Lang, & Marx, 2013](#), p. 598). Ranging a particular covariate from its minimum to its maximum past observed value, one might easily retrieve the impact on the estimated conditional distribution. Similarly, higher-order effects can be assessed.<sup>24</sup> This is useful in our case, as there are intuitively several candidate predictors that either alone or together ought to have an impact on the salmon price log return. Should these predictors be chosen in the variable selection process, one might try out corresponding scenario analyses.

Relative to an approach such as standard linear regression, quantile regression avoids several restrictive assumptions; for example, the error terms need no longer follow a specific type of distribution, neither are they required to be homoscedastic ([Fahrmeir et al., 2013](#), p. 598). This might be useful as the (log) returns of the NQSALMON visually seems to exhibit some volatility clustering, evident in [Figure 3c](#). This might again impact a regression model's residuals, violating the OLS assumption of homoscedasticity. Furthermore, in standard linear regression, the mean is modelled rather than respective quantiles. As the median (or other quantiles) are less sensitive to outliers, quantile regression might prove robust in this setting. Generally, [Koenker and Hallock \(2001\)](#) find that inference for quantile regression is, if anything, more robust than most other forms for inference commonly found in econometrics. That being said, as evident by [Figure 3a-3d](#), the empirical, weekly log return distribution of the index is not particularly heavy-tailed; however it is clear that monthly movements might be of greater magnitude.

Nonetheless, there are some drawbacks of quantile regression in our specific application of predicting the log return of NQSALMON, however we argue that there does not seem to be other appropriate frameworks in this regard. In applications, there is often a particular interest in regression quantiles that describe extreme observations in terms of covariates ([Fahrmeir et al., 2013](#), p. 598). For agents in energy markets such as that of the power market, modelling the tails is often more important than formulating central expectations ([Hagfors, Bunn, Kristoffersen, Staver, & Westgaard, 2016](#)). However, in our case, we only consider a sample size of  $n = 103$  log returns. Thus, trying to fit extreme quantiles, say at levels smaller than  $\sim 0.01$  or larger than  $\sim 0.99$ , is, to put it mildly, impractical. With

---

<sup>24</sup>By *higher-order* effects, we mean that the predictors themselves interact, for instance by having slight collinearity. Thus, the increase in one predictor might be correlated with an increase in another, and one has to take into account this joint effect in a regression.

such a small sample, considering tail dependencies could possibly require distributional assumptions. Thus, one moves into the area of "unknown unknowns" (Taleb, 2009), that is, we have neither observed nor are able to fit a particular distribution for extremes in the log returns of the NQSALMON. In the end, we argue that our approach with quantile regression is sound – we are primarily interested in a robust model for the more "central" expectations rather than the extremes, as there are few literary examples in general of a distribution-based prediction model for the returns of the price of salmon.

In this study, we employ a *linear* quantile model, which, by definition, may be unable to capture certain non-linear relationships (Landajo, De Andrés, & Lorca, 2008). One might consider to specify and fit parametric non-linear models for the relevant conditional quantiles to deal with a non-linear quantile regression problem (Joshua, Victor, & Fernandez-Val, 2013). However, it requires rich *a priori* knowledge to specify a functional form for each conditional quantile, which is scarce in our case. Equivalently, inherently non-linear models such as neural networks will usually require many degrees of freedom to make a good fit (Bishop, 2006, Chapter 5), which, as previously described, quickly may result in overfitting. Hence, we apply a linear quantile regression approach to model the price distribution for the NQSALMON.

#### *On Model Parameters: Choices and Challenges*

The model overall utilises a monthly time resolution, even though the salmon spot price is reported on a weekly basis. In particular, many important fundamental, exogenous variables are reported on a monthly basis. Furthermore, the signal-to-noise ratio might be increased with a courser time scale (Aas & Dimakos, 2004). Also, market participants are often interested in long-term price movements, and a monthly resolution might therefore suffice. However, such a resolution also yields equivalently few observations for which to estimate respective models. As discussed above, due to the 103 log return observations, the choice of monthly resolution makes us unable to assess more "extreme" quantiles. That being said, due to the fact that exogenous variables often are available only at course time scales, one would eventually have a strategy to "smooth" out such observations to obtain e.g. weekly estimates. The strength of a monthly time resolution is that it fits all explanatory variables, and does not depend on any smoothing assumptions.<sup>25</sup>

The horizon set  $h \in \{1, \dots, 12\} \equiv \mathcal{H}$  is deemed appropriate for the  $h$ -step ahead modelling, as it reflects the desired forecasting range for market participants. Since the production cycle is long, up to approximately three years from the spawning of smolt to harvest, slaughtering and processing (Marine Harvest, 2016), it makes sense to have a long-term picture of the plausible price paths of the particular commodity price. However, the chosen set of horizons effectively restricts the sample size slightly more. That is, to be able to fit the 12-month ahead price, the last twelve observations of the sample have to be cut off. Thus, as the last observation of the NQSALMON in our data set is of Jan. 2017, the sample which is used to fit the model has to end in Jan. 2016. From this, the whole range of horizons  $h \in \{1, \dots, 12\}$  can be assessed, whereas e.g. for Feb. 2016 the 12-month ahead value would not be available.

The model size is restricted to  $U = 8$  covariates to create parsimonious models – avoiding issues such as overfitting and other issues when applied in an out-of-sample context, as discussed previously in Section 5.3. Including too many variables in a linear quantile regression model might cause *overfitting*, and, thus, unsatisfactory out-of-sample performance. We could possibly have fitted even sparser quantile regression models, however, a part of our approach is to consider how the candidate predictors interact and jointly impact

<sup>25</sup>Smoothing monthly observations of exogenous variables to weekly observations could result in badly behaved log returns. For instance, linear smoothing would yield constant weekly log returns between each two respective, monthly data points – the intuition being that the (price) series derivative is constant.

the  $h$ -step ahead log return distribution.

Quantiles  $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\} \equiv \mathcal{T}$  are considered appropriate in approximating the  $h$ -step ahead log return distribution. As previously discussed, due to the limited number of observations, it does not make sense to model extreme quantiles in the quantile regression framework. Furthermore, due to the small sample size, specification procedures, such as Christoffersen's test for independence, might be particularly sensitive to violations at extreme quantiles.

An important aspect with respect to the quantiles is that the variable selection is done across quantiles for each horizon  $h \in \mathcal{H}$ , as outlined in [Algorithm 1](#). This is an important design choice, and has several motivations. Firstly, this approach avoids unnecessarily many different covariates ( $12 \cdot 8 = 96$  versus  $12 \cdot 8 \cdot 5 = 480$ ). Secondly, this introduces a constrained search for suitable variables, which has to a smaller extent been carried through in existing literature. Finally, the results will be slightly more interpretable in the context of scenario analyses and sensitivities with respect to the selected covariates; one might for instance consider how the coefficients  $\beta_{\tau}^{(h)}$  vary across quantiles  $\tau \in \mathcal{T}$  for some selected  $h$ -step ahead submodel.

#### *In-Sample Modelling Due to Lack of Data*

In addition to the reduction in data points due to the aforementioned horizons  $h \in \{1, \dots, 12\}$ , the amount of data points is furthermore reduced by the fact that not all past history is available for all the exogenous variables. Even though the NQSALMON ranges back to Jan. 1995, we only have data for e.g. *Smolt release, #individuals (Norway)*, ranging back to Jan. 2005.<sup>26</sup> Also, the usage of particularly large lags, for instance up to 17 months in the smolt release case, as per [Table 3](#), effectively shrinks the data set further, having the first observation in Jul. 2007.

A particular caveat of the few data points is the design choice to employ the whole, available data set to fit the model. That is, we do not perform any out-of-sample modelling. In particular, the results and robustness of tests for the quantile regression models' respective out-of-sample predictive power would be no better than the size of the out-of-sample data set. A common split of the in- and out-of-sample data is 80/20, thus hypothetically  $\sim .2 \cdot 103 \simeq 21$  data points would be available out of sample. For a "tail" quantile, such as  $\tau = 0.10$ , one would expect very few observations below, in this case  $\sim 2$ . Specification tests, such as Kupiec' unconditional coverage test, would in such cases be highly sensitive to small deviations in the number of out-of-sample data points that fit particular quantiles. We acknowledge that such prediction models are no better than they perform in an out-of-sample setting; also, it would be of interest to assess whether or not the GA is able to select candidate predictors that jointly yield a set of out-of-sample predictive quantile regression models. However, the methodological approach might prove interesting enough; also, one might draw some qualitative and quantitative conclusions based on the covariates and associated regression coefficients across quantiles. Ultimately, the prediction model is unquestionably functional in the context of forecasting. It predicts the salmon price 1- to 12-months ahead, based on the information available at the start of the period. Thus, it can be applied to forecast the salmon price, both out-of-sample and in actual forward-looking situations.

## 5.6 SPECIFICATION TESTS

After the selection of variables for all submodels, we perform various specification tests for the corresponding quantile regression models. At first, we consider the coverage tests

---

<sup>26</sup>As we do not utilise dummy variables in the regression, we essentially have monthly values from Jan. 2005 to Jan. 2016 in the observation set.



of Kupiec (1995) and Christoffersen (1998), to evaluate whether or not the observed exceedances at the quantiles of interest are in line with what is predicted by the respective regression. Note that the aforementioned tests are constructed for evaluating interval *forecasts* (Christoffersen, 1998). In this study, however, they are applied for evaluating the quantile regression models *in-sample*. Next, we apply the novel quantile regression equation specification error test (QRESET) by Otsu (2009) to assess whether or not there are non-linear effects in the quantile regressions that have not been accounted for.

#### Unconditional Coverage: The Coverage Rate and the Kupiec Test

Initially, we evaluate the *coverage rate* of each quantile regression model. This is essentially the proportion of observed exceedances *below* some predicted quantile  $\tau$  of the conditional distribution  $y | X$ . For instance, in a particular model which is fitted to  $n = |T| = 103$  observations with  $\tau = 0.10$ , one would expect the corresponding coverage rate, equivalent to  $0.10 \cdot 103 \simeq 10$  exceedances.

Our first statistical test for model specification is that of Kupiec (1995), which considers whether or not the quantile regression model provides the appropriate unconditional coverage. The test is likelihood-based, and compares the observed proportion of exceedances  $\pi_{\text{obs}}$  against the expected proportion of exceedances  $\pi_{\text{exp}} \equiv \tau$ . Now, using the notation of Alexander (2008a, p. 359), an *exceedance* is defined by the following indicator function:

$$I_t = \begin{cases} 1, & \text{if } y_t \text{ falls into a given interval, i.e. } y_t \text{ is an } \textit{exceedance} \\ 0, & \text{otherwise, i.e. } y_t \text{ is a } \textit{non-exceedance}, \end{cases} \quad (10)$$

where the given interval is given as all values *below* the predicted quantile.  $I_t$  is assumed to follow an i.i.d. Bernoulli process. The unconditional test statistic is given as:

$$-2 \ln(\text{LR}_{\text{uc}}) = -2 [n_0 \ln(1 - \pi_{\text{exp}}) + n_1 \ln(\pi_{\text{exp}}) - n_0 \ln(1 - \pi_{\text{obs}}) - n_1 \ln(\pi_{\text{obs}})] \stackrel{\text{asy}}{\sim} \chi_1^2, \quad (11)$$

where  $n_1$  is the number of exceedances, and  $n_0$  is the number of non-exceedances. Moreover,  $n = n_1 + n_0$  is the number of observations in the sample, corresponding to the total number of observations – that is,  $n = |T| = 103$ . Thus,  $\pi_{\text{obs}} = \frac{n_1}{n}$ .

#### Conditional Coverage: The Christoffersen Test

A drawback with the Kupiec test is that it does not take into account whether or not the quantile exceedances are related, that is, whether they tend to appear in clusters or are isolated. To account for this, we employ the conditional coverage test of Christoffersen (1998). Informally, it tests whether an exceedance today influences the probability of an exceedance tomorrow. It jointly tests unconditional coverage, as before, as well as independency in the exceedances. The test statistic can be written as

$$\begin{aligned} -2 \ln(\text{LR}_{\text{cc}}) = -2 \ln(\text{LR}_{\text{uc}} \cdot \text{LR}_{\text{ind}}) = -2 [n_0 \ln(1 - \pi_{\text{exp}}) + n_1 \ln(\pi_{\text{exp}}) \\ - n_{00} \ln(1 - \pi_{01}) - n_{01} \ln(\pi_{01}) - n_{10} \ln(1 - \pi_{11}) - n_{11} \ln(\pi_{11})] \stackrel{\text{asy}}{\sim} \chi_2^2, \end{aligned} \quad (12)$$

where

$$\pi_{01} = \frac{n_{01}}{n_{00} + n_{01}}, \quad \text{and} \quad \pi_{11} = \frac{n_{11}}{n_{10} + n_{11}}.$$

Here,  $n_{ij}$  represents the number of times an observation of indicator value  $i$  is followed by an observation of indicator value  $j$ . Thus, clustering in exceedances are identified by high proportions  $\pi_{11}$ . In small samples, which is particularly relevant for our case with  $n = 103$ , it might be likely that e.g.  $n_{11} = 0$ , making the test statistic in Eq. 12 being handled slightly differently. In particular, it can be shown that in the corresponding (log) likelihood

function, the term  $-n_{10} \ln(1 - \pi_{11}) + n_{11} \ln(\pi_{11})$  in the parentheses can be removed altogether (MathWorks, 2017), and the remaining, *reduced* likelihood ratio  $-2 \ln(\text{LR}_{\text{cc}})$  will still be well-defined. This equivalently holds for the case when  $n_{00} = 0$ , again implying that  $\pi_{01} = 1$ . In this case, the  $n_{00} \ln(1 - \pi_{01}) - n_{01} \ln(\pi_{01})$  term will be removed in Eq. 12.

#### *Testing for Non-Linearities: The QRESET Test*

As the Kupiec and Christoffersen tests only utilise information on past quantile violations, they might not have satisfactory capability in detecting *misspecified* models. To increase the power, Bunn, Andresen, Chen, and Westgaard (2013) suggest that one may want to consider whether or not violations can be predicted by including other data in the information set such as past returns and estimated volatility. Due to our small sample, we do not adapt this method, as incorporating long lags will yield correspondingly short data sets for testing. The same argument applies when we choose to disregard the dynamic quantile (DQ) test of Engle and Manganelli (2004). Thus, we do not conduct tests to investigate the areas in which the prediction model might be misspecified, and what variables might help to mitigate this misspecification. Rather, we employ a novel, functional form test to assess the validity of using linear quantile regression. This is the novel quantile regression equation specification error test (QRESET), introduced by Otsu (2009). This test can be used for determining the soundness of some particular functional specification of a quantile regression model. It is based on the regression equation specification error test (RESET) by Ramsey (1969). Except for using quantile regression estimators rather than least-squares estimators, the implementation of the QRESET is almost equivalent as that of the RESET. In particular, this novel test verifies whether or not a quantile regression is correctly specified in the sense of being linear in its covariates. Each variant of the test, denoted  $\text{QRESET}(p)$ , adds powers to the fitted dependent variables  $\hat{y}$  up to order  $p \geq 2$  and check the significance of those added covariates. For specific details regarding our implementation of the QRESET test, we refer to Appendix A.8.

There are several reasons for why this test fits our domain: Firstly, it is easier to implement than nonparametric tests, as it is free from choices of kernel function and bandwidth parameters (Otsu, 2009). Furthermore, in small sample sizes, typically less than 100 (recall that, in our case,  $|T| = 103$ ), the test may be more robust than the nonparametric kernel-based specification test for quantile regression introduced by Zheng (1998). That being said, there is no clear ordering in the suggested number of powers  $p$  to include of the fitted dependent variable. As the sample size increases, the differences among the QRESETs become insignificant (Otsu, 2009). However, the  $\text{QRESET}(2)$ , which correspondingly adds the covariate  $\hat{y}^2$ , is more suited than  $\text{QRESET}(3)$  and  $\text{QRESET}(4)$  for small samples of size  $\leq 200$ . Moreover, for polynomial misspecifications,  $\text{QRESET}(2)$  is generally more powerful than the other  $\text{QRESET}(p)$ s (Otsu, 2009). Thus,  $\text{QRESET}(2)$  is probably more applicable for evaluating the model specification in this study.

## RESULTS AND DISCUSSION

---

We structure our evaluation of the overall model in the chronological order of its components, which has been outlined in [Section 5.1](#). At first, we present relevant algorithmic results, selected model statistics, and overall performance of the genetic algorithm (GA) that is applied in the variable subset selection search. Next, we consider the performance of each submodel for each horizon  $h \in \mathcal{H}$ , and its corresponding quantile regressions, with respect to goodness-of-fit. That is, we evaluate each quantile regression's ability to explain the variation in the salmon price. We then proceed with a set of measures for determining the soundness of the specifications of the quantile regressions. Initially, this is done through evaluating the coefficient significances and determine if the quantile regressions can be considered significant in some general sense. Next, we examine whether there exist misspecifications – such as dependencies in the exceedances of the predicted quantiles or non-linearities that are unaccounted for. Finally, the selected covariates, with respective coefficients and significances, is discussed in context of the variable hypotheses that were outlined in [Chapter 3](#).

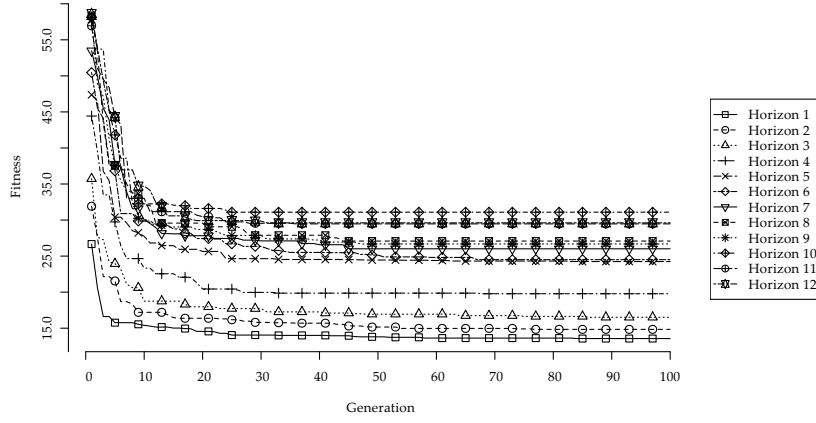
### 6.1 GENETIC ALGORITHM SEARCH EVALUATION

In this section, we present relevant results and statistics from running the genetic algorithm (GA). The intent of using variable selection is to evaluate  $p = 357$  candidate predictors and shrink this set down to parsimonious submodels with  $U = 8$  covariates across the twelve horizons. The GA, whose implementation details and parameters are provided in [Section 5.4](#) in the methodology, is run for each horizon  $h \in \mathcal{H}$  in accordance with [Table 4](#). The selected covariates for each submodel, together with coefficients and pseudo-R-squared in the corresponding regressions, can be found in [Tables 15-18](#) in [Appendix A.3](#).

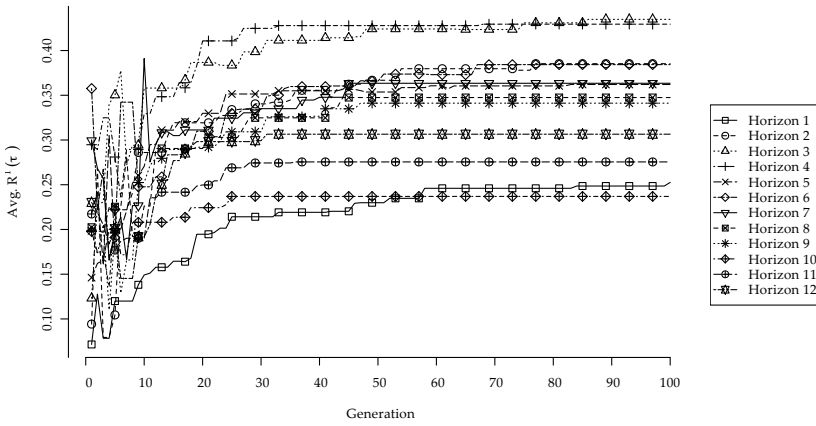
A visualisation of the per-generation performance for the runs of the GA, each run belonging to a particular horizon  $h \in \mathcal{H}$ , is presented in [Figure 6](#). In [Figure 6a](#), the development of the fitness value of the best-fit individual per generation is displayed. As specified in the parametrisation given in [Table 5](#), the GA is run for 200 generations for each horizon  $h$  – in line with what has been used as stopping criterion in the literature for variable selection ([Broadhurst et al., 1997](#)). However, evident from both plots in [Figure 6](#), the algorithm converges relatively quickly for all horizons. By *convergence*, we mean that as the generations proceed, the algorithm gets closer and closer to a specific value. As evident by [Figure 6a](#), the improvement (i.e., reduction) in fitness is generally small after approximately 50 generations, and the usage of the term *convergence* is therefore appropriate in this setting. Note that the lack of solution improvement not necessarily implies that a global optimum in the search space of candidate predictors is reached – the search might as well have found a local optimum. However, by experimentally running the algorithm multiple times, with slightly varying parameter settings, we are able to verify that the fitness values as plotted in [Figure 6a](#) generally are close to the global optimum. Consequently, the plot's horizontal axis is cut at 100 generations, as the improvement in the final 100 generations is minuscule. Lastly, in the context of variable selection for quantile regression, this study is a proof-of-concept of the usage of GAs. Thus, we let actual comparison with branch-and-bound, or greedy algorithms such as forward or backward selection, be left to further benchmarking studies.

Next, in [Figure 6b](#), the development in the  $\overline{R^1(\tau)} : \tau \in \mathcal{T}$  for the individual with the best (i.e., lowest) fitness in each generation is tracked for each horizon  $h \in \mathcal{H}$ . The  $R^1(\tau)$  metric, also referred to as *pseudo-R-squared*, creates a local measure for the goodness-of-fit for a particular quantile regression, by calculating a weighted sum of the absolute resid-





(a) Development in fitness



(b) Development in mean pseudo-R-squared for the corresponding submodel

Figure 6: Various metrics for the "best-fit" individual in each generation of the genetic algorithm, for each horizon.

uals over the in-sample period (Koenker & Machado, 1999). Generally, a higher pseudo-R-squared means that the regression is able to explain a larger amount of variation at the given quantile. For a detailed specification of the  $R^1(\tau)$ , see Appendix A.7. Now, as a submodel contains five quantile regressions, we have that  $\overline{R^1(\tau)}$  denotes the average across these quantiles. Back to Figure 6b, we clearly observe convergence also in this case. Furthermore, the fitness plots verify that lower fitness implies better overall submodel fit by the  $\overline{R^1(\tau)}$  metric. However, it is of importance to note that the initial noise in Figure 6b is due to the fact that infeasible submodels are fitted, which temporarily might increase the goodness-of-fit. By infeasibility, we refer to a solution whose respective subset of selected covariates violates one or more of the constraints in Table 4. The algorithm penalises (i.e., increases) the fitness of such infeasible submodels to a slight extent, thus the infeasible covariates are not present in later generations due to this impact on the fitness function. However, as initialisation is done at random, infeasible covariates might be present even in the best-fitness solution in earlier generations.

The apparent convergence of the implemented algorithm, given the current parametrisation, is verified by similar findings in the literature on GAs. Frost and Molt (1998) find good convergence properties of a GA considering  $p = 15$  candidate predictors for principal component regression (PCR), with the corresponding algorithm being able to solve the problem with a population size of 200 within 20 generations. For comparison, in this study,

we employ a population size of 230, a multiple of the number of available computational cores (23),<sup>27</sup> and experience more or less full convergence in approximately 50 generations.

Depending on the horizon, there is some variation in running time. Specifically, each run for a horizon  $h$  lasts from 30 minutes to approximately 1 hour. This variation is due to our structure on lags and time windows as defined in Table 3. In particular, we see that shorter horizons require more running time relative to longer horizons. Shorter horizons typically offer a larger subset of candidate predictors in accordance with the constraints outlined in Table 4. On the other hand, longer horizons will typically constrain the search space, i.e. the number of available predictors, significantly, so the overall number of candidate predictors in this case will be  $\ll 357$ . This results in longer horizons potentially yielding fewer, highly predictive combinations of covariates, and thus the submodels will generally have a poorer fit. Visually, we observe this phenomenon in Figure 6, where longer horizons have higher fitness values and correspondingly lower goodness-of-fit. On the other hand, one could expect a better fit with increasing horizon as e.g. the signal-to-noise ratio might improve accordingly.

## 6.2 IN-SAMPLE MODEL PERFORMANCE

### Model Performance

The goodness-of-fit metric  $R^1(\tau)$  was estimated for all quantile regressions.<sup>28</sup> For the results, we refer to Table 6. Here, from the lowermost row of the table, it appears that the performances of the *cross-models* for all quantiles are rather evenly distributed.<sup>29</sup> However, the cross-models are generally more successful in explaining the variation in the salmon price for the tails of the distribution. Specifically, the cross-model for quantile  $\tau = 0.10$  achieves the best performance, producing a pseudo-R-squared of 0.40.

Table 6: Aggregated results. Values for  $R^1(\tau)$  for the model.

Horizon	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$	Average
$h = 1$	0.25	0.23	0.23	0.24	0.27	<b>0.24</b>
$h = 2$	0.43	0.35	0.35	0.39	0.40	<b>0.39</b>
$h = 3$	0.48	0.44	0.42	0.40	0.43	<b>0.43</b>
$h = 4$	0.47	0.38	0.42	0.43	0.46	<b>0.43</b>
$h = 5$	0.39	0.32	0.35	0.40	0.42	<b>0.37</b>
$h = 6$	0.45	0.37	0.34	0.35	0.39	<b>0.38</b>
$h = 7$	0.42	0.33	0.32	0.37	0.38	<b>0.38</b>
$h = 8$	0.45	0.36	0.30	0.31	0.31	<b>0.35</b>
$h = 9$	0.37	0.35	0.30	0.34	0.34	<b>0.34</b>
$h = 10$	0.39	0.30	0.18	0.13	0.20	<b>0.24</b>
$h = 11$	0.34	0.31	0.29	0.21	0.23	<b>0.28</b>
$h = 12$	0.39	0.37	0.33	0.23	0.21	<b>0.31</b>
<b>Average</b>	<b>0.40</b>	<b>0.34</b>	<b>0.32</b>	<b>0.32</b>	<b>0.34</b>	<b>0.35</b>

Focusing on the rightmost column of the table, we observe that the submodels' performance in terms of pseudo-R-squared is comparatively better for horizons  $h \in \{2, 3, 4, 5, 6, 7\}$ .

<sup>27</sup>On a technical note, we run the algorithm on a Windows Server 2012 R2 system, employing two Intel® Xeon® X5650 2.67 GHz CPUs in which a total of 23 logical cores are utilised.

<sup>28</sup>We attempted to incorporate other goodness-of-fit metrics as well, such as the AIC and BIC. However, most metrics, including the AIC and BIC, can not be used to compare models of different data sets. Also, the same dependent variable have to be used for all the models (Hyndman, 2013). Hence, as our submodels utilise different subsets of covariates, and forecast over twelve horizons yielding twelve different dependent variables, we have not been able to identify any metrics other than the  $R^1(\tau)$  to compare the performance.

<sup>29</sup>By the term *cross-model* for some quantile  $\tau$ , we refer to the quantile regressions for a particular quantile  $\tau \in \mathcal{T}$  across all submodels. Therefore, each cross-model will contain twelve regressions for the quantile  $\tau$ , each such regression belonging to a particular submodel.

