Aaløkken, Maurits Mogenssøn
Sveinsson, Jørgen Andersen

# A Study of the Performance of Value-at-Risk Averaging and Expected Shortfall Averaging in the Nordic Power Futures Market

Master's thesis in Industrial Economics and Technology Management
Supervisor: Westgaard, Sjur
June 2019

Master's thesis

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

NTNU
Norwegian University of
Science and Technology

Aaløkken, Maurits Mogenssøn
Sveinsson, Jørgen Andersen

# A Study of the Performance of Value-at-Risk Averaging and Expected Shortfall Averaging in the Nordic Power Futures Market

Master's thesis in Industrial Economics and Technology Management
Supervisor: Westgaard, Sjur
June 2019

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

NTNU
Norwegian University of
Science and Technology

# Problem Description

The deregulation of the Nordic electricity market during the early 1990s was undertaken to create a more efficient market with increased security of supply. The highly volatile nature of the commodity strongly entails appropriate risk management measures to minimize the probability of huge losses. A recent reminder of the large fluctuations in the electricity market, is the case of the Norwegian power trader Einar Aas. However, despite several studies on risk management for other commodities, risk management within power trading, and especially within trading in the Nordic power futures market, has not been extensively researched.

In their 2014 paper, Nowotarski et al. (2014) strongly suggest further research on the topic of forecast averaging in the electricity spot price market, following the results of their study. To our knowledge, using an average of Value-at-Risk models and Expected Shortfall models respectively in the Nordic power futures market seems to be unexplored territory.

The purpose of this paper is threefold:

1. First, to compare the in-sample fit of well known univariate risk models, for both Value-at-Risk and Expected Shortfall, in the Nordic power futures market.
2. Second, to study the out-of-sample performance of the models.
3. Third, to investigate both the in-sample fit and the out-of-sample performance of equally weighted averages of the same risk models, to determine whether these simple average models are more adequate than the individual models alone for risk management in the Nordic power futures market.

# Preface

This thesis is original and independent work by Jørgen Andersen Sveinsson and Maurits Mogenssøn Aaløkken during the spring of 2019, and concludes our degree of Master of Science in Industrial Economics and Technology Management at the Norwegian University of Science and Technology (NTNU).

We would like to thank our supervisor, Professor Sjur Westgaard, at the Department of Industrial Economics and Technology Management at NTNU for his guidance and advice during the work of this thesis.

# Abstract

The purpose of this study is to investigate the performance of well known univariate risk models for both Value-at-Risk and Expected Shortfall in the highly volatile Nordic power futures market, and study whether simple averages of the same models used to calculate Value-at-Risk(VaR) and Expected Shortfall(ES) provide better results than the individual models.

The individual models used for calculating Value-at-Risk are normally-distributed GARCH, t-distributed GARCH, normally-distributed GJR-GARCH, t-distributed GJR-GARCH, quantile regression using normally-distributed GARCH, quantile regression using t-distributed GARCH, quantile regression using normally-distributed GJR-GARCH, quantile regression using t-distributed GJR-GARCH, quantile regression using RiskMetrics, RiskMetrics with Cornish Fisher and Filtered Historical Simulation using t-distributed GARCH. We use normally-distributed GARCH, t-distributed GARCH, RiskMetrics with Cornish Fisher and Filtered Historical Simulation using t-distributed GARCH to calculate Expected Shortfall.

The performance of the models used to predict Value-at-Risk is assessed using the Unconditional Coverage test of Kupiec and the Conditional Independence test of Christoffersen. The performance of the models used to predict Expected Shortfall is assessed using the backtesting procedure described by McNeil and Frey.

The results show that the simple average models chosen perform very well for Value-at-Risk, both in- and out-of-sample. The general tendency when backtesting VaR both in- and out-of-sample, is that the models including the quantile regression approach perform best among the individual models. T-distributed GARCH outperforms normally-distributed GARCH due to the fat tails of the Nordic power futures. Filtered Historical Simulation performs acceptable, while RiskMetrics with Cornish Fisher approximation is not able to correctly account for the fat tails. The simple average models perform just as good as the best individual models.

For the ES in-sample fit and out-of-sample backtesting, the conclusion is quite clear cut; the simple average models that includes Filtered Historical Simulation, GARCH with t-distribution and GARCH with normal-distribution perform very well. Among the individual models, GARCH with t-distribution is the only model performing acceptable. RiskMetrics with Cornish Fisher and GARCH-n tend to strongly underestimate the risk, while Filtered Historical Simulation tends to strongly overestimate the risk. The simple average models perform equally good or better than the best individual model.

## Sammendrag

I denne oppgaven undersøker vi hvordan kjente univariate risikomodeller brukt til Value-at-Risk(VaR) og Expected Shortfall(ES) presterer i det meget volatile markedet for fremtidskontrakter på strøm, for deretter å undersøke hvorvidt gjennomsnitt av forskjellige konstellasjoner av de samme modellene gir bedre resultater enn de individuelle modellene.

Modellene vi bruker for å kalkulere Value-at-Risk er normally-distributed GARCH, GARCH med t-distribution, normally-distributed GJR-GARCH, t-distributed GJR-GARCH, quantile regression ved bruk av normally-distributed GARCH, quantile regression ved bruk av t-distributed GARCH, quantile regression ved bruk av normally-distributed GJR-GARCH, quantile regression ved bruk av t-distributed GJR-GARCH, quantile regression ved bruk av RiskMetrics, RiskMetrics med Cornish Fisher og Filtered Historical Simulation ved bruk av t-distributed GARCH. Vi bruker normally-distributed GARCH, t-distributed GARCH, RiskMetrics med Cornish Fisher og Filtered Historical Simulation ved bruk av t-distributed GARCH for å kalkulere Expected Shortfall.

For å teste hvorvidt et gjennomsnitt av de utvalgte modellene gir bedre resultater enn det de samme modellene klarer på egenhånd, bruker vi Kupiecs Unconditional Coverage test og Christoffersens Conditional Independence test, mens prestasjonen til modellene brukt til å beregne Expected Shortfall, blir målt ved bruk av McNeil og Freys fremgangsmåte.

Resultatene våre viser at gjennomsnittsmodellene presterer veldig solid for Value-at-Risk, både for in-sample fit og for out-of-sample testing. Den generelle tendensen blant de individuelle modellene, er at modellene som inkluderer kvantilregresjon presterer best. T-distributed GARCH presterer bedre enn normally-distributed GARCH på grunn av de tykke halene i det nordiske markedet for fremtidskontrakter på strøm. Filtered Historical Simulation presterer greit, mens Cornish Fisher approksimasjonen til RiskMetrics ikke evner å ta riktig høyde for de tykke halene i det nordiske markedet for fremtidskontrakter på strøm. Gjennomsnittsmodellene presterer like godt som de beste individuelle modellene.

For Expected Shortfall in-sample fit og out-of-sample backtesting, er konklusjonen klar; gjennomsnittsmodellene som inkluderer Filtered Historical Simulation, t-distributed GARCH og normally-distributed GARCH presterer veldig bra. Blant de individuelle modellene, er t-distributed GARCH den eneste modellen som prester akseptabelt. RiskMetrics med Cornish Fisher og normally-distributed GARCH undervurderer risikoen i stor grad, mens Filtered Historical Simulation overvurderer risikoen i stor grad. Gjennomsnittsmodellene presterer like godt eller bedre enn den beste individuelle modellen.

# Contents

# List of Tables

# List of Figures

## Acronyms

**ADFtest** Augmented Dickey-Fueller test

**ARCH** Autoregressive Conditional Heteroskedasticity model

**ASL** Achieved Significance Level

**CAC 40** Cotation Assistée en Continu 40 Index

**CC** Conditional Coverage

**CF** Cornish-Fisher

**CI** Conditional Independence

**CVaR** Conditional Value-at-Risk

**DAX 30** Deutscher Aktienindex 30 Index

**EGARCH** Exponential Generalized Autoregressive Conditional Heteroskedasticity model

**EKurt** Excess Kurtosis

**EPF** Electricity Price Forecasting

**ES** Expected Shortfall

**ETL** Expected Tail Risk

**EVT** Extreme Value Theory

**EWMA** Exponentially Weighted Moving Average

**FHS** Filtered Historical Simulation

**FTSE100** Financial Times Stock Exchange 100 Index

**GARCH** Generalized Autoregressive Conditional Heteroskedasticity model

**GJR-GARCH** Glosten-Jagannathan-Runkle-GARCH

**HS** Historical Simulation

**JBtest** Jarque Bera test

**Nordic Power M1** Nordic front month Power future

**Nordic Power Q1** Nordic front quarter Power future

**Nordic Power Y1** Nordic front year Power future

**OMX** Office Max (stock symbol)

**P-value** Probability-value/calculated probability

**PF** Portfolio

**POT** Peak Over Threshold

**QQ-plot** Quantile-Quantile-plot

**QR** Quantile Regression

**RM** RiskMetrics

**S&P 500** Standard & Poor's 500 Index

**SD** Standard Deviation

**Skew** Skewness

**UC** Unconditional Coverage

**VaR** Value-at-Risk

# 1 Introduction

## 1.1 Motivation and objective

In their 2014 paper, Nowotarski et al. (2014) strongly suggest additional research in the direction of forecast averaging in the electricity spot price market, following the superior results in their study. This is further examined in the literature review. To our knowledge, using an average of Value-at-Risk models or an average of Expected Shortfall models in the Nordic power futures market, seems to be unexplored territory. By using several statistical models for calculating day ahead VaR and ES, we investigate whether average models provide better results compared to the individual models in the Nordic power market. The highly volatile nature of the Nordic power futures requires models that are able to react fast to the market conditions, and it will be interesting to see if the chosen individual models and simple-average models are able to account for these conditions adequately.

According to Vehviläinen and Keppo (2003), the deregulation of the electricity market has increased the risk of loss compared to the monopolistic market because of the highly volatile nature of the commodity. They further advocate that electricity has to be consumed at the same time as it is produced, and that electricity is highly volatile in comparison to other commodities because of the volatile market conditions. The highly volatile nature of the commodity strongly entails a requirement for risk management for actors affected by fluctuations in electricity prices and investors in the power futures market. A recent reminder of the large fluctuations in the electricity market, is the case of the Norwegian power trader Einar Aas. September 10th 2018, he lost about NOK 1.3 billion (approximately US$151 million) over the weekend (E24, 2018). NRK (2018) informs that Aas took a large position that was dependent on a decreasing spread between the German and Nordic power price, and that he could not cover all of the losses himself. This resulted in that various power companies were forced to share the loss of NOK 1 billion. To be able to take a position this big without having capital to cover the losses, raises the question whether one has paid enough attention to the risk the position was exposed to. Measures such as Value-at-Risk and Expected Shortfall are prevalent for the matter.

This study aims first, to investigate the in-sample fit of well known univariate risk models in the Nordic power futures market; second, to study the out-of-sample performance to see whether the models are adequate in a more realistic situation; and third, to investigate the performance of equally weighted averages of the same risk models. The risk models will be used to obtain Value-at-Risk and Expected Shortfall estimates for the 90%, 95% and 99% quantiles of the loss distribution for both long and short positions. We use all available daily data from Front Month-, Front Quarter- and Front Year futures contracts.

## 1.2 The Nordic Power Market

On January 1st 1996, Norway and Sweden established a common electricity market and power exchange named Nord Pool. Nord Pool became the world's first multinational exchange for trading and clearing financial power contracts (Nasdaq, 2018). The main motivation for deregulation in Norway, was to increase the efficiency in resource utilization. The investment behaviour when the market was regulated caused capacity to exceed demand considerably (Bye and Hope, 2005). The introduction of a deregulated electricity market has lead to

electricity exchanges similar to the financial market. Nord Pool was licensed as a regulated exchange and clearinghouse in 2002, and the clearing business was demerged into a separate company, Nord Pool Clearing ASA(Nasdaq, 2018).

Nord Pool Clearing was acquired by Nasdaq OMX in 2008, and the exchange switched name to Nasdaq OMX Commodities Europe in 2010 (Nasdaq, 2018). NasdaqCommodities (2018) inform that they list futures contracts for trading, and settlement of futures contracts involves both a daily mark-to-market settlement and a final spot reference cash settlement, after the contract reaches its expiry date. They further notify that mark-to-market settlement covers profit or loss from day-to-day changes in the daily closing price of each contract. Final settlement, which begins at delivery, covers the difference between the final closing price of the futures contract and the system price in the delivery period (NasdaqCommodities, 2018).

### 1.3 Value-at-Risk as a risk metric

Following the new banking regulations during the 1990s, banks were required to measure their risk as accurately as possible, and hold capital in proportion to this risk. Risk assessment developed rapidly because of this, and virtually all financial institutions began using some form of Value-at-Risk (VaR) as a risk metric (Alexander, 2008*b*). VaR is defined as the minimum loss you will experience some proportion of the time dependent on the VaR-quantile you look at. As an example, the day ahead $VaR_{t+1}^{0.95}$ should denote the minimum loss you will experience $(1-0.95) = 5\%$ of the tomorrows, indicating that you are 95% confident that you will not lose more. Value-at-Risk is currently the official measure of market risk contained in Basel III Capital Accord, and constitutes as the standard used to calculate the capital requirements in the banking system (BIS, 2018*a*). An obvious pitfall of VaR, is the lack of information about the tail risk. This was thoroughly highlighted during the 2008 financial crisis where the VaR models clearly underestimated the risk, suggesting that the crisis was a one in a 100.000 year event (BIS, 2018*b*). Despite of it's pitfalls, VaR remains a widely used risk metric for measuring risk in various financial markets.

### 1.4 Expected Shortfall as a risk metric

The shortcomings that became evident during the financial crisis led to several proposed changes by the Basel Committee on Banking Supervision for Basel IV, expected to enter into force in 2019 (PwC, 2016). One of the proposed changes is to use Expected Shortfall (ES) as risk measure for market risk instead of Value-at-Risk. While VaR is only concerned with the minimum loss one can expect for a proportion of the time dependent on the VaR-quantile chosen, Expected Shortfall accounts for extreme losses, and provides information about the tail of the distribution. ES is calculated by finding the expected value of tomorrow's loss, conditional on it being worse than the VaR. As an example, the day ahead $ES_{t+1}^{0.95}$ should denote the average loss you will experience $(1 - 0.95) = 5\%$ of the tomorrows. Expected Shortfall is more complicated to implement than Value-at-Risk, but should be used when possible by risk managers because extreme losses are much more likely to cause financial distress than moderate losses (Christoffersen, 2012).

### 1.5 Structure of the paper

The remainder of the study is organized as follows: In section 2 we review existing relevant research on VaR models, ES models and forecast averaging. In section 3, we present the sources of data and the data cleansing needed to use the data, before we describe the descriptive statistics of our data. In section 4, we present the theory and models we have chosen to use in our analyses, and the backtesting procedures. Ultimately we present and discuss our results in section 5, and in section 6 we conclude and suggest directions for further research. Appendix and bibliography is situated at the end of the paper.

# 2 Literature Review

In this section, we examine previous research on Value-at-Risk and Expected Shortfall. Developing a search strategy to filter the most relevant research done on ES and VaR enhances the probability of finding information that is relevant to our study, and reduces biased findings. We therefore conduct a systematic literature search with specific search phrases listed in table 1. However, the findings using the systematic search procedure regarding Value-at-Risk and Expected Shortfall in the Nordic power futures market, were very scarce, even though VaR and ES are extensively researched for a wide range of commodities. We therefore conduct a non-systematic search as well.

## 2.1 Procedure of the systematic literature search

The phrases relevant to our study are Expected Shortfall and synonyms, Value at Risk, average of ES and VaR respectively, and power futures, particularly Nordic power futures. In cases were there are thousands of hits, we narrow the search down by demanding that the phrases are used in the title of the papers. The information regarding keywords we are using and the number of hits obtained, along with the name of the papers that proves to be relevant, are presented in table 1. Google Scholar is the main search engine in this process,

| **Search phrase** | **Search date** | **#hits** | **Relevant papers** |
|---|---|---|---|
| ("Expected shortfall" OR"Expected tail loss" OR "ETL"OR "Conditional Value at Risk"OR "CVaR") ("Nordic Power futures" OR "Nordic Power futures") | 14.05.2019 | 2 | None |
| ("Expected shortfall" OR"Expected tail loss" OR "ETL"OR "Conditional Value at Risk"OR "CVaR") ("Power future" OR "Power futures") | 14.05.2019 | 63 | Westgaard et al. (2014), Dahlen et al. (2015) |
| allintitle: ("value-at-risk") ("average" OR "averaging") | 16.05.2019 | 33 | Gabrielsen et al. (2015) |
| allintitle: ("Expected shortfall" OR"Expected tail loss" OR "ETL"OR "Conditional Value at Risk"OR "CVaR") ("average" OR "averaging") | 20.05.2019 | 8 | None |
| ("Value at Risk") ("Nordic Power futures" OR "Nordic Power future") | 20.05.2019 | 2 | None |
| ("Value at Risk") ("Power futures" OR "Power future") | 20.05.2019 | 72 | Dahlen et al. (2015), Westgaard et al. (2014) |

Table 1: Table of search procedure for the systematic literature search.

and we evaluate only papers written in English. Note that Expected Shortfall, Conditional

Value-at-Risk (CVaR) and Expected Tail Loss (ETL) are used as synonyms by many, and all three terms are used in the search because of that. The abbreviations "ES" and "VaR" are mostly used in papers with absolutely no relevance to our research, and the terms are therefore not used here.

## 2.2 Papers of the systematic literature search

Westgaard et al. (2014) investigated empirical properties in the European energy futures markets, and discussed what risk models are applicable for different participants in these markets. One of the markets in question, is the Nordic power futures market. Westgaard et al. (2014) concluded that when comparing energy commodities with traditional assets, standard deviation, empirical VaR and Expected Shortfall(ES) are generally much higher for the former. Nordic power futures is among the commodities with the highest volatility, and volatility clustering occur during supply shocks (e.g because of a power station shutdown for a period) and demand shocks (e.g because of an abnormal cold winter). Westgaard et al. (2014) further advocated that one should be careful applying standard models from banks, i.e RiskMetrics and Historical Simulation, for energy commodities portfolios. This is because of the return distribution characteristics, which is neither normal nor constant over time. Proper VaR models need to capture the specific distribution and the changing correlation and dynamics. To evaluate the risk models one should assess both in-sample and out-of-sample VaR and ES performance, both in the univariate and multivariate case. They recommended further research for the different contracts analyzed in their paper.

Dahlen et al. (2015) examined whether it is possible to provide consistent results for different energy commodity futures, when calculating Value-at-Risk using non-estimation complex methods. They compared RiskMetrics, historical simulation, filtered historical simulation and quantile regression applied on crude oil, gas oil, natural gas, coal, carbon and electricity futures. The findings regarding European energy futures indicate that filtered historical simulation is an accurate and easy model that produce consistent results on both portfolios of energy futures, and on single energy futures contracts. The quantile regression approach performs good as well. Dahlen et al. (2015) conclude that the RiskMetrics approach performs poor because the normal distributed assumption about the returns is a simplification that does not work well for the heavy-tailed and skewed return of the energy futures data. The historical simulation perform poor as well because it is unable to capture the changing volatility.

Gabrielsen et al. (2015) proposed an exponential weighted moving average model in their paper, using a modified form of the Gram-Charlier density to estimate volatility, skewness and kurtosis over time. The proposed model were evaluated using 1-day and 10-day VaR forecasts, and using GARCH, historical simulation and filtered historical simulation to compare the performance. Unconditional, independence and conditional likelihood tests, in addition to the Basel II regulatory tests were used to measure the adequacy of the VaR forecasts. The backtesting were conducted on S&P 500, NASDAQ, FTSE100, DAX 30 and CAC 40. The results of the study were mixed, but Gabrielsen et al. (2015) emphasized that the exponential weighted moving average model performed as well as the GARCH model on average for both the in-sample and out-of-sample period.

## 2.3 Papers of the non-systematic literature search

Chan and Gray (2006) proposed an EVT-based model to forecast VaR for several international power markets, in addition to NordPool, using daily aggregated electricity spot prices. Their proposed AR-EGARCH-EVT model performs well, while a naive quantile estimator based on Historical Simulation which serves as a benchmark in their research, also performs surprisingly good for the data from NordPool. Chan and Gray suggested that this was due to the fact that the skewness and kurtosis in this market was notably lower than that of other markets in the study.

Botterud et al. (2010) studied the relationship between spot and futures prices in the Nord Pool electricity market, and found a close correlation between the two. When they compared the spot with a one-week ahead future, and a six-week ahead future, the latter tends to deviate more, although it follows the spot development most of the time. The correlation between the six-week ahead future and the spot is calculated at 0.97, while the one-week ahead future has a correlation of 0.98.

Nowotarski et al. (2014) presented a comprehensive empirical study where they evaluated the use of forecast averaging in the context of electricity prices, by using data from Nord Pool in addition to other markets. They introduced a method for producing average forecasts, where they essentially made point forecasts with different models, averaged them, and applied quantile regression to predict the quantiles of the distribution. Their results indicated that for spot electricity markets, forecast averaging provides superior results under normal market conditions, but fails to outperform alternative approaches of using an individual model in a more volatile environment or in the presence of price jumps and spikes. They strongly suggested additional research in this direction.

Steen et al. (2015) evaluated the performance of RiskMetrics-, Historical Simulation- and quantile regression in predicting Value-at-Risk for various commodities, not including power futures market. They concluded that the quantile regression outperforms RiskMetrics and Historical Simulation for the commodities in their study.

There exists extensive research on the use of GARCH-type models to model volatility (Weron, 2014, Füss et al., 2016, Aggarwal et al., 2009, Sheedy, 2008). Weron (2014) gave an overview of its use in electricity price forecasting (EPF). Füss et al. (2016) addressed GARCH in the context of hedge fund return volatility, and find it to be a superior measure of downside risk. Füss et al. (2008) concluded that GARCH-type models are the most qualified for VaR modelling in the commodity futures markets, when they compared eight different VaR models. However, they emphasized that the choice of VaR model should be dependent on the return series. Giot and Laurent (2003) assessed the performance of two ARCH-type models and RiskMetrics, in producing VaR predictions in six different commodity futures markets. They found the ARCH-type models to be better in predicting VaR than the traditional RiskMetrics approach. ARCH-type models are also popular in the electricity price forecasting literature Weron (2014). Garcia et al. (2005) found that their GARCH model outperforms the ARIMA model in predicting day ahead electricity prices.

In their 2006 study, Harmantzis et al. (2006) aimed to study the performance of several models for VaR and expected shortfall for heavy tailed returns using historical data for various currency exchange rates and stock market indices. The models used in the study are a model based on Generalized Pareto (peak over threshold (POT) technique of extreme value theory (EVT)), the Gaussian (Normal) model, the Stable Paretian models (symmetric

and skewed), and the model based on the historical (or empirical) approach. For the ES estimation, the study concludes that the Gaussian model underestimates ES, while the Stable Paretian model overestimates ES. The POT method and the historical method do give more correct estimations compared with the Gaussian model and the Stable Paretian model. For the VaR estimations, the study shows that fat tailed models outperform the non fat tailed models (POT in the case of 95 per cent confidence level Stable and SaS in the case of 99 per cent confidence level). Harmantzis et al. (2006) further emphasize that the Stable model should be preferred over the symmetric Stable for very heavytailed and non-symmetric data sets for VaR estimation.

In the paper of Ardia and Hoogerheide (2014), Estimation Frequency on one-day ahead 95% and 99% Value-at-Risk and expected shortfall forecasts are studied for various GARCH models, using daily returns from the S&P 500 index. They conclude that there are only marginally improvements of the performance of the GARCH equation using daily updates of the parameters, compared with weekly, monthly or quarterly updates. Ardia and Hoogerheide (2014) emphasize that the asymmetric GARCH model with non-parametric kernel density estimate (GJR-Kernel) performs well, while specifying a Student-t (or Gaussian) innovations' density yields substantially and significantly worse forecasts, especially for expected shortfall. The worst-performing model in the study is the Exponentially Weighted Moving Average (EWMA) of RiskMetrics approach. Ardia and Hoogerheide (2014) accentuate that the simpler models with daily updated parameters performs worse in the study than the more advanced model with infrequently updated parameters.

Zhu and Galbraith (2011) attempted to further extend the research on forecasting downside risk for asymmetric and heavy-tailed return distributions. By forecasting expected shortfall using general, asymmetric exponential power and Student-t distributions with separate parameters to control skewness and the thickness, Zhu and Galbraith (2011) seeks to answer whether the additional parameters provide discernible improvements of the forecasts, or whether the additional flexibility is unnecessary or is dominated at available sample sizes by the efficiency cost of estimating additional parameters. They conclude that the additional generality does improve the fit and forecasting power relative to more restricted specifications of the distribution of standardized innovations.

## 2.4 This study in the context of existing literature

The purpose of this study is to investigate VaR and ES performance of well known univariate risk models in the Nordic power futures market, and study whether equally weighted averages of the same risk models outperform the individual models. The scarce findings in the literature indicates that this particular field of study has not been extensively researched. We will assess both in-sample fit and out-of-sample performance, as Westgaard et al. (2014) emphasized the importance of. The comprehensive study of Nowotarski et al. (2014) regarding the use of forecast averaging in the context of electricity prices is the main motivation in this regard.

Dahlen et al. (2015) obtained promising results regarding filtered historical simulation and quantile regression for European energy futures, while RiskMetrics performed poor because of the normal distribution assumption. We will study if RiskMetrics with the Cornish Fisher approximation are able to correctly account for the distribution of the Nordic power futures in this study, in addition to studying filtered historical simulation. Various quantile regression approaches for VaR will also be assessed in this study.

Füss et al. (2008) concluded that GARCH-type models are the most qualified for VaR modelling in the commodity futures market, and we will use various GARCH models for both VaR and ES to study the performance for the Nordic power futures. Zhu and Galbraith (2011) emphasized that the additional parameters used to account for fat tails of a distribution provide discernible improvements when forecasting expected shortfall, and we will study if this is the case in this context as well.

The highly volatile nature of the Nordic power futures makes it particularly interesting to investigate whether the models are able to react fast to the volatility clustering described by Westgaard et al. (2014). We will backtest the VaR models using both unconditional coverage and conditional independence to assess the VaR-models ability to correctly estimate the risk, and react fast to sudden changes in volatility. The ES-models will be backtested using the approach of McNeil and Frey (2000, p.294).

# 3 Data and Descriptive Statistics

This section is devoted to describe the data used in the study, and the cleansing of the data. Thereafter we aim first, to present the descriptive statistics of Front Month-, Front Quarter- and Front Year Nordic power futures to form the basis for studying the in-sample fit of the models; and second, to present the descriptive statistics of the in-sample and out-of-sample data separately.

## 3.1 Source of data

The Nordic power futures data we use in this study is collected from montelnews.com, which is a company that provides data and analysis for professionals in the European energy markets. The data was collected on May 21st 2019. The contracts we use in our empirical study are NPE ENO Y1 (Front Year), NPE ENO Q1(Front Quarter) and NPE ENO M1(Front Month).

## 3.2 Data cleansing

When a futures contract expires, a new one is created proceeding the previous contract. The shifting from one contract to another, i.e the rollover of the contract, needs to be accounted for by not using the return at this time. The return between shifting contracts is misleading in reality because an investor would have to sell his position on the expiration day, and reinvest the position in the next futures contract the following day, and would thus not receive this return. By not accounting for the rollover, one would experience eventual spikes in the returns, falsely indicating impetuous volatility. Following the removal of the rollover-return, we calculated the daily log-returns, and performed the analysis using this data.

## 3.3 In-sample and Out-of-sample window

We conduct the study using in-sample data to estimate the models, and out-of-sample data to investigate the performance of the models. The use of an out-of-sample window when backtesting the models, mirrors a realistic situation and prevents biased conclusions on the basis of prospective volatility, contrary to using in-sample data.

In order to backtest the 99% quantile correctly, one should have at least 30 rejections. The scarce amount of observations available for the three contracts are thus not optimal in this regard, and we choose an in-sample window consisting of the first half of the total data series for each contract in order to minimize that pitfall. The out-of-sample performance will then be evaluated on the basis of the rest of the data at hand. The next step is to estimate the parameters of the models on the basis of the in-sample data. These parameters are then used to predict the volatility of tomorrow, and the VaR and ES are calculated using this volatility. The next step is to re-estimate the parameters using an expanding window by including the new information from the last day, in addition to all the information at hand from the past. The expanding window is preferred to the rolling window because of the sparse amount of observations available for the Nordic power futures. The same procedure

carries on to the last observation, and the performance of the model is backtested for the out-of-sample window to conclude whether the model performs well in a realistic situation or not.

Even though the out-of-sample performance of the models is the most important feature of the study, we will also conduct in-sample backtesting of the entire sample of each contract to investigate the in-sample fit of the models. Due to this we will first describe the whole data set, and then describe the data from the in- and out-of-sample windows.

### 3.4 Descriptive statistics of the entire data set

This subsection is devoted to present the descriptive statistics of the entire data set of Front Month-, Front Quarter- and Front Year Nordic power futures to prepare the study of the in-sample fit of the models. The periods for the different contracts in question are displayed in table 2.

| Contract | Start date | End date | Number of observations |
|---|---|---|---|
| Front Month | 8. April 2003 | 20. May 2019 | 3825 |
| Front Quarter | 8. September 2004 | 20. May 2019 | 3631 |
| Front Year | 30. December 1999 | 20. May 2019 | 4826 |

Table 2: Start date, end date and number of observations for the entire data set.

Table 3 shows descriptive statistics of the data used. The daily mean and median of the Nordic power futures are close to zero for all three contracts, while the daily standard deviation is 2.83%, 2.36%, and 1.61% for respectively Front Month-, Front Quarter- and Front Year contract. This entails an annual standard deviation of about 44.8%, 37.4%, and 25.5% using 251 trading days. The volatility is very high compared to e.g the S&P500, and implies high risk, especially for the shorter contracts. Nordic electricity market is among the commodities with the highest volatility, and volatility clustering occur during supply shocks, e.g because of a power station shutdown for a period, and demand shocks, e.g because of an abnormal cold winter (Westgaard et al., 2014). Such occasional shocks are demonstrated in figure 1, where spikes of the distribution are clearly evident.

| Summary statistics | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Contract | Mean[%] | Median[%] | SD[%] | Min[%] | Max[%] | Skew | EKurt | JB | ADF | N |
| Front Month | -0.09 | 0 | 2.83 | -16.71 | 21.51 | 0.13 | 3.52 | 0 | 0 | 3825 |
| Front Quarter | -0.03 | 0 | 2.36 | -15.65 | 13.35 | -0.17 | 2.84 | 0 | 0 | 3631 |
| Front Year | 0.03 | 0 | 1.61 | -12.26 | 16.35 | -0.04 | 6.86 | 0 | 0 | 4826 |

Table 3: The table lists descriptive statistics for all available historical data for Front Month-, Front Quarter-, and Front Year Nordic power futures. "Mean" refers to the mean logreturn, "Median" refers to the median logreturn, "SD" refers to the standard deviation of the logrerurns, "Min" and "Max" refers to the most extreme returns observed in either direction, "Skew" refers to skewness, "Ekurt" refers to excess kurtosis, "JB" refers to the p-value of the Jarque Bera test for normality, "ADF" refers to the Augmented Dickey-Fueller test and "N" is the number of data points.

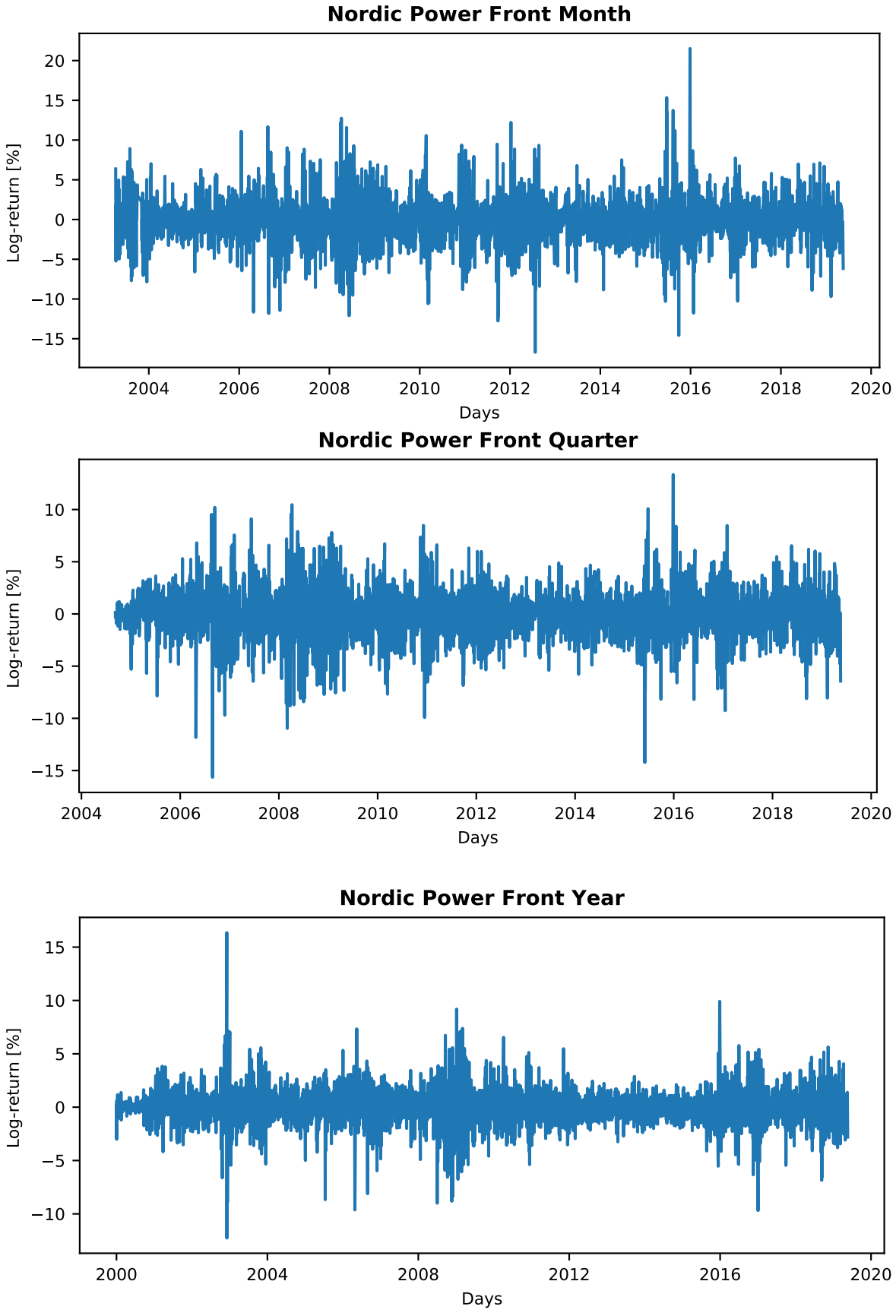



Figure 1: The figure presents logreturn plots of the three contracts in the entire range of data. By the end of 2015 one can see an example of volatility clustering both for the Front Month and Front Quarter contract

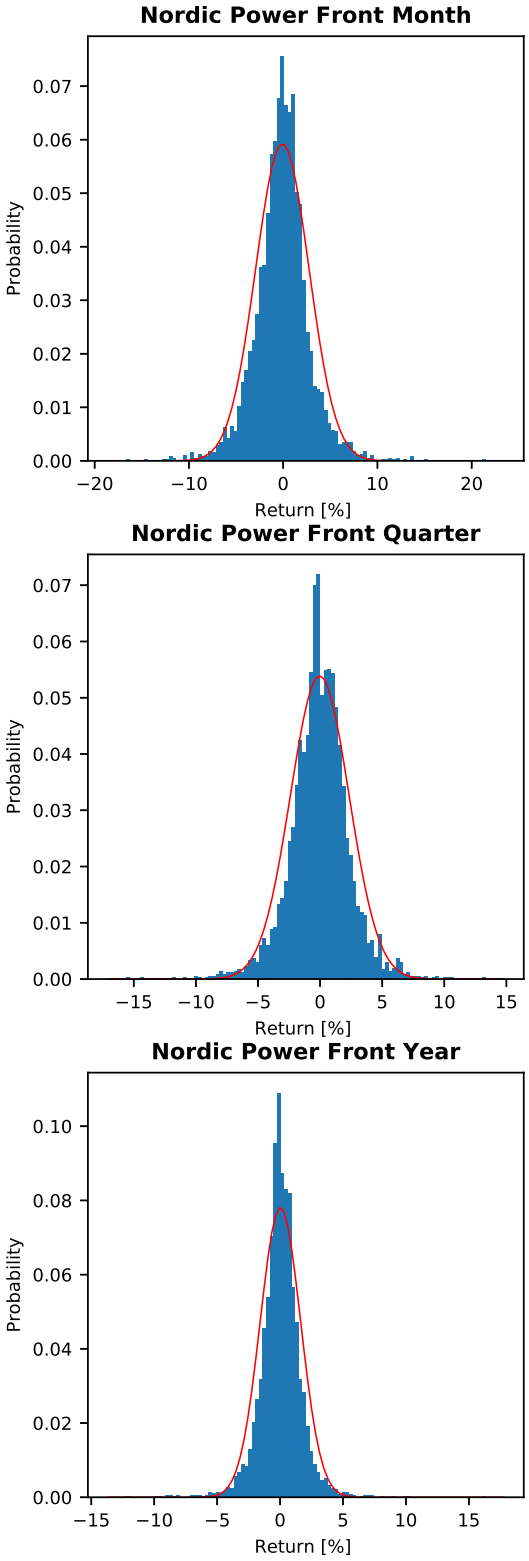



Figure 2: Histograms of returns with the corresponding normal probability density function. Data far out in the tails, and high density close to zero is evident, demonstrating the leptokurtic shape of the distribution