Also, worth noticing is the relatively poor  $R^1(\tau)$  produced by horizon  $h = 1$ . A possible explanation might be the presence of noise in the log returns of NQSALMON over the very short term, and a corresponding low signal-to-noise ratio. Moreover, the poor performance of the submodels for longer horizons, compared to the performance of the horizons  $h \in \{2, 3, 4, 5, 6, 7\}$ , might be a consequence of restrictions imposed by the lag structure. With reference to the discussion in [Chapter 3](#) and [Section 5.4](#), a richer variable search space is considered when the genetic algorithm (GA) selects covariates for submodels on shorter horizons. Longer horizons should intuitively yield a comparatively lower predictive power, as the variable search space for these horizons are more constrained.

#### Visualisation of In-Sample Predictive Power

[Figure 7](#) illustrates the in-sample predictive power of the submodels for all horizons  $h \in \{1, \dots, 12\}$ . See [Tables 15-18](#) in [Appendix A.3](#) for the covariates, with corresponding coefficients, employed by the twelve submodels. As evident from the figure, the submodels capture the fluctuations in the salmon price distribution reasonably well across the 12-month period, given the information available at the end of July 2013.

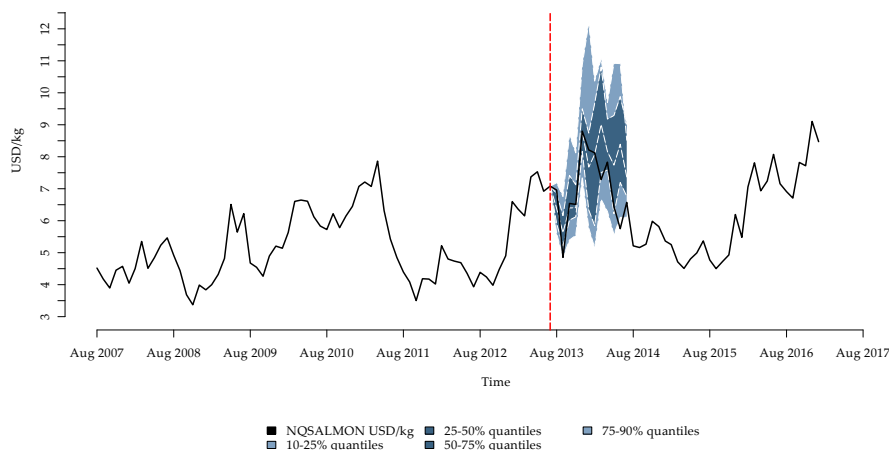


Figure 7: Visualisation of the in-sample predictive power of the 1- to 12-month ahead submodels. Based on the information available at July 31<sup>st</sup> 2013, the price distributions for August 31<sup>st</sup> 2013 through July 31<sup>st</sup> 2014 are predicted.

### 6.3 IN-SAMPLE MODEL SPECIFICATION

Several tests have been conducted to examine the soundness of the model specification. In particular, we test for the significance of the covariates' coefficients, the model's ability to provide the correct unconditional and conditional coverage, and whether it has a correct specification in terms of being linear in its covariates. In the following subsections, the corresponding test results are presented and discussed.

#### Significance of the Covariates' Coefficients

We assess the significance of the covariates' coefficients to further examine their predictive power and the robustness of the model. For a complete overview of the significance of the covariates in the model, see [Tables 15-18](#) in [Appendix A.3](#). Here, we present only aggregated results, as seen in [Table 7](#) and [Figure 8](#). [Table 7](#) illustrates the number of coefficients that are significant for the model at the 1% level, from which we notice that 111 out of 540 are significant at the 1% level. [Figure 8](#) provides a detailed representation of the significance of the coefficients for the different cross-models.

From the lowermost row of the [Table 7](#), we observe that the significance of the coeffi-

Table 7: Aggregated results. Number of occurrences of coefficient  $p$ -values less than 0.01 (including intercept), for the model.

Horizon	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$	Sum
$h = 1$	0	2	0	2	1	5
$h = 2$	2	5	3	3	0	13
$h = 3$	0	4	3	2	0	9
$h = 4$	2	2	2	2	2	10
$h = 5$	1	1	3	5	0	10
$h = 6$	1	3	3	2	0	9
$h = 7$	1	4	1	6	0	12
$h = 8$	1	3	1	2	0	7
$h = 9$	0	4	3	3	0	10
$h = 10$	3	3	2	1	1	10
$h = 11$	2	3	2	1	1	9
$h = 12$	1	4	2	0	0	7
<b>Sum</b>	<b>14</b>	<b>38</b>	<b>25</b>	<b>29</b>	<b>5</b>	<b>111 of 540<sup>1</sup></b>

1) 9 coefficients  $\cdot$  5 quantiles  $\cdot$  12 horizons = 540. Of the 111, intercepts constitute 30.

coefficients is greatest around the central quantiles, that is, for  $\tau \in \{0.25, 0.50, 0.75\}$ . Conversely, the occurrence of coefficients significant at the 1% level is much less frequent at the tails, which likely is a consequence of few observations in our data sets, as discussed in Section 5.5. We have 103 log return observations to base the model on, implying on average  $\sim 10$  exceedances and non-exceedances for the cross-models for  $\tau \in \{0.10, 0.90\}$ . From Figure 8, it is clear that the tails not only have few occurrences of coefficients significant at the 1% level, but at the 5% and 10% levels as well.

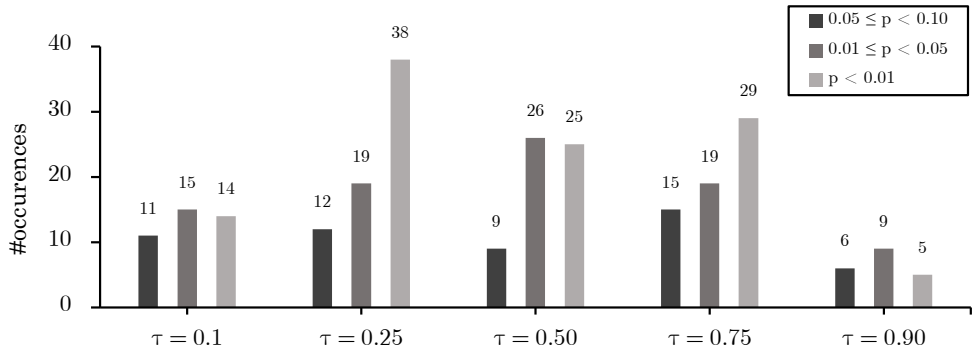


Figure 8: Aggregated results. Distribution of  $p$ -values for the coefficients of the five cross-models.

Examining the significance of the coefficients over the submodels, we observe from the rightmost column of Table 7 that significances at the 1% level are relatively evenly distributed. However, the submodels for horizons  $h \in \{1, 8, 12\}$  compare negatively to the others. In particular, for horizon  $h = 1$ , only five coefficients are significant at the 1% level. This is somewhat counterintuitive, as this submodel is the only one that is allowed to select its covariates from the entire search space. As mentioned in the previous section, a possible explanation might be the presence of noise in the salmon price over the very short term. On the other hand, horizon  $h = 2$  stands out as the one with the highest number of coefficients with significance at the 1% level. When examining this submodel more thoroughly, focusing on the occurrence of coefficients significant at the 5% and 10% levels from Table 15 in Appendix A.3, we observe that the submodel has a higher number of coefficients at these significance levels than the average submodel. This is also the case for the submodels for horizons  $h \in \{3, 4\}$ , which, indeed, indicates that the larger variable

search space available for the shorter horizons increase their ability to predict the variation in the salmon price distribution.

### Model Specification Test Results

Here, the results from the Kupiec, Christoffersen, and QRESET tests are presented. See Section 5.6 for the motivation behind including these tests, and a description of how they are implemented. The complete test results, together with coverage rates for the model, are given in Table 19 in Appendix A.4. In line with the previously presented results, we only include the aggregated test results here, as seen in Table 8 and Table 9.

Table 8: Aggregated specification test results for the cross-models. The coverage rate displays the average for each cross-model, while the numbers assigned to the tests correspond to the number of quantile regressions in the cross-model that pass the test under consideration. That is, their test statistics yield no significance for  $p < 0.1$ . Thus,  $optimal\ score = |\mathcal{H}| = 12$ .

Quantile	Average coverage rate	Kupiec test	Christoffersen test	QRESET(2)	QRESET(3)	QRESET(4)
$\tau = 0.10$	0.10	10	11	12	10	12
$\tau = 0.25$	0.25	12	5	12	0	1
$\tau = 0.50$	0.50	12	1	10	0	3
$\tau = 0.75$	0.74	12	5	2	2	5
$\tau = 0.90$	0.89	12	4	11	12	12
Average <sup>1</sup>	–	97%	43%	78%	40%	55%

1) The proportion of quantile regressions that pass the tests for the overall model.

Table 9: Aggregated specification test results for the submodels. The coverage rate displays the average for each submodel over all quantile regressions. The numbers assigned to the tests correspond to the number of quantile regressions that, for each submodel, pass the test under consideration. That is, their test statistics yield no significance for  $p < 0.1$ . Thus,  $optimal\ score = |\mathcal{T}| = 5$ .

Horizon	Average coverage rate	Kupiec test	Christoffersen test	QRESET(2)	QRESET(3)	QRESET(4)
$h = 1$	0.49	5	3	4	3	4
$h = 2$	0.51	5	4	4	2	3
$h = 3$	0.51	5	2	5	3	3
$h = 4$	0.50	5	3	4	1	2
$h = 5$	0.51	5	2	5	2	3
$h = 6$	0.50	4	2	4	2	4
$h = 7$	0.47	5	2	4	2	3
$h = 8$	0.49	5	3	4	2	2
$h = 9$	0.49	5	2	4	2	3
$h = 10$	0.49	4	1	2	1	2
$h = 11$	0.49	5	1	4	2	2
$h = 12$	0.50	5	1	3	2	2
Average <sup>1</sup>	–	97%	43%	78%	40%	55%

1) The proportion of quantile regressions that pass the tests for the overall model.

### Coverage Rate and the Unconditional Test of Kupiec

Generally, the model produces satisfactory coverage rates, especially for exceedances at central quantiles. By considering Table 8, it is clear that the coverage rates for the cross-models are in agreement with the nominal coverage rates that are to be fitted. For the central quantiles  $\tau \in \{0.25, 0.50, 0.75\}$ , the variation in the submodels' coverage rates is low. In other words, each predicted interval is fairly in line with the corresponding quantile, irrespective of the submodel under consideration. However, the average coverage rates are somewhat misleading for the cross-models for quantiles  $\tau \in \{0.10, 0.90\}$ , as the coverage rates vary considerably across the submodels. Evident from Table 19 in Appendix A.4, for quantile  $\tau = 0.10$ , the coverage rates range from 0.049 to 0.155, and it is merely a coincidence that the average coverage rate equals 0.10 for the corresponding cross-model. For the  $\tau = 0.90$  quantile, the coverage rates vary from 0.854 to 0.932. To explain the results

for the cross-models for these two quantiles, we refer to the discussion in [Section 5.5](#). The number of observations in the log returns is limited to 103, which undoubtedly inhibits the model's ability to capture the tail properties of the corresponding distributions.

Despite the relatively wide range of coverage rates produced by the submodels for the quantiles  $\tau \in \{0.10, 0.90\}$ , the results from the Kupiec test are generally insignificant. As seen from [Table 8](#), all cross-models achieve satisfactory unconditional coverage, except for two instances for quantile  $\tau = 0.10$ . Now, if we consider [Table 9](#), which presents the test results for the submodels, it is clear that the two instances of incorrect unconditional coverage occur for the submodels for horizons  $h \in \{6, 10\}$ .

Note that, although the submodels appear to provide the correct unconditional coverage, this is no more than a minimum requirement. This follows from the formulation of linear quantile regression, explained in [Section 5.2](#). Specifically, the objective function effectively weights the data observations to be distributed above and below the regression line in the same proportion as the modelled quantile of interest.

#### *The Conditional Test of Christoffersen*

Of all the 60 quantile regression models under consideration, the Christoffersen test is significant for more than a half at the 10% level, as depicted by the lowermost row in [Table 8](#). As evident by the squared log returns in [Figure 3c](#), the log returns of the NQSALMON have been subject to occasional volatility shifts in the past, with periods of raised volatility. The fact that the Christoffersen test proves significant across so many submodels justifies that the current model is unable to capture higher-order moment dynamics of the NQSALMON log returns.

By [Table 19](#) in the [Appendix A.4](#), it is apparent that the Christoffersen test is significant at the 1% level for 12 out of 60, i.e. 20%, of the quantiles. In all these cases, however, the Kupiec test passes. Thus, the unconditional coverage distribution does not seem to change over the sample. Correspondingly, the conditional coverage test fails because the independence test fails. According to [Alexander \(2008a, p. 360\)](#), this is a common case when backtesting e.g. value-at-risk (VaR) or other quantile models that are unable to predict conditional volatility.

The failure of the respective Christoffersen tests can at first be investigated through considering quantiles. Some particular quantiles, across all submodels, are more frequently yielding test violations. Specifically, the median quantile is significant at all horizons  $h \neq 1$ . Generally, we see that the predicted median log return tracks the realised log return quite well, however the pattern in exceedances is striking, as visualised by [Figure 9](#). The figure displays the median quantile predictions in the particular case of  $h = 10$ . An exceedance is signified by a grey bar in the figure. It is clear that the exceedances in the plot exhibit a high degree of clustering, with the  $\pi_{11}$  rate being 0.71. Equivalent patterns can be found in the same set of median quantiles. However, there are longer periods in which the predictions overestimate the returns, and longer periods in which they provide underestimations. This might hint at various return regimes, which possibly could be modelled with Markov switching ([Hamilton, 1989](#)). Another explanation might be more intuitive: In distributions without fat tails, one might encounter many observations around the mean, which should be close to the median as long as the distribution is not severely skewed. In this case, an exceedance will happen whenever an observation is just slightly below the median, which could happen quite often.

Along the second explanatory dimension are the various horizons. Generally, the number of quantiles that violate the Christoffersen test is higher for the longer horizons. This is particularly true for horizons  $h \geq 10$ , of which every submodel has only one quantile that passes the Christoffersen test at the 10% level, as evident from [Table 9](#). Also, by observing



Table 19 in Appendix A.4, there are no test results significant at the 1% level for horizons  $h \leq 3$ .

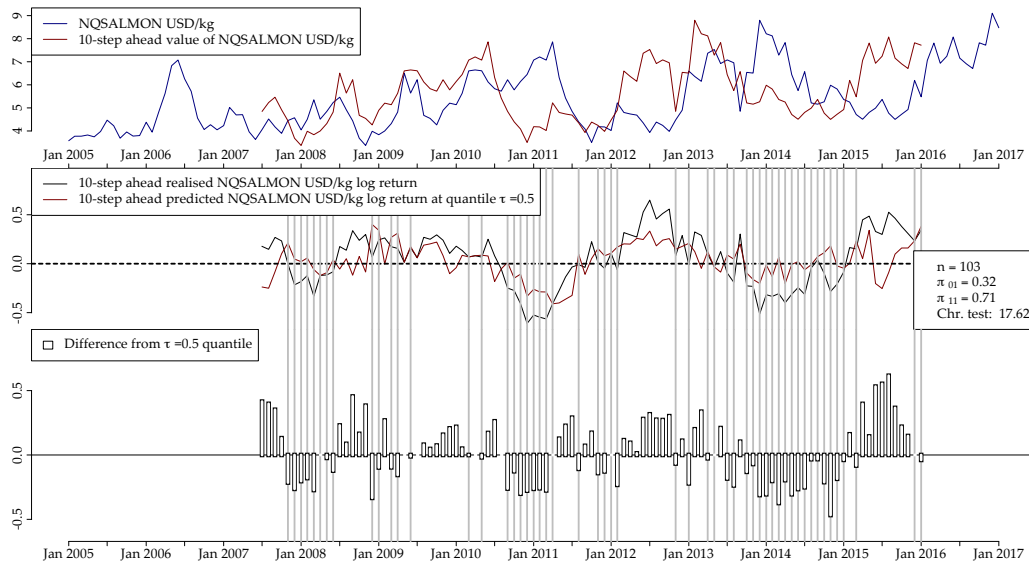


Figure 9: In-sample exceedance plot at horizon  $h = 10$  and quantile  $\tau = 0.50$ .<sup>30</sup>

The broad conclusion is that the submodels at longer horizons are worse in the sense that they yield clustering in the exceedances. However, this is to some extent unsurprising. In the value-at-risk (VaR) literature, which is somewhat analogue to quantile modelling in the context of risk management, much interest has been put in long-term VaR forecasts. Initial attempts have used the simple square-root rule to scale up the 1-step ahead VaR with the square root of the horizon of interest (Down, Blake, & Cairns, 2004), that is  $\text{VaR}(h) = \text{VaR}(1)\sqrt{h}$  (the analogue holds for quantile forecasting). However, this obviously does not account for volatility, which also generally is much more difficult to model many steps ahead (Christoffersen, 1998, p. 109). That being said, Down et al. (2004) relatively successively employ a standard quantile formula over a long-term horizon to estimate VaR, not considering volatilities along the timesteps. We employ a similar approach; however, in our case there is in fact clustering in the exceedances. This indicate that volatility, and potentially higher-order statistical moments, might be influential after all, particularly at longer horizons.

In order to better capture volatility fluctuations, we tried to include a volatility parameter as an additional covariate. This parameter was added on top of the eight covariates that already were selected by the GA for each submodel, resulting in nine covariates (plus intercept) in total. In particular, we used the log returns history of the NQSALMON to construct the time-varying volatility estimate, firstly by an exponentially weighted moving average under the RiskMetrics™ settings (J.P. Morgan, 1996), and secondly through the exponential GARCH (EGARCH) model by Nelson (1991). This resulted in slightly improved goodness-of-fit metrics, however, the results of the Christoffersen test remained more or less unchanged. Our preliminary efforts suggest that adding a volatility parameter is not a sufficient improvement. This is in line with the results of Oglend and Sikveland (2008b), which state that the predictive power in the salmon price volatility (with respect

<sup>30</sup>The first panel displays the NQSALMON and the 10-step ahead NQSALMON time series. The second panel displays the corresponding, realised 10-step ahead log return of the NQSALMON time series, as well as the predicted median ( $\tau = 0.50$ ) of the 10-step ahead log return distribution. The third panel display the absolute difference between the realised and the predicted median 10-step ahead log returns. Finally, each grey bar indicates an exceedance, in which the realised log return is lower than the predicted median.

to future volatility) is negligible over horizons of one month or above. Hence, they indicate that the benefit from including volatility parameters in our case, predicting the price 1- to 12-months ahead, is minimal. According to Christoffersen (1998), clustered exceedances are a consequence of the model failing to account for the dynamics of higher-order moments. Therefore, other ways of mitigating the effect of clustered exceedances, although not explored in this study, could be to incorporate time series of higher moments such as skewness and kurtosis, and eventually through regime switching as previously proposed.

The consequence of the failure of Christoffersen’s test is obvious: In the current setup, there will be some periods in which the model fails to predict the correct quantiles several months in a row. Furthermore, as we have seen, the median prediction as well as longer horizons are particular issues. In the context of risk management, the longer horizons’ respective quantile estimates should be judged carefully – as we have seen, they tend to overstate or understate the realised long-term returns. Hence, although our model may be correct on average, as indicated by the Kupiec test for unconditional coverage, it fails to properly incorporate times of volatile returns.

#### *Linear Specification: The QRESET Test*

We conduct three different QRESET( $p$ ) tests on the submodels, that is, the QRESET(2), QRESET(3), and QRESET(4). When synthesising the results from the three tests, we choose to put particular emphasis on the  $p \in \{2, 4\}$  tests. According to Otsu (2009), the QRESET(2) and QRESET(4) tests are the most powerful in determining polynomial and non-polynomial misspecification, respectively. Also, recalling the discussion from Section 5.6, the QRESET(2) test should be given the most attention, as this is comparatively more suitable for data sets with few observations. We also include the  $p = 3$  test to give additional support to the two others; however, this particular test does not have any comparative advantage in determining misspecifications.

If we put particular emphasis on the results from the QRESET(2), due to our small data set, we find that  $\sim 78\%$ , i.e., the majority, of all quantiles are correctly specified in terms of being approximated by quantile regressions that have a linear functional form. Nevertheless, the results are not entirely unanimous, as the  $p = 2$  test is significant for the majority of the submodels for the  $\tau = 0.75$  quantile. Therefore, we argue that the given QRESET test results indicate that some non-linear effects are unaccounted for in the overall model. However, for other quantiles  $\tau \in \{0.10, 0.25, 0.50, 0.90\}$ , linear quantile regression appears to be a sound functional form to estimate the multi-step ahead distributions.

For an overview of the cross-models that behave well under the linear quantile regression specification, according to the respective QRESET tests, we refer to Table 8. Considering the  $p \in \{2, 4\}$  tests, which, as discussed, are the most relevant, the tables can be summed up as follows:

- i)  $\tau = 0.10$ . The submodels are properly specified. The results do not indicate that the submodels would benefit from fitting neither higher-order polynomials nor other non-linear combinations, as indicated by QRESET(2) and QRESET(4), respectively.
- ii)  $\tau \in \{0.25, 0.50\}$ . The results indicate no need to fit higher-order polynomials, given by QRESET(2), but they suggest the use of other non-linear combinations, suggested by QRESET(4).
- iii)  $\tau = 0.75$ . Most submodels are misspecified according to both QRESET(2) and QRESET(4). These submodels might benefit from fitting higher-order polynomials or other non-linear combinations.
- iv)  $\tau = 0.90$ . All submodels are properly specified, apart from one. The results do not indicate that the submodels would benefit from fitting nei-



ther higher-order polynomials nor other non-linear combinations, as indicated by QRESET(2) and QRESET(4), respectively.

As the above results indicate, the cross-model for quantile  $\tau = 0.10$  is the only one with no types of linear misspecifications. Overall, according to the  $p = 2$  test, the cross-models for quantiles  $\tau \in \{0.10, 0.25, 0.50, 0.90\}$  are correctly specified, while the  $\tau = 0.75$  cross-model is possibly incorrectly specified. This is also supported by the results of the QRESET(3) and QRESET(4) for  $\tau = 0.75$ .

Evident from [Table 9](#), there is not a very consistent pattern regarding which submodels that generally yield correctly specified linear quantile regressions. If anything, we observe a slightly higher occurrence of linear misspecifications for longer horizons. As discussed in the previous section, the search space of candidate predictors generally shrinks with increasing horizons. Thus, at longer horizons, the variable selection procedure is likely to output subsets of predictors that provide a worse fit to the data. Ostensibly, such lesser-fitting subsets may yield regression model misbehaviour – as observed in terms of unexplained non-linear effects indicated by the QRESET tests.

As with the RESET, a drawback with the QRESET is that it provides no clear direction on how to proceed if the model is rejected ([Wooldridge, 2011](#), p. 304). That being said, there are many possible approaches in which to accommodate possible non-linear effects in the covariates of the quantile regressions for the above, failing  $\tau = 0.75$  quantile. Firstly, the regressions of interest might be multiplicatively non-linear, in which a sensible solution could be to transform the data into logarithms ([Brooks, 2008](#), p. 175). However, in our case, the data is already stationarised through logarithmic differencing, as thoroughly discussed in [Section 4.5](#). Thus, another logarithmic transformation is probably not sufficient. Another approach might be to introduce a polynomial formulation of the regression equation, as indicated by the relatively frequent significance of the QRESET(2) test for the  $\tau = 0.75$  quantile.

Going beyond quantile regression, one might consider to specify and fit other parametric non-linear models ([Joshua et al., 2013](#)). With reference to the discussion in [Section 5.5](#), sufficient *a priori* knowledge is needed to specify a functional form for the conditional quantile – however, we lack such knowledge in our case. Later efforts could incorporate more complex, non-linear models such as neural networks or support vector machines (support vector machine (SVM)s) ([Bishop, 2006](#), Chapters 5 and 7); nonetheless, the difficulty of such approaches lies in their often unintuitive interpretation. The advantage of quantile regression in particular is the ability to read out and analyse the coefficients, and perform subsequent sensitivity analyses, in a straightforward way.

In total, we argue that a linear quantile regression still might be considered a reasonable functional form for approximating the density of the log returns of the salmon price. That being said, the QRESET test statistics of this study indicate that more work can be done with respect to modelling the quantiles in a more accurate way.

#### 6.4 KEY EXPLANATORY VARIABLES

In total, the twelve submodels utilise 17 of the 25 variables available in the database, as can be seen from [Table 11](#). Of the eight variables that are not utilised in the submodels, five represent consumption of salmon. Thus, none of the consumption-related variables from the database are included in the model. The last three variables that have not been utilised are harvest volumes for Norway, global harvest volumes exclusive of Norway, and the price of trout. Moreover, across the twelve models, time series for standing biomass of both salmon and trout, feed consumption, the price of alternative proteins, and the currency pair USD/EUR are generally assigned the largest coefficients and greatest significances. See [Tables 15-18](#) in [Appendix A.3](#) for a complete overview of the covariates' coefficients and signifi-

cances across the modelling horizons. Also, note that, due to the log transformation of all time series, as discussed in [Chapter 4.5](#), all results presented here refer to the relationship between the *log returns* of the covariates and the NQSALMON.

As discussed previously in this study, the variable selection is less constrained by the lag structures for shorter horizons relative to longer horizons. Consequently, the shorter horizons tend to return the best goodness-of-fit when predicting the salmon price, as discussed in [Section 6.2](#). Also, they appear to exhibit the highest coefficient significances, as elaborated on in [Section 6.3](#). Hence, when pointing out the most important variables for predicting the NQSALMON, the shorter horizons are probably the most insightful. In particular, we demonstrate in the previous sections that the submodels for horizons  $h \in \{2, 3, 4\}$  yield best performance, both with respect to goodness-of-fit and statistical significance of their covariates. However, we also suggest that horizon  $h = 1$  provided limited explanatory power for the salmon price, possibly due to short term noise in the log returns. Accordingly, we now proceed with an analysis of the explanatory variables employed by the submodels for  $h \in \{2, 3, 4\}$ .

By looking at [Tables 15-18](#), among the most important covariates over these three horizons, in terms of significance and coefficient sizes, are the prices of alternative proteins, standing biomass of salmon, sea temperatures, and feed consumption. The lags and time windows employed for these covariates are rather consistent over the three horizons. Prices of alternative proteins employ 6 and 9 months lags and MoM time windows, standing biomass of salmon employs 3 and 6 months lag and MoM time window, and sea temperature employs 6 months lag and MoM time window.<sup>31</sup> Feed consumption is less consistent in its appearance, employing 3-4 months lag, as well as both MoM and YoY time windows.

Holding on to the most important covariates for the three horizons  $h \in \{2, 3, 4\}$ , and thereby the most important covariates for predicting the NQSALMON, it is evident from [Table 11](#) on p. 55 that their regression coefficient signs deviate considerably from expectations. At first, we know that prices of alternative proteins exhibit positive correlation with the NQSALMON, as by [Table 2](#). Therefore, based on direct effects alone, we would expect these covariates to have positive coefficients. However, of their 24 appearances over the twelve horizons, there is only one instance of a positive coefficient. Hence, our model suggests that increasing prices of alternative proteins do *not* give support to higher salmon prices, which contradicts an assumption that is widely accepted by the industry. Moreover, standing biomass of salmon has positive coefficient signs for 12 out of 14 appearances. This implies that increased future supply of salmon is expected to impose upwards price pressure, which is highly unlikely. An unexpected relationship is observed for sea temperatures as well, with the coefficient sign being positive for three out of five appearances. This indicates that increased temperatures, which speed up the growth rate of the fish and thereby increase the production rate, contribute to higher prices of salmon. Again, this is highly unlikely in reality. However, feed consumption exhibits a strictly negative relationship with the NQSALMON, which is in line with our and the industry's assumptions. Overall, of the 17 covariates employed for all submodels, only three of them have consistent relationships with the NQSALMON that comply with our assumptions.