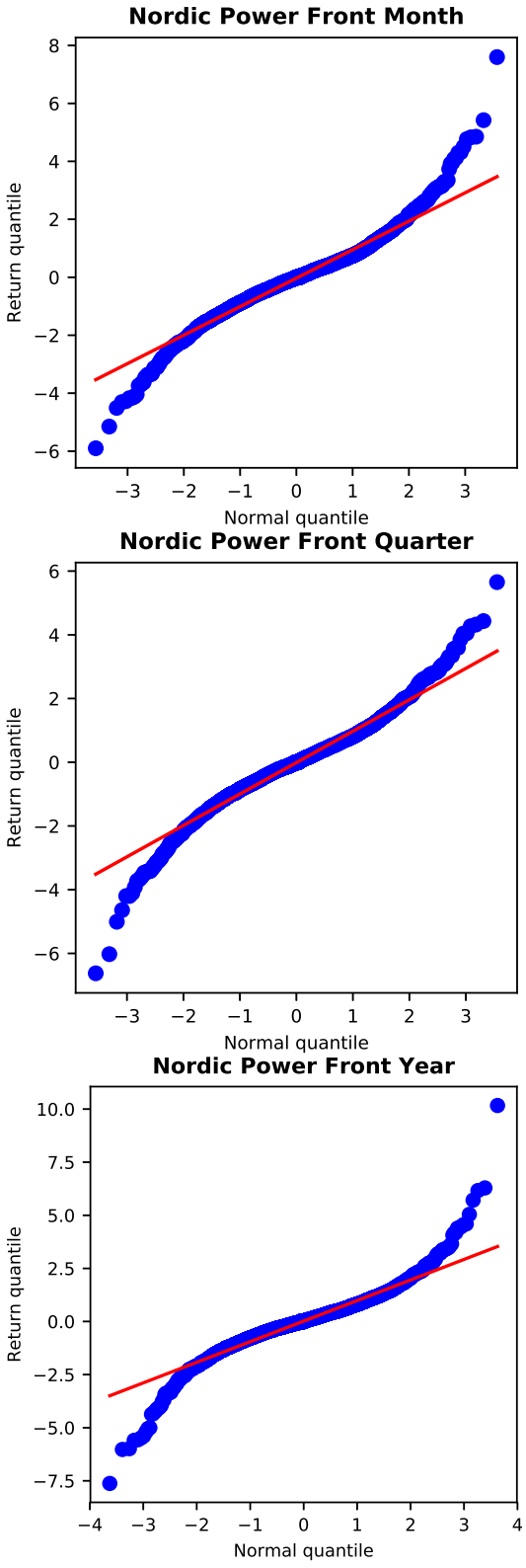



Figure 3: QQ-plots of logreturn quantiles against normal quantiles. The straight line indicates where normally distributed data would hit. The utmost values on both sides deviates from the line, entailing fat tails, and non-normally distributed data

Table 3 exhibits low skewnesses of 0.13, -0.17, and -0.04 for Front Month-, Front Quarter- and Front Year contract respectively. This implies that the distributions are somewhat symmetric, i.e the tails on both side of the mean balance each other out. The symmetry of the distributions is displayed in figure 2. The excess kurtosises are 3.52, 2.84, and 6.86 for Front Month-, Front Quarter- and Front Year contract, entailing fat tails. Fat tails implies more extreme values on each side of the mean compared to a normal distribution. This is shown in figure 2. Here one can clearly see more data far out in the tails, and higher density closer to zero compared to the red line, which indicate a normal distribution with standard deviation and mean from the empirical data. The histogram demonstrates the leptokurtic form that fat tails exhibit. The QQ-plots illustrated in figure 3 supports the assumption that the distributions are non-normal. A QQ-plot is a graphical method used to compare two distributions by plotting the quantile of one plot in relation to another. A normal distributed set of data would exhibit scatter plots following the red line. Figure 3 demonstrates plots with more extreme values at the outermost left and right side compared to the red line, confirming the fat tails previously explained.

The fact that the three contracts in question are far from normally distributed, is further demonstrated in the Jarque-Bera test displayed in table 3. All three contracts show p-values very close to zero. This indicates that the data is non-normally distributed. The Augmented Dickey-Fuller test (ADF) shown in the same table, demonstrates p-values of zero as well. ADF tests the null hypothesis that a unit root is present in a time series. A p-value close to zero indicates that the process is stationary, i.e that parameters such as mean and variance do not change over time. This is an important assumption in time-series analysis.

Table 4, 5, 6, 7, 8 and 9 displays the empirical Value-at-Risk and Expected Shortfall for Front Month., Front Quarter- and Front Year contract. The tables show that the empirical VaR and ES have fluctuated considerably throughout the years for all quantiles. It is also evident that the empirical VaR and ES is far from symmetric when comparing the long and short positions for individual years. 2008 seems to be the year with the highest empirical VaR and ES for the Front Quarter contract, while 2002 is the year with highest empirical VaR and ES for the Front Year contract. For the Front Month contract the empirical VaR is highest in 2008, while the empirical ES is highest in 2015.

| Empirical VaR, Front Month Contract | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| VaR | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 |
| 99% (long) | -7.40 | -3.70 | -5.29 | -10.72 | -7.27 | -9.45 | -4.88 | -8.33 | -7.86 |
| 95% (long) | -5.64 | -2.50 | -3.56 | -4.77 | -4.79 | -5.59 | -4.08 | -3.96 | -5.41 |
| 90% (long) | -4.50 | -2.16 | -2.52 | -3.49 | -3.81 | -4.97 | -2.95 | -2.92 | -3.74 |
| 90% (short) | -3.83 | -1.64 | -2.94 | -3.03 | -3.22 | -4.26 | -2.85 | -3.52 | -2.63 |
| 95% (short) | -4.99 | -2.15 | -4.00 | -4.53 | -5.65 | -6.38 | -3.97 | -5.54 | -3.64 |
| 99% (short) | -6.79 | -5.03 | -5.54 | -7.73 | -8.40 | -10.68 | -6.14 | -8.18 | -7.79 |
| VaR | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | |
| 99% (long) | -8.67 | -6.17 | -4.65 | -9.21 | -6.61 | -5.73 | -6.47 | -6.58 | |
| 95% (long) | -6.08 | -3.20 | -3.54 | -6.16 | -4.21 | -3.93 | -3.79 | -4.50 | |
| 90% (long) | -4.13 | -2.46 | -2.98 | -4.02 | -2.50 | -2.75 | -2.67 | -3.77 | |
| 90% (short) | -3.68 | -2.08 | -2.21 | -3.67 | -3.31 | -2.33 | -3.34 | -2.29 | |
| 95% (short) | -4.60 | -3.12 | -3.14 | -5.17 | -4.41 | -2.95 | -4.58 | -3.48 | |
| 99% (short) | -8.87 | -4.78 | -5.14 | -13.71 | -6.56 | -5.31 | -6.87 | -4.60 | |

Table 4: The table shows 90%, 95% and 99% empirical VaR, for both long and short positions. The numbers are log-returns in percent. The empirical VaR is obtained applying the percentile function in excel on returns ranging one year at a time. Note that empirical VaR for 2003 is from 8. April 2003, and empirical VaR for 2019 is to 20. May 2019 because of the start and end date of the contract

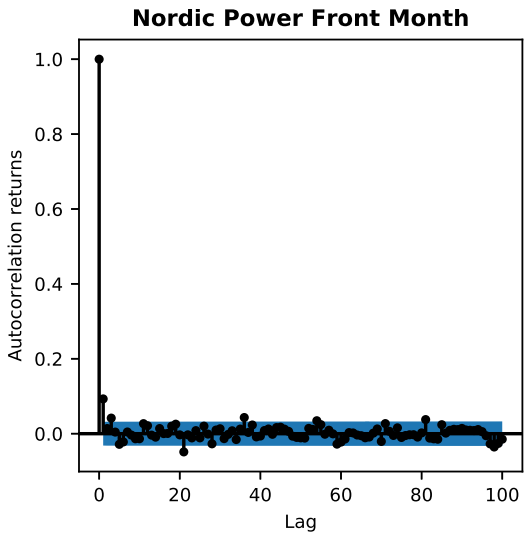
| Empirical VaR, Front Quarter Contract | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| VaR | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 |
| 99% (long) | -1.58 | -5.06 | -9.82 | -5.63 | -8.64 | -6.85 | -6.67 | -5.84 |
| 95% (long) | -1.24 | -2.73 | -4.76 | -4.04 | -6.58 | -4.03 | -4.15 | -3.78 |
| 90% (long) | -0.93 | -1.90 | -2.82 | -2.77 | -4.63 | -2.85 | -2.85 | -3.00 |
| 90% (short) | -0.61 | -2.20 | -2.92 | -2.85 | -4.15 | -3.19 | -3.07 | -2.13 |
| 95% (short) | -0.95 | -2.70 | -4.04 | -4.07 | -5.58 | -4.35 | -4.08 | -2.88 |
| 99% (short) | -1.14 | -3.41 | -7.16 | -6.76 | -7.57 | -6.57 | -6.25 | -6.05 |
| VaR | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 |
| 99% (long) | -4.92 | -3.82 | -3.66 | -6.78 | -6.88 | -4.71 | -5.25 | -6.58 |
| 95% (long) | -3.39 | -2.57 | -2.97 | -3.87 | -3.60 | -3.29 | -3.25 | -4.08 |
| 90% (long) | -2.39 | -1.93 | -2.30 | -3.10 | -2.51 | -2.16 | -2.50 | -3.50 |
| 90% (short) | -2.21 | -1.61 | -1.99 | -2.17 | -3.36 | -2.26 | -2.94 | -2.19 |
| 95% (short) | -3.05 | -2.39 | -2.50 | -3.62 | -4.08 | -3.00 | -4.07 | -3.96 |
| 99% (short) | -4.45 | -4.14 | -4.17 | -7.52 | -5.77 | -4.56 | -5.92 | -4.42 |

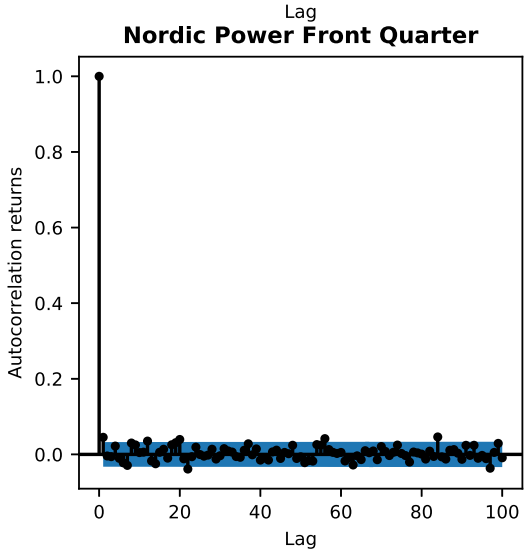
Table 5: The table shows 90%, 95% and 99% empirical VaR, for both long and short positions. Numbers are log-returns in percent. The empirical VaR is obtained applying the percentile function in excel on returns ranging one year at a time. Note that empirical VaR for 2004 is from 8. September 2004, and empirical VaR for 2019 is to 20. May 2019 because of the start and end date of the contract

| Empirical VaR, Front Year Contract | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| VaR | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 |
| 99% (long) | -2.17 | -3.32 | -6.85 | -4.90 | -2.19 | -4.16 | -7.54 | -3.45 | -7.46 | -5.40 |
| 95% (long) | -0.95 | -1.98 | -2.26 | -3.26 | -1.69 | -2.14 | -2.78 | -1.98 | -3.84 | -3.71 |
| 90% (long) | -0.59 | -1.46 | -1.40 | -2.36 | -1.45 | -1.57 | -2.17 | -1.41 | -3.01 | -3.06 |
| 90% (short) | -0.77 | -1.73 | -1.82 | -2.08 | -1.28 | -1.80 | -2.10 | -1.61 | -2.35 | -2.63 |
| 95% (short) | -0.97 | -2.32 | -3.41 | -2.81 | -1.72 | -2.25 | -3.05 | -1.96 | -3.15 | -4.27 |
| 99% (short) | -1.55 | -3.47 | -7.71 | -5.23 | -2.70 | -2.85 | -4.09 | -2.68 | -4.98 | -6.36 |
| VaR | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 |
| 99% (long) | -3.60 | -3.30 | -2.18 | -2.44 | -1.82 | -3.56 | -4.90 | -4.76 | -4.15 | -3.48 |
| 95% (long) | -2.72 | -2.16 | -1.50 | -1.64 | -1.50 | -2.12 | -3.04 | -2.09 | -2.85 | -3.16 |
| 90% (long) | -1.90 | -1.72 | -1.28 | -1.17 | -1.12 | -1.64 | -2.28 | -1.62 | -1.88 | -2.63 |
| 90% (short) | -2.12 | -1.25 | -1.16 | -1.12 | -1.08 | -0.93 | -2.37 | -1.69 | -2.34 | -2.19 |
| 95% (short) | -2.95 | -1.81 | -1.58 | -1.47 | -1.27 | -1.46 | -3.44 | -2.06 | -2.89 | -3.11 |
| 99% (short) | -4.34 | -3.04 | -2.16 | -2.20 | -1.89 | -2.37 | -4.92 | -3.39 | -4.57 | -4.08 |

Table 6: The table shows 90%, 95% and 99% empirical VaR, for both long and short positions. Numbers are log-returns in percent. The empirical VaR is obtained applying the percentile function in excel on returns ranging one year at a time. Note that empirical VaR for 2019 is to 20. May 2019 because of the end date of the contract

The autocorrelation of log-returns and squared log-returns for the three contracts are demonstrated in figure 4 and 5. A positive autocorrelation indicates that we can predict something about the future, and for our data, a positive autocorrelation of squared log-returns are evident, while there are no signs of autocorrelation for the log-returns. This is more or less as expected, because it is a stylized fact of most financial assets (Christoffersen, 2012), and entails that variance forecasting is applicable for our data.



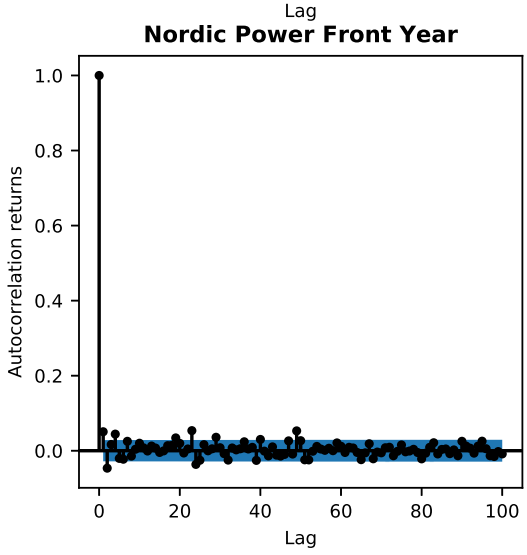
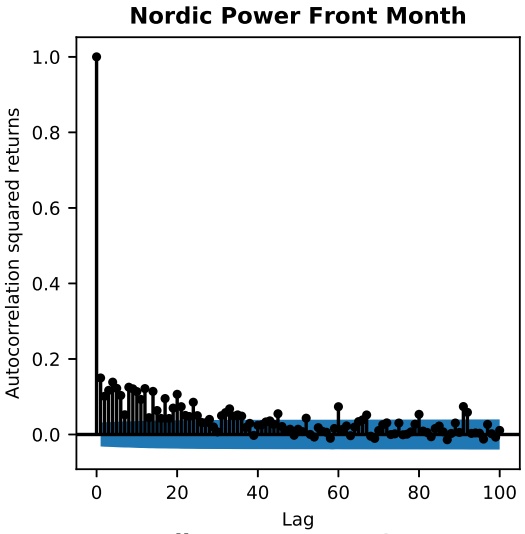
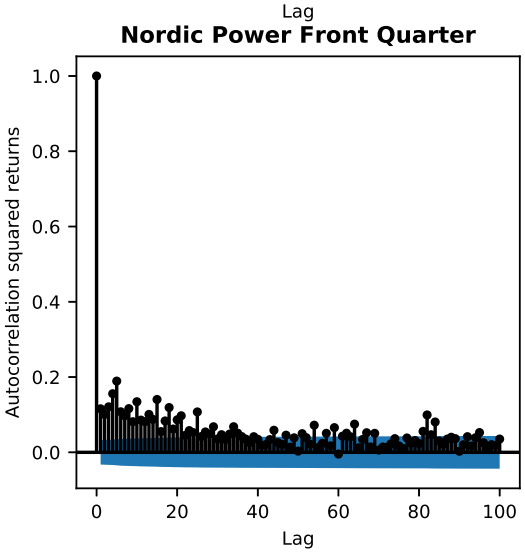




Figure 4: Autocorrelation of logreturns. No or insignificant signs of autocorrelation