An explanation of the unexpected coefficient signs might be the presence of multicollinearity effects in the submodels. According to [Kutner et al. \(2005\)](#), multicollinearity is associated with information overlap among the covariates. In turn, this makes it nearly impossible to distinguish between the individual contributions of the covariates. However, in general, this does not inhibit the overall model's ability to obtain a good fit ([Kutner et al., 2005](#)). In particular, if the coefficients appear stable both in terms of sign and magnitude, it is likely that multicollinearity effects have limited impact on the overall model fit. As seen

---

<sup>31</sup>Time series for the price of alternative proteins employ MoM time windows by design, i.e. they are not allowed to use YoY time windows.

from Table 11, for the coefficient signs, and Tables 15-18, for the coefficient magnitudes, the calculated regression coefficients appear relatively stable across all horizons. This indicates a proper overall fit judging from multicollinearity effects alone.

To further investigate the potential presence of multicollinearity in the submodels, and the implications for the interpretation of the regression coefficients, we build cross-correlation tables for the covariates of all the submodels. The cross-correlation measure does not enable us to confirm, nor disregard, the presence of multicollinearity in the submodels. However, high cross-correlation between the covariates is certainly indicative of the presence of multicollinearity effects in the submodels. Table 20 and Table 21 in Appendix A.5 shows cross-correlations between the covariates employed over the twelve horizons. Here, the respective log returns are lagged and transformed according to each covariate’s lag and time window size, as chosen by the GA. Also, for illustrative purposes, we include the results for horizon  $h = 1$  in Table 10 below.

Table 10: Cross-correlation between log returns for horizon  $h = 1$ .

TIME SERIES, UNIT	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$
$\beta_1$ NQSALMON, USD/kg (lag 1, MoM)	1.00							
$\beta_2$ Standing biomass, Tonnes (Norway) (lag 3, MoM)	-0.03	1.00						
$\beta_3$ Feed consumption, Tonnes (Norway) (lag 2, YoY)	-0.11	0.11	1.00					
$\beta_4$ Harvest volume of trout, Tonnes (Norway) (lag 9, MoM)	0.04	-0.15	-0.03	1.00				
$\beta_5$ Sea lice occurrence, #Lice/fish (Norway) (lag 12, MoM)	-0.11	-0.25	-0.08	-0.01	1.00			
$\beta_6$ Sea temperature, Degrees celsius (Norway) (lag 6, MoM)	0.09	0.91	0.04	-0.19	-0.26	1.00		
$\beta_7$ Shrimp price, US cents/pound (lag 6, MoM)	0.16	0.04	-0.19	0.16	-0.10	0.06	1.00	
$\beta_8$ Average harvest weight, kg (Norway) (lag 1, YoY)	-0.10	-0.06	-0.13	0.02	-0.05	-0.03	0.03	1.00

In general, the figures in the cross-correlation tables are favourable, with some exceptions. All the horizons have at least one instance where the correlation between two covariates is greater than or equal to 0.34. Here, naturally, variables that are to provide the same type of information, such as the supply of salmon from Norwegian producers, or the supply of trout, tend to be most frequently represented. For instance, we observe that sea temperatures in Norwegian waters are closely correlated with the standing biomass for Norwegian producers (0.91). Also, we note that standing biomass of trout, measured by the number of individuals, is closely correlated with standing biomass of trout measured in terms of tonnes (0.76 and above). The two time series for standing biomass of trout are together employed by seven out of twelve submodels, and almost exclusively with opposite coefficient signs. This indicates that there is no use in trying to interpret the proposed coefficients for the standing biomass of trout.

One could incorporate various measures to obtain more interpretable coefficients. Firstly, it is likely that our model would benefit from removing some of the variables from the database, based on the criterion of reducing the cross-correlation between the covariates employed by a given submodel. In turn, this is likely to reduce the alleged effect of multicollinearity in the submodels. Another solution is to constrain the variable selection procedure to only select covariates with low cross-correlation.<sup>32</sup> Given that this suffice for removing multicollinearity effects in the submodels, this approach is probably more appropriate compared to the aforementioned alternative of removing covariates based on a cross-correlation table. If we were to remove explanatory variables from the database based only on inspection of cross-correlation tables, we would not be able to effectively measure the impact of the removals on the submodels’ predictive power. The genetic algorithm, on the other hand, is able to run hundreds of iterations to ensure that both the cross-correlation constraint and the predictive power of the submodels are taken into account when disre-

<sup>32</sup>Optimally, we would utilise a rule that disregards covariates based on multicollinearity effects in the submodels, and not direct cross-correlation effects. However, unlike for ordinary least squares regression, where a straightforward implementation of the variance inflation factor could have been used, we have not been able to identify its quantile regression counterpart.

garding the covariates. A third solution is to aggregate groups of intercorrelated variables, effectively reducing the number of covariates to a smaller set of uncorrelated variables. Here, principal component analysis (PCA) could have been used, as described in (Jolliffe, 2002). However, this particular study seeks to identify which exogenous variables that affect the price of salmon, and not merely components of them. Hence, the use of PCA is deemed impractical for our purpose.

Table 11: Overview of the covariates, with corresponding lag, time window, and coefficient sign, employed for each submodel. Empty cells for a particular submodel (for some horizon  $h$ ) imply that use of the variable is not allowed due to lag constraints.

TIME SERIES, UNIT	LAG <sup>1</sup> , TIME WINDOW												COEFFICIENT <sup>2</sup>		
	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$	$h = 7$	$h = 8$	$h = 9$	$h = 10$	$h = 11$	$h = 12$	Hypothesis	#+	#-
NQSALMON, USD/kg	1, 1	2, 1											+	0	2
Standing biomass, #Individuals (Norway)	-	-	3, 1	9, 12	-	9, 1	9, 1	9, 1	9, 1				-	6	0
Standing biomass, Tonnes (Norway)	3, 1	-	3, 1	6, 1	6, 1	9, 12	9, 1	9, 1	9, 1				-	6	2
Feed consumption, Tonnes (Norway)	2, 12	3, 1	3, 1	4, 12									-	0	4
Smolt release, #Individuals (Norway)	-	-	-	-	-	-	17, 1	17, 1	-	15, 12	16, 1	15, 12	-	1	4
Harvest volume, Tonnes (Norway)	-												-	0	0
Standing biomass of trout, #Individuals (Norway)	-	-	6, 12	9, 12	6, 12	6, 12	9, 12	9, 12	9, 12	12, 12	12, 12	12, 12	-	0	10
Standing biomass of trout, Tonnes (Norway)	-	9, 12	-	-	6, 12	-	9, 12	9, 12	9, 12	12, 12	12, 12	12, 12	-	7	1
Harvest volume of trout, Tonnes (Norway)	9, 1	9, 12	9, 12	12, 12	-	12, 12	-	-	-	-	-	-	-	5	0
Sea lice occurrence, #Lice/fish (Norway)	12, 1	-	-	-	12, 1	12, 1	-	-	-	12, 1	12, 12	12, 12	+	2	4
Sea lice treatments, % of fish being treated (Norway)	-	-	-	-	-	-	12, 12	12, 12	12, 12	12, 12	12, 12	12, 12	+	0	6
Sea temperature, Degrees celsius (Norway)	6, 1	6, 1	6, 1	-	6, 12	6, 1							-	3	2
Meat price index, Index	-	6, 1	-	6, 1	9, 1	6, 1	-	-	-	12, 1	-	-	+	1	4
Poultry index, Index	-	-	-	-	6, 1	-	-	-	-	-	-	-	+	0	1
Beef price, US cents/pound	-	-	6, 1	6, 1	6, 1	9, 1	9, 1	9, 1	9, 1	12, 1	12, 1	12, 1	+	0	10
Shrimp price, US cents/pound	6, 1	9, 1	-	9, 1	-	-	12, 1	12, 1	12, 1	-	12, 1	12, 1	+	0	8
Currency pair, USD/EUR	-	6, 1	9, 1	-	-	-	-	-	15, 1	12, 1	12, 1	12, 1	-	5	1
Trout price, NOK/kg (Norway)	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0
Average harvest weight, kg (Norway)	1, 12												+	0	1
Consumption of Atl. Salmon, 1000 Tonnes (EU)	-	-	-										-	0	0
Consumption of Atl. Salmon, 1000 Tonnes (US)	-	-	-										-	0	0
Consumption of Atl. Salmon, 1000 Tonnes (Russia)	-	-	-										-	0	0
Consumption of Atl. Salmon, 1000 Tonnes (Japan)	-	-	-										-	0	0
Consumption of Atl. Salmon, 1000 Tonnes (Emerging markets)	-	-	-										-	0	0
Harvest volume, 1000 Tonnes (Global excl. Norway)	-	-	-	-	-	-							-	0	0

1) All lags are denoted relative to the horizon being forecasted. To get the number of lags relative to the forecasters point of view, subtract by  $h$  months for any horizon  $h$ .

2) Expected vs. employed coefficient sign. The two columns on the right hand side show the frequency of a particular coefficient sign, measured over all horizons.

## SCENARIO ANALYSIS

An important area within risk management is the analysis and prediction of extreme events that subsequently may cause significant price movements. Due to the stability of the regression covariate coefficients' signs and magnitude over the twelve horizons, and the absence of correlation between most of the covariates, we regard the proposed model as sufficiently robust to carry out a scenario analysis. Here, we provide an example of a scenario analysis by creating a conditional distribution of the salmon price. In particular, we investigate the impact on the NQSALMON from changes in *Feed consumption, Tonnes (Norway)*. This explanatory variable is chosen based on the following criteria: 1) It exhibits limited correlation with other covariates, 2) its coefficients are relatively significant and large in magnitude when feed consumption is present as a covariate in the quantile regressions, 3) its coefficient signs are in line with the industry's assumptions, and 4) it is utilised by the model across a wide range of submodels.

In the analysis, we create a base scenario using the information available at February 2016 for all explanatory variables other than the feed consumption. Moreover, we employ the submodel for horizon  $h = 4$ , as this is among the best performing submodels both in terms of goodness-of-fit and specification tests. Hence, as we predict the price 4 months ahead, we are effectively forecasting the price for June 2016. Furthermore, in this submodel, feed consumption is represented with a YoY time window and a 4-month lag relative to the horizon being forecasted. We let the log returns for feed consumption take on values ranging from their minimum to maximum based on historical values in our sample data set, i.e. from -0.3179 to 0.3883.<sup>33</sup> Now, the observation for feed consumption 16 months (4-month lag and 12-month time window) prior to June 2016 was 86 702 tonnes. Due to the given sample minimum and maximum log returns, we let the 4-month lagged value, prior to June 2016, for feed consumption range from 63 090 tonnes to 127 839 tonnes. [Table 12](#) provides an overview of the input data used in the scenario analysis for the different covariates applied by the submodel for  $h = 4$ . Also, note that the price of salmon was USD 7.06 as of February 2016.

Table 12: Data used in the scenario analysis.

VARIABLE, UNIT <sup>1</sup>	PAST OBSERVATION <sup>2</sup>	CURRENT OBSERVATION <sup>3</sup>	LOG RETURN
Standing biomass, #Individuals (Norway) (lag 9, YoY)	418 788	415 558	-0.0077
Standing biomass, Tonnes (Norway) (lag 6, MoM)	724 344	721 597	-0.0038
Feed consumption, Tonnes (Norway) (lag 4, YoY)	86 702	63 090 – 127 839	-0.3179 – 0.3883
Standing biomass of trout, #Individuals (Norway) (lag 9, YoY)	25 592	26 004	-0.0160
Harvest volume of trout, Tonnes (Norway) (lag 12, YoY)	5 354	6 706	0.2252
Meat price index, Index (lag 6, MoM)	154.6	150.0	-0.0300
Beef price, US cents/pound (lag 6, MoM)	184.6	177.1	-0.0414
Shrimp price, US cents/pound (lag 9, MoM)	15.9	15.9	0.0000

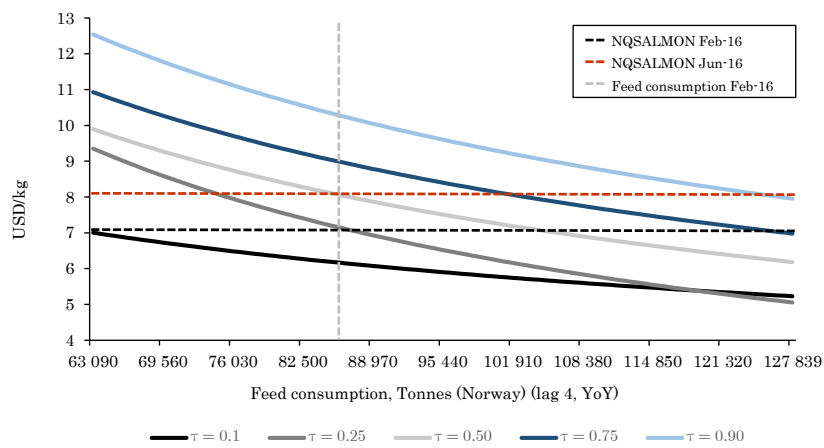
1) All lags are denoted relative to the price point being forecasted, that is, 4-months ahead. To get the number of lags relative to the forecaster's point of view, subtract by 4 months.

2) Past observation adjusted for lag structure *and* time windows. Example: For feed consumption, the observation corresponds to that of 16 months (4 + 12) prior to the point being forecasted.

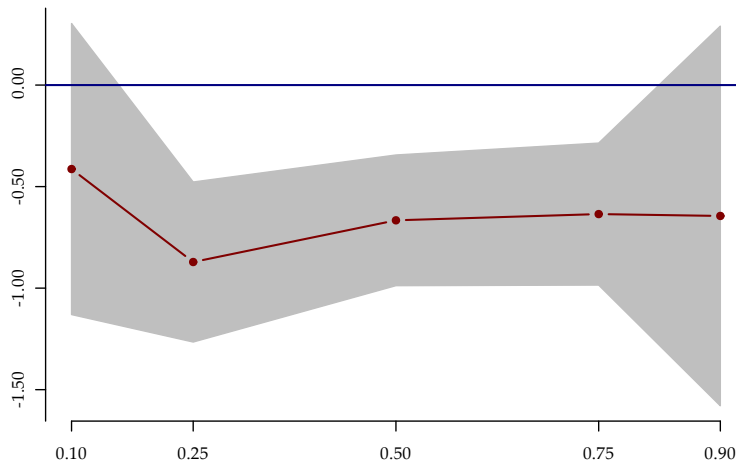
3) Current observation adjusted for lag structure. Example: For feed consumption, the observation corresponds to that of 4 months prior to the point being forecasted, i.e. today's observation from the forecaster's perspective.

<sup>33</sup>These correspond to arithmetic returns of -27.2% and 47.4%, respectively.

Figure 10 displays the results from the scenario analysis, and a corresponding coefficient sensitivity plot for feed consumption for all quantiles  $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$ . Focusing on Figure 10a, it is clear that the price of salmon has an inverse relationship with feed consumption for all the quantiles of the distribution. This is in line with the expectations, as increased feed consumption year-over-year generally is expected to increase the harvest volumes year-over-year.<sup>34</sup> Accordingly, the marketed volumes increase, which implies that the price of salmon should decrease given that the demand and the other regression covariates remain unchanged. Also, recognise that the realised price for June 2016, as depicted by the red line in the figure, crosses the line representing the median price at the exact same point as it crosses the line for the February 2016 feed consumption. This is the observation of feed consumption 4-month lagged *relative* to the June 2016 spot price, which is to be predicted. Clearly, the forecasted median quantile is aligned with the realised price. That being said, no general conclusions with respect to the predicted price distribution can be made only based on this single data point.



(a) The price distribution for the NQSALMON conditional on changes in feed consumption.



(b) Coefficient sensitivity across the quantiles of the predicted log return distribution of the NQSALMON.<sup>35</sup>

Figure 10: Results from the scenario analysis for *Feed consumption, Tonnes (Norway)*, based on the submodel for horizon  $h = 4$ , along with the coefficients employed for feed consumption.

<sup>34</sup>See Figure 11 in Appendix A.1 and Marine Harvest (2016) for the relationship between harvest volumes and the price of salmon, and Chapter 3 for the relationship between feed consumption and harvest volumes.

<sup>35</sup>The red line denotes coefficient values, while the grey band expresses the adjusted standard errors of the coefficient estimates. The band is calculated with the bootstrap by Parzen, Wei, and Ying (1994), with 1000 replications.



It should be noted that the regression line for the quantile  $\tau = 0.25$  crosses the line for the quantile  $\tau = 0.10$  in [Figure 10a](#). This can, obviously, not be explained by fundamental relationships. However, as elaborated upon in the previous chapter, the submodels struggle to correctly capture the properties of the tails, likely due to few observations in the data sets. This is also reflected by the significance of the coefficients for feed consumption, as can be seen from [Table 16](#) in [Appendix A.3](#). Here, for horizon  $h = 4$ , the coefficients for feed consumption are significant at the 1% level only for the quantiles  $\tau \in \{0.25, 0.50, 0.75\}$ . Hence, we are not able to create proper scenario analyses of the price of salmon, conditional on changes in feed consumption, for the tails of the price distribution.

We observe from [Figure 10b](#) that the coefficients for *Feed consumption, Tonnes (Norway)* generally are negative across all quantiles. Only at the lower- and uppermost quantiles are the coefficients' standard errors crossing the zero line. Therefore, we can conclude that all quantiles are skewed negatively when there are positive shocks to the feed consumption. Ignoring the coefficient for the  $\tau = 0.10$  quantile, which is insignificant due to its high  $p$ -value, we observe a trend in which the coefficient values slightly increase with quantiles. Therefore, sufficiently large positive changes in the feed consumption tend to skew the lower quantiles, e.g.  $\tau = 0.25$ , into more negative returns, and thus we should observe correspondingly higher price drops. *Mutatis mutandis*, the opposite holds for declines in feed consumption – in such a case, the  $\tau = 0.25$  quantile will be driven in positive direction, suggesting relatively less frequent observations of price drops.

There are several possible explanations of the fact that *Feed consumption, Tonnes (Norway)* to a larger extent drives lower-quantile realisations of salmon price returns. As a direct effect, feed consumption is an indicator of future harvest volumes. As salmon is a perishable product whose price is highly sensitive to the balance between supply and consumer demand ([Marine Harvest, 2016](#)), increased future volumes will, all else equal, drive supply and reduce prices. However, this does not fully describe the observed impact on the lower  $\tau = 0.25$  quantile. As positive shocks to feed consumption will impose a negative skew on the predicted price distribution, one could derive that prices generally are close to consumers' willingness to pay. Thus, only a slight oversupply in the market will force retailers to dump the perishable product at significantly lower prices, an event which according to our lag structure will be incurred four months after the observed year-over-year increase in feed consumption. This will propagate directly to producers who are required to accept reduced prices in the spot market.

For producers, the above interpretations suggest that the combination of high spot prices and observed increases in feed consumption (on a market-wide level) might yield a higher chance of spot price drops in the relatively short term. This is a relevant result, particularly for larger producers. They could, for instance, time harvests and inventory to be able to increase output in the window of higher spot prices before a potential price drop (in, say, four months). Correspondingly, they could engage in futures hedging activities to mitigate the increased downside risk of the spot price. This also fits the conclusions of [Asche et al. \(2016b\)](#), who find that the futures market for salmon is immature. Furthermore, they conclude that futures lack a price discovery role, and that the spot price leads the futures prices of salmon. Combined with our findings, the downside risk might therefore not be reflected in the futures prices. As a result, futures might possibly be bought at cheaper prices. It should be noted that financial hedging is particularly relevant due to the farmers' long production cycles, inability to build up inventories, and corresponding lacking capability to significantly adjust short-term production volumes and spot market sales. Thus, futures hedging and price distribution insights, such as the sensitivity with respect to feed consumption, could prove highly useful.



## CONCLUSION

---

### 8.1 MOTIVATION, FINDINGS, AND REMARKS

The growth in output from salmon farming, together with increasing commercialisation and financialisation of the aquaculture industry, motivate prediction and increased understanding of the Atlantic salmon spot price. The need for financing, risk management such as hedging, and salmon spot trading increasingly require accurate models for long-term predictions of the price path and corresponding distributional properties. Furthermore, operational considerations, e.g. harvest timing, inventory management, and entries into new markets will need equivalent scrutiny of the development of the spot price of salmon. Participants in this commodity market will need two interlinked insights in order to succeed: Firstly, accurate expectations of the price, both in a short- and long-term perspective. This also entails an understanding of the uncertainty in such estimates. Secondly, awareness of what drivers are important for the salmon spot price across various time frames.

In this study, we choose the NQSALMON index, denoted in USD, to represent the spot price of Atlantic salmon, and make a prediction model of its log return distribution in a multi-step ahead framework. In particular, we model 1- to 12-month ahead returns – i.e., 12 horizons. Returns are used due to their favourable statistical properties in context of prediction, however, the conversion to corresponding prices is straightforward. To approximate this density, we employ linear quantile regressions across a pre-selected set of quantiles  $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$ . We utilise a set of 25 causally related variables (including the NQSALMON), such as average sea temperatures, standing biomass, and prices of alternative proteins. These are transformed into month-over-month (MoM) or year-over-year (YoY) log returns, and lagged in accordance with market participants' judgement of how and when they ought to impact the spot price. These combinations of monthly and yearly returns, together with a restricted lag structure, generate a large set of candidate predictors for the multi-step ahead density estimation. In context of quantile regression, we extend the current methodology on variable selection by employing a genetic algorithm (GA) for this purpose, in order to determine the best subsets of size  $U = 8$  of these candidate predictors. Each selected subset is fixed for a particular horizon, i.e., a submodel, and is used to construct regressions for each quantile  $\tau$ .

Generally, we find a good model fit in the sample in terms of explained variation across the quantiles. The GA is able to select quantile regressions whose pseudo-R-squared averaged over the quantiles of each horizon range from  $\sim 0.24 - 0.43$ . That being said, we see that the shorter horizons generally yield better goodness-of-fit scores. We attribute the lesser scores on longer horizons with two factors: To begin with, the availability of candidate predictors is larger for shorter horizons due to the aforementioned lag structure. Thus, the variable selection algorithm has a larger search space and more combinations to assess, possibly yielding better explanatory power. Also, log return dynamics such as its higher-order statistical moments, e.g. variance, skewness, and kurtosis, may yield difficulties in explaining the variation. These moments are generally hard to predict and obviously impact the (conditional) distribution (Christoffersen, 1998, p. 109).

In terms of the variable search, the GA shows promising properties and results. It quickly converges to near-optimal variable subsets across all  $h$ -step ahead horizons, usually after 30-50 iterations (i.e., generations), in line with previous experience by e.g. Frost and Molt (1998) who applied the same methodology for principal component regression (PCR). The variable selection problem is inherently combinatorial, with exhaustive search being computationally intractable. We find that the GA-aided variable selection exhibits

promising properties, and it should be given further attention in the context of quantile regression.

We find that the GA is able to select subsets whose corresponding regressions in many cases are correctly specified; however, the overall model has some slight misbehaviour. Firstly, a large share of the selected subsets yield covariates that prove significant in the quantile regressions. Most covariates are highly significant for central quantiles of the distribution, i.e. the 0.25, 0.50, and 0.75 quantiles. There is, however, no clear pattern with respect to coefficient significances on shorter versus longer horizons. Regarding unconditional coverage, the quantile regressions give satisfactory results under the Kupiec test in the sample, being able to "split" the log returns observations across the respective quantiles. However, under the Christoffersen test we find that these exceedances generally are not independent – they tend to exhibit clustering, particularly at longer horizons. We argue that this clustering might be due to aforementioned, uncaptured higher-order dynamics in the log returns of the NQSALMON. We try to mitigate the clustering in exceedances by incorporating a time-varying volatility of the NQSALMON log return in addition to the eight covariates, however this proves unsuccessful. Looking at the 10-, 11-, and 12-month ahead horizons, the clustering in exceedances is particularly present, with highly significant Christoffersen test statistics. For almost all horizons, the median quantile is particularly unfavourable. For this quantile, the regressions have larger stretches of exceedances and non-exceedances, which indicate regime-like dynamics in the NQSALMON. Therefore, long-term price predictions will either underestimate or overestimate the quantiles. Finally, we run a quantile regression equation specification error test (QRESET) to assess whether or not non-linear effects are present.  $\sim 78\%$  of the models pass the QRESET(2) test (i.e. insignificant at the 10% level), this particular test being the most reliable QRESET-variant due to our small sample size. Emphasising these test scores, we argue that linear quantile regression is a sound functional form to approximate multi-step ahead distribution of the salmon spot price.

There are some patterns in the selected explanatory variables. Specifically, 17 out of the 25 explanatory variables are present in the twelve submodels. However, no consumption-related figures are included. This indicates that variables for consumed volumes might lack predictive power for the salmon spot price. In general, the most common categories of selected variables are the prices of alternative proteins, standing biomass, sea temperature, and feed consumption. Even though they ought to have a causal relationship with respect to the price, they sometimes have deviating coefficient signs relative to the expectation. Such deviations might be a result of collinearity between variables and corresponding information overlaps (Kutner et al., 2005). This phenomenon can be a result of seasonality found in some explanatory variables. To mitigate collinearity, techniques such as PCA could be applied on the covariates; however, each resulting principal component derived from the original data might not have an intuitive interpretation. In this study, we choose to use a straightforward application of quantile regression because of its interpretability, which should be of importance for the industry participants.

## 8.2 FURTHER RESEARCH

The combination of variable selection and quantile modelling could certainly be improved with more fundamental time series. Even though we model a particular spot price index, the NQSALMON, with several data series only pertaining to Norway, salmon farming is a global industry. Correspondingly, we could add variables representing other geographical areas, such as biomass from the major producers in Chile and Scotland. This could, in turn, yield better goodness-of-fit.

A general-to-specific approach could also be utilised, more in line with Campos et al.

(2005). In particular, the same variable selection procedure with the described genetic algorithm (GA) could be run with a lesser-constrained search space, with e.g. variables lagged more freely. This could certainly yield interesting results; however, such an approach could be characterised as data mining (Hoover & Perez, 2000).

Due to the possibility of misspecification, indicated by clustered exceedances at some quantiles, one could consider the application of non-linear models. At the median, the exceedances seem to have a regime-like structure, thus e.g. Markov switching approaches as by Hamilton (1989) could be interesting. According to Christoffersen (1998), clustering in exceedances are a result of the model failing to account for the dynamics of higher-order statistical moments. Therefore, the incorporation of volatility, skewness, and kurtosis in e.g. a non-linear quantile regression could prove useful. Even though the QRESET(2) results generally favour the current linear quantile regression formulation, one might possibly apply inherently non-linear regression models, such as neural networks and support vector machines (SVMs), to yield even better results with respect to model specification.

We believe the novel approach with GA-assisted variable selection for quantile regression can be further investigated. For instance, it could be adapted to other domains, such as in density modelling of commodities, stock indices, etc. For multi-step ahead prediction, one could also try to run the variable selection algorithm for a fixed quantile, rather than for a fixed horizon – as has been done in this study. Also, more rigorous benchmarking studies of various variable selection algorithms for quantile regression could be considered.

In the end, the application of our methodology in the context of risk management should be straightforward. Correspondingly, more complex value-at-risk (VaR) insights could be derived.