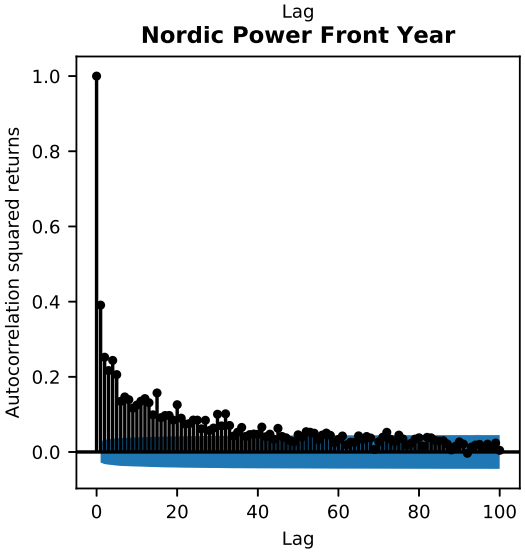




Figure 5: Autocorrelation of squared returns. There is significant autocorrelation for the 25-50 first lags, in all three contracts

| Empirical ES, Front Month Contract | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ES | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 |
| 99% (long) | -7.77 | -4.19 | -5.96 | -11.64 | -8.00 | -10.37 | -5.93 | -9.97 | -10.95 |
| 95% (long) | -6.73 | -3.29 | -4.55 | -8.06 | -6.33 | -8.19 | -4.84 | -6.76 | -7.62 |
| 90% (long) | -5.91 | -2.82 | -3.83 | -6.13 | -5.26 | -6.72 | -4.13 | -5.11 | -6.06 |
| 90% (short) | .5.37 | -2.76 | -4.23 | -5.30 | -5.92 | -6.99 | -4.31 | -6.03 | -5.06 |
| 95% (short) | -6.32 | -3.64 | -4.97 | -6.72 | -7.43 | -8.68 | -5.14 | -7.18 | -6.93 |
| 99% (short) | -8.10 | -5.94 | -5.85 | -10.42 | -8.80 | -12.16 | -6.57 | -9.39 | -8.72 |
| ES | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | |
| 99% (long) | -11.98 | -6.96 | -6.48 | -11.46 | -8.58 | -7.41 | -7.60 | -9.70 | |
| 95% (long) | -8.14 | -4.71 | -4.74 | -7.97 | -5.96 | -5.16 | -5.51 | -5.98 | |
| 90% (long) | -6.52 | -3.74 | -4.02 | -6.53 | -4.80 | -4.23 | -4.41 | -5.14 | |
| 90% (short) | -5.76 | -3.37 | -3.57 | -7.57 | -4.87 | -3.53 | -4.91 | -3.72 | |
| 95% (short) | -7.40 | -4.18 | -4.54 | -10.92 | -5.90 | -4.37 | -5.67 | -4.23 | |
| 99% (short) | -10.14 | -5.78 | -6.51 | -16.86 | -7.73 | -6.08 | -7.03 | -4.74 | |

Table 7: The table shows 90%, 95% and 99% empirical ES, for both long and short positions. Numbers are log-returns in percent. The empirical ES is obtained by calculating the average loss of losses larger than the empirical VaR of table 4. Note that empirical ES for 2003 is from 8. April 2003, and empirical ES for 2019 is to 20. May 2019 because of the start and end date of the contract

| Empirical ES, Front Quarter Contract | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ES | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 |
| 99% (long) | -1.95 | -6.28 | -12.47 | -6.06 | -9.49 | -7.37 | -8.10 | -6.42 |
| 95% (long) | -1.58 | -4.14 | -7.48 | -5.06 | -8.07 | -5.61 | -5.81 | -4.94 |
| 90% (long) | -1.35 | -3.27 | -5.57 | -4.23 | -6.83 | -4.54 | -4.80 | -4.16 |
| 90% (short) | -0.96 | -2.83 | -4.88 | -4.83 | -5.94 | -4.70 | -4.43 | -3.70 |
| 95% (short) | -1.07 | -3.14 | -6.15 | -6.13 | -7.04 | -5.60 | -5.34 | -4.81 |
| 99% (short) | -1.18 | -3.62 | -9.00 | -7.84 | -9.31 | -7.23 | -7.53 | -6.38 |
| ES | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 |
| 99% (long) | -5.25 | -4.51 | -4.79 | -9.95 | -7.50 | -7.02 | -6.55 | -8.06 |
| 95% (long) | -4.35 | -3.46 | -3.61 | -5.95 | -5.55 | -4.80 | -4.59 | -5.50 |
| 90% (long) | -3.65 | -2.88 | -3.12 | -4.74 | -4.30 | -3.72 | -3.73 | -4.64 |
| 90% (short) | -3.30 | -2.62 | -2.88 | -4.67 | -4.56 | -3.44 | -4.34 | -3.68 |
| 95% (short) | -4.02 | -3.32 | -3.51 | -6.41 | -5.24 | -4.18 | -5.20 | -4.30 |
| 99% (short) | -5.59 | -4.66 | -4.42 | -10.44 | -6.81 | -6.39 | -6.26 | -4.84 |

Table 8: The table shows 90%, 95% and 99% empirical ES, for both long and short positions. Numbers are log-returns in percent. The empirical ES is obtained by calculating the average loss of losses larger than the empirical VaR of table 5. Note that empirical ES for 2004 is from 8. September 2004, and empirical ES for 2019 is to 20. May 2019 because of the start and end date of the contract

### 3.5 Descriptive statistics of the in-sample and out-of-sample window

This subsection is devoted to present the descriptive statistics of the in-sample and out-of-sample data for Front Month-, Front Quarter- and Front Year Nordic power futures. The periods for the different contracts in question are displayed in table 10. The summary statistics of the in-sample and out-of-sample data are displayed in table 11 and 12, and the logreturn plots of the three contracts are illustrated in figure 6.

Table 11 displays the summary statistics for the in-sample data. The mean and median are approximately zero for all three contracts, and the daily standard deviation is 2.88%, 2.56% and 1.75% for Front Month, Front Quarter and Fronth Year respectively. This entails an annual standard deviation of 45.63%, 40.56% and 27.73%, given 251 trading days, which is

| Empirical ES, Front Year Contract | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| ES | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 |
| 99% (long) | -2.59 | -3.69 | -9.41 | -5.40 | -2.47 | -5.96 | -8.59 | -4.07 | -8.72 | -6.08 |
| 95% (long) | -1.63 | -2.88 | -4.69 | -4.11 | -2.04 | -3.44 | -5.05 | -2.86 | -5.84 | -4.88 |
| 90% (long) | -1.22 | -2.33 | -3.30 | -3.40 | -1.80 | -2.65 | -3.78 | -2.31 | -4.63 | -4.14 |
| 90% (short) | -1.12 | -2.55 | -4.45 | -3.37 | -1.90 | -2.36 | -3.28 | -2.13 | -3.48 | -4.48 |
| 95% (short) | -1.33 | -2.99 | -6.20 | -4.28 | -2.29 | -2.69 | -3.92 | -2.40 | -4.21 | -5.43 |
| 99% (short) | -1.66 | -3.75 | -11.53 | -6.02 | -3.07 | -3.19 | -5.66 | -3.11 | -5.92 | -7.89 |
| ES | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 |
| 99% (long) | -4.39 | -3.55 | -2.53 | -2.81 | -1.95 | -4.50 | -5.57 | -6.75 | -5.86 | -3.80 |
| 95% (long) | -3.44 | -2.80 | -1.95 | -2.18 | -1.69 | -3.01 | -4.27 | -3.85 | -3.82 | -3.48 |
| 90% (long) | -2.89 | -2.35 | -1.68 | -1.78 | -1.52 | -2.46 | -3.47 | -2.88 | -3.13 | -3.15 |
| 90% (short) | -3.24 | -2.08 | -1.67 | -1.57 | -1.45 | -2.05 | -3.60 | -2.49 | -3.31 | -3.14 |
| 95% (short) | -3.93 | -2.64 | -1.93 | -1.88 | -1.69 | -2.82 | -4.32 | -3.02 | -3.86 | -3.70 |
| 99% (short) | -5.47 | -4.12 | -2.41 | -2.50 | -2.00 | -5.79 | -5.36 | -4.48 | -5.17 | -4.29 |

Table 9: The table shows 90%, 95% and 99% empirical ES, for both long and short positions. Numbers are log-returns in percent. The empirical ES is obtained by calculating the average loss of losses larger than the empirical VaR of table 6. Note that empirical ES for 2019 is to 20. May 2019 because of the end date of the contract

| | In-sample | | | Out-of-sample | | |
|---|---|---|---|---|---|---|
| Contract | Start date | End date | #obs. | Start date | End date | #obs. |
| Front Month | 8. April 2003 | 13. May 2011 | 1912 | 16. May 2011 | 20. May 2019 | 1913 |
| Front Quarter | 8. Sept. 2004 | 4. Jan. 2012 | 1815 | 5. Jan. 2012 | 20. May 2019 | 1816 |
| Front Year | 30. Dec. 1999 | 17. Sept. 2009 | 2413 | 18. Sept. 2009 | 20. May 2019 | 2413 |

Table 10: Start date, end date and number of observations for in-sample and out-of-sample data

quite substantial compared to e.g the stock market. The biggest losses are 12.10%, 15.65% and 12.26% for Front Month, Front Quarter and Front Year respectively, while the highest positive returns are 12.74%, 10.47% and 16.35%. Comparing the in-sample data with the entire dataset displayed in table 3, reveals that the biggest loss of Front Quarter and Front Year appears in the in-sample data, in addition to the highest positive return for Front Year. This is illustrated in figure 6 as well.

The skewnesses of the data are 0.11, -0.23 and 0.08 for Front Month, Front Quarter and Front Year respectively, while the excess kurtosises are 1.95, 2.48 and 8.15 for the three contracts. This entails that the contracts have somewhat symmetric returns, with a leptokurtic form. This is displayed in the histograms in figure 14 in the appendix. Especially the Front Year contract exhibits very fat tails. The Jarque-Bera test results in p-values of zero for all three contracts. This further demonstrates that the in-sample data are not normally distributed. The Augmented Dickey-Fueller test values of zero indicates stationarity.

| Summary statistics in-sample data | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Contract | Mean[%] | Median[%] | SD[%] | Min[%] | Max[%] | Skew | EKurt | JB | ADF | N |
| Front Month | -0.09 | -0.10 | 2.88 | -12.10 | 12.74 | 0.11 | 1.95 | 0 | 0 | 1912 |
| Front Quarter | -0.05 | 0 | 2.56 | -15.65 | 10.47 | -0.23 | 2.48 | 0 | 0 | 1815 |
| Front Year | 0.04 | 0.05 | 1.75 | -12.26 | 16.35 | 0.08 | 8.15 | 0 | 0 | 2413 |

Table 11: The table lists descriptive statistics for in-sample data for Front Month-, Front Quarter-, and Front Year Nordic power futures contract. "Mean" refers to the mean logreturn, "Median" refers to the median logreturn, "SD" refers to the standard deviation of the logrerurns, "Min" and "Max" refers to the most extreme returns observed in either direction, "Skew" refers to skewness, "Ekurt" refers to excess kurtosis, "JB" refers to the p-value of the Jarque Bera test for normality, "ADF" refers to the Augmented Dickey-Fueller test and "N" is the number of data points

Table 12 displays the statistics for the out-of-sample data. The mean and median of the Front Month, Front Quarter and Front Year are close to zero for the out-of-sample data as well. The standard deviation of 2.78%, 2.15% and 1.46% entails annual standard deviations of 44.04%, 34.06% and 23.13% for Front Month, Front Quarter and Front Year respectively, using 251 trading days. The standard deviations for the three contracts for the out-of-sample data are thus smaller than the standard deviation for the in-sample data, indicating that the Nordic power futures were more volatile in the 2000s than the 2010s. The biggest losses of the out-of-sample data are 16.71%, 14.21% and 9.69% for Front Month, Front Quarter and Front Year respectively, while the biggest positive returns are 21.51%, 13.35% and 9.92%. The biggest loss of the Front Month, and the biggest positive return of the Front Month and Front Quarter data are thus appearing in the out-of-sample window.

The skewnesses of the out-of-sample data for Front Month, Front Quarter and Front Year are 0.16, -0.05 and 0.01 respectively, while the excess kurtosises are 5.31, 3.05 and 3.26. The returns are thus quite symmetrical, but the fat tails are evident, entailing a leptokurtic form of the returns. This is displayed in the histograms in figure 15 in the appendix. As with the in-sample data, the Jarque-Bera test results in p-values of zero for all three contracts. This demonstrates that the out-of-sample data are not normally distributed. The Augmented Dickey-Fueller test values of zero indicates stationarity.

| Summary statistics out-of-sample data | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Contract | Mean[%] | Median[%] | SD[%] | Min[%] | Max[%] | Skew | EKurt | JB | ADF | N |
| Front Month | -0.08 | 0 | 2.78 | -16.71 | 21.51 | 0.16 | 5.31 | 0 | 0 | 1913 |
| Front Quarter | -0.01 | 0 | 2.15 | -14.23 | 13.35 | -0.05 | 3.05 | 0 | 0 | 1816 |
| Front Year | 0.01 | 0 | 1.46 | -9.69 | 9.92 | 0.01 | 3.26 | 0 | 0 | 2413 |

Table 12: The table lists descriptive statistics for out-of-sample data for Front Month-, Front Quarter-, and Front Year Nordic power futures contract. "Mean" refers to the mean logreturn, "Median" refers to the median logreturn, "SD" refers to the standard deviation of the logrerurns, "Min" and "Max" refers to the most extreme returns observed in either direction, "Skew" refers to skewness, "Ekurt" refers to excess kurtosis, "JB" refers to the p-value of the Jarque Bera test for normality, "ADF" refers to the Augmented Dickey-Fueller test and "N" is the number of data points

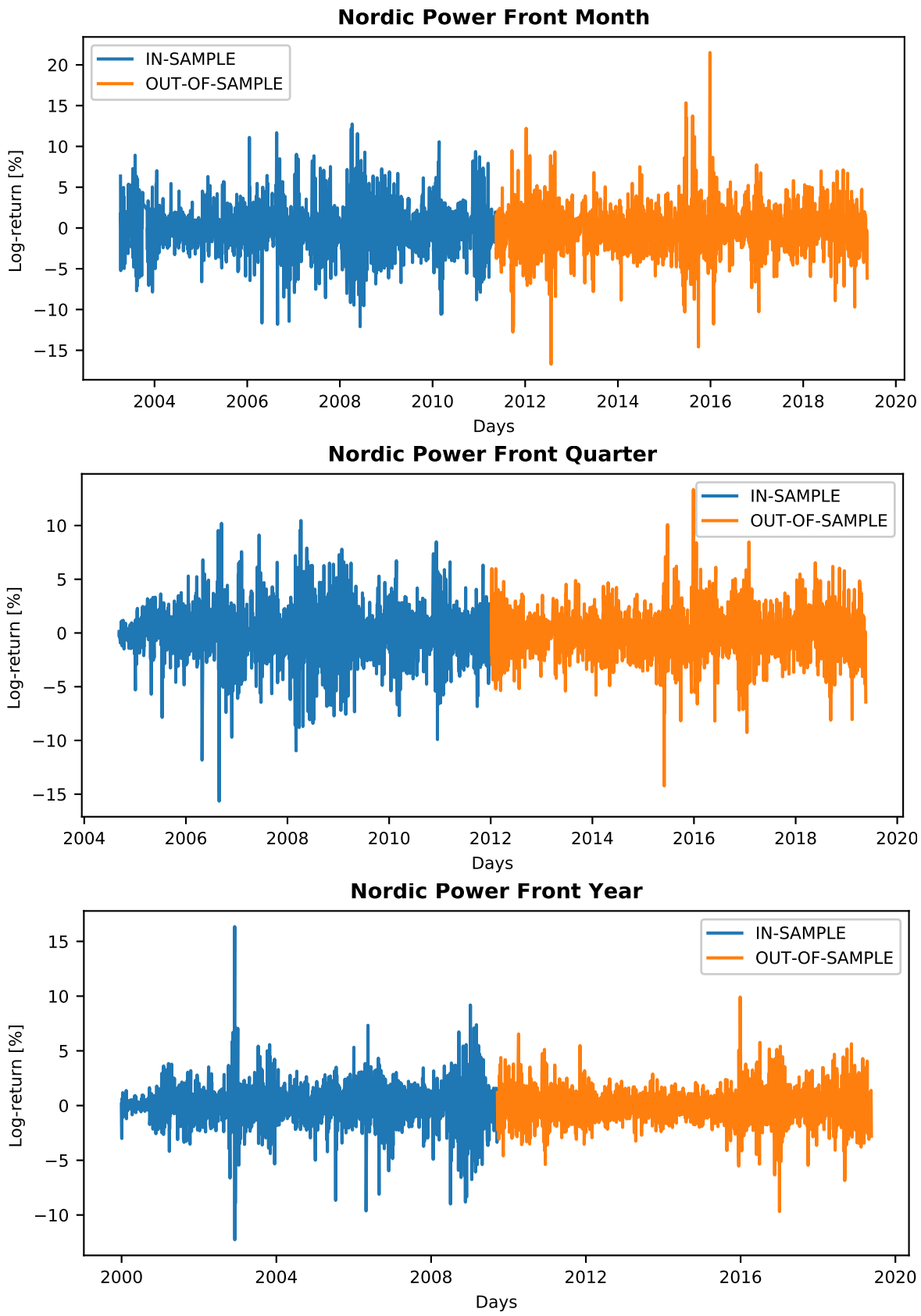



Figure 6: Logreturn plots for the whole data series of the Front Month-, Front Quarter- and Front Year contract. Blue indicates the chosen in-sample window, while orange indicates the out-of-sample window. The periods of the windows for each contract are listed in table 10

# 4 Methodology

In this section we present relevant theory and models used in our research. Note that abbreviations of these models used throughout the rest of the paper will be introduced in the following paragraph.

There are several different models to estimate VaR and ES. It is evident from the data in table 3, and figure 2 and 3 that the distributions are non-normal. This needs to be accounted for when when choosing models for forecasting VaR and ES. We have chosen to focus on some of the most common models applied on VaR and ES that account for the characteristics described in section 3, and we will combine these to see whether a combination of the VaR-models is useful or not. RiskMetrics is one of the most common models for forecasting dynamic variance. To study how well this model performs in relation to other more sophisticated models, using the Cornish-Fisher approach to account for the fat-tailed distribution of the Nordic power futures, is interesting. We denote this approach as RM-CF. RiskMetrics is a simplified version of the GARCH-model, that avoids parameter estimation. In order to account for time-varying volatility, we will apply some GARCH-type models, that previously have shown to perform well in volatile periods in the stock markets (Sheedy, 2008). For all the GARCH models we apply, we use one ARCH-lag and one GARCH-lag. We will study the performance of normal-distributed GARCH(1,1)(GARCH-n), t-distributed GARCH(1,1)(GARCH-t) and a t-distributed GJR-GARCH(1,1)(GJR-GARCH-t) model to see how important the leverage effect and the chosen distribution is for the performance of the models. A Filtered Historical Simulation-approach (FHS) is also chosen to see if a combination of the model-free approach of Historical Simulation, and a model-based approach proves successful. Here, we will use the GARCH-t-model to compute the dynamic variance needed. Quantile regression is also chosen because of its superior performance compared to Historical Simulation and RiskMetrics for various commodities in the paper by Steen et al. (2015). We will use their proposed QR-model using RiskMetrics(QR-RM) to compute the dynamic variance needed. To account for the findings in section 3, we intend to investigate the results of quantile regression using the dynamic variance of the t-distributed GARCH(QR-GARCH-t) as well. Note that quantile regression will only be used for calculating VaR. In the following section we will define and explain VaR and ES, the models chosen and the procedure for model evaluation.

## 4.1 Value-at-Risk

Value-at-Risk (VaR) is a risk measure that aims to find the loss that will be exceeded only $p * 100\%$ of the time in the next K trading days. This can be denoted as

$$Pr(-R_{PF} > VaR) = p \tag{1}$$

where $R_{PF}$ is the log return (Christoffersen, 2012).

Christoffersen (2012) claimed that Value-at-Risk has become the industry benchmark for risk calculation, because it captures one of the most important aspects of risk; the probability of loosing a predefined amount. He further advocate that the ease of implementation and the fact that it is easy to understand, are other factors of it's popularity.

VaR does however have some limitations. VaR only cares about the risk of experiencing a loss

that will be exceeded $p * 100\%$ of the time, not the magnitude of what the losses experienced in that period includes. Value-at-Risk for returns with fat-tailed distributions tends to be underestimated because of this, and it is often these extreme values that risk managers want to avoid. Other methods such as Expected Shortfall should thus be considered for such cases in order to get information about the tail of the distribution as well.

## 4.2 Expected Shortfall

Alexander (2008*b*) defines Expected Shortfall (ES), also called conditional VaR, as the expected loss given that the loss exceeds the VaR. The ES provides more information than VaR, because while VaR is only concerned with the minimum loss one can expect, Expected Shortfall gives information about the expected loss when the VaR is exceeded. This is especially an important feature for distributions with fat tails where extreme values occur more often. Alexander (2008*b*, p. 344) use the following notation for the $100\alpha\%$ daily ES measured at time t:

$$ES_{1,\alpha,t} = -E_t(Y_{t+1}|Y_{t+1} < -VaR_{1,\alpha,t}) \tag{2}$$

$Y_{t+1}$ denotes the realized daily return of the portfolio from time t to time t+1, and the ES is thus expressed as a proportion of the portfolio's value.

## 4.3 Models used to forecast VaR and ES

This section is devoted to describe the models used in this study to compute Value-at-Risk and Expected Shortfall. We will start off with RiskMetrics and the Cornish Fisher approximation, followed by the various GARCH-models used, the Filtered Historical Simulation, and finally quantile regression.

### 4.3.1 RiskMetrics with Cornish-Fisher

The volatility of tomorrow using RiskMetrics can be updated using:

$$\sigma^2_{t+1} = \lambda\sigma^2_t + (1-\lambda)R^2_t \tag{3}$$

where $R_t$ is the log return of the $R_t = ln(A_t/A_{t-1})$ (Christoffersen, 2012). The value of the parameter $\lambda$ is determined by an optimization procedure, and using this on a widely diversified portfolio, the value $\lambda = 0.94$ produces the best backtesting results (Mina et al., 2001). Estimates were quite similar across assets, so $\lambda = 0.94$ is simply set for every asset (Christoffersen, 2012).
RiskMetrics has several advantages compared to other models for forecasting volatility. The ease of calculation is obvious compared to comparable models such as GARCH because there is no need to calculate parameters. Christoffersen (2012) informed that RiskMetrics does, unlike GARCH, ignore the fact that the long-run variance tends to be relatively stable over time. He further emphasizes that recent observations count more than previous observations, and this makes RiskMetrics far more able to react fast to sudden changes than e.g Historical Simulation does. RiskMetrics does however assume some distribution in order to calculate VaR. For non-normal distributed data, the Cornish-Fischer approach offers a simple alternative to calculate VaR, taking skewness($\zeta_1$) and excess kurtosis($\zeta_2$) of the data into account

(Christoffersen, 2012). VaR using RiskMetrics and Cornish-Fischer can be denoted as

$$VaR_{t+1}^{p} = -\sigma_{PF,t+1} CF_{p}^{-1} \tag{4}$$

assuming $z_{t+1} = \frac{R_{PF,t+1}}{\sigma_{PF,t+1}} \overset{\text{i.i.d.}}{\sim} D(0,1)$ (Christoffersen, 2012). $z_t$ denotes the standardized returns, and D(0,1) denotes a distribution with mean equal to 0 and variance equal to 1. Christoffersen (2012) denoted $CF_p^{-1}$ as

$$CF_{p}^{-1} = \Phi_{p}^{-1} + \frac{\zeta_1}{6}[(\Phi_{p}^{-1})^2 - 1] + \frac{\zeta_2}{24}[(\Phi_{p}^{-1})^3 - 3\Phi_{p}^{-1}] - \frac{\zeta_1^2}{36}[2(\Phi_{p}^{-1})^3 - 5\Phi_{p}^{-1}] \tag{5}$$

where $\Phi_p^{-1}$ is the p-th quantile of the normal distribution, and $\zeta_1$ is the skewness and $\zeta_2$ is the excess kurtosis of the standardized returns, $z_{t+1}$. Christoffersen (2012) denoted Expected Shortfall using RiskMetrics with Cornish-Fisher as

$$ES_{t+1}^{p} = -\sigma_{PF,t+1} ES_{CF}(p) \tag{6}$$

where

$$ES_{CF}(p) = \frac{-\phi(CF_{p}^{-1})}{p}\left[1 + \frac{\zeta_1}{6}(CF_{p}^{-1})^3 + \frac{\zeta_2}{24}\left[(CF_{p}^{-1})^4 - 2(CF_{p}^{-1})^2 - 1\right]\right] \tag{7}$$

Here, $\phi(\cdot)$ denotes the density function.

### 4.3.2 GARCH(1,1)

Bollerslev (1986) presented the generalized autoregressive conditional heteroskedasticity model of dynamic variance (GARCH). The model is more sophisticated than RiskMetrics, but requires parameter estimation. The simplest form of the model, GARCH(1,1), is denoted as

$$\sigma_{t+1}^2 = \omega + \alpha R_t^2 + \beta \sigma_t^2 \tag{8}$$

where $\alpha + \beta < 1$. Here, we assume that the standardized returns are normally distributed:

$$R_t = \sigma_t z_t, \quad where\ z_t \sim i.i.d.N(0,1) \tag{9}$$

The three parameters in this model, $\alpha$, $\beta$, $\omega$, needs to be estimated using the maximum likelihood estimation (Christoffersen, 2012, pp. 70-75), given by:

$$MaxlnL = Max \sum_{t=1}^{T}\left[-\frac{1}{2}ln(2\pi) - \frac{1}{2}ln(\sigma_t^2) - \frac{1}{2} - \frac{R_t^2}{\sigma_t^2}\right] \tag{10}$$

Christoffersen (2012) denoted Value-at-Risk using the dynamic variance of GARCH as:

$$VaR_{t+1}^{p} = -\sigma_{PF,t+1}\Phi_{p}^{-1} \tag{11}$$

Christoffersen (2012) denoted Expected Shortfall using the dynamic variance of GARCH as:

$$ES_{t+1}^{p} = \sigma_{PF,t+1}\frac{\phi(\Phi_{p}^{-1})}{p} \tag{12}$$

where $\Phi$ denotes the cumulative density function of the standard normal distribution, and $\phi$ indicates the density function.

A GARCH model with Student's t distribution catches attributes such as fatter tails and pronounced peak in the distribution of $z_t$ better than the GARCH-model presented above. Christoffersen (2012, pp. 128-131) presented this model in the following way. First we need to specify our model returns:

$$R_{PF,t} = \sigma_{PF,t} z_t, \quad where\ z_t \overset{\text{i.i.d.}}{\sim} \tilde{t}(d) \tag{13}$$

To estimate the parameters used in this model, $\alpha$, $\beta$, $\omega$ and $d$, we must maximize the log-likelihood of the sample returns, given by:

$$lnL_2 = \sum_{t=1}^{T} ln(f(R_{PF,t}; d)) \tag{14}$$

where $d$ denotes degrees of freedom. $f(R_{PF,t}; d)$ is denoted as:

$$f(R_{PF,t}; d) = \frac{C(d)}{\sigma_{PF,t}} \left[ \frac{1 + \left( \frac{R_{PF,t}}{\sigma_{PF,t}} \right)^2}{d-2} \right]^{-\frac{1+d}{2}} \tag{15}$$

$C(d)$ is given by:

$$C(d) = \frac{\Gamma((d+1)/2)}{\Gamma(d/2)\sqrt{\pi(d-2)}} \tag{16}$$

A Value-at-Risk model using the dynamic variance of GARCH with Student's t distribution is denoted as:

$$VaR^p_{t+1} = -\sigma_{PF,t+1} \tilde{t}_p^{-1}(d) \tag{17}$$

where $\tilde{t}_p^{-1}$ is the pth quantile of the $\tilde{t}(d)$ distribution. The Expected Shortfall when the mean is zero is denoted by Alexander (2008*b*) as:

$$ES^p_{t+1} = p^{-1}(d-1)^{-1}(d-2+\tilde{t}_p^{-1}(d)^2) f_d(\tilde{t}_p^{-1}(d)) \sigma_{PF,t+1} \tag{18}$$

where

$$f_d(\tilde{t}_p^{-1}(d)) = ((d-2)\pi)^{-\frac{1}{2}} \Gamma\Big(\frac{d}{2}\Big)^{-1} \Gamma\Big(\frac{d+1}{2}\Big) (1 + (d-2)^{-1} \tilde{t}_p^{-1}(d)^2)^{-\frac{1+d}{2}} \tag{19}$$

The GARCH with Student's t VaR and Expected Shortfall can be calculated by using the relationship explained by Christoffersen (2012, p. 132), denoted as:

$$\tilde{t}_p^{-1}(d) = \sqrt{\frac{d-2}{d}} t_p^{-1}(d) \tag{20}$$

Christoffersen (2012) described an extension of the GARCH model, referred to as GJR-GARCH, that captures the leverage effect by defining an indicator variable $I_t$:

$$I_t = \begin{cases} 1 \text{ if } \mathrm{R}_t < 0 \\ 0 \text{ if } \mathrm{R}_t \geq 0 \end{cases} \tag{21}$$

According to him, the variance dynamics can then be specified as

$$\sigma^2_{t+1} = \omega + \alpha R_t^2 + \alpha\theta I_t R_t^2 + \beta \sigma_t^2 \tag{22}$$

### 4.3.3 Filtered Historical Simulation

Christoffersen (2012) described Historical Simulation (HS) as one of the most common models for estimating VaR. The main reason may be the ease of implementation that HS offers, according to him. There is no need of estimating any parameters, and the only consideration that has to be made, is the sample length. HS does not rely on any particular parametric model, and Christoffersen (2012) emphasized that this model-free nature prevents any pitfalls a model-dependent model may fall in if the parametric model is poor. The fact that one needs to set the sample length, is however a drawback. A small sample may not include enough large losses to calculate the VaR with precision. The advantage is that the most recent observations carry a larger weight, and the model will thus react faster to shocks, according to Christoffersen (2012). He notified that the total empirical distribution of returns with sample length m can then be interpreted as $\{R_{t+1-\tau}\}_{\tau=1}^{m}$. The VaR is then:

$$VaR_{t+1}^{p} = -Percentile(\{R_{t+1-\tau}\}_{\tau=1}^{m}, 100p) \tag{23}$$

where p is the 100pth percentile. Christoffersen(2012) proposed a Filtered Historical Simulation model that aims to combine the best of the model-free approach of Historical Simulation and a model-based method of dynamic variance (Christoffersen, 2012, pp. 124-126). The VaR can be calculated by:

$$VaR_{t+1}^{p} = -\sigma_{PF,t+1} Percentile(\{\hat{z}_{t+1-\tau}\}_{\tau=1}^{m}, 100p) \tag{24}$$

where he described $\hat{z}_{t+1-\tau}$ as the standardized returns denoted by

$$\hat{z}_{t+1-\tau} = \frac{R_{PF,t+1-\tau}}{\sigma_{PF,t+1-\tau}} \tag{25}$$

and $\sigma_{PF,t+1}$ is the dynamic variance. Christoffersen (2012) denoted the Expected Shortfall as:

$$ES_{t+1}^{p} = -\frac{\sigma_{PF,t+1}}{p \cdot m} \sum_{i=1}^{m} \hat{z}_{t+1-\tau} \cdot \mathbf{1}(\hat{z}_{t+1-\tau} < -Percentile\{\{\hat{z}_{t+1-\tau}\}_{\tau=1}^{m}, 100p\}) \tag{26}$$

where **1** is an indicator function which returns 0 if the argument is false, and 1 if the argument is true.

### 4.3.4 Quantile regression

Alexander (2008*a*) informed that quantile regression aims to compute a set of regression curves which corresponds to quantiles of of the conditional distribution of the dependent variable. He denoted the quantile regression model as

$$Y_t^q = \alpha^q + \beta^q X_t + \epsilon_t^q \tag{27}$$

where Y is the dependent variable, q is the quantile, and $\epsilon$ is an independent and identically distributed error term that depends on the independent variable X. $\alpha$ and $\beta$ are constant parameters that need to be estimated using the optimization problem denoted using the notation of Alexander (2008*a*, pp. 305-306):

$$\min_{\alpha,\beta} \sum_{t=1}^{T} (q - \mathbf{1}_{Y_t \leq \alpha + \beta X_t})(Y_t - (\alpha + \beta X_t)) \tag{28}$$

where

$$\mathbf{1}_{Y_t \leq \alpha + \beta X_t} = \begin{cases} 1 \text{ if } \mathrm{Y}_t \leq \alpha + \beta X_t, \\ 0 \text{ otherwise} \end{cases} \tag{29}$$

Steen et al. (2015, p. 67) suggested that the VaR is expressed as

$$VaR_t^q|\sigma_{t-1} = \hat{\alpha}_t^q + \hat{\beta}_t^q \sigma_{t-1} + \epsilon_t^q|\sigma_{t-1} \tag{30}$$

where a unique set of regression parameters ($\alpha$, $\beta$) can be obtained for each quantile of interest.

## 4.4 Model testing and evaluation

"When backtesting the risk model, we construct a sequence $\{I_{t+1}\}_{t=1}^T$ across $T$ days indicating when past violations occurred"(Christoffersen, 2012, p.301). A rightly specified $VaR$-model, when compared to the actual losses, should produce the same fraction of violations as the specified $VaR$-level p. Additionally the sequence of violations $\{I_{t+1}\}_{t=1}^T$ should be independent and identically distributed. (Christoffersen, 2012, pp.301-302). Christoffersen (2012, p.301) defined the sequence of violations as:

$$I_{t+1} = \begin{cases} 1, \text{ if -R}_{t+1} > VaR_{t+1}^p \\ 0, \text{ if -R}_{t+1} \leq VaR_{t+1}^p \end{cases} \tag{31}$$

meaning $I_{t+1}$ takes the value 1 if a violation is recorded, and 0 if not. The two null hypothesis we need to test are:

$$H_0 : \frac{\sum_{t=1}^T I_{t+1}}{T} = (1-p) \tag{32}$$

$$H_0 : I_{t+1} \sim i.i.d. Bernoulli(1-p) \tag{33}$$

where p is the VaR-level under consideration and $I_{t+1}$ is defined as in eq.(31). To test for the null hypothesis in eq.(32) we use the unconditional coverage test introduced by Kupiec (1995). The null hypothesis in eq.(33) will be tested applying the conditional independence test introduced by Christoffersen (1998).

### 4.4.1 Unconditional coverage: The Kupiec Test

The unconditional coverage test of Kupiec, tests if the predicted VaR produces the promised fraction of violations specified by the VaR-level. If the fraction of violations is significantly different at a specified significance level, the null hypothesis is rejected. The unconditional coverage hypothesis is tested using a likelihood ratio test. Using similar notation to that of Christoffersen (2012, p.303) we have:

$$LR_{uc} = -2ln\left[\frac{p^{T_0} \cdot (1-p)^{T_1}}{(1-\frac{T_1}{T})^{T_0} \cdot (\frac{T_1}{T})^{T_1}}\right] \sim \chi_1^2 \tag{34}$$

where $T_1 = \sum_{t=1}^T I_{t+1}$, $T_0 = T - T_1$, $p$ is the VaR-level and $LR_{uc}$ is asymtotically distributed as $\chi_1^2$. From our test statistic $LR_{uc}$ we can then easily calculate the p-value:

$$P - value \equiv 1 - F_{\chi_1^2}(LR_{uc}) \tag{35}$$

where $F_{\chi_1^2}(*)$ denotes the cumulative density function of a $\chi_1^2$ variable (Christoffersen, 2012, p.303). If the $P$-value is smaller than the significance level we chose, the null hypothesis from eq.(32) is rejected.

### 4.4.2 Conditional independence

Christoffersen (1998) introduced "The LR test of Independence". He estimated a first-order Markov chain on the sequence $\{I_{t+1}\}_{t=1}^{T}$ and tested the hypothesis that an exceedance is independent from the previous exceedance. From the standard result in Hoel(1954 as cited in (Christoffersen, 1998, p.846)), the likelihood ratio test of independence can be written as:

$$LR_{ind} = -2ln\left[\frac{(1-\frac{T_1}{T})^{T_0}\cdot(\frac{T_1}{T})^{T_1}}{(1-\frac{T_{01}}{T_{00}+T_{01}})^{T_{00}}\cdot(\frac{T_{01}}{T_{00}+T_{01}})^{T_{01}}\cdot(1-\frac{T_{11}}{T_{11}+T_{10}})^{T_{10}}\cdot(\frac{T_{11}}{T_{11}+T_{10}})^{T_{11}}}\right] \sim \chi_1^2 \tag{36}$$

where $T_{ij}, i, j = 0, 1$ is the number of observations with a $j$ following an $i$. $LR_{ind}$ is also asymtotically distributed as $\chi_1^2$ (Christoffersen, 1998, p.846). The $P-value$ is calculated as in eq.(35) with $LR_{ind}$ instead of $LR_{uc}$. In the case of $T_{11} = 0$ we substitute the denominator in eq.(36) with
$(1-\frac{T_{01}}{T_{00}+T_{01}})^{T_{00}}\cdot(\frac{T_{01}}{T_{00}+T_{01}})^{T_{01}}$ (Christoffersen, 2012, p.306).