## APPENDIX

## A.1 FACTORS AFFECTING THE SALMON PRICE

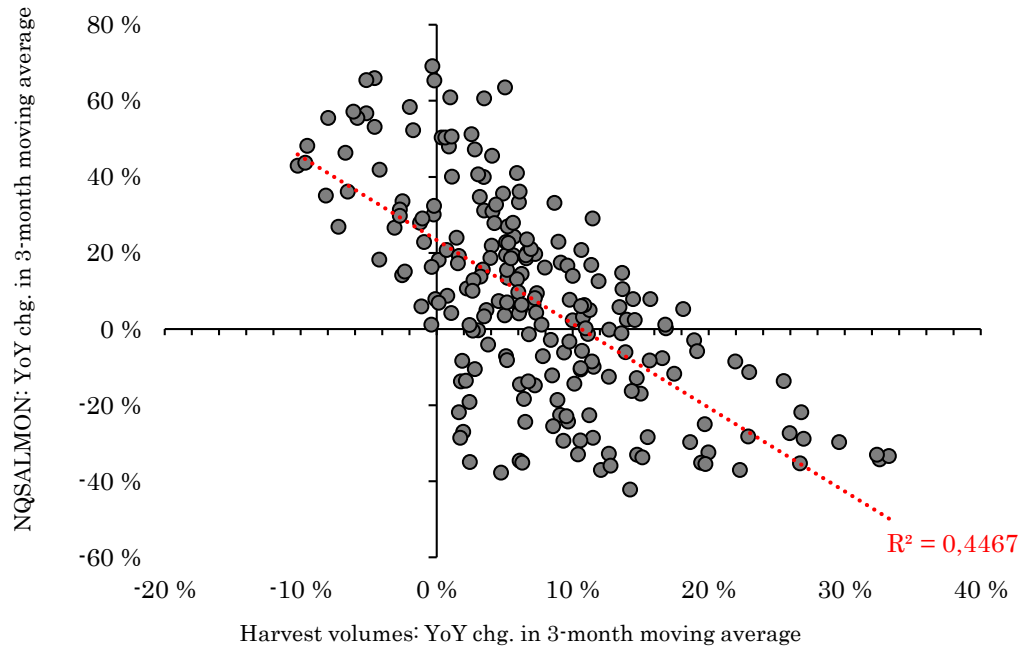


Figure 11: Changes in the NQSALMON vs. changes in global harvest volumes.

## A.2 DATA SERIES USED IN THE MODEL

### Time Series Overview

On the the following two pages, we provide an overview of all time series that have been used. The table includes the salmon price itself, NQSALMON, and all variables in the database.

Table 13: Complete overview of the database of time series. Lags and time windows are based on the [Chapter 3](#) discussion, and not the candidate predictor analysis that follows in [Chapter 4](#).

TIME SERIES, UNIT	DESCRIPTION	LAG(S) <sup>1</sup>	TIME WINDOW	DATA SOURCE	RESOLUTION <sup>2</sup>
NQSALMON, USD/kg	Volume-weighted average of weekly reported sales prices of fresh Atlantic Superior Salmon, head-on-gutted	1-2	MoM	The NASDAQ Group	Weekly
Standing biomass, #Individuals (Norway)	Sum of the biomass of large salmon (typically less than 3 months until harvest) in all cages in Norwegian waters	3,6,9	YoY	Directorate of Fisheries	Monthly
Standing biomass, Tonnes (Norway)	Sum of the biomass of large salmon (typically less than 3 months until harvest) in all cages in Norwegian waters	3,6,9	YoY	Directorate of Fisheries	Monthly
Feed consumption, Tonnes (Norway)	The amount of feed consumed by all the salmon in cages in Norwegian waters	2-4	YoY	Directorate of Fisheries	Monthly
Smolt release, #Individuals (Norway)	The transfer of smolt (salmon with weight of 60-100 grams) from land-based production facilities to cages in seawater	15-17	YoY	Directorate of Fisheries	Monthly
Harvest volume, Tonnes (Norway)	Salmon taken out from the cages and prepared for sale	1	YoY	Directorate of Fisheries	Monthly
Standing biomass of trout, #Individuals (Norway)	Sum of the biomass of trout in all cages in Norwegian waters	6,9,12	YoY	Directorate of Fisheries	Monthly
Standing biomass of trout, Tonnes (Norway)	Sum of the biomass of trout in all cages in Norwegian waters	6,9,12	YoY	Directorate of Fisheries	Monthly
Harvest volume of trout, Tonnes (Norway)	Trout taken out from the cages and prepared for sale	3,6,9,12	YoY	Directorate of Fisheries	Monthly
Sea lice occurrence, #Lice/fish (Norway)	Avg. number of sea lice per salmon	3,12	YoY	Lusedata	Weekly
Sea lice treatments, % of fish being treated (Norway)	The share of all salmon being treated for sea lice	3,12	YoY	Lusedata	Weekly
Sea temperature, Degrees celsius (Norway)	Avg. sea temperature in Norwegian waters	3,6	YoY	Lusedata	Weekly
Meat price index, Index	Index composed of four other indices: poultry, pig, bovine, and ovine	3,6,9,12	MoM	FAO	Monthly
Poultry index, Index	Index based on export values for broiler cuts (US, export price) and chicken (Brazil, export price)	3,6,9,12	MoM	FAO	Monthly
Beef price, US cents/pound	Beef, Australian and New Zealand 85% lean fores (US import price)	3,6,9,12	MoM	Quandl	Monthly
Shrimp price, US cents/pound	Mexican shrimp. Wholesale price New York	3,6,9,12	MoM	Quandl	Monthly
Currency pair, USD/EUR	-	6,9,12,15	MoM	Oanda	Daily
Trout price, NOK/kg (Norway)	Price of trout farmed in Norway	3,6,9,12	MoM	Seafood.no	Weekly
Average harvest weight, kg (Norway)	Harvest volume divided by the number of salmon harvested	1	YoY	Directorate of Fisheries	Monthly
Consumption of Atl. Salmon, 1000 Tonnes (EU)	Total consumption of Atlantic salmon in the EU	1-3	YoY	Marine Harvest Group	Monthly
Consumption of Atl. Salmon, 1000 Tonnes (US)	Total consumption of Atlantic salmon in the US	1-3	YoY	Marine Harvest Group	Monthly
Consumption of Atl. Salmon, 1000 Tonnes (Russia)	Total consumption of Atlantic salmon in Russia	1-3	YoY	Marine Harvest Group	Monthly
Consumption of Atl. Salmon, 1000 Tonnes (Japan)	Total consumption of Atlantic salmon in Japan	1-3	YoY	Marine Harvest Group	Monthly
Consumption of Atl. Salmon, 1000 Tonnes (Emerging markets)	Total consumption of Atlantic salmon in Emerging markets	1-3	YoY	Marine Harvest Group	Monthly
Harvest volume, 1000 Tonnes (Global excl. Norway)	Salmon taken out from the cages and prepared for sale	3,6	YoY	Marine Harvest Group	Monthly

1) Lags available for the GA to choose from. All lags are denoted in months.

2) Resolution as retrieved. All time series have been converted to monthly observations.

Arithmetic Returns

Next, descriptive statistics for the monthly *arithmetic* returns of the respective series are provided in Table 14. Note that arithmetic returns only are provided for reference; we use log returns for the modelling part.

Table 14: Descriptive statistics of the arithmetic returns of the data series, time period: Jul. 2007 – Jan. 2017.

Series name	GENERAL INFORMATION			DESCRIPTIVE STATISTICS						TESTS			Ljung-Box $\rho = 6$
	N	Mean %	#	Std. dev. %	Min %	Max %	Skew	Ex. kurt.	NQSALMON corr.	JB	DF	ADF	
$\Delta\%$ [NQSALMON, USD/kg]	115	1.51	12.75	30.22	35.19	0.53	0.39	1.00	6.19	-12.13	-7.46	8.05	
$\Delta\%$ [Standing biomass, #Individuals (Norway)]	115	0.45	5.77	-6.95	20.22	1.17	1.35	-0.19	34.90	-7.06	-9.32	42.13	
$\Delta\%$ [Standing biomass, Tonnes (Norway)]	115	0.67	4.85	-7.24	11.12	0.50	-0.89	-0.21	8.52	-4.02	-6.34	214.06	
$\Delta\%$ [Feed consumption, Tonnes (Norway)]	115	2.29	20.36	-29.01	46.50	0.51	-0.95	-0.20	9.38	-4.34	-4.82	251.85	
$\Delta\%$ [Small release, #Individuals (Norway)]	115	631.00	5913.88	-100.00	63443.79	10.56	109.61	0.06	59707.96	-10.63	-9.43	0.09	
$\Delta\%$ [Harvest volume, Tonnes (Norway)]	115	1.10	13.10	-34.12	33.46	0.27	-0.10	-0.23	1.46	-11.89	-9.17	22.72	
$\Delta\%$ [Standing biomass of trout, #Individuals (Norway)]	115	-0.34	5.18	-11.55	15.97	0.51	0.07	-0.12	5.04	-7.80	-7.41	34.47	
$\Delta\%$ [Standing biomass of trout, Tonnes (Norway)]	115	-0.29	5.95	-13.96	13.03	0.34	-0.37	-0.19	2.89	-6.34	-5.50	69.96	
$\Delta\%$ [Harvest volume of trout, Tonnes (Norway)]	115	2.79	25.61	-55.95	74.77	0.31	-0.20	-0.14	1.97	-12.45	-9.06	16.52	
$\Delta\%$ [Sea lice occurrence, #Lice/fish (Norway)]	115	8.03	44.25	-64.84	158.15	1.08	1.18	-0.27	29.28	-6.71	-5.73	63.53	
$\Delta\%$ [Sea lice treatments, % of fish being treated (Norway)]	115	13.13	84.88	-90.82	787.58	6.80	58.69	-0.10	17388.87	-12.64	-7.68	8.70	
$\Delta\%$ [Sea temperature, Degrees celsius (Norway)]	115	1.54	20.98	-27.83	62.14	0.64	-0.68	-0.15	10.07	-3.53	-7.82	256.57	
$\Delta\%$ [Meat price index, Index]	115	0.19	2.86	-8.68	6.16	-0.72	0.58	0.09	74.19	-6.05	-4.21	74.19	
$\Delta\%$ [Poultry index, Index]	115	0.04	3.93	-13.87	12.04	-0.19	2.18	0.03	23.37	-9.89	-5.46	13.08	
$\Delta\%$ [Beef price, US cents/pound]	115	0.46	4.85	-16.44	19.71	0.29	2.84	0.14	40.34	-7.86	-7.38	19.41	
$\Delta\%$ [Shrimp price, US cents/pound]	115	0.31	5.06	-21.59	20.19	0.70	7.71	-0.03	294.50	-7.85	-6.17	12.50	
$\Delta\%$ [Currency pair, USD/EUR]	115	0.23	2.48	-6.10	8.13	0.36	0.64	-0.25	4.48	-8.04	-6.61	10.45	
$\Delta\%$ [Trout price, NOK/kg (Norway)]	115	1.29	8.99	-24.10	26.97	0.20	0.41	0.30	1.60	-9.64	-8.50	10.82	
$\Delta\%$ [Average harvest weight, kg (Norway)]	115	-0.85	10.69	-100.00	25.59	-6.93	63.11	0.04	20001.90	-5.36	-4.76	0.88	
$\Delta\%$ [Consumption of Atl. Salmon, 1000 Tonnes (EU)]	115	0.12	14.70	-100.00	30.00	-2.71	17.19	-0.07	1556.82	-8.90	-7.28	17.21	
$\Delta\%$ [Consumption of Atl. Salmon, 1000 Tonnes (US)]	115	-0.26	13.46	-100.00	32.24	-3.24	24.67	-0.07	3117.28	-13.60	-7.85	13.06	
$\Delta\%$ [Consumption of Atl. Salmon, 1000 Tonnes (Russia)]	115	1.94	23.36	-100.00	56.76	-0.77	2.34	0.01	37.57	-10.47	-8.34	5.53	
$\Delta\%$ [Consumption of Atl. Salmon, 1000 Tonnes (Japan)]	115	1.57	20.72	-100.00	48.00	-0.99	4.10	0.12	99.15	-14.38	-9.74	15.21	
$\Delta\%$ [Consumption of Atl. Salmon, 1000 Tonnes (Emerging markets)]	115	0.43	13.05	-100.00	22.62	-3.95	28.67	0.00	4237.27	-9.77	-7.57	8.79	
$\Delta\%$ [Harvest volume, 1000 Tonnes (Global, excl. Norway)]	115	-0.15	11.62	-100.00	23.11	-5.49	45.78	-0.07	10620.19	-6.84	-4.83	0.87	

Critical values: JB:  $\chi^2_{6, \alpha=10\%} > 4.61$ ;  $\chi^2_{2, \alpha=5\%} > 5.99$ ;  $\chi^2_{2, \alpha=1\%} > 9.21$   
 DE, ADF:  $\tau_{\alpha=10\%} < -1.62$ ;  $\tau_{\alpha=5\%} < -1.95$ ;  $\tau_{\alpha=1\%} < -2.58$   
 Ljung-Box:  $\chi^2_{6, \alpha=10\%} > 10.64$ ;  $\chi^2_{6, \alpha=5\%} > 12.59$ ;  $\chi^2_{6, \alpha=1\%} > 16.81$

## Descriptive Figures

Here we provide descriptive figures for the time series that have been used in the modelling. In particular, plots of both the time series, corresponding log returns, and log returns histograms are included in the plots below.

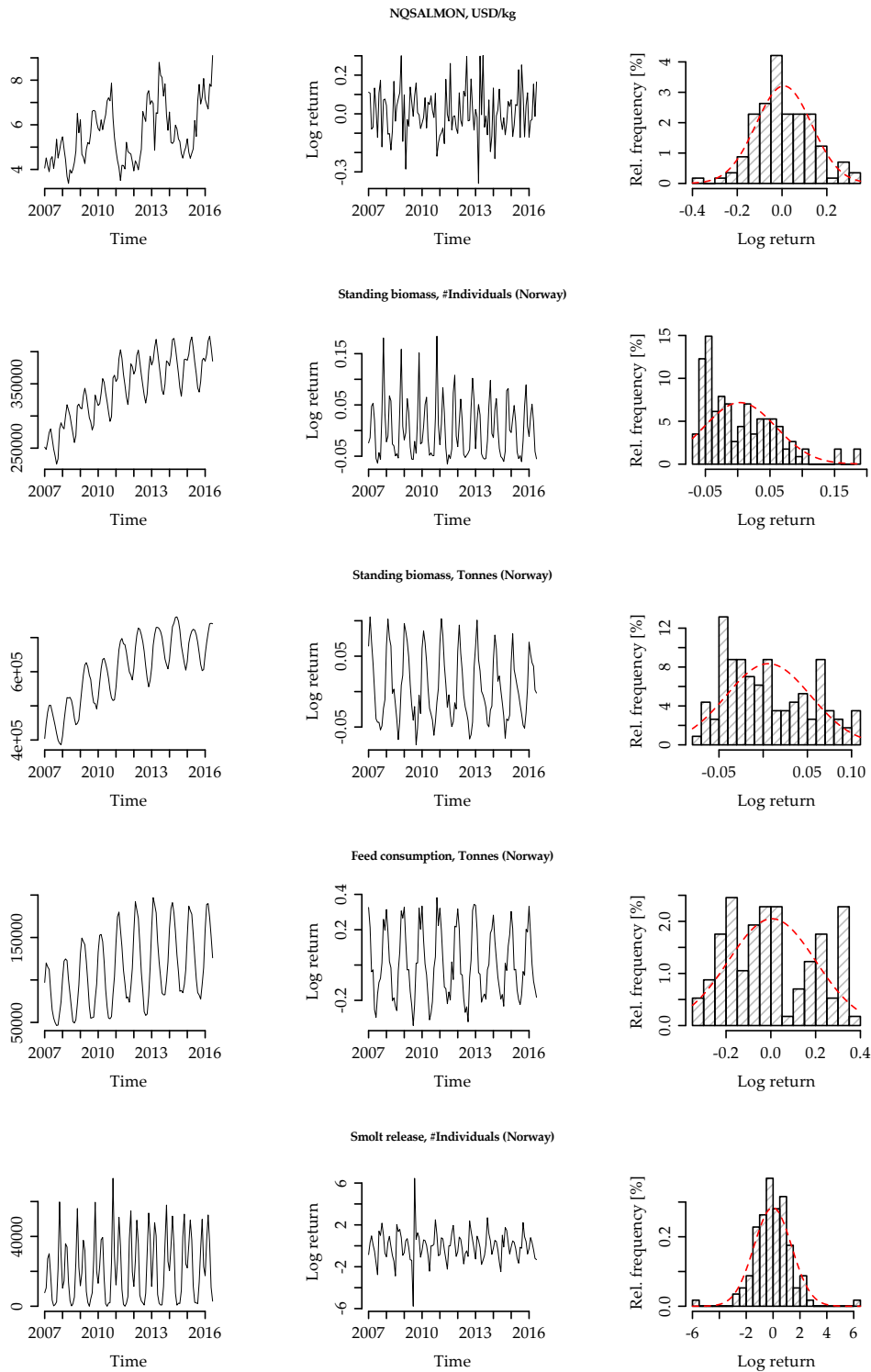


Figure 12: Plot of the explanatory variable time series (p. 1).

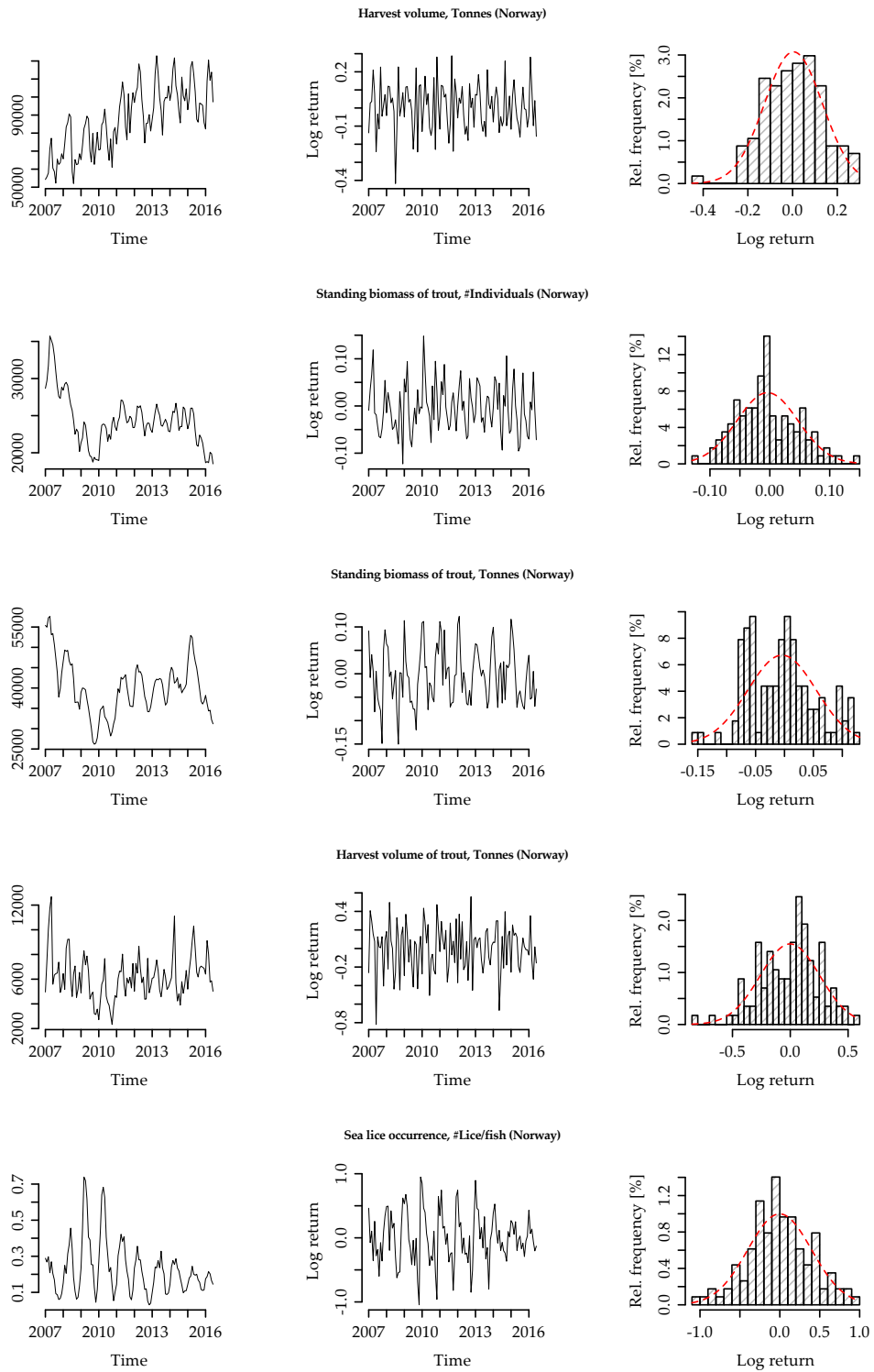


Figure 13: Plot of the explanatory variable time series (p. 2).



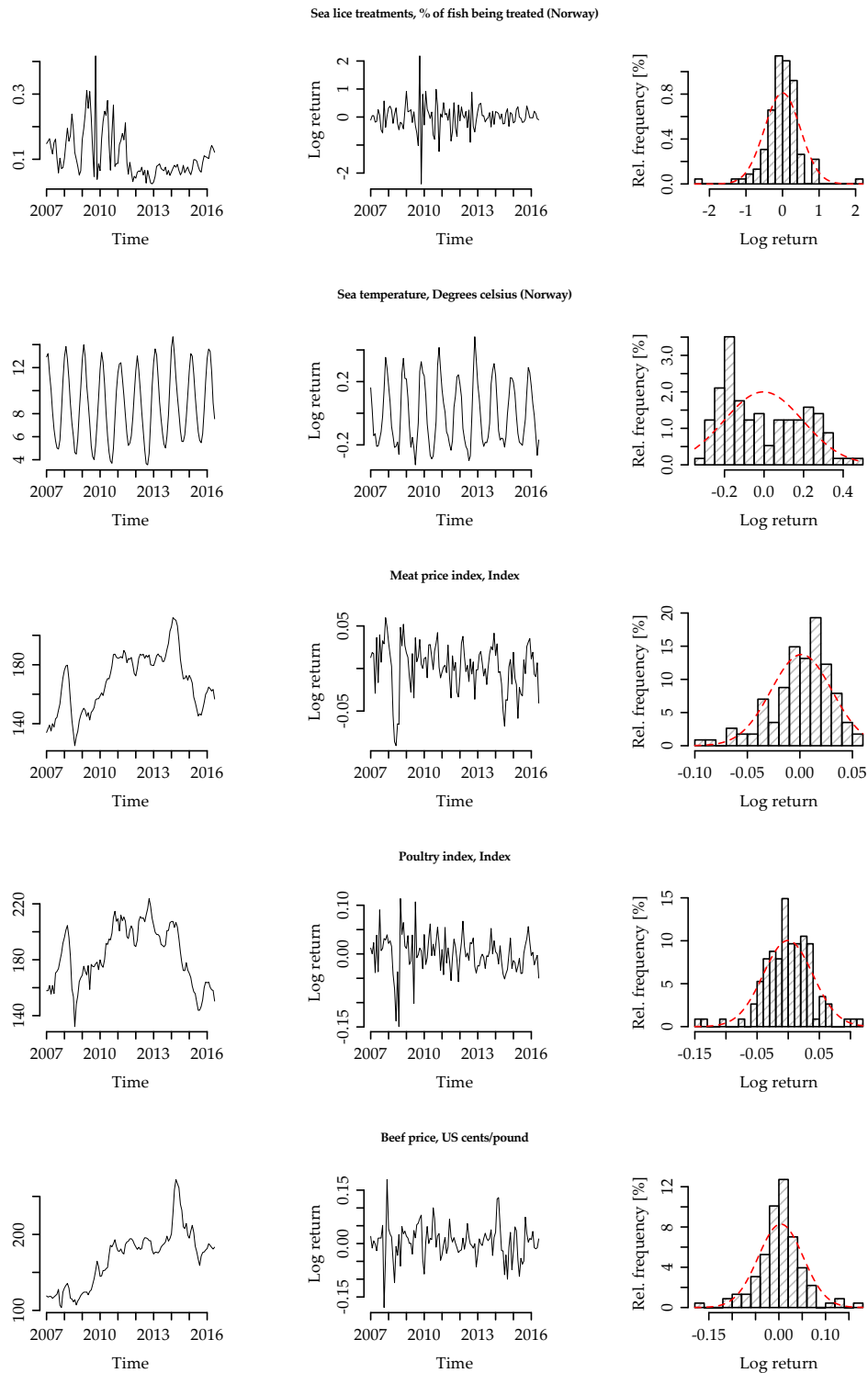


Figure 14: Plot of the explanatory variable time series (p. 3).

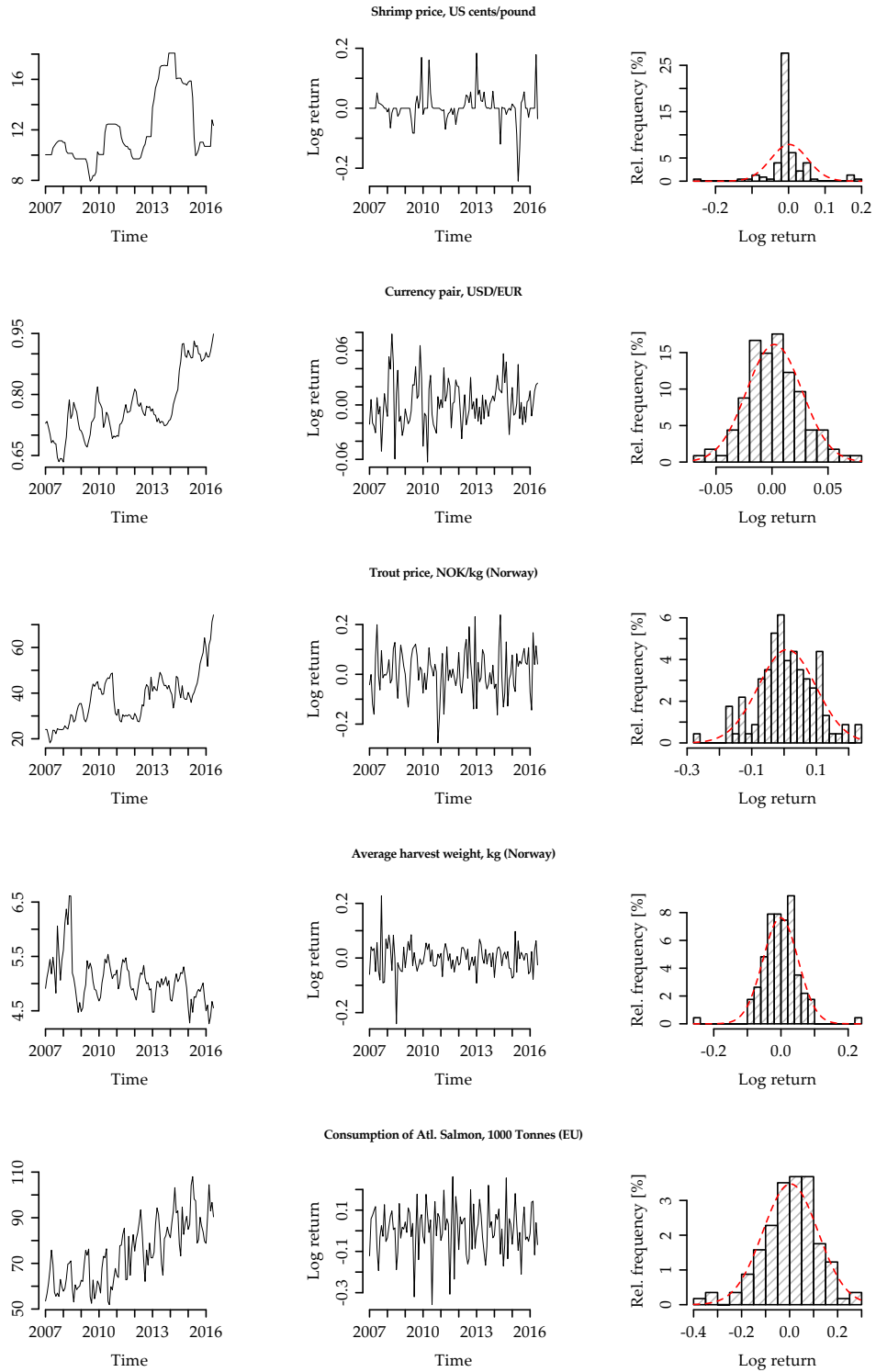


Figure 15: Plot of the explanatory variable time series (p. 4).

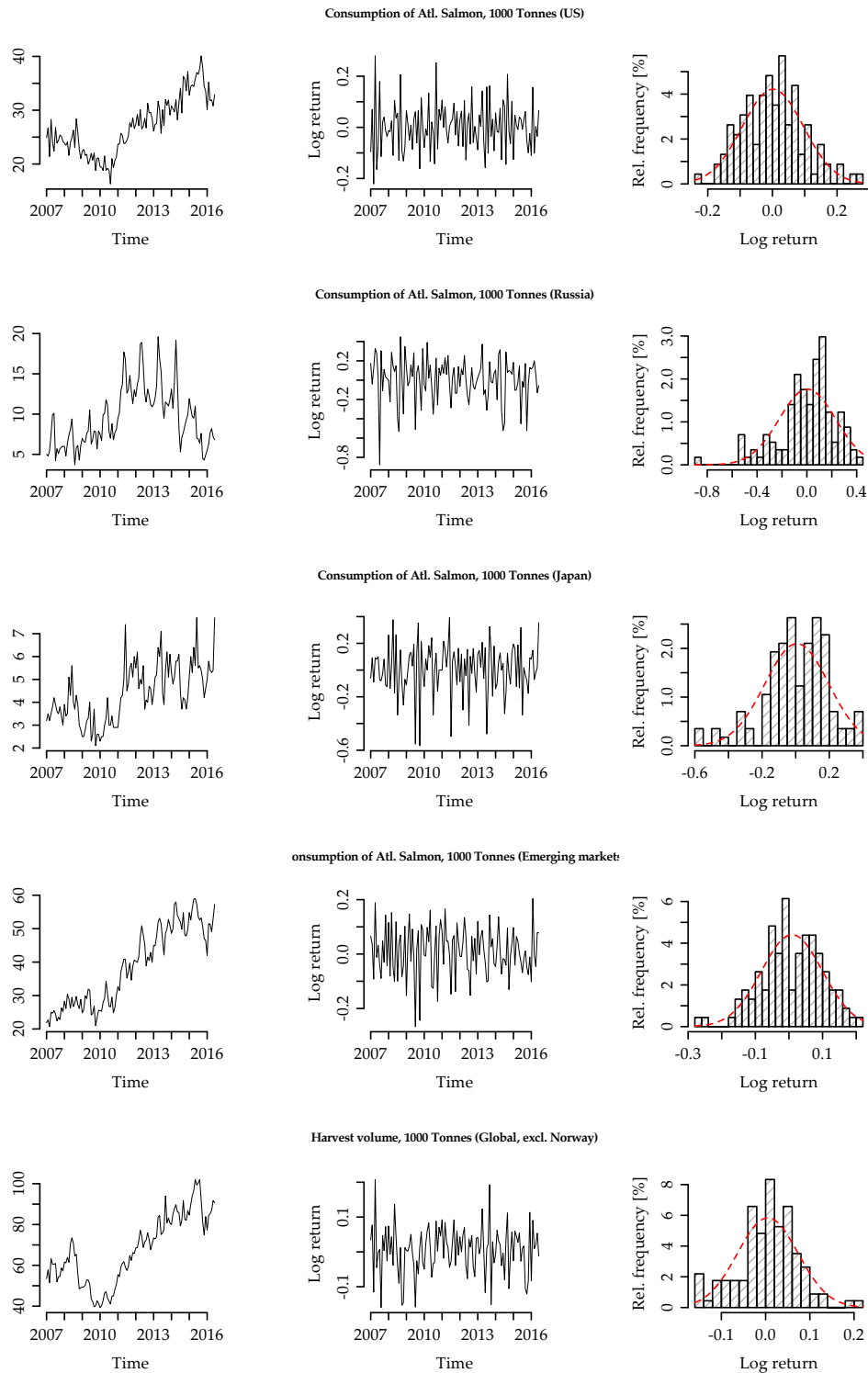


Figure 16: Plot of the explanatory variable time series (p. 5).

### Cross-Correlation with Respect to Lag and Time Window

As per the discussion in [Section 4.5](#), we provide surface plots of the cross-correlation function between the  $h$ -step ahead NQSALMON log return and each explanatory lagged  $l$  times and set to some time window  $\delta$ .

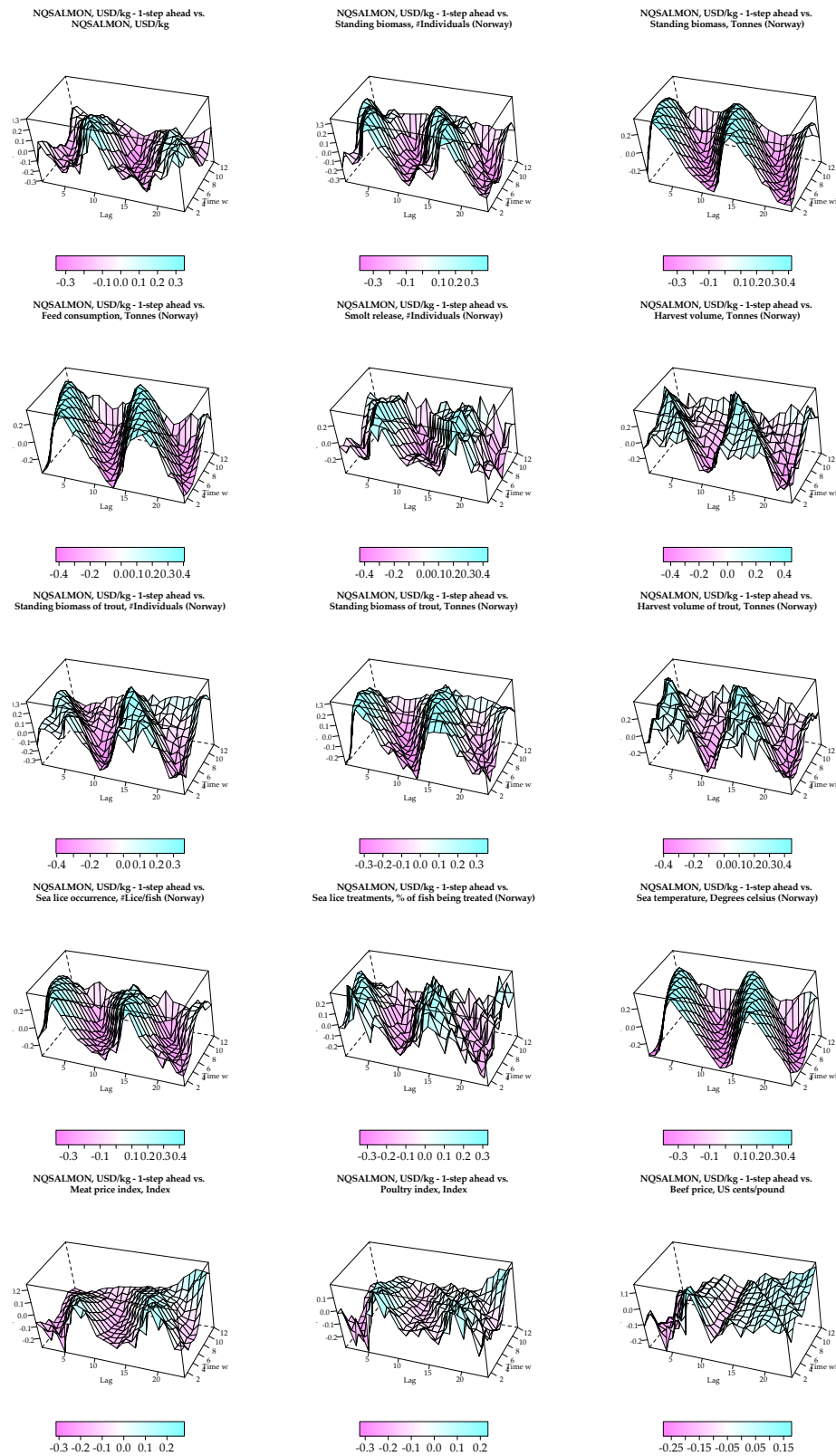


Figure 17: 1-month ahead Pearson correlation (p. 1).

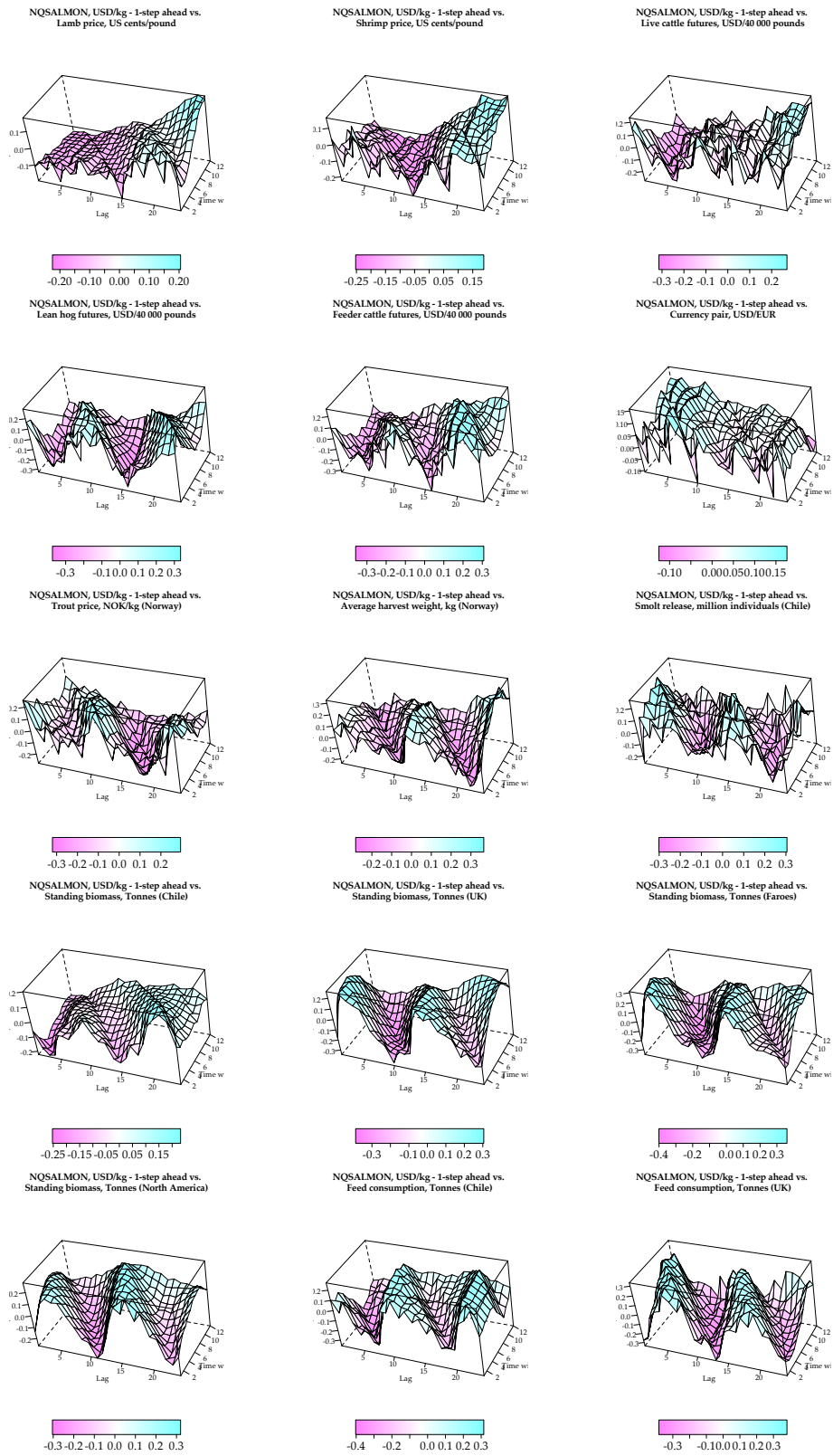


Figure 18: 1-month ahead Pearson correlation (p. 2).

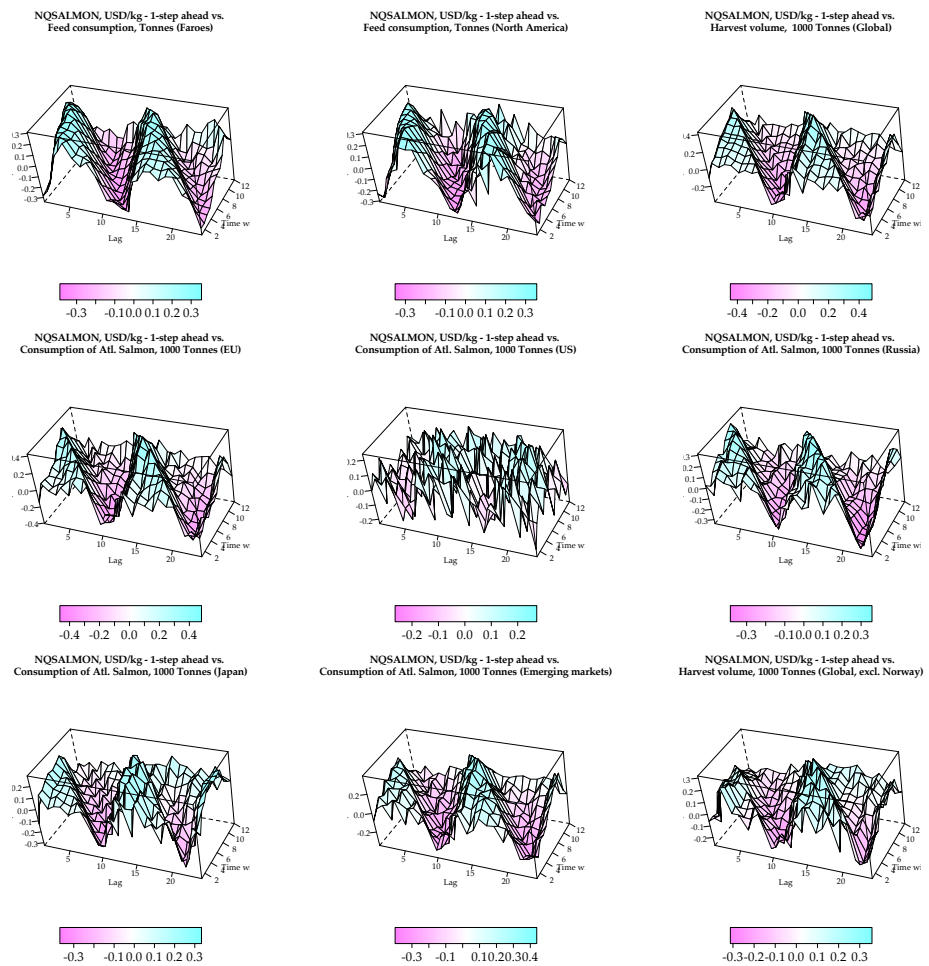


Figure 19: 1-month ahead Pearson correlation (p. 3).

### A.3 IN-SAMPLE QUANTILE REGRESSION RESULTS

The overall model consists of in total  $|\mathcal{H}| \cdot |\mathcal{T}| = 12 \cdot 5 = 60$  quantile regressions. On the following four pages, we present the respective regressions with their covariates (fixed for all submodels, including intercept in every case), with coefficients and corresponding significance levels. Also, for each regression the goodness-of-fit metric pseudo-R-squared (i.e.,  $R^1(\tau)$ ) is given.

Table 15: Coefficients overview of in-sample quantile regressions across horizons  $h \in \{1, \dots, 12\}$  (p. 1 of 4). <sup>1)</sup>

Horizon 1		$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
$\beta_0$	(Intercept)	-0.102** (-2.205)	-0.030* (-1.801)	0.011 (0.611)	0.099*** (4.652)	0.160*** (2.725)
$\beta_1$	NQSALMON, USD/kg (lag 1, MoM)	-0.137 (-0.682)	-0.285** (-2.094)	-0.250* (-1.765)	-0.374** (-2.402)	-0.283 (-1.067)
$\beta_2$	Standing biomass, Tonnes (Norway) (lag 3, MoM)	-1.026 (-0.928)	-1.831** (-2.087)	-0.943 (-0.999)	-1.173 (-1.123)	-1.642 (-0.797)
$\beta_3$	Feed consumption, Tonnes (Norway) (lag 2, YoY)	-0.184 (-0.784)	-0.183 (-1.608)	-0.243** (-2.128)	-0.348* (-1.876)	-0.334 (-0.908)
$\beta_4$	Harvest volume of trout, Tonnes (Norway) (lag 9, MoM)	0.101 (0.741)	0.145** (2.433)	0.119** (2.419)	0.198*** (3.043)	0.258* (1.877)
$\beta_5$	Sea lice occurrence, #Lice/fish (Norway) (lag 12, MoM)	-0.115 (-1.407)	-0.052 (-1.318)	-0.064 (-1.392)	-0.102** (-2.187)	-0.068 (-0.760)
$\beta_6$	Sea temperature, Degrees celsius (Norway) (lag 6, MoM)	0.397 (1.548)	0.648*** (2.812)	0.416 (1.543)	0.428* (1.717)	0.502 (1.138)
$\beta_7$	Shrimp price, US cents/pound (lag 6, MoM)	-1.273* (-1.870)	-1.066*** (-3.086)	-0.783* (-1.960)	-0.576 (-1.080)	-0.397 (-0.430)
$\beta_8$	Average harvest weight, kg (Norway) (lag 1, YoY)	-0.046 (-0.157)	-0.244 (-1.506)	-0.270 (-1.386)	-0.282 (-1.124)	-0.111 (-0.257)
$R^1(\tau)$		0.247	0.233	0.230	0.236	0.269

Horizon 2		$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
$\beta_0$	(Intercept)	-0.114** (-2.313)	-0.066*** (-3.142)	0.037** (2.036)	0.096*** (5.701)	0.155** (2.382)
$\beta_1$	NQSALMON, USD/kg (lag 2, MoM)	-0.335 (-1.558)	-0.364** (-2.414)	-0.429*** (-2.658)	-0.356** (-2.005)	-0.167 (-0.518)
$\beta_2$	Feed consumption, Tonnes (Norway) (lag 3, MoM)	-0.553*** (-2.753)	-0.539*** (-4.666)	-0.590*** (-5.745)	-0.478*** (-5.961)	-0.446** (-2.314)
$\beta_3$	Standing biomass of trout, Tonnes (Norway) (lag 9, YoY)	-0.453* (-1.871)	-0.336** (-2.084)	-0.061 (-0.493)	-0.038 (-0.434)	0.091 (0.432)
$\beta_4$	Harvest volume of trout, Tonnes (Norway) (lag 9, YoY)	0.247* (1.878)	0.186*** (2.633)	0.095 (1.418)	0.139** (2.087)	0.135 (1.021)
$\beta_5$	Sea temperature, Degrees celsius (Norway) (lag 6, MoM)	0.609*** (2.922)	0.539*** (4.246)	0.423*** (4.136)	0.362*** (4.056)	0.291* (1.934)
$\beta_6$	Meat price index, Index (lag 6, MoM)	-2.885** (-2.541)	-2.946*** (-4.133)	-1.732** (-2.454)	-1.450* (-1.688)	-2.626** (-2.076)
$\beta_7$	Shrimp price, US cents/pound (lag 9, MoM)	-1.080 (-1.366)	-0.251 (-0.420)	-0.801* (-1.862)	-0.824** (-2.550)	-0.739 (-0.920)
$\beta_8$	Currency pair, USD/EUR (lag 6, MoM)	-0.649 (-0.575)	-1.294 (-1.618)	-0.647 (-0.997)	-0.935 (-1.552)	-0.809 (-0.502)
$R^1(\tau)$		0.431	0.347	0.353	0.393	0.404

Horizon 3		$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
$\beta_0$	(Intercept)	-0.137** (-2.080)	-0.044 (-1.598)	0.026 (1.409)	0.102*** (4.026)	0.195* (1.926)
$\beta_1$	Standing biomass, #Individuals (Norway) (lag 3, MoM)	0.923 (1.111)	0.539 (1.242)	0.522 (1.131)	0.342 (0.676)	0.281 (0.173)
$\beta_2$	Standing biomass, Tonnes (Norway) (lag 3, MoM)	-3.352* (-1.892)	-3.174*** (-2.743)	-2.068 (-1.657)	-1.216 (-0.831)	-2.567 (-1.078)
$\beta_3$	Feed consumption, Tonnes (Norway) (lag 3, MoM)	-0.555* (-1.901)	-0.568*** (-4.075)	-0.701*** (-6.276)	-0.492*** (-2.967)	-0.444 (-1.146)
$\beta_4$	Standing biomass of trout, #Individuals (Norway) (lag 6, YoY)	-0.468 (-1.321)	-0.375* (-1.933)	-0.270** (-2.147)	-0.357** (-2.307)	-0.420 (-1.298)
$\beta_5$	Harvest volume of trout, Tonnes (Norway) (lag 9, YoY)	0.202 (1.208)	0.194** (2.150)	0.185*** (2.789)	0.135* (1.850)	0.126 (1.010)
$\beta_6$	Sea temperature, Degrees celsius (Norway) (lag 6, MoM)	0.999** (2.430)	0.961*** (4.126)	0.814*** (3.167)	0.595* (1.767)	1.045** (2.040)
$\beta_7$	Beef price, US cents/pound (lag 6, MoM)	-0.242 (-0.422)	-1.154*** (-2.711)	-1.180** (-2.414)	-1.505** (-2.097)	-2.270** (-2.036)
$\beta_8$	Currency pair, USD/EUR (lag 9, MoM)	1.724 (1.031)	1.798** (2.178)	1.218 (1.573)	1.230 (1.223)	0.217 (0.146)
$R^1(\tau)$		0.484	0.438	0.423	0.400	0.428

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

1) In parantheses are the adjusted standard errors, which account for possible misspecifications. These are calculated with the bootstrap by Parzen et al. (1994), with 1000 replications. Also, note that all lags are denoted relative to the horizon being forecasted.



Table 16: Coefficients overview of in-sample quantile regressions across horizons  $h \in \{1, \dots, 12\}$  (p. 2 of 4).

<b>Horizon 4</b>		$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
$\beta_0$	(Intercept)	-0.233*** (-3.481)	-0.092* (-1.918)	0.013 (0.372)	0.113** (2.316)	0.224*** (2.664)
$\beta_1$	Standing biomass, #Individuals (Norway) (lag 3, YoY)	0.820 (0.851)	0.881 (1.176)	1.101** (2.007)	0.645 (1.541)	0.344 (0.334)
$\beta_2$	Standing biomass, Tonnes (Norway) (lag 6, MoM)	3.057*** (3.438)	2.645*** (4.208)	2.655*** (5.389)	2.929*** (5.762)	2.901*** (2.831)
$\beta_3$	Feed consumption, Tonnes (Norway) (lag 4, YoY)	-0.414 (-0.956)	-0.871*** (-3.735)	-0.666*** (-3.620)	-0.635*** (-2.871)	-0.645 (-1.471)
$\beta_4$	Standing biomass of trout, #Individuals (Norway) (lag 9, YoY)	-0.434 (-1.197)	-0.252 (-1.047)	-0.274** (-2.190)	-0.182 (-1.231)	0.039 (0.081)
$\beta_5$	Harvest volume of trout, Tonnes (Norway) (lag 12, YoY)	-0.009 (-0.070)	0.046 (0.429)	0.164* (1.946)	0.162* (1.948)	0.092 (0.527)
$\beta_6$	Meat price index, Index (lag 6, MoM)	-3.058** (-2.220)	-2.456** (-2.482)	-1.038 (-1.228)	-1.733* (-1.665)	-2.630 (-1.455)
$\beta_7$	Beef price, US cents/pound (lag 6, MoM)	-0.693 (-1.124)	-0.799 (-1.241)	-1.636** (-2.437)	-1.268* (-1.821)	-1.423 (-1.374)
$\beta_8$	Shrimp price, US cents/pound (lag 9, MoM)	-1.107 (-1.026)	-1.217 (-1.628)	-1.237** (-2.158)	-1.063* (-1.958)	-1.284 (-1.272)
$R^1(\tau)$		0.473	0.378	0.424	0.425	0.461

<b>Horizon 5</b>		$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
$\beta_0$	(Intercept)	-0.253*** (-3.168)	-0.111*** (-2.837)	0.027 (1.071)	0.115*** (3.942)	0.223 (0.989)
$\beta_1$	Standing biomass, Tonnes (Norway) (lag 6, MoM)	3.592** (2.399)	2.374** (2.191)	2.519*** (3.567)	2.665*** (4.086)	1.938 (0.667)
$\beta_2$	Standing biomass of trout, #Individuals (Norway) (lag 6, YoY)	-0.506 (-0.818)	-0.878** (-2.144)	-1.339*** (-3.657)	-1.614*** (-3.898)	-1.230 (-1.171)
$\beta_3$	Standing biomass of trout, Tonnes (Norway) (lag 6, YoY)	0.322 (0.495)	0.474 (1.245)	0.722** (2.635)	1.042*** (3.853)	1.064 (1.495)
$\beta_4$	Sea lice occurrence, #Lice/fish (Norway) (lag 12, MoM)	-0.044 (-0.236)	-0.108 (-0.723)	-0.101 (-1.515)	-0.132** (-2.286)	-0.141 (-0.588)
$\beta_5$	Sea temperature, Degrees celsius (Norway) (lag 6, YoY)	-0.279 (-0.800)	-0.355 (-1.348)	-0.472** (-2.178)	-0.573*** (-2.841)	-0.340 (-0.552)
$\beta_6$	Meat price index, Index (lag 9, MoM)	-2.042 (-0.975)	-2.737* (-1.808)	-2.893** (-2.085)	-2.006 (-1.599)	-1.608 (-0.780)
$\beta_7$	Poultry index, Index (lag 6, MoM)	-1.656 (-1.098)	-2.384** (-2.020)	-0.869 (-0.907)	-0.825 (-1.229)	-0.698 (-0.315)
$\beta_8$	Beef price, US cents/pound (lag 6, MoM)	-1.481 (-1.368)	-0.749 (-1.023)	-0.670 (-1.159)	-0.476 (-0.676)	-1.189 (-0.815)
$R^1(\tau)$		0.385	0.317	0.347	0.398	0.417

<b>Horizon 6</b>		$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
$\beta_0$	(Intercept)	-0.293*** (-3.783)	-0.184*** (-3.435)	-0.045 (-0.983)	0.088 (1.444)	0.295** (2.088)
$\beta_1$	Standing biomass, #Individuals (Norway) (lag 9, MoM)	0.157 (0.187)	0.250 (0.366)	1.444** (2.484)	1.527* (1.678)	2.236 (1.588)
$\beta_2$	Standing biomass, Tonnes (Norway) (lag 9, YoY)	1.265 (1.547)	1.045* (1.831)	1.007** (2.004)	0.846 (1.344)	-0.287 (-0.269)
$\beta_3$	Standing biomass of trout, #Individuals (Norway) (lag 6, YoY)	-0.472* (-1.689)	-0.484** (-2.505)	-0.509** (-2.268)	-0.192 (-0.520)	-0.680 (-1.168)
$\beta_4$	Harvest volume of trout, Tonnes (Norway) (lag 12, YoY)	0.265* (1.754)	0.167* (1.786)	0.126 (1.101)	0.191 (1.615)	0.228 (1.367)
$\beta_5$	Sea lice occurrence, #Lice/fish (Norway) (lag 12, MoM)	-0.166 (-1.495)	-0.213*** (-3.020)	-0.272*** (-3.994)	-0.281*** (-2.755)	-0.264 (-1.522)
$\beta_6$	Sea temperature, Degrees celsius (Norway) (lag 6, MoM)	-0.099 (-0.324)	-0.193 (-1.043)	-0.286** (-2.224)	-0.406** (-2.206)	-0.226 (-0.551)
$\beta_7$	Meat price index, Index (lag 6, MoM)	-3.393* (-1.686)	-2.728*** (-2.673)	-2.748*** (-3.321)	-2.587* (-1.939)	-3.000 (-1.305)
$\beta_8$	Beef price, US cents/pound (lag 9, MoM)	-0.580 (-0.808)	-1.128 (-1.604)	-2.249*** (-2.825)	-2.163*** (-2.640)	-2.394* (-1.736)
$R^1(\tau)$		0.451	0.365	0.341	0.349	0.393

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 17: Coefficients overview of in-sample quantile regressions across horizons  $h \in \{1, \dots, 12\}$  (p. 3 of 4).