### 4.4.3 Backtesting Expected Shortfall

As previously presented, the backtesting of Value-at-Risk is only concerned with the number of exceedances of the VaR-barrier, and whether these exceedances are independent or not. For the backtesting of ES, we are also concerned about the actual size of the exceedance when the exceedance occur. McNeil and Frey (2000, p.294)developed a method for backtesting Expected Shortfall. The methodology is based on a time series of exceedance residuals and hypothesis testing using bootstrap. The hypothesis testing can be conducted using standardised or unstandardised residuals with similar results according to McNeil and Frey (2000). The first step in backtesting the Expected Shortfall is to extract the residuals:

$$\{\varepsilon_{t+1} : t \in T, -R_{t+1} > VaR_{1,\alpha,t}\}, \text{where } \varepsilon_{t+1} = -R_{t+1} - ES_{1,\alpha,t} \tag{37}$$

An adequate ES-model manages to make estimates that are closely related to the actual residuals one experience when using the model. We conduct a one-sided hypothesis test with the null hypothesis being; the mean of the residuals, denoted $\mu_\varepsilon$, equal to zero. The alternative being that the mean is greater than zero. If the null is rejected, the risk is systematically underestimated.

$$\begin{aligned} &\mathrm{H}_0 : \mu_\varepsilon = \mu_0 \\ &\mathrm{H}_1 : \mu_\varepsilon > \mu_0 \end{aligned} \tag{38}$$

Here, $\mu_0$ equals zero. The alternative hypothesis is set to the mean being greater than zero since this is the most likely direction of failure (McNeil and Frey, 2000). The hypothesis testing is conducting using a bootstrap test that makes no assumption about the underlying distribution of the residuals. For ES-calculations with few residuals, a Bootstrap estimation similar to that of Efron and Tibshirani (1993, p.224), is useful in this regard. The procedure can be summarized in the following fashion:

1. Use the total sample of n residuals, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)$, and calculate the approximate distribution of the test statistic $t(\boldsymbol{\varepsilon}) = \frac{\bar{\varepsilon}-\mu_0}{\bar{\sigma}/\sqrt{n}}$, where $\bar{\varepsilon}$ and $\bar{\sigma}$ is the mean and standard deviation of $\boldsymbol{\varepsilon}$.

2. Calculate the empirical distribution of the points: $\tilde{\varepsilon}_i = \varepsilon_i - \bar{\varepsilon} + \mu_0$

3. Sample B number of times $\tilde{\varepsilon}_1^*$, ..., $\tilde{\varepsilon}_n^*$ with replacement from $\tilde{\varepsilon}_1$, ..., $\tilde{\varepsilon}_n$

4. For each sample, compute: $t(\tilde{\boldsymbol{\varepsilon}}^*) = \frac{\bar{\tilde{\varepsilon}}^* - \mu_0}{\tilde{\sigma}^*/\sqrt{n}}$

5. Estimate the achieved significance level: $\widehat{ASL}_{boot} = \#\{t((\tilde{\boldsymbol{\varepsilon}}^{*\boldsymbol{b}}) \geq t(\boldsymbol{\varepsilon})\}/\mathrm{B}$

The achieved significance level (ASL) is defined to be "the probability of observing at a least that large a value when the null hypothesis is true" (Efron and Tibshirani, 1993, p.203). As Efron and Tibshirani (1993) further advocates, this means that the smaller the value of ASL, the stronger the evidence against the null hypothesis, $H_0$. A large ASL thus indicates that the model probably does not underestimate the risk. Since underestimation of risk is the most important feature of an ES-model to disclose, the ASL provides vital information when comparing different ES-models. The backtesting procedure of McNeil and Frey (2000) does however not consider adequately whether the ES-model overestimates the risk. A model that exaggerates the risk greatly will therefore perform very well using the backtesting procedure of McNeil and Frey (2000), even though such a model obviously fails to provide the correct Expected Shortfall. To obtain a more holistic view of the model's performance, we use the bootstrapping method to construct a simulated distribution of the residuals at hand. An empirical analysis of the simulated distribution add valuable information to the model evaluation, in addition to the information the backtesting procedure of McNeil and Frey (2000) provides.

### 4.5 Value-at-Risk-averaging and Expected Shortfall-averaging approach

Following the calculation of the Value-at-Risk quantiles of the individual models presented above, we will compute equally weighted averages of these quantiles, and study the performance of these. We will also study equally weighted averages of the expected shortfall quantiles for the models chosen in a similar fashion as we will do with the VaR-averaging. The execution of these simple average-models is further elaborated in section 4.6.

### 4.6 Test procedures

The models presented were chosen to study in-sample fit and out-of-sample performance of Value-at-Risk and Expected Shortfall. For the out-of-sample calculations, we used the first half of all the data available to study the in-sample performance, and chose the composition of the simple average models based on the in-sample results, before we tested the out-of-sample performance for the last half of the data available. Value-at-Risk, Expected Shortfall and dynamic variance for all models were coded using Python. To estimate the GARCH-parameters, we used the function "archmodel" from the "ARCH" library. To estimate the parameters from the quantile regression, we used the "QuantReg" function from "statsmodels" library. We used "random.choice" from "numpy" library for the bootstrapping and backtesting of the Expected Shortfall. The backtesting of VaR were conducted using the "varbacktest" function from the Econometric Toolbox in Matlab. The results are presented and interpreted in section 5.

# 5 Results and discussion

This section is divided into two parts; the first part aims to study the in-sample fit. The second part aims to study the out-of-sample performance, as Westgaard et al. (2014) emphasize the importance of.

1. The first part, section 5.2, aims to study the in-sample fit for both Value-at-Risk and Expected Shortfall of the entire data set for all individual models.

   The result of the in-sample fit backtesting of the VaR-models is presented under section 5.2.1 in table 16, 17 and 18, and the results are summarized in table 15. The in-sample fit backtesting performance of the ES-models is presented under section 5.3.2 in table 20, 22 and 24, and the results are summarized in table 19. GARCH-parameters and QR-parameters used to estimate the volatility and VaR/ES for the in-sample fit are displayed in the appendix.

2. The second part, section 5.3, aims to study the out-of-sample performance of the models.

   First, we conduct the in-sample tests of the first half of the dataset. The results from the VaR-backtests are presented under section 5.3.1 in table 27, 28, 29, and summarized in table 26.

   The in-sample backtests of the individual ES-models are presented under section 5.3.2 in table 31, 33, 35, and summarized in table 30.

   The choice of simple average models is based on the in-sample performance, and presented in section 5.3.1, with shortenings of the models summarized in section 5.1 table 13.

   Next, we assess the out-of-sample performance. The results of the VaR-backtesting are presented under section 5.3.3 in table 38, 39, 40, and summarized in table 37.

   The out-of-sample results of the backtesting of the ES-models are presented under section 5.3.4 in table 42, 44, 46, and summarized in table 41.

VaR-backtesting for the different models and simple average models used to predict ES will also be included. This is important because an ES-model is more or less worthless if the underlying VaR-model fails to provide the promised fraction of violations specified by the VaR-level, or if clustering of violations occur to a significant extent.

## 5.1 Simple average models

The constellations of the simple average models are based on in-sample results in section 5.3.1 and 5.3.2 for VaR and ES respectively. The choice of simple average models is justified in the same section. The shortenings used for the simple average models are summarized in table 13 and 14.

| Shortenings of VaR-simple average models | |
|---|---|
| Shortening | Models |
| "All above" | GARCH-n, GARCH-t, GJR-GARCH-n, GJR-GARCH-t, QR-GARCH-n, QR-GARCH-t, QR-GJR-GARCH-n, QR-GJR-GARCH-t, QR-RM, RM-CF and FHS |
| "Average1" | GARCH-t, GJR-GARCH-t, QR-GARCH-t, QR-GJR-GARCH-t, QR-RM and FHS |
| "Average2" | GARCH-t, QR-GARCH-t and FHS |
| "Average3" | QR-GARCH-t and GARCH-t |
| "Average4" | QR-GARCH-t and FHS |

Table 13: Shortenings of the VaR-simple average models. The left column show the shortenings we will use hereafter. In the right column, the models used in the corresponding average models to the left are listed

| Shortenings of ES-simple average models | |
|---|---|
| Shortening | Models |
| $G_t/G_n$/FHS/RM-CF | GARCH-t, GARCH-n, FHS, RM-CF |
| $G_t/G_n$/FHS | GARCH-t, GARCH-n, FHS |
| $G_t$/FHS | GARCH-t, FHS |
| $G_n$/FHS | GARCH-n, FHS |

Table 14: Shortenings of the ES-simple average models. The left column show the shortenings we will use hereafter. In the right column, the models used in the corresponding average models to the left are listed

## 5.2 In-sample fit of the entire dataset

This section is devoted to study the performance of the in-sample fit of the VaR- and ES-models on the entire data set. The performance of the in-sample fit for VaR is presented in section 5.2.1, and the performance of the in-sample fit for ES is presented in section 5.2.2.

### 5.2.1 In-sample fit VaR Results of the entire dataset

The results following the in-sample fit backtesting of the VaR-models are presented in table 16, 17 and 18. The results are summarized in table 15. Note that we will include the test statistics for the simple-average models in this part, even though the constellation of these models are decided in section 5.3.

The Value-at-Risk backtesting performed at the 11 individual models on the Front Month, Front Quarter and Front Year contract, reveal that QR-GARCH-t and QR-GJR-GARCH-n are superior in terms of number of test rejections, both with a total of 5 rejections at the 5% significance level. GJR-GARCH-t and QR-GARCH-n both have 6 rejections, while GARCH-t has 7 rejections, and QR-GJR-GARCH-t has 8 rejections. QR-RM and FHS have 10 and 11 rejections respectively, while GARCH-n and GJR-GARCH-n both have 15 rejections. RM-CF has 19 rejections, and performs very poor in this regard compared to the other VaR-models.

The general tendency of the in-sample fit test is that the quantile regression approach seems to perform well for the highly volatile Nordic power future contracts. QR-RM provides better results than the RM-CF approach. The GARCH-t models seems to capture the fat tails previously described in table 3, and figure 2 and 3, in section 3, and generally outperforms the GARCH-n approaches because of this. The QQ-plots presented in figure 7 and 8 further confirms the superiority of the GARCH-t approach compared to the GARCH-n approach for the Nordic power futures.

The seven best individual models have very few test rejections for unconditional coverage, suggesting that the models do predict a Value-at-Risk level capable of catching the correct amount of rejections for the 5% significance level. The models do however not withstand the conditional independence test as well as the kupiec test. It seems like the models underestimate risk when volatility is increasing fast, resulting in too low estimates of Value-at-Risk in these periods, and thus consecutive exceedances. An illustration is provided in figure 9. The descriptive statistics of the Nordic power futures in question, presented in section 3, revealed the highly volatile nature of the commodity, and that volatility clustering does appear occasionally. The models presented for the in-sample fit test are thus only partly capable of accounting for the volatility clustering of Nordic power futures.

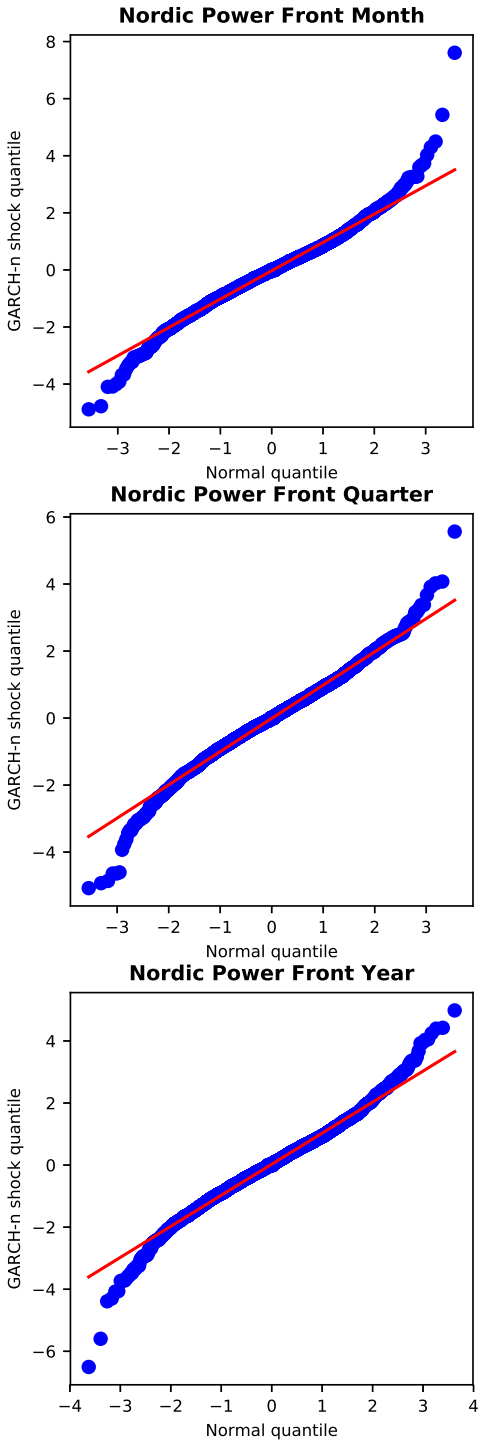



Figure 7: QQ-plot of GARCH-n shock quantiles against normal quantiles. The straight line indicates where normally distributed data would hit. The utmost values on both sides of all QQ-plots confirms more extreme values than normal quantiles entail, and thus fat tails

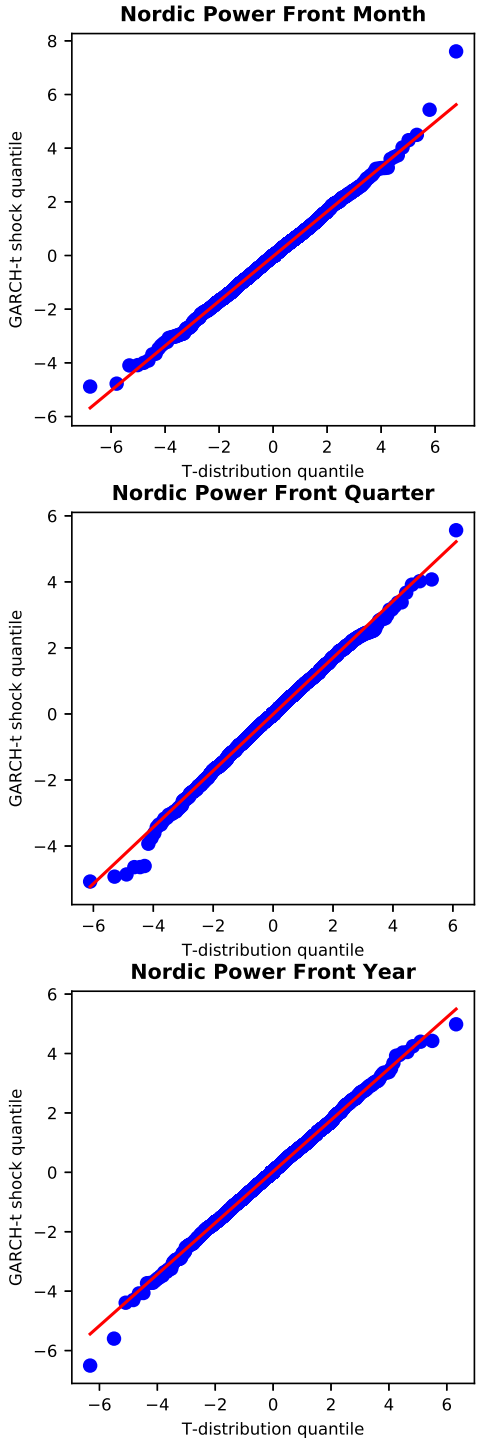



Figure 8: QQ-plot of GARCH-t shock quantiles against t-distribution quantiles. The straight line indicates where t-distributed data would hit. The shape indicates that the t-distribution fits the tails better than the normal distribution does

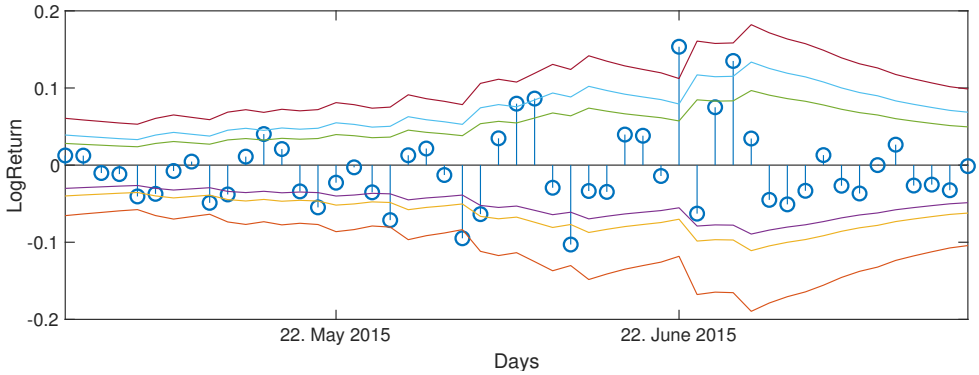



Figure 9: This is a period of volatility clustering in the Front Month contract during the summer of 2015. The six lines represent VaR estimates from the QR-GARCH-t model. The orange and red lines are 99% long and short VaR, the yellow and turquoise lines are 95% long and short VaR and the purple and green lines are 90% long and short VaR. The logreturns are illustrated as stems. We see that when volatility increase, the model does not manage to update the VaR fast enough, resulting in occasional consecutive exceedances for the 90% VaR and the 95% VaR

The simple average models perform very well, even though neither is able to outperform the two best individual models. Average2 and Average4 have a total of 6 rejections. Average1 and Average3 have a total of 7 rejections, while "All above" has 8 rejections. Neither of the models have any UC-rejections, indicating that the simple average models do produce the promised fraction of violations. The rejections of CI indicates that the simple average models tend to underestimate risk when volatility is increasing fast, resulting in too low estimates of Value-at-Risk in these periods, and thus clustering of exceedances. Figure 9 illustrates this phenomenon.

| Total in-sample fit test rejections for VaR | | | | |
|---|---|---|---|---|
| Rank | Model | CI | UC | Sum |
| 1 | QR-GARCH-t | 5 | 0 | 5 |
| 1 | QR-GJR-GARCH-n | 5 | 0 | 5 |
| 3 | Average2 | 6 | 0 | 6 |
| 3 | Average4 | 6 | 0 | 6 |
| 3 | GJR-GARCH-t | 5 | 1 | 6 |
| 3 | QR-GARCH-n | 6 | 0 | 6 |
| 7 | Average1 | 7 | 0 | 7 |
| 7 | Average3 | 7 | 0 | 7 |
| 7 | GARCH-t | 6 | 1 | 7 |
| 10 | All above | 8 | 0 | 8 |
| 10 | QR-GJR-GARCH-t | 8 | 0 | 8 |
| 12 | QR-RM | 10 | 0 | 10 |
| 13 | FHS | 8 | 3 | 11 |
| 14 | GARCH-n | 7 | 8 | 15 |
| 14 | GJR-GARCH-n | 8 | 7 | 15 |
| 16 | RM-CF | 10 | 9 | 19 |