<b>Horizon 7</b>		$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
$\beta_0$	(Intercept)	-0.269*** (-3.048)	-0.146*** (-3.514)	0.028 (0.717)	0.168*** (4.879)	0.265** (2.399)
$\beta_1$	Standing biomass, #Individuals (Norway) (lag 9, MoM)	0.328 (0.270)	0.430 (0.466)	1.494** (1.989)	1.879*** (2.866)	1.839 (1.372)
$\beta_2$	Standing biomass, Tonnes (Norway) (lag 9, MoM)	3.073** (2.128)	2.651** (2.588)	1.723** (2.371)	0.954 (1.425)	0.846 (0.570)
$\beta_3$	Smolt release, #Individuals (Norway) (lag 17, MoM)	-0.041 (-1.280)	-0.057*** (-2.665)	-0.042* (-1.685)	-0.065*** (-2.674)	-0.068 (-1.297)
$\beta_4$	Standing biomass of trout, #Individuals (Norway) (lag 9, YoY)	-1.583** (-2.526)	-1.140*** (-2.800)	-0.896** (-2.043)	-0.436 (-0.932)	-0.434 (-0.610)
$\beta_5$	Standing biomass of trout, Tonnes (Norway) (lag 9, YoY)	1.025** (2.072)	0.558* (1.664)	0.515 (1.503)	0.170 (0.476)	0.189 (0.337)
$\beta_6$	Sea lice treatments, % of fish being treated (Norway) (lag 12, YoY)	-0.119 (-0.795)	-0.126 (-1.177)	-0.132* (-1.941)	-0.130*** (-3.034)	-0.107 (-1.328)
$\beta_7$	Beef price, US cents/pound (lag 9, MoM)	-1.167 (-1.231)	-1.630** (-2.166)	-1.184 (-1.562)	-1.956*** (-3.162)	-2.374** (-1.993)
$\beta_8$	Shrimp price, US cents/pound (lag 12, MoM)	-0.750 (-0.620)	-2.070*** (-2.889)	-2.557*** (-3.940)	-3.426*** (-3.334)	-3.892** (-2.109)
$R^1(\tau)$		0.423	0.328	0.320	0.371	0.375

<b>Horizon 8</b>		$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
$\beta_0$	(Intercept)	-0.269*** (-2.927)	-0.153*** (-3.608)	0.003 (0.082)	0.156*** (3.462)	0.284 (1.271)
$\beta_1$	Standing biomass, #Individuals (Norway) (lag 9, MoM)	0.384 (0.360)	1.027 (1.363)	1.419** (2.116)	1.675** (1.996)	0.982 (0.376)
$\beta_2$	Standing biomass, Tonnes (Norway) (lag 9, MoM)	3.756** (2.566)	2.517*** (2.744)	1.475** (2.296)	1.176* (1.814)	1.695 (0.835)
$\beta_3$	Smolt release, #Individuals (Norway) (lag 17, MoM)	-0.044 (-1.165)	-0.050** (-2.133)	-0.039* (-1.725)	-0.045 (-1.270)	0.002 (0.030)
$\beta_4$	Standing biomass of trout, #Individuals (Norway) (lag 9, YoY)	-1.105** (-2.009)	-1.223*** (-3.826)	-1.330*** (-2.988)	-0.943 (-1.629)	-0.449 (-0.406)
$\beta_5$	Standing biomass of trout, Tonnes (Norway) (lag 9, YoY)	0.409 (0.674)	0.620* (1.762)	0.751** (2.296)	0.505 (1.443)	0.480 (0.600)
$\beta_6$	Sea lice treatments, % of fish being treated (Norway) (lag 12, YoY)	-0.068 (-0.419)	-0.109 (-1.070)	-0.156** (-2.393)	-0.133** (-2.417)	-0.089 (-0.911)
$\beta_7$	Beef price, US cents/pound (lag 9, MoM)	-1.734** (-2.133)	-1.619** (-2.389)	-1.307 (-1.493)	-2.268*** (-2.682)	-1.894 (-1.402)
$\beta_8$	Shrimp price, US cents/pound (lag 12, MoM)	-1.829 (-1.189)	-1.878* (-1.891)	-1.039 (-1.325)	-1.611** (-2.033)	-1.977 (-1.058)
$R^1(\tau)$		0.449	0.361	0.302	0.314	0.312

<b>Horizon 9</b>		$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
$\beta_0$	(Intercept)	-0.241** (-2.576)	-0.141*** (-3.740)	0.002 (0.055)	0.136*** (3.461)	0.209* (1.719)
$\beta_1$	Standing biomass, #Individuals (Norway) (lag 9, MoM)	0.336 (0.396)	0.555 (0.991)	1.268* (1.953)	0.918 (1.082)	1.027 (0.671)
$\beta_2$	Standing biomass, Tonnes (Norway) (lag 9, MoM)	1.261 (0.748)	2.435*** (3.055)	1.852*** (2.742)	1.682** (2.457)	1.567 (1.091)
$\beta_3$	Standing biomass of trout, #Individuals (Norway) (lag 9, YoY)	-1.696** (-2.628)	-1.635*** (-4.875)	-1.824*** (-4.271)	-1.719*** (-2.863)	-1.754* (-1.935)
$\beta_4$	Standing biomass of trout, Tonnes (Norway) (lag 9, YoY)	0.848 (1.174)	1.046*** (2.726)	0.888*** (2.730)	0.828** (2.327)	0.900 (1.330)
$\beta_5$	Sea lice treatments, % of fish being treated (Norway) (lag 12, YoY)	-0.098 (-0.528)	-0.195* (-1.788)	-0.120 (-1.573)	-0.147*** (-2.668)	-0.134 (-1.451)
$\beta_6$	Beef price, US cents/pound (lag 9, MoM)	-1.436 (-1.249)	-1.315* (-1.818)	-1.284* (-1.983)	-1.146* (-1.671)	-1.699 (-1.531)
$\beta_7$	Shrimp price, US cents/pound (lag 12, MoM)	0.475 (0.271)	-0.517 (-0.552)	-0.977 (-1.352)	-1.365** (-2.175)	-1.338 (-1.018)
$\beta_8$	Currency pair, USD/EUR (lag 15, MoM)	2.063 (0.855)	1.625 (1.221)	1.362 (1.065)	2.985* (1.796)	4.357 (1.511)
$R^1(\tau)$		0.372	0.350	0.305	0.337	0.341

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 18: Coefficients overview of in-sample quantile regressions across horizons  $h \in \{1, \dots, 12\}$  (p. 4 of 4).

<b>Horizon 10</b>		$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
$\beta_0$	(Intercept)	-0.302*** (-4.345)	-0.205*** (-5.529)	-0.019 (-0.369)	0.209*** (4.156)	0.367*** (3.666)
$\beta_1$	Smolt release, #Individuals (Norway) (lag 15, YoY)	-0.046 (-0.326)	-0.137 (-1.349)	-0.150 (-1.628)	-0.053 (-0.662)	-0.084 (-1.026)
$\beta_2$	Standing biomass of trout, #Individuals (Norway) (lag 12, YoY)	-2.306*** (-2.709)	-2.376*** (-4.427)	-1.647*** (-3.242)	-0.721 (-1.069)	-0.112 (-0.109)
$\beta_3$	Standing biomass of trout, Tonnes (Norway) (lag 12, YoY)	1.275*** (3.143)	1.320*** (4.275)	1.155*** (2.804)	0.590 (0.972)	0.075 (0.084)
$\beta_4$	Sea lice occurrence, #Lice/fish (Norway) (lag 12, MoM)	-0.026 (-0.143)	0.125 (1.019)	0.125 (1.233)	0.102 (1.387)	0.040 (0.299)
$\beta_5$	Sea lice treatments, % of fish being treated (Norway) (lag 12, YoY)	-0.190 (-1.414)	-0.223** (-2.166)	-0.134 (-1.486)	-0.136* (-1.689)	-0.125 (-0.950)
$\beta_6$	Meat price index, Index (lag 12, MoM)	1.320 (0.588)	1.220 (0.803)	-0.175 (-0.110)	-1.234 (-0.914)	0.219 (0.095)
$\beta_7$	Beef price, US cents/pound (lag 12, MoM)	-0.919 (-1.119)	-0.786 (-1.271)	-0.325 (-0.337)	-0.687 (-0.588)	-2.374 (-1.316)
$\beta_8$	Currency pair, USD/EUR (lag 12, MoM)	3.721* (1.815)	3.929** (2.309)	3.155 (1.380)	1.840 (0.850)	0.838 (0.296)
$R^1(\tau)$		0.388	0.297	0.176	0.128	0.196

<b>Horizon 11</b>		$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
$\beta_0$	(Intercept)	-0.284*** (-3.657)	-0.180*** (-4.199)	0.000 (-0.011)	0.185*** (3.283)	0.339*** (3.791)
$\beta_1$	Smolt release, #Individuals (Norway) (lag 16, MoM)	0.033 (0.711)	0.061* (1.911)	0.041* (1.846)	0.014 (0.610)	0.014 (0.428)
$\beta_2$	Standing biomass of trout, #Individuals (Norway) (lag 12, YoY)	-1.990*** (-2.864)	-2.324*** (-4.572)	-1.526*** (-3.224)	-0.717 (-0.950)	-0.563 (-0.528)
$\beta_3$	Standing biomass of trout, Tonnes (Norway) (lag 12, YoY)	0.981* (1.787)	1.452*** (4.034)	0.852** (2.524)	0.390 (0.708)	0.438 (0.554)
$\beta_4$	Sea lice occurrence, #Lice/fish (Norway) (lag 12, YoY)	-0.086 (-0.528)	-0.146 (-1.132)	0.016 (0.161)	0.083 (0.917)	0.031 (0.201)
$\beta_5$	Sea lice treatments, % of fish being treated (Norway) (lag 12, YoY)	-0.076 (-0.334)	-0.157 (-1.036)	-0.255*** (-3.022)	-0.208** (-2.622)	-0.075 (-0.656)
$\beta_6$	Beef price, US cents/pound (lag 12, MoM)	-0.831 (-0.856)	-0.399 (-0.542)	-1.020 (-1.244)	-1.548 (-1.644)	-1.492 (-1.359)
$\beta_7$	Shrimp price, US cents/pound (lag 12, MoM)	-0.050 (-0.054)	-1.067 (-1.402)	-1.469** (-2.035)	-1.741 (-1.465)	-1.599 (-0.849)
$\beta_8$	Currency pair, USD/EUR (lag 12, MoM)	2.986 (1.365)	3.068** (2.287)	1.594 (1.251)	-0.005 (-0.003)	0.915 (0.367)
$R^1(\tau)$		0.343	0.311	0.291	0.205	0.227

<b>Horizon 12</b>		$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
$\beta_0$	(Intercept)	-0.254*** (-3.170)	-0.164*** (-4.508)	-0.024 (-0.632)	0.131* (1.969)	0.315 (1.505)
$\beta_1$	Smolt release, #Individuals (Norway) (lag 15, YoY)	0.057 (0.419)	-0.125 (-1.042)	-0.079 (-0.838)	-0.137* (-1.825)	-0.096 (-0.866)
$\beta_2$	Standing biomass of trout, #Individuals (Norway) (lag 12, YoY)	-1.858** (-2.469)	-2.073*** (-5.732)	-2.019*** (-4.530)	-2.185** (-2.455)	-0.919 (-0.759)
$\beta_3$	Standing biomass of trout, Tonnes (Norway) (lag 12, YoY)	1.156* (1.879)	1.220*** (4.706)	1.116*** (3.427)	1.177** (1.997)	0.735 (0.695)
$\beta_4$	Sea lice occurrence, #Lice/fish (Norway) (lag 12, YoY)	-0.192 (-1.392)	-0.037 (-0.354)	-0.058 (-0.533)	-0.016 (-0.120)	0.158 (0.869)
$\beta_5$	Sea lice treatments, % of fish being treated (Norway) (lag 12, YoY)	-0.038 (-0.186)	-0.225* (-1.788)	-0.193** (-2.555)	-0.208** (-2.321)	-0.234 (-1.549)
$\beta_6$	Beef price, US cents/pound (lag 12, MoM)	-0.938 (-0.929)	-0.742 (-0.953)	-0.833 (-1.018)	-1.152 (-1.220)	-1.419 (-0.739)
$\beta_7$	Shrimp price, US cents/pound (lag 12, MoM)	-0.592 (-0.678)	-0.381 (-0.675)	-0.621 (-1.024)	-0.296 (-0.341)	-0.670 (-0.394)
$\beta_8$	Currency pair, USD/EUR (lag 12, MoM)	2.716 (1.627)	3.747*** (3.183)	1.889 (1.144)	1.051 (0.502)	-0.515 (-0.154)
$R^1(\tau)$		0.394	0.369	0.326	0.234	0.209

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

#### A.4 IN-SAMPLE SPECIFICATION TEST RESULTS

In Table 19, we present the results from various specification tests that are run on the quantile regression models. The Kupiec, Christoffersen, and QRESET tests are described in Section 5.6. Implementation specifics of the QRESET, with a detailed procedure in accordance with Otsu (2009), are given in Appendix A.6.

Table 19: Specification test results from the in-sample quantile regressions across horizons  $k \in \{1, \dots, 12\}$ .

(a) Horizons 1 –6													
Horizon 1		$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$	Horizon 2		$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
Coverage rate	0.087	0.252	0.476	0.738	0.874	0.874	Coverage rate	0.107	0.243	0.515	0.757	0.913	0.913
Kupiec test	0.190	0.003	0.243	0.080	0.732	0.732	Kupiec test	0.052	0.029	0.087	0.029	0.190	0.190
Christoffersen test	2.118†	1.087	1.589	4.807*	8.439**†	8.439**†	Christoffersen test	0.888	2.826	4.689*	2.929	1.916†	1.916†
QRESET(2)	-1.380	-0.957	1.147	2.327**	1.492	1.492	QRESET(2)	-0.806	-0.655	0.519	2.016**	0.954	0.954
QRESET(3)	2.022	8.259**	12.527***	1.811	1.322	1.322	QRESET(3)	1.552	10.273**	15.476***	4.908*	1.373	1.373
QRESET(4)	1.502	10.931**	4.264	2.900	0.584	0.584	QRESET(4)	0.491	6.959*	9.877**	3.150	0.588	0.588
(b) Horizons 7 –12													
Horizon 7		$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$	Horizon 8		$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
Coverage rate	0.078	0.223	0.476	0.738	0.854	0.854	Coverage rate	0.068	0.262	0.476	0.728	0.932	0.932
Kupiec test	0.613	0.401	0.243	0.080	2.120	2.120	Kupiec test	1.309	0.080	0.243	0.257	1.309	1.309
Christoffersen test	2.139*	5.201*	6.270**	12.025***	3.506	3.506	Christoffersen test	1.966	2.711	15.962***	3.511	7.547**	7.547**
QRESET(2)	-0.582	-0.593	0.321	2.325**	1.499	1.499	QRESET(2)	-0.962	-1.147	0.174	2.233*	1.616	1.616
QRESET(3)	1.922	10.089**	33.002***	17.509**	1.569	1.569	QRESET(3)	2.143	11.163***	18.700**	17.430**	3.510	3.510
QRESET(4)	1.374	14.733***	16.546**	4.809	0.672	0.672	QRESET(4)	1.731	20.932***	9.688*	12.542***	1.222	1.222
Horizon 9		$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$	Horizon 10		$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
Coverage rate	0.087	0.243	0.485	0.709	0.903	0.903	Coverage rate	0.087	0.243	0.485	0.709	0.903	0.903
Kupiec test	0.190	0.029	0.087	0.904	0.010	0.010	Kupiec test	0.190	0.029	0.087	0.904	0.010	0.010
Christoffersen test	2.047	2.826	15.908***	4.774*	6.125**	6.125**	Christoffersen test	2.047	2.826	15.908***	4.774*	6.125**	6.125**
QRESET(2)	-1.167	-0.684	0.759	2.010*	1.658	1.658	QRESET(2)	-1.167	-0.684	0.759	2.010*	1.658	1.658
QRESET(3)	2.369	16.088**	9.688**	17.091***	4.051	4.051	QRESET(3)	2.369	16.088**	9.688**	17.091***	4.051	4.051
QRESET(4)	1.386	20.125***	5.200	14.616***	0.594	0.594	QRESET(4)	1.386	20.125***	5.200	14.616***	0.594	0.594
Horizon 11		$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$	Horizon 12		$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
Coverage rate	0.049	0.272	0.515	0.718	0.874	0.874	Coverage rate	0.107	0.243	0.485	0.757	0.874	0.874
Kupiec test	3.670*	0.257	0.087	0.533	0.732	0.732	Kupiec test	0.052	0.029	0.087	0.029	0.732	0.732
Christoffersen test	4.286†	5.275*	17.621***	26.166***	11.855***	11.855***	Christoffersen test	0.317	9.655**	10.363**	21.850***	11.855**	11.855**
QRESET(2)	-0.395	-0.810	1.675*	3.161**	2.047**	2.047**	QRESET(2)	-0.852	-1.163	0.803	3.490**	1.554	1.554
QRESET(3)	4.716*	16.869**	22.944**	10.578***	4.267	4.267	QRESET(3)	4.053	23.888**	19.879**	19.572**	2.247	2.247
QRESET(4)	2.672	16.477**	15.121**	19.991***	0.690	0.690	QRESET(4)	1.930	27.037***	22.025**	21.133***	1.357	1.357
Horizon 11		$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$	Horizon 12		$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
Coverage rate	0.107	0.243	0.485	0.757	0.874	0.874	Coverage rate	0.136	0.262	0.495	0.738	0.874	0.874
Kupiec test	0.052	0.029	0.087	0.029	0.732	0.732	Kupiec test	1.343	0.080	0.010	0.080	0.732	0.732
Christoffersen test	0.317	9.655**	10.363**	21.850***	11.855**	11.855**	Christoffersen test	1.641	6.431**	6.169**	16.401**	5.041*	5.041*
QRESET(2)	-0.852	-1.163	0.803	3.490**	1.554	1.554	QRESET(2)	-0.497	0.186	1.844*	3.764**	1.399	1.399
QRESET(3)	4.053	23.888**	19.879**	19.572**	2.247	2.247	QRESET(3)	3.389	41.886***	35.326**	24.832**	2.112	2.112
QRESET(4)	1.930	27.037***	22.025**	21.133***	1.357	1.357	QRESET(4)	3.144	46.445***	36.504***	36.792***	0.505	0.505

Note: †p<0.1; \*p<0.05; \*\*p<0.01; ‡ $\tau_{0.1} = 0$ ; † $\tau_{0.1} = 1$

Critical values:  $|q_{0.5, \tau=0.5, 0.0}| > 1.66$ ;  $|q_{0.5, \tau=0.25, 0}| > 1.99$ ;  $|q_{0.5, \tau=0.5, 0.5}| > 2.63$   
 $\chi^2_{1, \alpha=10\%} > 2.71$ ;  $\chi^2_{1, \alpha=5\%} > 3.84$ ;  $\chi^2_{1, \alpha=1\%} > 6.63$   
 $\chi^2_{2, \alpha=10\%} > 4.61$ ;  $\chi^2_{2, \alpha=5\%} > 5.99$ ;  $\chi^2_{2, \alpha=1\%} > 9.21$   
 $\chi^2_{3, \alpha=10\%} > 6.25$ ;  $\chi^2_{3, \alpha=5\%} > 7.81$ ;  $\chi^2_{3, \alpha=1\%} > 11.34$

## A.5 CROSS-CORRELATION TABLES FOR THE EXPLANATORY VARIABLES

Below we present cross-correlation tables for the log returns (lagged and at time windows MoM or YoY, as indicated) of the covariates utilised in the quantile regressions across the given horizons.

Table 20: Cross-correlation between log returns,  $h \in \{1, \dots, 6\}$ .

$h = 1$	TIME SERIES, UNIT	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$
$\beta_1$	NQSALMON, USD/kg (lag 1, MoM)	1.00							
$\beta_2$	Standing biomass, Tonnes (Norway) (lag 3, MoM)	-0.03	1.00						
$\beta_3$	Feed consumption, Tonnes (Norway) (lag 2, YoY)	-0.11	0.11	1.00					
$\beta_4$	Harvest volume of trout, Tonnes (Norway) (lag 9, MoM)	0.04	-0.15	-0.03	1.00				
$\beta_5$	Sea lice occurrence, #Lice/fish (Norway) (lag 12, MoM)	-0.11	-0.25	-0.08	-0.01	1.00			
$\beta_6$	Sea temperature, Degrees celsius (Norway) (lag 6, MoM)	0.09	0.91	0.04	-0.19	-0.26	1.00		
$\beta_7$	Shrimp price, US cents/pound (lag 6, MoM)	0.16	0.04	-0.19	0.16	-0.10	0.06	1.00	
$\beta_8$	Average harvest weight, kg (Norway) (lag 1, YoY)	-0.10	-0.06	-0.13	0.02	-0.05	-0.03	0.03	1.00

$h = 2$	TIME SERIES, UNIT	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$
$\beta_1$	NQSALMON, USD/kg (lag 2, MoM)	1.00							
$\beta_2$	Feed consumption, Tonnes (Norway) (lag 3, MoM)	-0.31	1.00						
$\beta_3$	Standing biomass of trout, Tonnes (Norway) (lag 9, YoY)	0.00	0.02	1.00					
$\beta_4$	Harvest volume of trout, Tonnes (Norway) (lag 9, YoY)	0.13	0.03	0.52	1.00				
$\beta_5$	Sea temperature, Degrees celsius (Norway) (lag 6, MoM)	-0.11	0.29	0.02	0.00	1.00			
$\beta_6$	Meat price index, Index (lag 6, MoM)	-0.07	0.16	0.08	0.06	0.34	1.00		
$\beta_7$	Shrimp price, US cents/pound (lag 9, MoM)	-0.08	-0.02	-0.07	-0.14	-0.02	0.11	1.00	
$\beta_8$	Currency pair, USD/EUR (lag 6, MoM)	-0.08	-0.03	-0.08	0.03	-0.07	-0.36	-0.11	1.00

$h = 3$	TIME SERIES, UNIT	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$
$\beta_1$	Standing biomass, #Individuals (Norway) (lag 3, MoM)	1.00							
$\beta_2$	Standing biomass, Tonnes (Norway) (lag 3, MoM)	0.21	1.00						
$\beta_3$	Feed consumption, Tonnes (Norway) (lag 3, MoM)	0.52	0.42	1.00					
$\beta_4$	Standing biomass of trout, #Individuals (Norway) (lag 6, YoY)	0.04	0.12	0.05	1.00				
$\beta_5$	Harvest volume of trout, Tonnes (Norway) (lag 9, YoY)	-0.03	0.00	-0.03	0.16	1.00			
$\beta_6$	Sea temperature, Degrees celsius (Norway) (lag 6, MoM)	0.11	0.91	0.29	0.14	0.01	1.00		
$\beta_7$	Beef price, US cents/pound (lag 6, MoM)	0.01	-0.01	0.04	0.00	0.01	-0.04	1.00	
$\beta_8$	Currency pair, USD/EUR (lag 9, MoM)	-0.02	-0.04	0.07	-0.09	-0.16	0.00	-0.17	1.00

$h = 4$	TIME SERIES, UNIT	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$
$\beta_1$	Standing biomass, #Individuals (Norway) (lag 9, YoY)	1.00							
$\beta_2$	Standing biomass, Tonnes (Norway) (lag 6, MoM)	0.06	1.00						
$\beta_3$	Feed consumption, Tonnes (Norway) (lag 4, YoY)	0.44	0.12	1.00					
$\beta_4$	Standing biomass of trout, #Individuals (Norway) (lag 9, YoY)	0.11	0.10	-0.04	1.00				
$\beta_5$	Harvest volume of trout, Tonnes (Norway) (lag 12, YoY)	0.18	-0.02	-0.01	0.18	1.00			
$\beta_6$	Meat price index, Index (lag 6, MoM)	0.09	0.01	-0.05	0.05	-0.01	1.00		
$\beta_7$	Beef price, US cents/pound (lag 6, MoM)	-0.02	0.05	0.00	-0.09	-0.16	0.30	1.00	
$\beta_8$	Shrimp price, US cents/pound (lag 9, MoM)	-0.14	0.04	-0.06	0.08	-0.13	0.11	0.07	1.00

$h = 5$	TIME SERIES, UNIT	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$
$\beta_1$	Standing biomass, Tonnes (Norway) (lag 6, MoM)	1.00							
$\beta_2$	Standing biomass of trout, #Individuals (Norway) (lag 6, YoY)	-0.03	1.00						
$\beta_3$	Standing biomass of trout, Tonnes (Norway) (lag 6, YoY)	0.01	0.78	1.00					
$\beta_4$	Sea lice occurrence, #Lice/fish (Norway) (lag 12, MoM)	-0.50	-0.08	-0.02	1.00				
$\beta_5$	Sea temperature, Degrees celsius (Norway) (lag 6, YoY)	0.05	0.11	0.22	0.04	1.00			
$\beta_6$	Meat price index, Index (lag 9, MoM)	0.36	0.15	0.00	-0.21	-0.04	1.00		
$\beta_7$	Poultry index, Index (lag 6, MoM)	-0.07	0.12	-0.01	-0.15	-0.15	-0.01	1.00	
$\beta_8$	Beef price, US cents/pound (lag 6, MoM)	0.05	0.02	-0.08	0.15	-0.02	0.15	0.15	1.00

$h = 6$	TIME SERIES, UNIT	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$
$\beta_1$	Standing biomass, #Individuals (Norway) (lag 9, MoM)	1.00							
$\beta_2$	Standing biomass, Tonnes (Norway) (lag 9, YoY)	0.06	1.00						
$\beta_3$	Standing biomass of trout, #Individuals (Norway) (lag 6, YoY)	0.02	-0.23	1.00					
$\beta_4$	Harvest volume of trout, Tonnes (Norway) (lag 12, YoY)	-0.02	0.16	0.01	1.00				
$\beta_5$	Sea lice occurrence, #Lice/fish (Norway) (lag 12, MoM)	-0.08	0.01	-0.08	-0.04	1.00			
$\beta_6$	Sea temperature, Degrees celsius (Norway) (lag 12, MoM)	-0.18	0.04	0.08	-0.07	-0.25	1.00		
$\beta_7$	Meat price index, Index (lag 6, MoM)	-0.07	0.20	0.17	-0.02	-0.17	0.34	1.00	
$\beta_8$	Beef price, US cents/pound (lag 9, MoM)	-0.06	0.08	0.03	0.01	0.02	0.06	-0.02	1.00

Table 21: Cross-correlation between log returns,  $h \in \{7, \dots, 12\}$ .

$h = 7$	TIME SERIES, UNIT	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$
$\beta_1$	Standing biomass, #Individuals (Norway) (lag 9, MoM)	1.00							
$\beta_2$	Standing biomass, Tonnes (Norway) (lag 9, MoM)	0.19	1.00						
$\beta_3$	Smolt release, #Individuals (Norway) (lag 17, MoM)	0.02	-0.35	1.00					
$\beta_4$	Standing biomass of trout, #Individuals (Norway) (lag 9, YoY)	0.07	-0.04	-0.02	1.00				
$\beta_5$	Standing biomass of trout, Tonnes (Norway) (lag 9, YoY)	0.02	0.00	-0.01	0.80	1.00			
$\beta_6$	Sea lice treatments, % of fish being treated (Norway) (lag 12, YoY)	0.02	0.08	-0.10	-0.05	-0.02	1.00		
$\beta_7$	Beef price, US cents/pound (lag 9, MoM)	0.08	0.05	-0.08	0.01	-0.08	-0.01	1.00	
$\beta_8$	Shrimp price, US cents/pound (lag 12, MoM)	0.07	0.04	0.23	0.17	0.00	0.00	0.07	1.00

$h = 8$	TIME SERIES, UNIT	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$
$\beta_1$	Standing biomass, #Individuals (Norway) (lag 9, MoM)	1.00							
$\beta_2$	Standing biomass, Tonnes (Norway) (lag 9, MoM)	0.21	1.00						
$\beta_3$	Smolt release, #Individuals (Norway) (lag 17, MoM)	-0.01	-0.35	1.00					
$\beta_4$	Standing biomass of trout, #Individuals (Norway) (lag 9, YoY)	0.02	-0.03	-0.05	1.00				
$\beta_5$	Standing biomass of trout, Tonnes (Norway) (lag 9, YoY)	-0.03	0.01	-0.02	0.78	1.00			
$\beta_6$	Sea lice treatments, % of fish being treated (Norway) (lag 12, YoY)	0.01	0.08	-0.11	-0.07	-0.03	1.00		
$\beta_7$	Beef price, US cents/pound (lag 9, MoM)	0.09	0.05	-0.08	0.02	-0.08	-0.01	1.00	
$\beta_8$	Shrimp price, US cents/pound (lag 12, MoM)	0.05	0.04	-0.24	0.16	-0.02	0.00	0.07	1.00

$h = 9$	TIME SERIES, UNIT	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$
$\beta_1$	Standing biomass, #Individuals (Norway) (lag 9, MoM)	1.00							
$\beta_2$	Standing biomass, Tonnes (Norway) (lag 9, MoM)	0.21	1.00						
$\beta_3$	Standing biomass of trout, #Individuals (Norway) (lag 9, YoY)	0.03	0.00	1.00					
$\beta_4$	Standing biomass of trout, Tonnes (Norway) (lag 9, YoY)	-0.03	0.03	0.76	1.00				
$\beta_5$	Sea lice treatments, % of fish being treated (Norway) (lag 12, YoY)	0.00	0.10	-0.13	-0.07	1.00			
$\beta_6$	Beef price, US cents/pound (lag 9, MoM)	0.10	0.06	0.03	-0.08	-0.01	1.00		
$\beta_7$	Shrimp price, US cents/pound (lag 12, MoM)	0.07	0.05	0.17	-0.03	-0.02	0.10	1.00	
$\beta_8$	Currency pair, USD/EUR (lag 15, MoM)	-0.02	-0.04	0.01	0.05	0.02	-0.05	-0.08	1.00

$h = 10$	TIME SERIES, UNIT	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$
$\beta_1$	Smolt release, #Individuals (Norway) (lag 15, YoY)	1.00							
$\beta_2$	Standing biomass of trout, #Individuals (Norway) (lag 12, YoY)	-0.01	1.00						
$\beta_3$	Standing biomass of trout, Tonnes (Norway) (lag 12, YoY)	0.08	0.80	1.00					
$\beta_4$	Sea lice occurrence, #Lice/fish (Norway) (lag 12, MoM)	0.23	-0.05	-0.05	1.00				
$\beta_5$	Sea lice treatments, % of fish being treated (Norway) (lag 12, YoY)	-0.21	-0.10	-0.14	-0.03	1.00			
$\beta_6$	Meat price index, Index (lag 12, MoM)	-0.06	0.20	0.01	0.02	-0.04	1.00		
$\beta_7$	Beef price, US cents/pound (lag 12, MoM)	-0.04	0.01	-0.08	-0.10	-0.09	0.30	1.00	
$\beta_8$	Currency pair, USD/EUR (lag 12, MoM)	0.06	-0.10	-0.12	-0.12	0.10	-0.36	-0.07	1.00

$h = 11$	TIME SERIES, UNIT	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$
$\beta_1$	Smolt release, #Individuals (Norway) (lag 16, MoM)	1.00							
$\beta_2$	Standing biomass of trout, #Individuals (Norway) (lag 12, YoY)	0.04	1.00						
$\beta_3$	Standing biomass of trout, Tonnes (Norway) (lag 12, YoY)	0.04	0.78	1.00					
$\beta_4$	Sea lice occurrence, #Lice/fish (Norway) (lag 12, YoY)	0.05	-0.33	-0.20	1.00				
$\beta_5$	Sea lice treatments, % of fish being treated (Norway) (lag 12, YoY)	0.06	-0.14	-0.17	0.49	1.00			
$\beta_6$	Beef price, US cents/pound (lag 12, MoM)	0.09	0.02	-0.08	0.12	-0.09	1.00		
$\beta_7$	Shrimp price, US cents/pound (lag 12, MoM)	0.19	0.08	-0.15	-0.05	-0.06	0.16	1.00	
$\beta_8$	Currency pair, USD/EUR (lag 12, MoM)	-0.02	-0.10	-0.12	-0.02	0.10	-0.06	-0.08	1.00

$h = 12$	TIME SERIES, UNIT	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$
$\beta_1$	Smolt release, #Individuals (Norway) (lag 15, YoY)	1.00							
$\beta_2$	Standing biomass of trout, #Individuals (Norway) (lag 12, YoY)	-0.04	1.00						
$\beta_3$	Standing biomass of trout, Tonnes (Norway) (lag 12, YoY)	0.06	0.76	1.00					
$\beta_4$	Sea lice occurrence, #Lice/fish (Norway) (lag 12, YoY)	-0.05	-0.35	-0.22	1.00				
$\beta_5$	Sea lice treatments, % of fish being treated (Norway) (lag 12, YoY)	-0.22	-0.17	-0.18	0.48	1.00			
$\beta_6$	Beef price, US cents/pound (lag 12, MoM)	-0.04	0.03	-0.08	0.12	-0.10	1.00		
$\beta_7$	Shrimp price, US cents/pound (lag 12, MoM)	0.05	0.09	-0.15	-0.05	-0.06	0.15	1.00	
$\beta_8$	Currency pair, USD/EUR (lag 12, MoM)	0.06	-0.11	-0.13	-0.02	0.10	-0.06	-0.08	1.00

## A.6 TESTS FOR STATISTICAL PROPERTIES

Four different statistical tests are conducted to examine the properties of the dependent variable and the explanatory variables. This include the *Jarque-Bera test*, the *Dickey-Fuller test*, the *(Augmented) Dickey-Fuller test*, and the *Ljung-Box Q-Test* to evaluate the descriptive statistics for the variables. A corresponding description for each test is given in the following section.

### *Jarque-Bera (JB) Test*

The JB test applies to any random variable whenever an assumption of normality can be justified (Jarque & Bera, 1980). It is based on the estimates of the sample skewness  $\hat{\tau}$ , and the sample kurtosis  $\hat{\kappa}$ . In accordance with Alexander (2008b, p. 158), the test statistic is expressed as:

$$JB = \frac{T-k}{6} \left( \hat{\tau} + \frac{1}{4}(\hat{\kappa} - 3)^2 \right) \stackrel{\text{asy}}{\sim} \chi_2^2,$$

where  $T$  is the number of observations and  $k$  is the number of regressors. The hypotheses may be summarised as follows:

- $H_0$  : Both skewness and excess kurtosis are zero, matching the properties of the normal distribution.
- $H_1$  : Either one or both of the skewness and excess kurtosis differs from zero, i.e., the sample data is non-normally distributed.

In accordance with the critical values of the  $\chi_2^2$  distribution, the null hypothesis is rejected at a 1% significance level whenever  $JB > 9.21$ .

### *Dickey-Fuller (DF) Test*

The DF test is a basic *unit root test* that can be used to determine whether or not a time series is stationary (Dickey & Fuller, 1979). Consider the time series  $\{x_t\}_{t=1}^T$ , apply the DF regression given as (Alexander, 2008a, p. 217):

$$\Delta x_t = \alpha + \beta x_{t-1} + \varepsilon_t.$$

The test statistic is the *t-ratio* of  $\hat{\beta}$ , and it is a one-sided test for:

- $H_0$  :  $\beta = 0$ ,
- $H_1$  :  $\beta < 0$ .

Under  $H_0$ ,  $\{x_t\}_{t=1}^T \sim I(1)$  - equivalently, the time series is non-stationary. Under the alternative hypothesis,  $H_1$ ,  $\{x_t\}_{t=1}^T \sim I(0)$  - that is, the time series is stationary.  $H_0$  is rejected if the *t-ratio* of  $\hat{\beta}$  is more negative than the critical value at a given significance level. The critical value of the *t-ratio* of  $\hat{\beta}$  is somewhat larger than the corresponding *t-statistic*. In particular, this depends on the sample size. Note that the ordinary DF test is biased if there is autocorrelation in the residuals of the DF regression.

### *Augmented Dickey-Fuller (ADF) Test*

In contrast to the DF test, the ADF test removes potential autocorrelation in the residuals from the DF regression by including a number of lagged dependent variables. The *order* for the test is denoted  $\rho$ , where  $\rho$  refers to the number of lags included for the dependent variable (Alexander, 2008a, p. 218). The test is obtained from the following auxiliary regression:

$$\Delta x_t = \alpha + \beta x_{t-1} + \gamma_1 \Delta x_{t-1} + \dots + \gamma_\rho \Delta x_{t-\rho} + \varepsilon_t.$$

The test statistic and hypotheses are the same as for the standard DF test. However, the critical values will differ depending on the number of lags  $\rho$  included in the auxiliary regression. In this paper, all the ADF tests are calculated with a single lag, i.e.  $\rho = 1$ , as the tests do not yield notably different statistics for longer lags.

### *Ljung-Box Q-Test*

The Ljung-Box Q-test is a *portmanteau test* that evaluates the presence for autocorrelation in the time series. Thus, the test considers for autocorrelation across several lags  $\rho$ , rather than evaluating autocorrelation effects at a particular lag, for a given time series. The null hypothesis states that no autocorrelation is present, while the alternative hypothesis is more loosely specified. According to [Ljung and Box \(1978\)](#), the test statistic is given as:

$$Q_{LB} = T(T+2) \sum_{k=1}^{\rho} \frac{r_k^2}{T-k},$$

where  $T$  is the sample size,  $r_k$  is the sample autocorrelation at lag  $k$ , and  $\rho$  is the number of lags being tested. Under the null hypothesis, the  $Q_{LB}$ -statistic follows a  $\chi_\rho^2$  distribution. Generally, large values of  $Q_{LB}$  indicate autocorrelation in the time series. Formally, the hypotheses are:

$H_0$  : No autocorrelation is present, i.e., the data is independently distributed.

$H_1$  : Autocorrelation is present, i.e., the data is not independently distributed.



## A.7 MODEL PERFORMANCE MEASURE

We apply the pseudo-R-squared metric, denoted  $R^1(\tau)$ , to assess the goodness-of-fit of the submodels.

**Goodness-of-Fit for Quantile Regression:**  $R^1(\tau)$

$R^1(\tau)$  measures the relative success of quantile regression models at a specific quantile, and thus creates a *local measure* for the goodness-of-fit for a particular quantile (Koenker & Machado, 1999). In contrast, the conventional goodness-of-fit measures *R-squared*, denoted  $R^2$ , and *adjusted R-squared*, denoted  $R_{adj}^2$ , are *global measures* that estimates model performance over the entire conditional distribution.

Using a similar notation as in Section 5.2, consider the following linear model for a conditional quantile regression model for a given quantile  $\tau$  and price observation  $y_t$ :

$$F_{y_t}^{-1}(\tau | \mathbf{x}_t) = \mathbf{x}_t \boldsymbol{\beta}_\tau + F_\varepsilon^{-1}(\tau) = \mathbf{x}_{t1} \boldsymbol{\beta}_{\tau1} + \mathbf{x}_{t2} \boldsymbol{\beta}_{\tau2} + F_\varepsilon^{-1}(\tau),$$

where  $\hat{\boldsymbol{\beta}}_\tau$  denotes the least squares estimator of the model  $F_{y_t}^{-1}(\tau | \mathbf{x}_t)$ , and  $\tilde{\boldsymbol{\beta}}_\tau = (\tilde{\boldsymbol{\beta}}_{\tau1}, \mathbf{0})$  denotes the estimator under the  $q$ -dimensional linear restriction that  $\mathbf{H}_0 : \boldsymbol{\beta}_{\tau2} = \mathbf{0}$ . We may denote two minimising function; for an unconstrained problem, denoted  $\hat{V}(\tau)$ , and for a corresponding constrained problem, denoted  $\tilde{V}(\tau)$ . Moreover, let  $\hat{\boldsymbol{\beta}}_\tau$  and  $\tilde{\boldsymbol{\beta}}_\tau$  denote the corresponding minimisers for the unconstrained and constrained problem, respectively. Then, according to Koenker and Machado (1999), the pseudo-R-squared is given as:

$$R^1(\tau) = 1 - \frac{\hat{V}(\tau)}{\tilde{V}(\tau)} \quad : \quad \hat{V}(\tau) = \min_{\{\mathbf{b} \in \mathbb{R}^p\}} \sum_{t=1}^T \rho_\tau(y_t - \mathbf{x}_t \mathbf{b}), \quad \tilde{V}(\tau) = \min_{\{\mathbf{b} \in \mathbb{R}^{p-q}\}} \sum_{t=1}^T \rho_\tau(y_t - \mathbf{x}_{1t} \mathbf{b}_1).$$

Here,  $\rho_\tau$  is specified as in Eq. 9 in Section 5.2. Equivalent to  $R^2$  and  $R_{adj}^2$ ,  $R^1(\tau)$  lies between 0 and 1, where large values indicate a better fit than small values. For a complete discussion of  $R^1(\tau)$ , see Koenker and Machado (1999).

**Regression Equation Specification Error Test for Quantile Regression (QRESET)**

Based on the RESET introduced by Ramsey (1969), the regression equation specification error test for quantile regression (QRESET) by Otsu (2009) evaluates whether or not a quantile regression model is linear in its covariates. The test is an auxiliary regression that adds powers to the fitted dependent variables  $\hat{y}$  up to order  $p \geq 2$  and check the significance of those added covariates. Thus, a particular version of the test is denoted QRESET( $p$ ).

While the RESET focuses on the misspecification for the conditional mean function  $\mathbb{E}[y|X]$ , the QRESET( $p$ ) focuses on the conditional quantile function  $F^{-1}(\tau | X)$ . A deviation from the linear functional form might be estimated by performing a Taylor expansion around  $X = \mathbf{0}$ , and subsequently investigating the significance of higher-order polynomials of  $X$  resulting from the Taylor expansion. However, when the dimension of  $X$  is large, it is impractical to include all the components that appear in the  $p^{\text{th}}$  order polynomial of the Taylor expansion (Otsu, 2009). This problem may be prevented by applying the approach from RESET – substituting the effect of the  $p^{\text{th}}$  order polynomials by the  $p^{\text{th}}$  power of the fitted dependent variable  $\hat{y}^p = (X\hat{\beta}_\tau)^p$ . Using the notation of Otsu (2009), this is written as:

$$\text{approximate } F^{-1}(\tau | X) \text{ by } X\beta + (\Omega_p)^\top \alpha_p^\tau,$$

where  $\Omega_p = (\hat{y}^2, \dots, \hat{y}^p)^\top$  and  $\alpha_{\tau p}$  is a  $(p - 1) \times 1$  parameter vector for  $\tau$ . Based on this approximation, the QRESET( $p$ ) for the linear specification against the  $p$ -th order polynomial is defined as the following joint significance test:

$$H_0 : \alpha_{\tau p} = 0.$$

The procedure employed to implement the QRESET( $p$ ) in this study is specified as follows.

- 1) Compute  $\hat{\beta}_\tau$  by Eq. 9 in Section 5.2, and set  $\hat{y}_t = \mathbf{x}_t \hat{\beta}_\tau$  for  $t = 1, \dots, T$
- 2) Using  $\{\hat{y}_t\}_{t=1}^T$ , compute

$$(\hat{\beta}_{\tau p}, \hat{\alpha}_{\tau p}) = \arg \min_{(b, a) \in \mathbb{R}^{n+1} \times \mathbb{R}^{p-1}} \frac{1}{T} \sum_{t=1}^T \rho_\tau(y_t - \mathbf{x}_t b - (\Omega_{pt})^\top a),$$

and estimate the asymptotic variance of  $\hat{\alpha}_{\tau p}$ , where  $\rho_\tau$  is defined equivalent to Eq. 9 in Section 5.2.

- 3) Using  $\hat{\alpha}_{\tau p}$  and its asymptotic variance estimator, denoted by  $\hat{V}_{\tau p}$ , the following Wald test statistic is computed for step 4)

$$W = \hat{\alpha}_{\tau p} (\hat{V}_{\tau p})^{-1} \hat{\alpha}_{\tau p}.$$

Thereafter, conduct the significance test by the  $\chi_{p-1}^2$  critical value. If  $p = 2$ , then we can apply the t-value for  $\hat{\alpha}_{\tau 2}$ .

- 4) If  $H_0$  is rejected, there is evidence that the current linear functional form for the quantile regression is a misspecification.

The QRESET is computationally cheap, as the test only involves one additional quantile regression. However, the QRESET may provide lower significance when the polynomial approximation is not exact (Otsu, 2009). On a side note, we employ the bootstrap method of Parzen et al. (1994) to compute the asymptotic variance estimator  $(\hat{V}_{\tau p})^{-1}$ , using  $n = 1000$  replications.

## A.9 DATA SOURCES

A variety of online sources have been used for data collection. In every case, we have tried to find reliable and publicly available sources. In the following, brief descriptions of the employed data sources are presented.

### **fiskeridir.no**

The Directorate of Fisheries is a part of the Norwegian government. It promotes sustainable and user-oriented management of Norway's marine resources and the marine environment. They monitor the Norwegian salmon farming industry closely, and collect large amounts of data. We use their services to obtain important data on supply indicators, such as smolt release, feed consumption, standing biomass, and harvest volume.

### **fao.org**

The Food and Agriculture Organization of the United Nations, dating back to 1945, is an international agency that leads international efforts to defeat hunger. We use their services to obtain price indices for meat.

### **lusedata.no**

Lusedata provides data on the sea lice levels and sea temperatures in Norwegian waters, both on a county and nationwide level. Their mission is to aid the Norwegian aquaculture industry against high levels of sea lice by providing easily accessible statistics, knowledge, and guidance on the subject. Typically, salmon farming companies and research communities access their data for risk management and research purposes. The data is reported on a weekly or monthly basis to The Norwegian Food Safety Authority (Sjømat Norge, 2016). Therefore, we have used `lusedata.no` to collect data on sea lice levels, sea lice treatments, and sea temperatures.

### **marineharvest.com**

Marine Harvest is the world's largest producer of Atlantic salmon. In their quarterly and annual reports, they include key information about global harvest volumes and consumption of salmon. We have used these reports to obtain data for consumption of salmon split by region and global harvest volumes (excl. Norway).

### **nasdaqomx.com**

Nasdaq is a global financial services provider, founded in 1971. It is a product of a merge between the U.S.-based NASDAQ, and OMX, a leading exchange operator in the Nordics. We have used Nasdaq to obtain time series for the NQSALMON.

### **oanda.com**

OANDA is a foreign exchange company operating since 1996. On `oanda.com`, a range of different services is provided, including currency conversion, retail foreign exchange trading, foreign currency transfers, and currency exchange information (Oanda, 2016). Due to OANDA's comprehensive and publicly accessible data, we have used `oanda.com` to collect foreign exchange data for the exchange rates NOK/EUR.

### **quandl.com**

Founded in 2012, Quandl is a platform providing financial and economic data, primarily for professionals in hedge funds, asset management, and investment banks. The data is sourced from over 500 publishers including the UN, CLS Group, and Zacks. Quandl's data

sets are easily accessible at `quandl.com`, and through an API via programming languages such as R, Python, and Matlab. Most of the offered data sets at Quandl are publicly available. We have used `quandl.com` for collecting data series on shrimp and beef prices.

#### **seafood.no**

The Norwegian Seafood Council is wholly owned by the Norwegian Directorate of Fisheries. One of its objectives is to increase the value of Norwegian seafood resources, mainly through the development of new markets. We have used it to collect data on trout prices.

## A.10 BACKGROUND: IMPORTANT ASPECTS OF THE SALMON FARMING INDUSTRY

### *Important Producers and Production Growth*

The salmon farming space consists of a few dominating producing countries, as several unique conditions have to be in place to achieve profitable production. Among the most important requirements are stable and relatively low sea temperatures, in addition to sheltered coast lines. Accordingly, the most significant salmon producing countries are Norway, Chile, Canada, Scotland, and the Faroes (Solibakke, 2012). Of the five, Norway and Chile are the most important producers, producing 54 % and 23% of the volumes of farmed Atlantic salmon, respectively.

During the last decades, aquaculture has been the world's fastest growing animal-based food sector (FAO, 2017b). This is primarily due to rapid productivity growth, which has reduced production costs as knowledge has been transferred from the terrestrial sector (Anderson, 2002; Asche, Hansen, Tveteras, & Tveteras, 2009; Kumar & Engle, 2016). Within aquaculture, salmon is one of the most successful species in terms of production growth, outperforming the average production growth in the sector (Brækkan & Thyholdt, 2014), and it is the second most valuable species after shrimp (FAO, 2017b). In numbers, the aquaculture production of Atlantic salmon has increased from a few thousand tons in 1980 to beyond two million tons in 2016 (FAO, 2017b). This growth is partly illustrated in Figure 20, presenting global harvest volumes of Atlantic salmon from 1998 through 2016. Going forward, a number of authors indicate that salmon production will continue to increase, as productivity and demand are expected to continue to grow (Kobayashi et al., 2015; Torrissen et al., 2011). However, a further increase in production also requires a certain advance in technology, as the salmon farming industry has reached a level of production where biological boundaries are being pushed (Marine Harvest, 2016).

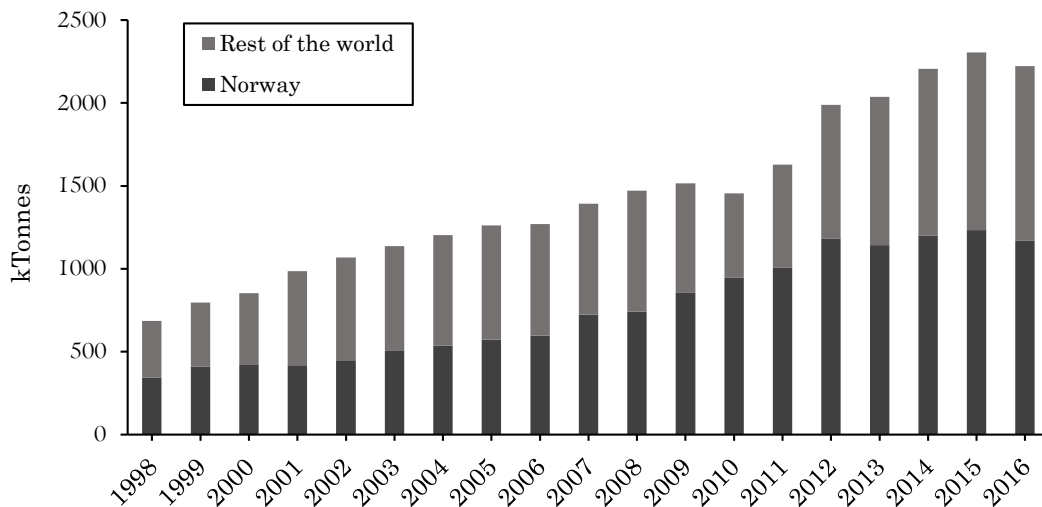


Figure 20: Historical harvest volumes of Atlantic salmon.

### *Externalities and Their Impact on Production Growth*

During the last decade, the salmon farming industry has been struggling with frequent occurrence of sea lice and diseases. The most notable event, with respect to impact on global harvest volumes of Atlantic salmon, is probably the disease outbreak in Chile in 2008. This outbreak destroyed two-thirds of the Chilean production from 2009 through 2011, and had a significant impact on global production volumes, as seen in Figure 20. Another notable event, also in Chile, resulted in Chilean production volumes falling close to 20% YoY in 2016 (Terazono, 2016). This particular event was caused by a natural toxic algae bloom, and was

not a consequence of farmers overburdening the surrounding environment. Nevertheless, the two aforementioned incidents underpin how fragile the salmon farming industry is to external factors.

To better meet the risks associated with sea lice and diseases, the industry is continuously developing technologies and procedures to prevent large production losses. Vaccines were introduced in the late 1980s, and have reduced the need for antibiotics. Also, there has been an increased use of farm monitoring technologies, in addition to specialists such as veterinarians and marine biologists (Asche & Sikveland, 2015). Furthermore, the Norwegian Government grants *development licenses* to projects with focus on innovations that can solve the *environmental issues* faced by the industry (Directorate of Fisheries, 2016).<sup>36</sup>

#### *Regulations for Sustainable Growth of the Industry*

In addition to encouraging innovation, governments are regulating the production through license systems. The aim is to achieve better control on parasite and disease outbreaks, and establish a basis for sustainable development of the industry. While similar production technology is used across the salmon farming countries, the government efforts designed to mitigate externalities differ significantly, and so do the effects from the efforts (Gibbs, 2009; Osmundsen, Almklov, & Tveterås, 2017). In the UK and Canada, strict regulations have been imposed with rather unsuccessful results. In comparison, Chile has a significantly more liberal regulation regime. However, Chilean farmers have also been struggling with disease outbreaks that likely could have been avoided (Asche et al., 2009). Over the period from 2005 to 2015, Norway's production volume has increased by 115%, followed by Chile, the UK, and Canada with production growth of 53%, 38%, and 25%, respectively. Clearly, Norwegian regulations have been the most successful, as all of the countries have had increased production as their first priority (Osmundsen et al., 2017).<sup>37</sup>

#### *The Production Process and Supply Chain of Salmon*

The salmon farming production cycle typically lasts two to three years, depending on the attractiveness of local conditions. During the *smoltification* the first 10 to 16 months, salmon eggs are fertilised and spawned, before they are transferred to fresh water. Here, the newly spawned egg, that is a *fry*, becomes a *parr*, and subsequently a *smolt*. Next, the *smolts* are transferred to seawater cages, where they grow for 14 to 22 months and typically gain four to five kg (Sandvold & Tveterås, 2014). Finally, the salmon is moved to a factory for slaughtering, gutting, and filleting to become a marketable product. The marketed products range from affordable private label products to expensive branded products. This diversity has made salmon available to a vast group of consumers, with wide-ranging preferences and income. In turn, this has led to an increased demand for salmon (Landazuri-Tveteraas, Asche, Gordon, & Tveteraas, 2017).

Regarding the supply chain for salmon, there have been profound changes during the past decades. The market has become global, and the focus from retailers on efficient logistics has resulted in increased coordination upstream and a growing number of large production companies (Asche, Roll, Sandvold, Sørvig, & Zhang, 2013; Straume, 2014). Further-

---

<sup>36</sup>*Development license*: Gives salmon farmers the opportunity to increase their production in return for innovations that benefit the entire industry.

*Environmental issues*: A particular problem for Norwegian salmon farmers is the lack of suitable locations for farming. A high concentration of farms in a given area increases the probability for high sea lice levels and disease outbreaks. Hence, most applications for development licenses that have been submitted to the Directorate of Fisheries concern innovation of the farming facilities, such as submerged, closed, and offshore facilities (Directorate of Fisheries, 2017b).