Table 15: In-sample fit VaR test rejections for the conditional independence test(CI) and the kupiec unconditional coverage test(UC) for the entire dataset. The average-models are listed in table 13

| In-sample fit backtesting results for Value-at-Risk for Front Month Nordic power futures | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | VaR-levels | | | | | | | | | | | | Rejects |
| | Long positions | | | | | | Short positions | | | | | | at |
| | 99% | | 95% | | 90% | | 90% | | 95% | | 99% | | 5% |
| Model | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | sign. |
| GARCH-n | 0.18 | 0.00 | 0.67 | 0.80 | 0.03 | 0.33 | 0.01 | 0.00 | 0.02 | 0.19 | 0.62 | 0.10 | 5/12 |
| GARCH-t | 0.35 | 0.48 | 0.24 | 0.11 | 0.00 | 0.18 | 0.02 | 0.12 | 0.04 | 0.77 | 0.48 | 0.32 | 3/12 |
| GJR-GARCH-n | 0.18 | 0.00 | 0.69 | 0.86 | 0.03 | 0.35 | 0.01 | 0.00 | 0.02 | 0.19 | 0.62 | 0.10 | 5/12 |
| GJR-GARCH-t | 0.36 | 0.59 | 0.34 | 0.15 | 0.00 | 0.19 | 0.02 | 0.11 | 0.04 | 0.77 | 0.48 | 0.32 | 3/12 |
| QR-GARCH-n | 0.39 | 0.84 | 0.98 | 0.99 | 0.01 | 0.93 | 0.03 | 0.85 | 0.01 | 0.92 | 0.38 | 0.97 | 3/12 |
| QR-GARCH-t | 0.39 | 0.84 | 0.98 | 0.99 | 0.01 | 0.93 | 0.03 | 0.85 | 0.01 | 0.92 | 0.38 | 0.97 | 3/12 |
| QR-GJR-GARCH-n | 0.39 | 0.84 | 0.98 | 0.99 | 0.01 | 0.89 | 0.03 | 0.85 | 0.01 | 0.92 | 0.38 | 0.97 | 3/12 |
| QR-GJR-GARCH-t | 0.39 | 0.84 | 0.98 | 0.99 | 0.01 | 0.93 | 0.03 | 0.85 | 0.01 | 0.86 | 0.38 | 0.97 | 3/12 |
| QR-RM | 0.41 | 0.71 | 0.11 | 0.86 | 0.00 | 0.93 | 0.00 | 0.93 | 0.00 | 0.92 | 0.40 | 0.84 | 3/12 |
| RM-CF | 0.47 | 0.41 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.06 | 0.00 | 0.19 | 0.60 | 0.01 | 7/12 |
| FHS | 0.24 | 0.02 | 0.71 | 0.17 | 0.01 | 0.76 | 0.00 | 0.49 | 0.20 | 0.44 | 0.57 | 0.18 | 3/12 |
| All above | 0.32 | 0.31 | 0.50 | 0.44 | 0.00 | 0.45 | 0.01 | 0.38 | 0.01 | 0.89 | 0.43 | 0.64 | 3/12 |
| Average1 | 0.35 | 0.48 | 0.67 | 0.53 | 0.00 | 0.49 | 0.02 | 0.63 | 0.02 | 0.95 | 0.43 | 0.64 | 3/12 |
| Average2 | 0.32 | 0.31 | 0.99 | 0.39 | 0.01 | 0.49 | 0.01 | 0.72 | 0.07 | 0.80 | 0.45 | 0.52 | 2/12 |
| Average3 | 0.36 | 0.59 | 0.93 | 0.53 | 0.00 | 0.52 | 0.02 | 0.48 | 0.02 | 0.89 | 0.45 | 0.52 | 3/12 |
| Average4 | 0.32 | 0.31 | 0.90 | 0.58 | 0.01 | 0.56 | 0.01 | 0.56 | 0.12 | 0.44 | 0.41 | 0.90 | 2/12 |

Table 16: In-sample fit P-values for the entire dataset of the conditional independence test(CI) and the Kupiec unconditional coverage test(UC). Red indicate that the null hypothesis of conditional independence or unconditional coverage has been rejected at the 5% significance level. The average-models are listed in table 13. The period of data is stated in table 2

| In-sample fit backtesting results for Value-at-Risk for Front Quarter Nordic power futures | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | VaR-levels | | | | | | | | | | | | Rejects |
| | Long positions | | | | | | Short positions | | | | | | at |
| | 99% | | 95% | | 90% | | 90% | | 95% | | 99% | | 5% |
| Model | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | sign. |
| GARCH-n | 0.87 | 0.00 | 0.63 | 0.93 | 0.81 | 0.09 | 0.00 | 0.00 | 0.04 | 0.38 | 0.13 | 0.09 | 4/12 |
| GARCH-t | 0.49 | 0.30 | 0.69 | 0.19 | 0.92 | 0.61 | 0.04 | 0.32 | 0.10 | 0.81 | 0.63 | 0.01 | 2/12 |
| GJR-GARCH-n | 0.07 | 0.00 | 0.65 | 0.87 | 0.95 | 0.13 | 0.00 | 0.00 | 0.15 | 0.26 | 0.32 | 0.23 | 3/12 |
| GJR-GARCH-t | 0.51 | 0.23 | 0.72 | 0.22 | 0.82 | 0.40 | 0.04 | 0.45 | 0.09 | 0.75 | 0.64 | 0.01 | 2/12 |
| QR-GARCH-n | 0.34 | 0.97 | 0.38 | 0.44 | 0.92 | 0.43 | 0.04 | 0.78 | 0.06 | 0.64 | 0.45 | 0.48 | 1/12 |
| QR-GARCH-t | 0.05 | 0.84 | 0.40 | 0.48 | 0.89 | 0.40 | 0.02 | 0.74 | 0.06 | 0.64 | 0.43 | 0.59 | 1/12 |
| QR-GJR-GARCH-n | 0.05 | 0.84 | 0.67 | 0.53 | 0.97 | 0.40 | 0.06 | 0.78 | 0.10 | 0.70 | 0.45 | 0.48 | 0/12 |
| QR-GJR-GARCH-t | 0.00 | 0.97 | 0.42 | 0.53 | 0.97 | 0.40 | 0.04 | 0.78 | 0.03 | 0.64 | 0.45 | 0.48 | 3/12 |
| QR-RM | 0.01 | 0.84 | 0.97 | 0.59 | 0.96 | 0.57 | 0.07 | 0.74 | 0.00 | 0.64 | 0.43 | 0.59 | 2/12 |
| RM-CF | 0.20 | 0.16 | 0.41 | 0.14 | 0.88 | 0.00 | 0.01 | 0.00 | 0.02 | 0.00 | 0.54 | 0.11 | 5/12 |
| FHS | 0.15 | 0.04 | 0.90 | 0.11 | 0.94 | 0.31 | 0.02 | 0.57 | 0.02 | 0.48 | 0.58 | 0.13 | 3/12 |
| All above | 0.37 | 0.84 | 0.57 | 0.35 | 0.69 | 0.69 | 0.00 | 0.60 | 0.07 | 0.94 | 0.42 | 0.89 | 1/12 |
| Average1 | 0.37 | 0.84 | 0.57 | 0.35 | 0.78 | 0.61 | 0.06 | 0.77 | 0.04 | 0.82 | 0.46 | 0.50 | 1/12 |
| Average2 | 0.41 | 0.59 | 0.80 | 0.31 | 0.46 | 0.53 | 0.02 | 0.77 | 0.01 | 0.64 | 0.46 | 0.50 | 2/12 |
| Average3 | 0.39 | 0.71 | 0.30 | 0.25 | 0.93 | 0.43 | 0.04 | 0.68 | 0.05 | 0.87 | 0.48 | 0.39 | 1/12 |
| Average4 | 0.41 | 0.59 | 0.80 | 0.31 | 0.31 | 0.65 | 0.02 | 0.95 | 0.01 | 0.53 | 0.38 | 0.71 | 2/12 |

Table 17: In-sample fit P-values for the entire dataset of the conditional independence test(CI) and the Kupiec unconditional coverage test(UC). Red indicate that the null hypothesis of conditional independence or unconditional coverage has been rejected at the 5% significance level. The average-models are listed in table 13. The period of data is stated in table 2

| In-sample fit backtesting results for Value-at-Risk for Front Year Nordic power futures | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | VaR-levels | | | | | | | | | | | | Rejects |
| | Long positions | | | | | | Short positions | | | | | | at |
| | 99% | | 95% | | 90% | | 90% | | 95% | | 99% | | 5% |
| Model | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | sign. |
| GARCH-n | 0.28 | 0.02 | 0.16 | 0.20 | 0.30 | 0.00 | 0.07 | 0.03 | 0.00 | 0.22 | 0.01 | 0.00 | 6/12 |
| GARCH-t | 0.40 | 0.57 | 0.13 | 0.83 | 0.14 | 0.86 | 0.08 | 0.27 | 0.01 | 0.68 | 0.00 | 0.45 | 2/12 |
| GJR-GARCH-n | 0.32 | 0.01 | 0.04 | 0.17 | 0.05 | 0.01 | 0.03 | 0.05 | 0.02 | 0.35 | 0.04 | 0.00 | 7/12 |
| GJR-GARCH-t | 0.11 | 0.74 | 0.06 | 0.54 | 0.09 | 0.78 | 0.12 | 0.32 | 0.01 | 0.83 | 0.16 | 0.29 | 1/12 |
| QR-GARCH-n | 0.10 | 0.85 | 0.26 | 0.63 | 0.28 | 0.54 | 0.07 | 0.54 | 0.02 | 0.54 | 0.02 | 0.54 | 2/12 |
| QR-GARCH-t | 0.10 | 0.85 | 0.16 | 0.63 | 0.23 | 0.51 | 0.06 | 0.58 | 0.02 | 0.54 | 0.12 | 0.54 | 1/12 |
| QR-GJR-GARCH-n | 0.10 | 0.85 | 0.03 | 0.63 | 0.23 | 0.51 | 0.12 | 0.58 | 0.02 | 0.54 | 0.12 | 0.54 | 2/12 |
| QR-GJR-GARCH-t | 0.10 | 0.85 | 0.03 | 0.63 | 0.28 | 0.54 | 0.08 | 0.61 | 0.02 | 0.58 | 0.12 | 0.54 | 2/12 |
| QR-RM | 0.01 | 0.85 | 0.03 | 0.68 | 0.03 | 0.54 | 0.01 | 0.75 | 0.03 | 0.68 | 0.11 | 0.74 | 5/12 |
| RM-CF | 0.31 | 0.18 | 0.04 | 0.41 | 0.07 | 0.01 | 0.02 | 0 | 0.00 | 0.01 | 0.02 | 0.63 | 7/12 |
| FHS | 0.74 | 0.11 | 0.04 | 0.22 | 0.02 | 0.30 | 0.08 | 0.48 | 0.01 | 0.34 | 0.02 | 0.01 | 5/12 |
| All above | 0.09 | 0.97 | 0.03 | 0.68 | 0.13 | 0.78 | 0.03 | 0.61 | 0.01 | 0.84 | 0.02 | 0.45 | 4/12 |
| Average1 | 0.09 | 0.97 | 0.03 | 0.54 | 0.14 | 0.98 | 0.03 | 0.54 | 0.01 | 0.73 | 0.11 | 0.63 | 3/12 |
| Average2 | 0.47 | 0.91 | 0.13 | 0.78 | 0.16 | 0.58 | 0.06 | 0.45 | 0.01 | 0.73 | 0.00 | 0.18 | 2/12 |
| Average3 | 0.43 | 0.68 | 0.14 | 0.73 | 0.13 | 0.98 | 0.02 | 0.54 | 0.01 | 0.78 | 0.00 | 0.45 | 3/12 |
| Average4 | 0.56 | 0.63 | 0.08 | 0.38 | 0.14 | 0.48 | 0.07 | 0.51 | 0.01 | 0.58 | 0.00 | 0.08 | 2/12 |

Table 18: In-sample fit P-values for the entire dataset of the conditional independence test(CI) and the Kupiec unconditional coverage test(UC). Red indicate that the null hypothesis of conditional independence or unconditional coverage has been rejected at the 5% significance level. The average-models are listed in table 13. The period of data is stated in table 2

### 5.2.2 In-sample fit ES Results of the entire dataset

The in-sample fit backtesting performance of the ES-models is presented in this section, in table 20, 22 and 24. The VaR-backtesting results for the same models are presented in table 21, 23 and 25. The results are summarized in table 19.

The backtesting of McNeil and Frey (2000, p.294), reveals that FHS and GARCH-t are by far the best models in predicting the Expected Shortfall, when comparing the 4 individual models chosen for calculating Expected Shortfall on Front Month, Front Quarter and Front Year data. FHS and GARCH-t produced a total of 0 and 1 rejection respectively. GARCH-n and RM-CF perform horrible with a total of 17 and 18 rejections respectively. The results indicate that FHS and GARCH-t manage to account for the fat tails adequately, previously described under section 3.4 in table 3.

The VaR backtesting of the individual ES-models shows that GARCH-t and FHS perform best with a total of 9 and 11 rejections respectively, while GARCH-n and RM-CF perform poor with a total of 15 and 19 rejections respectively. This further emphasize that GARCH-n and RM-CF fail to accurately estimate the risk of the highly volatile Nordic power futures contracts. GARCH-t manages to produce the promised fraction of violations specified by the VaR-level, and provides adequate ES-predictions, but it has many conditional independence rejections, due to the consecutive VaR-exceedences previously illustrated in figure 9. Hence, the model fails to react fast to sudden large changes in volatility. The same tendency is apparent for FHS. The many UC-, ES- and CI- rejections for GARCH-n and RM-CF, reveal that these models; fail to produce the promised fraction of violations specified, fail to provide adequate ES-predictions and fail to react fast to sudden large changes in volatility. Hence, making these models next to useless for Nordic power futures risk management.

Out of the simple average models, only $G_t/G_n$/FHS/RM-CF performs very poor with 18 ES-rejections and 8 VaR-rejections. $G_t/G_n$/FHS, $G_t$/FHS and $G_n$/FHS have 0 ES-rejections and 5, 7 and 9 VaR-rejections respectively. These simple average models are therefore the best performing ES-models for the in-sample fit of the entire data set, when taking both ES-rejections and VaR-rejections into account.

| In-sample test rejections for ES | | | |
|---|---|---|---|
| Rank | Model | ES | VaR |
| 1 | $G_t/FHS$ | 0 | 7 |
| 1 | $G_t/G_n/FHS$ | 0 | 5 |
| 1 | $G_n/FHS$ | 0 | 9 |
| 1 | FHS | 0 | 11 |
| 5 | GARCH-t | 1 | 9 |
| 6 | GARCH-n | 17 | 15 |
| 7 | $G_t/G_n/FHS$/RM-CF | 18 | 8 |
| 7 | RM-CF | 18 | 19 |

Table 19: Total number of in-sample fit test rejections for ES, and total VaR backtesting test rejections. The rank is based on ES-rejections. $G_t$ is GARCH-t and $G_n$ is GARCH-n

| In-sample fit for Expected Shortfall ASL-values for Front Month Nordic power futures | | | | | | | |
|---|---|---|---|---|---|---|---|
| | ES-levels | | | | | | Rejects |
| | Long positions | | | Short positions | | | at |
| | 99% | 95% | 90% | 90% | 95% | 99% | 5% |
| GARCH-n | 0 | 0 | 0 | 0 | 0 | 0 | 6/6 |
| GARCH-t | 0.779 | 0.687 | 0.522 | 0.36 | 0.567 | 0.234 | 0/6 |
| RM-CF | 0 | 0 | 0 | 0 | 0 | 0 | 6/6 |
| FHS | 1 | 1 | 0.964 | 0.987 | 1 | 1 | 0/6 |
| $G_t/G_n$/FHS/RM-CF | 0 | 0 | 0 | 0 | 0 | 0 | 6/6 |
| $G_t/G_n$/FHS | 1 | 0.55 | 0.372 | 0.272 | 0.66 | 0.942 | 0/6 |
| $G_t$/FHS | 1 | 0.982 | 0.874 | 0.78 | 0.986 | 0.997 | 0/6 |
| $G_n$/FHS | 1 | 0.548 | 0.314 | 0.197 | 0.748 | 0.999 | 0/6 |

Table 20: In-sample fit Expected Shortfall ASL-values for the entire dataset of the models tested for Front Month Nordic power futures. Red indicate that the null hypothesis has been rejected at the 5% significance level. $G_t$ is GARCH-t and $G_n$ is GARCH-n

| In-sample fit backtesting results for Value-at-Risk for Front Month Nordic power futures | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | VaR-levels | | | | | | | | | | | | Rejects |
| | Long positions | | | | | | Short positions | | | | | | at |
| | 99% | | 95% | | 90% | | 90% | | 95% | | 99% | | 5% |
| Model | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | sign. |
| GARCH-n | 0.18 | 0.00 | 0.67 | 0.80 | 0.03 | 0.33 | 0.01 | 0.00 | 0.02 | 0.19 | 0.62 | 0.10 | 5/12 |
| GARCH-t | 0.35 | 0.48 | 0.24 | 0.11 | 0.00 | 0.18 | 0.02 | 0.12 | 0.04 | 0.77 | 0.48 | 0.32 | 3/12 |
| RM-CF | 0.47 | 0.41 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.06 | 0.00 | 0.19 | 0.60 | 0.01 | 7/12 |
| FHS | 0.24 | 0.02 | 0.71 | 0.17 | 0.01 | 0.76 | 0.00 | 0.49 | 0.20 | 0.44 | 0.57 | 0.18 | 3/12 |
| $G_t/G_n$/FHS/RM-CF | 0.34 | 0.39 | 0.49 | 0.17 | 0.00 | 0.08 | 0.02 | 0.27 | 0.02 | 0.80 | 0.48 | 0.32 | 3/12 |
| $G_t/G_n$/FHS | 0.31 | 0.24 | 0.96 | 0.85 | 0.06 | 0.85 | 0.01 | 0.08 | 0.13 | 0.71 | 0.37 | 0.71 | 1/12 |
| $G_t$/FHS | 0.32 | 0.31 | 0.40 | 0.28 | 0.00 | 0.33 | 0.01 | 0.56 | 0.15 | 0.69 | 0.43 | 0.64 | 2/12 |
| $G_n$/FHS | 0.26 | 0.05 | 0.87 | 0.63 | 0.04 | 0.85 | 0.01 | 0.04 | 0.07 | 0.77 | 0.30 | 0.18 | 3/12 |