<sup>37</sup>It is important to note that there are several other factors affecting the production growth, such as technological innovations, the cost of production factors, and market access (Anderson, 2002; Asche, 2008; Kumar & Engle, 2016). However, the strong growth seen within Norwegian salmon farming does at the very least indicate that Norwegian regulations have been somewhat successful in the past.

more, more advanced transaction methods have been introduced, such as options and futures contracts, making salmon the species with one of the most diverse transaction modes compared to other seafoods (Straume, 2014).

#### *Financial Contracts*

Along with the development of more advanced transaction methods, Fish Pool has been established. This is an international marketplace for buying and selling of financial salmon contracts, and its majority owner is Oslo Stock Exchange, with a stake of 97%. As of 2011, the participants at Fish Pool consisted of 32% financial players, whereas 68% of the participants were a part of the salmon farming value chain (Rolland & Johnsen, 2014). Although Fish Pool aims to be the preferred global exchange for price hedging of fish and seafood products, its 2015 trading volumes totalled only 55 500 tonnes, or 2.4% of the total harvest volume that particular year (Fish Pool ASA, 2015). According to a study carried out by Rolland and Johnsen (2014), many salmon farmers hesitate to engage in contracts through Fish Pool. This doubt is supposedly attributed to low liquidity and high spreads, combined with too little knowledge of financial markets among the salmon farmers.

## REFERENCES

---

- Aas, K., & Dimakos, X. K. (2004). *Statistical modelling of financial time series: An introduction* (Tech. Rep.). Norwegian Computing Center (NR). Retrieved from <https://www.nr.no/files/samba/bff/SAMBA0804.pdf>
- Acosta-González, E., & Fernández-Rodríguez, F. (2007). Model selection via genetic algorithms illustrated with cross-country growth data. *Empirical Economics*, 33(2), 313–337. doi: 10.1007/s00181-006-0104-3
- Akram, Q. F. (2009). Commodity prices, interest rates and the dollar. *Energy Economics*, 31(6), 838–851. Retrieved from <http://dx.doi.org/10.1016/j.eneco.2009.05.016> doi: 10.1016/j.eneco.2009.05.016
- Alexander, C. (2008a). *Market Risk Analysis, Practical Financial Econometrics (The Wiley Finance Series) (Volume II)*. West Sussex, England: John Wiley & Sons Ltd.
- Alexander, C. (2008b). *Market Risk Analysis, Quantitative Methods in Finance (The Wiley Finance Series) (Volume I)*. West Sussex, England: John Wiley & Sons Ltd.
- Anderson, J. (2002). Aquaculture and the Future: Why Fisheries Economists Should Care. *Marine Resource Economics*, 17(2), 133–151. doi: 10.2307/42629357
- Ankamah-Yeboah, I., Nielsen, M., & Nielsen, R. (2016). Price formation of the salmon aquaculture futures market. *Aquaculture Economics and Management*, 7305(September), 24. doi: 10.1080/13657305.2016.1189014
- Asche, F. (2008). Farming the sea. *Marine Resource Economics*, 23(4), 527–547. doi: 10.2307/42629678
- Asche, F., Dahl, R. E., & Steen, M. (2015). Price Volatility in Seafood Markets: Farmed vs. Wild Fish. *Aquaculture Economics & Management*, 19(3), 316–335. Retrieved from <http://www.tandfonline.com/doi/full/10.1080/13657305.2015.1057879> doi: 10.1080/13657305.2015.1057879
- Asche, F., Hansen, H., Tveteras, R., & Tveteras, S. (2009). The Salmon Disease Crisis in Chile. *Marine Resource Economics*, 24(4), 405–411. doi: 10.5950/0738-1360-24.4.405
- Asche, F., & Misund, B. (2016). Hedging Efficiency of Atlantic Salmon Futures. *Aquaculture Economics & Management*, 20(4), 368–381. Retrieved from <http://www.tandfonline.com/doi/pdf/10.1080/13657305.2016.1212123> doi: 10.1080/13657305.2016.1212123
- Asche, F., Misund, B., & Oglend, A. (2016a). Determinants of the Atlantic salmon futures risk premium. *Journal of Commodity Markets*, 2(1), 6–17. doi: 10.1016/j.jcomm.2016.07.001
- Asche, F., Misund, B., & Oglend, A. (2016b). The Spot-Forward Relationship in the Atlantic Salmon Market. *Aquaculture Economics & Management*, 20(2), 222–234. Retrieved from <http://dx.doi.org/10.1080/13657305.2016.1156192> doi: 10.1080/13657305.2016.1156192
- 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. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/13657305.2013.812156> doi: 10.1080/13657305.2013.812156
- Asche, F., & Sikveland, M. (2015). The Behavior of Operating Earnings in the Norwegian Salmon Farming Industry. *Aquaculture Economics & Management*, 19(3), 301–315. Retrieved from <http://www.tandfonline.com/doi/full/10.1080/13657305.2015.1057880> doi: 10.1080/13657305.2015.1057880
- Ates, A. M., Lusk, J. L., & Brorsen, B. W. (2016). *Forecasting Meat Prices using Consumer Expectations from the Food Demand Survey (FoodS)* (Tech. Rep.). Southern Agricultural Economics Association. Retrieved from <http://econpapers.repec.org/>



- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)* (Vol. 4) (No. 356). Secaucus, NJ, USA: Springer-Verlag New York, Inc. Retrieved from <http://www.library.wisc.edu/selectedtocs/bg0137.pdf> doi: 10.1641/B580519
- Brækkan, E. H., & Thyholdt, S. B. (2014). The bumpy road of demand growth — an application to Atlantic salmon. *Marine Resource Economics*, 29(4), 339–350. doi: 10.1086/678927
- Broadhurst, D., Goodacre, R., Jones, A., Rowland, J. J., & Douglas, B. K. (1997). Genetic algorithms as a method for variable selection in multiple linear regression and partial least squares regression, with applications to pyrolysis mass spectrometry. *Analytica Chimica Acta*, 348, 71–86. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0003267097000652> doi: [http://dx.doi.org/10.1016/S0003-2670\(97\)00065-2](http://dx.doi.org/10.1016/S0003-2670(97)00065-2)
- Brooks, C. (2008). *Introductory Econometrics for Finance* (2nd ed.). New York City, New York: Cambridge University Press.
- Bunn, D., Andresen, A., Chen, D., & Westgaard, S. (2013). Analysis and forecasting of electricity price risks with quantile factor models. *The Energy Journal*, 37(1), 1–34. Retrieved from [http://www.ceem-dauphine.org/assets/dropbox/Derek\\_BUNN.pdf](http://www.ceem-dauphine.org/assets/dropbox/Derek_BUNN.pdf) doi: DOI:10.5547/01956574.37.1.dbun
- Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1996). *The Econometrics of Financial Markets*. Princeton University Press.
- Campos, J., Ericsson, N. R., & Hendry, D. F. (2005). General-to-Specific Modeling: An Overview and Selected Bibliography. *International Finance Discussion Papers*, 838, 1–94. Retrieved from <https://www.federalreserve.gov/pubs/ifdp/2005/838/ifdp838.pdf> doi: 10.1.1.119.8314
- Chatfield, C. (2000). *Time-Series Forecasting*. doi: 10.1198/tech.2002.s718
- Christoffersen, P. F. (1998). Evaluating Interval Forecasts. *International Economic Review*, 39(4), 841–862. Retrieved from <http://ideas.repec.org/a/ier/iecrev/v39y1998i4p841-62.html> doi: 10.2307/2527341
- Coleman, L. (2012). Explaining crude oil prices using fundamental measures. *Energy Policy*, 40(1), 318–324. Retrieved from <http://dx.doi.org/10.1016/j.enpol.2011.10.012> doi: 10.1016/j.enpol.2011.10.012
- Dahl, R. E., & Oglend, A. (2014). Fish Price Volatility. *Marine Resource Economics*, 29(4), 305–322. Retrieved from <http://www.jstor.org/stable/10.1086/678925> doi: 10.1086/678925
- 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(366), 427. Retrieved from <http://www.jstor.org/stable/2286348?origin=crossref> doi: 10.2307/2286348
- Diebold, F. X., Gunther, T. A., & Tay, A. S. (1997). Evaluating Density Forecasts. *International Economic Review*, 39, 863–883. doi: 10.2307/2527342
- Directorate of Fisheries. (2016). *Utviklingstillatelser*. Retrieved 2017-03-22, from <http://www.fiskeridir.no/Akvakultur/Tildeling-og-tillatelser/Saertillatelser/Utviklingstillatelser>
- Directorate of Fisheries. (2017a). *Biomassestatistikk*. Retrieved 2017-03-22, from <http://www.fiskeridir.no/Akvakultur/Statistikk-akvakultur/Biomassestatistikk>
- Directorate of Fisheries. (2017b). *Oversikt over søknader om utviklingstillatelser 10. mars 2017*. Retrieved 2017-03-22, from <http://www.fiskeridir.no/Akvakultur/Tildeling-og-tillatelser/Saertillatelser/>

- Down, K., Blake, D., & Cairns, A. (2004). Long-Term Value at Risk. *The Journal of Risk Finance*, 5(2), 52–57. Retrieved from <http://www.emeraldinsight.com/doi/10.1108/eb022986> doi: 10.1108/eb022986
- Engle, R. F., & Manganelli, S. (2004). CAViaR. *Journal of Business & Economic Statistics*, 22(4), 367–381. Retrieved from <http://www.tandfonline.com/doi/abs/10.1198/073500104000000370> doi: 10.1198/073500104000000370
- Fahrmeir, L., Kneib, T., Lang, S., & Marx, B. (2013). *Regression: Models, Methods and Applications*. Berlin, Heidelberg: Springer-Verlag. Retrieved from <http://link.springer.com/book/10.1007%2F978-3-642-34333-9> doi: 10.1007/978-3-642-34333-9
- FAO. (2017a). *FAO Fisheries & Aquaculture - Cultured Aquatic Species Information Programme - Oncorhynchus mykiss (Walbaum, 1792)*. Retrieved 2017-05-06, from [http://www.fao.org/fishery/culturedspecies/Oncorhynchus\\_mykiss/en](http://www.fao.org/fishery/culturedspecies/Oncorhynchus_mykiss/en)
- FAO. (2017b). *FAO Fisheries & Aquaculture - Species Fact Sheets - Salmo salar (Linnaeus, 1758)*. Retrieved 2017-02-10, from <http://www.fao.org/fishery/species/2929/en>
- Fish Pool ASA. (2015). *Årsrapport 2015 Fish Pool ASA*. Retrieved from <http://fishpool.eu/wp-content/uploads/2014/03/{Å}rsrapport-2015-Fish-Pool-ASA.pdf>
- Fish Pool ASA. (2017). *Nasdaq Salmon Index*. Retrieved 2017-02-22, from <http://fishpool.eu/price-information/reports/nasdaq-salmon-index-archive/>
- Frost, V., & Molt, K. (1998). Use of a genetic algorithm for factor selection in principal component regression. *Journal of Near Infrared Spectroscopy*, 6, A185– A191. Retrieved from <http://journals.sagepub.com/doi/abs/10.1255/jnirs.192> doi: <https://doi.org/10.1255/jnirs.192>
- Gen, M., & Cheng, R. (1999). *Genetic Algorithms and Engineering Optimization* (1st ed.). John Wiley & Sons, Inc.
- Gibbs, M. T. (2009). Resilience: What is it and what does it mean for marine policymakers? *Marine Policy*, 33(2), 322–331. doi: 10.1016/j.marpol.2008.08.001
- Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning* (1st ed.). Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc. Retrieved from <http://dl.acm.org/citation.cfm?id=534133>
- Gu, G., & Anderson, J. (1995). Deseasonalized state-space time series forecasting with application to the US salmon market. *Marine Resource Economics*, 10(2), 171–185. Retrieved from <https://www.jstor.org/stable/42629109> doi: 10.1086/521238
- Guttormsen, A. (2008). Faustmann in the Sea: Optimal Rotation in Aquaculture. *Marine Resource Economics*, 23, 401–410.
- Guttormsen, A. G. (1999). Forecasting weekly salmon prices: Risk management in fish farming. *Aquaculture Economics & Management*, 3(2), 159–166. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/13657309909380242> doi: 10.1080/13657309909380242
- Hagfors, L. I., Bunn, D., Kristoffersen, E., Staver, T. T., & Westgaard, S. (2016). Modeling the UK electricity price distributions using quantile regression. *Energy*, 102, 231–243. Retrieved from <http://dx.doi.org/10.1016/j.energy.2016.02.025> doi: 10.1016/j.energy.2016.02.025
- Hamilton, J. D. (1989). *A new approach to the economic analysis of nonstationary time series* (Vol. 57) (No. 2). doi: 10.2307/1912559
- Härdle, W. K., & Simar, L. (2014). *Applied Multivariate Statistical Analysis* (4th ed.). Berlin, Heidelberg: Springer-Verlag. doi: 10.1007/978-3-662-45171-7

- Harrell, F. E. (2015). *Regression Modeling Strategies* (2nd ed.). New York: Springer-Verlag. Retrieved from <http://www.springer.com/gb/book/9781441929181> doi: 10.1007/978-1-4757-3462-1
- Harri, A., Nalley, L., & Hudson, D. (2009). The relationship between oil, exchange rates, and commodity prices. *Journal of Agricultural and Applied Economics*, 2(August), 501–510. Retrieved from [http://www.researchgate.net/publication/46534254\\_The\\_Relationship\\_between\\_Oil\\_Exchange\\_Rates\\_and\\_Commodity\\_Prices/file/d912f512392af7eb13.pdf](http://www.researchgate.net/publication/46534254_The_Relationship_between_Oil_Exchange_Rates_and_Commodity_Prices/file/d912f512392af7eb13.pdf) doi: 10.1017/S1074070800002959
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning* (2nd ed., Vol. 1). Springer-Verlag. Retrieved from <http://www.springer.com/gp/book/9780387848570> doi: 10.1007/b94608
- Hellerstein, J. M. (2008). Quantitative Data Cleaning for Large Databases. *United Nations Economic Commission for Europe*, 42. Retrieved from <http://db.cs.berkeley.edu/jmh/papers/cleaning-unece.pdf>
- Holland, J. H. (1992). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. Cambridge, MA, USA: MIT Press.
- Hoover, K. D., & Perez, S. J. (2000). Three attitudes towards data mining. *Journal of Economic Methodology*, 7(2), 195–210. doi: 10.1080/13501780050045083
- Hyndman, R. (2013). *Facts and fallacies of the AIC*. Retrieved 2017-05-20, from <https://robjhyndman.com/hyndsight/aic/>
- Jarque, C. M., & Bera, A. K. (1980). Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals. *Economics Letters*, 6(3), 255–259. doi: 10.1016/0165-1765(80)90024-5
- Jolliffe, I. T. (2002). Principal Component Analysis, Second Edition. *Encyclopedia of Statistics in Behavioral Science*, 30(3), 487. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/0470013192.bsa501/full> doi: 10.2307/1270093
- Joshua, I. A., Victor, C., & Fernandez-Val, I. (2013). Quantile Regression under Misspecification, with an Application to the U . S . Wage Structure. *Working Papers*, 74(2), 539–563. doi: 10.1111/j.1468-0262.2006.00671.x
- J.P. Morgan. (1996). *RiskMetrics - Technical Document* (Tech. Rep.). New York City, New York: Morgan Guaranty Trust Company; Reuters Ltd.
- Kobayashi, M., Msangi, S., Batka, M., Vannuccini, S., Dey, M. M., & Anderson, J. L. (2015). Fish to 2030: The Role and Opportunity for Aquaculture. *Aquaculture Economics & Management*, 19(3), 282–300. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/13657305.2015.994240> doi: 10.1080/13657305.2015.994240
- Koenker, R. (2017). *Quantile regression in R: A Vignette* (Vol. 15; Tech. Rep.). Retrieved from <https://cran.r-project.org/web/packages/quantreg/vignettes/rq.pdf>
- Koenker, R., & Bassett, G. (1978). Regression Quantiles. *Econometrica*, 46(1), 33–50. Retrieved from <http://www.jstor.org/stable/pdf/1913643.pdf>
- Koenker, R., & Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, 15(4), 143–156. doi: 10.1257/jep.15.4.143
- Koenker, R., & Machado, J. a. F. (1999). Goodness of Fit and Related Inference Processes for Quantile Regression. *Journal of the American Statistical Association*, 94(November 2014), 1296–1310. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/01621459.1999.10473882> doi: 10.1080/01621459.1999.10473882
- Kowalski, C. (2016). *Inverse Relationship Between the Dollar and Commodities*. Retrieved 2017-03-22, from <https://www.thebalance.com/how-the-dollar-impacts>

-commodity-prices-809294

- 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. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/23308249.2015.1112357> doi: 10.1080/23308249.2015.1112357
- Kupiec, P. (1995). Techniques for Verifying the Accuracy of Risk Measurement Models. *The Journal of Derivatives*, Vol. 3, 73–84. doi: 10.3905/jod.1995.407942
- Kutner, M. H., Nachtsheim, C., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models*. Retrieved from [http://books.google.fr/books?id=0xqCAAACAAJ{&}dq=intitle:Applied+linear+statistical+models+djvu{&}hl={&}cd=1{&}source=gbs\\_api](http://books.google.fr/books?id=0xqCAAACAAJ{&}dq=intitle:Applied+linear+statistical+models+djvu{&}hl={&}cd=1{&}source=gbs_api)
- Landajo, M., De Andrés, J., & Lorca, P. (2008). Measuring firm performance by using linear and non-parametric quantile regressions. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 57(2), 227–250. doi: 10.1111/j.1467-9876.2007.00610.x
- Landazuri-Tveteraas, U., Asche, F., Gordon, D., & Tveteraas, S. (2017). Farmed fish to supermarket: Testing for price leadership and price transmission in the salmon supply chain. *Aquaculture Economics and Management*, 0(0), 1–19. Retrieved from <http://dx.doi.org/10.1080/13657305.2017.1284943> doi: 10.1080/13657305.2017.1284943
- Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297–303. Retrieved from <http://biomet.oxfordjournals.org/content/65/2/297.abstract> doi: 10.1093/biomet/65.2.297
- Mao, K. Z. (2004). Orthogonal Forward Selection and Backward Elimination Algorithms for Feature Subset Selection. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 34(1), 629–634. doi: 10.1109/TSMCB.2002.804363
- Marine Harvest. (2016). *Salmon Farming Industry Handbook 2016* (Tech. Rep.). Marine Harvest ASA. Retrieved from <http://www.marineharvest.com/globalassets/investors/handbook/2016-salmon-industry-handbook-final.pdf>
- MathWorks. (2017). *Conditional coverage independence test for value-at-risk (VaR) back-testing*. Retrieved 2017-05-15, from <https://se.mathworks.com/help/risk/cci.html>
- Mikosch, T. (1998). *Elementary Stochastic Calculus with Finance in View*. Singapore: World Scientific Publishing Co. Pte. Ltd.
- Milne, R. (2013, aug). *Norway sees Liu Xiaobo's Nobel Prize hurt salmon exports to China*. London, United Kingdom. Retrieved from <https://www.ft.com/content/ab456776-05b0-11e3-8ed5-00144feab7de>
- Morana, C. (2001). A semiparametric approach to short-term oil price forecasting. *Energy Economics*, 23(3), 325–338. doi: 10.1016/S0140-9883(00)00075-X
- Nau, R. F. (2017). *Stationarity and Differencing*. Retrieved 2017-05-13, from <https://people.duke.edu/~rnau/411diff.htm>
- Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, 59(2), 347–370. Retrieved from <https://www.jstor.org/stable/pdf/2938260.pdf>
- Oanda. (2016). *OANDA*. Retrieved from <http://www.oanda.com/>
- Oglend, A. (2013). Recent trends in salmon price volatility. *Aquaculture Economics & Management*, 17(3), 281–299. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/13657305.2013.812155> doi: 10.1080/13657305.2013.812155
- Oglend, A., & Sikveland, M. (2008a). The behaviour of salmon price volatility. *Marine Resource Economics*, 23(4), 507–526. Retrieved from <http://www.jstor.org/stable/42629677>.
- Oglend, A., & Sikveland, M. (2008b). Investigating Salmon Price Volatility. *IIFET 2008 Viet-*

- nam Proceedings*, 1–12. Retrieved from <https://core.ac.uk/download/pdf/10197226.pdf>
- Oreski, S., & Oreski, G. (2014). Genetic algorithm-based heuristic for feature selection in credit risk assessment. *Expert Systems with Applications*, 41(4 PART 2), 2052–2064. Retrieved from <http://dx.doi.org/10.1016/j.eswa.2013.09.004> doi: 10.1016/j.eswa.2013.09.004
- 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. Retrieved from <https://www.tandfonline.com/doi/full/10.1080/13657305.2017.1262476> doi: 10.1080/13657305.2017.1262476
- Otsu, T. (2009). RESET for quantile regression. *Test*, 18(2), 381–391. doi: 10.1007/s11749-008-0097-7
- Parzen, M., Wei, L. T., & Ying, Z. (1994). A resampling method based on pivotal estimating functions. *Oxford Journals*, 81(2), 341–350. Retrieved from <http://www.jstor.org/stable/pdf/2336964.pdf>
- Politi, M., Millot, N., & Chakraborti, A. (2012). The near-extreme density of intraday log-returns. *Physica A: Statistical Mechanics and its Applications*, 391(1-2), 147–155. doi: 10.1016/j.physa.2011.05.029
- Ramsey, J. B. (1969). Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 31(2), 350–371. Retrieved from <https://www.jstor.org/stable/2984219> doi: 6/96/58267
- Ratner, B. (2010). Variable selection methods in regression: Ignorable problem, outing notable solution. *Journal of Targeting, Measurement and Analysis for Marketing*, 18(1), 65–75. Retrieved from <http://www.palgrave-journals.com/doi/10.1057/jt.2009.26> doi: 10.1057/jt.2009.26
- Reklaitis, V. (2014, aug). *What Russian sanctions mean for salmon prices*. New York City, New York. Retrieved from <http://www.marketwatch.com/story/what-russian-sanctions-mean-for-salmon-prices-2014-08-08>
- Ripegut, H. (2017). *Derfor blir laksen du kjøper mye dyrere i år*. Retrieved from <http://www.nettavisen.no/na24/derfor-blir-laksen-du-kjper-mye-dyrere-i-ar/3423304681.html>
- Rolland, D., & Johnsen, A. (2014). *Opsjonshandel i Fish Pool*. Retrieved from <https://brage.bibsys.no/xmlui/bitstream/handle/11250/222845/Masterthesis98.pdf?sequence=1&isAllowed=y>
- Ruczinski, I. (2005). Linear Models with R. *Hypertension*, 67(2), 285–299. Retrieved from <http://doi.wiley.com/10.1111/j.1467-9868.2005.00502.x> doi: 10.1111/j.1467-9868.2005.00502.x
- Saghaian, S. H. (2010). The Impact of the Oil Sector on Commodity Prices: Correlation or Causation? *Journal of Agricultural and Applied Economics*, 42(03), 477–485. Retrieved from [http://www.journals.cambridge.org/abstract\\_S1074070800003667](http://www.journals.cambridge.org/abstract_S1074070800003667) doi: 10.1017/S1074070800003667
- Sandvold, H. N., & Tveterås, R. (2014). Innovation and Productivity Growth in Norwegian Production of Juvenile Salmonids. *Aquaculture Economics & Management*, 18(2), 149–168. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/13657305.2014.903313> doi: 10.1080/13657305.2014.903313
- Shen, S., Jiang, H., & Zhang, T. (2012). *Stock market forecasting using machine learning algorithms*. Retrieved from <http://cs229.stanford.edu/proj2012/ShenJiangZhang-StockMarketForecastingusingMachineLearningAlgorithms.pdf>
- Siedlecki, W., & Sklansky, J. (1989). A note on genetic algorithms for large-scale feature



selection. *Pattern Recognition Letters*, 10(5), 335–347. doi: 10.1016/0167-8655(89)90037

-8

- Sjømat Norge. (2016). *Lusedata*. Retrieved 2016-12-14, from <http://lusedata.no/>
- Solibakke, P. B. (2012). Scientific Stochastic Volatility Models for the Salmon Forward Market: Forecasting (Un-)Conditional Moments. *Aquaculture Economics & Management*, 16(3), 222–249. doi: 10.1080/13657305.2012.704618
- Straume, H.-M. (2014). Currency Invoicing in Norwegian Salmon Export. *Marine Resource Economics*, 29(4), 000–000. Retrieved from <http://www.bioone.org/doi/abs/10.1086/678930> doi: 10.1086/678930
- Taleb, N. N. (2009). *The Black Swan: The Impact of the Highly Improbable*. Random House.
- Tay, A. S., & Wallis, K. F. (1999). Density Forecasting : A Survey. *Journal of Forecasting*, 19(2000), 124–143.
- Terazono, E. (2016). *Chile's salmon sector hit by toxic algae bloom*. Retrieved from <https://www.ft.com/content/1978a158-12ac-11e6-91da-096d89bd2173>
- The Nasdaq Group Inc. (2016). *Rules for the Construction, Maintenance and Use of the NASDAQ Salmon Index - NQSALMON* (Tech. Rep.). Retrieved from <https://salmonprice.nasdaqomxtrader.com/NQSALMONRules160425.pdf>
- Tibshirani, R. (1996). *Regression Selection and Shrinkage via the Lasso* (Vol. 58) (No. 1). Retrieved from <http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.35.7574> doi: 10.2307/2346178
- Torrissen, O., Olsen, R. E., Toresen, R., Hemre, G. I., Tacon, A. G. J., Asche, F., ... Lall, S. (2011). Atlantic Salmon (*Salmo salar*): The 'Super-Chicken' of the Sea? *Reviews in Fisheries Science*, 19(3), 257–278. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/10641262.2011.597890> doi: 10.1080/10641262.2011.597890
- Tveteras, S., & Asche, F. (2008). International fish trade and exchange rates: an application to the trade with salmon and fishmeal. *Applied Economics*, 40(January 2015), 1745–1755. doi: 10.1080/00036840600905134
- Vukina, T., & Anderson, J. (1994). Price forecasting with state-space models of nonstationary time series: case of the Japanese salmon market. *Computers & Mathematics with Applications*, 27(5), 45–62. Retrieved from <http://www.sciencedirect.com/science/article/pii/0898122194900752> doi: 10.1016/0898-1221(94)90075-2
- Wooldridge, J. M. (2011). *Introductory Econometrics: A Modern Approach* (4th ed., Vol. 120-121). Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/21708159> doi: 10.1016/j.jconhyd.2010.08.009
- Wu, Y., & Liu, Y. (2009). Variable Selection in Quantile Regression. *Statistica Sinica*, 19, 801–817. Retrieved from <http://www.unc.edu/~yfliu/papers/qr2.pdf>
- Ye, M., Zyren, J., & Shore, J. (2005). A monthly crude oil spot price forecasting model using relative inventories. *International Journal of Forecasting*, 21(3), 491–501. doi: 10.1016/j.ijforecast.2005.01.001
- Zhao, J. H., Dong, Z. Y., Xu, Z., & Wong, K. P. (2008). A statistical approach for interval forecasting of the electricity price. *IEEE Transactions on Power Systems*, 23(2), 267–276. doi: 10.1109/TPWRS.2008.919309
- Zheng, J. X. (1998). A Consistent Nonparametric Test of Parametric Regression Models under Conditional Quantile Restrictions. *Cambridge University Press*, 14(1), 123–138.
- Zhou, M., Yan, Z., Ni, Y., Li, G. L., & Nie, Y. (2006). Electricity price forecasting with confidence-interval estimation through an extended ARIMA approach. *IEE Proceedings-Generation, Transmission and ...*, 151(3), 201–212. Retrieved from [http://digital-library.theiet.org/content/journals/10.1049/ip-gtd\\_20040098](http://digital-library.theiet.org/content/journals/10.1049/ip-gtd_20040098) doi: 10.1049/ip-gtd