Table 21: In-sample fit P-values for the entire dataset of the conditional independence test(CI) and the Kupiec unconditional coverage test(UC). Red indicate that the null hypothesis of conditional independence or unconditional coverage has been rejected at the 5% significance level. $G_t$ is GARCH-t and $G_n$ is GARCH-n

| In-sample fit Expected Shortfall ASL-values for Front Quarter Nordic power futures | | | | | | | |
|---|---|---|---|---|---|---|---|
| | ES-levels | | | | | | Rejects |
| | Long positions | | | Short positions | | | at |
| | 99% | 95% | 90% | 90% | 95% | 99% | 5% |
| GARCH-n | 0 | 0 | 0 | 0.005 | 0.008 | 0.067 | 5/6 |
| GARCH-t | 0.395 | 0.41 | 0.113 | 0.736 | 0.884 | 0.137 | 0/6 |
| RM-CF | 0 | 0 | 0 | 0 | 0 | 0 | 6/6 |
| FHS | 1 | 1 | 0.988 | 0.992 | 1 | 1 | 0/6 |
| $G_t/G_n$/FHS/RM-CF | 0 | 0 | 0 | 0 | 0 | 0 | 6/6 |
| $G_t/G_n$/FHS | 0.999 | 0.549 | 0.104 | 0.37 | 0.879 | 0.974 | 0/6 |
| $G_t$/FHS | 1 | 0.976 | 0.658 | 0.933 | 0.998 | 1 | 0/6 |
| $G_n$/FHS | 1 | 0.632 | 0.128 | 0.31 | 0.902 | 1 | 0/6 |

Table 22: In-sample fit Expected Shortfall ASL-values for the entire dataset of the models tested for Front Quarter Nordic power futures. Red indicate that the null hypothesis has been rejected at the 5% significance level. $G_t$ is GARCH-t and $G_n$ is GARCH-n

| In-sample fit backtesting results for Value-at-Risk for Front Quarter Nordic power futures | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | VaR-levels | | | | | | | | | | | | Rejects |
| | Long positions | | | | | | Short positions | | | | | | at |
| | 99% | | 95% | | 90% | | 90% | | 95% | | 99% | | 5% |
| Model | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | sign. |
| GARCH-n | 0.87 | 0.00 | 0.63 | 0.93 | 0.81 | 0.09 | 0.00 | 0.00 | 0.05 | 0.38 | 0.13 | 0.09 | 4/12 |
| GARCH-t | 0.49 | 0.30 | 0.69 | 0.19 | 0.92 | 0.61 | 0.04 | 0.32 | 0.10 | 0.81 | 0.63 | 0.01 | 2/12 |
| RM-CF | 0.20 | 0.16 | 0.41 | 0.14 | 0.88 | 0.00 | 0.01 | 0.00 | 0.02 | 0.00 | 0.54 | 0.11 | 5/12 |
| FHS | 0.15 | 0.04 | 0.90 | 0.11 | 0.94 | 0.31 | 0.02 | 0.57 | 0.02 | 0.48 | 0.58 | 0.13 | 3/12 |
| $G_t/G_n$/FHS/RM-CF | 0.37 | 0.84 | 0.83 | 0.35 | 0.56 | 0.34 | 0.04 | 0.68 | 0.01 | 0.82 | 0.48 | 0.39 | 2/12 |
| $G_t/G_n$/FHS | 0.54 | 0.17 | 0.88 | 0.44 | 0.35 | 0.77 | 0.02 | 0.06 | 0.09 | 0.75 | 0.46 | 0.50 | 1/12 |
| $G_t$/FHS | 0.49 | 0.30 | 0.72 | 0.22 | 0.26 | 0.53 | 0.01 | 0.73 | 0.01 | 0.64 | 0.48 | 0.39 | 2/12 |
| $G_n$/FHS | 0.17 | 0.03 | 0.97 | 0.59 | 0.36 | 0.52 | 0.01 | 0.05 | 0.04 | 0.81 | 0.47 | 0.38 | 3/12 |

Table 23: In-sample fit P-values for the entire dataset of the conditional independence test(CI) and the Kupiec unconditional coverage test(UC). Red indicate that the null hypothesis of conditional independence or unconditional coverage has been rejected at the 5% significance level. $G_t$ is GARCH-t and $G_n$ is GARCH-n

| In-sample fit Expected Shortfall ASL-values for Front Year Nordic power futures | | | | | | | |
|---|---|---|---|---|---|---|---|
| | ES-levels | | | | | | Rejects |
| | Long positions | | | Short positions | | | at |
| | 99% | 95% | 90% | 90% | 95% | 99% | 5% |
| GARCH-n | 0 | 0 | 0 | 0 | 0 | 0 | 6/6 |
| GARCH-t | 0.018 | 0.409 | 0.546 | 0.542 | 0.118 | 0.471 | 1/6 |
| RM-CF | 0 | 0 | 0 | 0 | 0 | 0 | 6/6 |
| FHS | 1 | 1 | 0.999 | 0.997 | 1 | 1 | 0/6 |
| $G_t/G_n$/FHS/RM-CF | 0 | 0 | 0 | 0 | 0 | 0 | 6/6 |
| $G_t/G_n$/FHS | 1 | 0.562 | 0.507 | 0.584 | 0.335 | 1 | 0/6 |
| $G_t$/FHS | 1 | 0.98 | 0.96 | 0.943 | 0.956 | 1 | 0/6 |
| $G_n$/FHS | 1 | 0.768 | 0.641 | 0.516 | 0.479 | 1 | 0/6 |

Table 24: In-sample fit Expected Shortfall ASL-values for the entire dataset of the models tested for Front Year Nordic power futures. Red indicate that the null hypothesis has been rejected at the 5% significance level. $G_t$ is GARCH-t and $G_n$ is GARCH-n

| In-sample fit backtesting results for Value-at-Risk for Front Year Nordic power futures | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | VaR-levels | | | | | | | | | | | | Rejects |
| | Long positions | | | | | | Short positions | | | | | | at |
| | 99% | | 95% | | 90% | | 90% | | 95% | | 99% | | 5% |
| Model | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | sign. |
| GARCH-n | 0.28 | 0.02 | 0.16 | 0.20 | 0.30 | 0.00 | 0.07 | 0.03 | 0.00 | 0.22 | 0.01 | 0.00 | 6/12 |
| GARCH-t | 0.40 | 0.57 | 0.13 | 0.83 | 0.14 | 0.86 | 0.08 | 0.27 | 0.01 | 0.68 | 0.00 | 0.45 | 2/12 |
| RM-CF | 0.31 | 0.18 | 0.04 | 0.41 | 0.07 | 0.01 | 0.02 | 0 | 0.00 | 0.01 | 0.02 | 0.63 | 7/12 |
| FHS | 0.74 | 0.11 | 0.04 | 0.22 | 0.02 | 0.30 | 0.08 | 0.48 | 0.01 | 0.34 | 0.02 | 0.01 | 5/12 |
| $G_t/G_n$/FHS/RM-CF | 0.38 | 0.47 | 0.03 | 0.94 | 0.11 | 0.90 | 0.05 | 0.39 | 0.02 | 0.58 | 0.01 | 0.11 | 3/12 |
| $G_t/G_n$/FHS | 0.67 | 0.23 | 0.12 | 0.69 | 0.11 | 0.38 | 0.06 | 0.95 | 0.01 | 0.94 | 0.00 | 0.04 | 3/12 |
| $G_t$/FHS | 0.58 | 0.54 | 0.08 | 0.78 | 0.12 | 0.72 | 0.09 | 0.42 | 0.02 | 0.58 | 0.00 | 0.04 | 3/12 |
| $G_n$/FHS | 0.70 | 0.18 | 0.14 | 0.80 | 0.15 | 0.38 | 0.07 | 0.60 | 0.02 | 0.99 | 0.00 | 0.01 | 3/12 |

Table 25: In-sample fit P-values for the entire dataset of the conditional independence test(CI) and the Kupiec unconditional coverage test(UC). Red indicate that the null hypothesis of conditional independence or unconditional coverage has been rejected at the 5% significance level. $G_t$ is GARCH-t and $G_n$ is GARCH-n

## 5.3 In-sample and out-of-sample test

This section is devoted to; study the in-sample performance of the models on the first 50% of each data set, decide which simple-average model constellations that seem applicable in the out-of-sample test, and finally evaluate the out-of-sample performance for the last 50% of the observations for all individual models and simple-average models chosen. The in-sample performance for VaR and ES is presented in section 5.3.1 and 5.3.2, and the out-of-sample performance for VaR and ES is presented in section 5.3.3 and 5.3.4.

### 5.3.1 In-sample VaR Results

The in-sample VaR performance of the models for the first 50% of each data set is presented in this section, in table 27, 28, 29, and summarized in table 26.

The test results from the in-sample backtest show that the QR-GJR-GARCH-n, GJR-GARCH-t and QR-GARCH-n perform best with a total of 4 rejections each. QR-GJR-GARCH-t has a total of 5 rejections, while GARCH-t has a total of 6 rejections. QR-GARCH-t, QR-RM and FHS have 7 rejections each. GARCH-n has 12 rejections, while GJR-GARCH-n and RM-CF have 15 rejections each.

The tendency of the in-sample results is similar to the results of the in-sample fit assessment presented in section 5.2; the quantile regression approach performs well. Generally, the GARCH-t approaches have fewer rejections than the GARCH-n approaches, because the GARCH-t approaches account for the fat tails of the distributions, previously described under section 3.5, in table 11. FHS has quite few rejections as well, with a total of 7. GARCH-n, GJR-GARCH-n and RM-CF perform very poor, especially in the unconditional coverage test, with a total of 7, 7 and 9 rejections respectively. As in the in-sample fit test, this means that GARCH-n, GJR-GARCH-n and RM-CF are not capable of predicting a VaR-level that yield the correct amount of rejections at the 5% significance level.

The in-sample backtesting results of the individual models were the basis for the constellations of the simple average models. These models, along with the shortenings used in this paper, are summarized in table 13. The "All above" average model is based on a simple average of all individual models, and it is useful as a reference point. The "Average1"-model does not include GARCH-n approaches or the RM-CF. The GARCH-n approaches are eliminated because these models generally perform poor because they are unable to account for the fat tails of the Nordic power futures. RM-CF is eliminated because it performs extremely poor.

The "Average2"-model includes GARCH-t, QR-GARCH-t and FHS. GARCH-t and GJR-GARCH-t tend to have very similar p-values because the leverage effect of GJR-GARCH-t is close to negligible. QR-GARCH-t and QR-GJR-GARCH-t have the same tendency. Using two similar models in a simple average approach is redundant, and we choose to eliminate both GJR-GARCH-t and QR-GJR-GARCH-t. We rather want to include FHS than the QR-RM approach, since it performs well, and since the model is very different from the GARCH-t and QR-GARCH-t approach. To study the simple average of two well performing models, we include QR-GARCH-t and GARCH-t in the "Average3"-model, and QR-GARCH-t and FHS in the "Average4"-model.

The simple average models perform very well for the in-sample test, with 5, 5, 5, 6 and 6

rejections for Average1, Average2, Average4, Average3 and "All above" respectively. The general tendency is thus that the models do not outperform the best individual models in terms of number of rejections at the 5% significance level, but outperform most of the models. As with the in-sample fit of the entire dataset, neither of the models have any UC-rejections. This indicates that the simple average models do produce the promised fraction of violations. The rejections of CI indicate that the simple average models tend to underestimate risk when volatility is increasing fast, resulting in too low estimates of Value-at-Risk in these periods, and thus clustering of exceedances. Figure 9 in section 5.2 illustrates concecutive exceedances when volatility is clustering.

| Total in-sample test rejections for VaR | | | | |
|---|---|---|---|---|
| Rank | Model | CI | UC | Sum |
| 1 | QR-GJR-GARCH-n | 4 | 0 | 4 |
| 1 | GJR-GARCH-t | 4 | 0 | 4 |
| 1 | QR-GARCH-n | 4 | 0 | 4 |
| 4 | Average1 | 5 | 0 | 5 |
| 4 | Average2 | 5 | 0 | 5 |
| 4 | Average4 | 5 | 0 | 5 |
| 4 | QR-GJR-GARCH-t | 5 | 0 | 5 |
| 8 | All above | 6 | 0 | 6 |
| 8 | Average3 | 6 | 0 | 6 |
| 8 | GARCH-t | 6 | 0 | 6 |
| 11 | QR-GARCH-t | 6 | 1 | 7 |
| 11 | QR-RM | 7 | 0 | 7 |
| 11 | FHS | 7 | 0 | 7 |
| 14 | GARCH-n | 5 | 7 | 12 |
| 15 | GJR-GARCH-n | 8 | 7 | 15 |
| 15 | RM-CF | 6 | 9 | 15 |

Table 26: Total number of in-sample rejections of the first 50% of the dataset for the conditional independence test(CI) and the kupiec unconditional coverage test(UC) for VaR. The average-models are the same as listed in table 13

| In-sample backtesting results for Value-at-Risk for Front Month Nordic power futures | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | VaR-levels | | | | | | | | | | | | Rejects |
| | Long positions | | | | | | Short positions | | | | | | at |
| | 99% | | 95% | | 90% | | 90% | | 95% | | 99% | | 5% |
| Model | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | sign. |
| GARCH-n | 0.36 | 0.03 | 0.63 | 0.90 | 0.07 | 0.55 | 0.00 | 0.03 | 0.02 | 0.66 | 0.44 | 0.21 | 4/12 |
| GARCH-t | 0.60 | 0.69 | 0.79 | 0.75 | 0.03 | 0.38 | 0.00 | 0.50 | 0.02 | 0.27 | 0.65 | 0.35 | 3/12 |
| GJR-GARCH-n | 0.38 | 0.06 | 0.63 | 0.90 | 0.08 | 0.61 | 0.00 | 0.03 | 0.24 | 0.75 | 0.46 | 0.30 | 2/12 |
| GJR-GARCH-t | 0.60 | 0.69 | 0.75 | 0.83 | 0.03 | 0.38 | 0.00 | 0.50 | 0.02 | 0.32 | 0.58 | 0.88 | 3/12 |
| QR-GARCH-n | 0.53 | 0.74 | 0.75 | 0.83 | 0.06 | 0.95 | 0.00 | 0.76 | 0.04 | 0.83 | 0.55 | 0.93 | 2/12 |
| QR-GARCH-t | 0.53 | 0.74 | 0.75 | 0.83 | 0.04 | 0.95 | 0.00 | 0.76 | 0.00 | 0.83 | 0.55 | 0.93 | 3/12 |
| QR-GJR-GARCH-n | 0.53 | 0.74 | 0.79 | 0.75 | 0.02 | 0.88 | 0.00 | 0.82 | 0.24 | 0.75 | 0.55 | 0.93 | 2/12 |
| QR-GJR-GARCH-t | 0.53 | 0.74 | 0.79 | 0.75 | 0.02 | 0.95 | 0.00 | 0.82 | 0.11 | 0.75 | 0.55 | 0.93 | 2/12 |
| QR-RM | 0.21 | 0.57 | 0.75 | 0.83 | 0.27 | 0.95 | 0.00 | 0.76 | 0.00 | 0.83 | 0.55 | 0.93 | 2/12 |
| RM-CF | 0.68 | 0.23 | 0.62 | 0.01 | 0.00 | 0.00 | 0.00 | 0.23 | 0.00 | 0.27 | 0.73 | 0.08 | 5/12 |
| FHS | 0.44 | 0.21 | 0.98 | 0.90 | 0.03 | 0.58 | 0.00 | 0.64 | 0.01 | 0.16 | 0.49 | 0.42 | 3/12 |
| All above | 0.58 | 0.88 | 0.83 | 0.66 | 0.02 | 0.70 | 0.00 | 0.73 | 0.01 | 0.44 | 0.58 | 0.88 | 3/12 |
| Average1 | 0.58 | 0.88 | 0.75 | 0.83 | 0.02 | 0.64 | 0.00 | 0.95 | 0.03 | 0.51 | 0.60 | 0.69 | 3/12 |
| Average2 | 0.55 | 0.93 | 0.75 | 0.83 | 0.05 | 0.53 | 0.00 | 0.95 | 0.03 | 0.38 | 0.60 | 0.69 | 2/12 |
| Average3 | 0.58 | 0.88 | 0.71 | 0.92 | 0.04 | 0.58 | 0.00 | 0.92 | 0.01 | 0.51 | 0.60 | 0.69 | 3/12 |
| Average4 | 0.53 | 0.74 | 0.59 | 0.81 | 0.07 | 0.64 | 0.00 | 0.88 | 0.03 | 0.44 | 0.53 | 0.74 | 2/12 |

Table 27: In-sample P-values of the first 50% of the dataset for the conditional independence test(CI) and the Kupiec unconditional coverage test(UC). Red indicate that the null hypothesis of conditional independence or unconditional coverage has been rejected at the 5% significance level. The average-models are listed in table 13. The period of data is stated in table 10

| In-sample backtesting results for Value-at-Risk for Front Quarter Nordic power futures | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | VaR-levels | | | | | | | | | | | | Rejects |
| | Long positions | | | | | | Short positions | | | | | | at |
| | 99% | | 95% | | 90% | | 90% | | 95% | | 99% | | 5% |
| Model | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | sign. |
| GARCH-n | 0.49 | 0.01 | 0.84 | 0.59 | 0.60 | 0.52 | 0.00 | 0.04 | 0.02 | 0.88 | 0.23 | 0.41 | 4/12 |
| GARCH-t | 0.26 | 0.29 | 0.64 | 0.27 | 0.59 | 0.59 | 0.03 | 0.70 | 0.04 | 0.51 | 0.75 | 0.07 | 2/12 |
| GJR-GARCH-n | 0.12 | 0.00 | 0.80 | 0.51 | 0.46 | 0.58 | 0.01 | 0.08 | 0.09 | 0.70 | 0.57 | 0.93 | 2/12 |
| GJR-GARCH-t | 0.26 | 0.29 | 0.57 | 0.18 | 0.28 | 0.23 | 0.06 | 0.83 | 0.03 | 0.67 | 0.75 | 0.07 | 1/12 |
| QR-GARCH-n | 0.15 | 0.93 | 0.72 | 0.38 | 0.45 | 0.34 | 0.03 | 0.65 | 0.07 | 0.32 | 0.49 | 0.41 | 1/12 |
| QR-GARCH-t | 0.01 | 0.93 | 0.72 | 0.38 | 0.65 | 0.30 | 0.04 | 0.53 | 0.02 | 0.44 | 0.20 | 0.56 | 3/12 |
| QR-GJR-GARCH-n | 0.15 | 0.93 | 0.68 | 0.32 | 0.13 | 0.34 | 0.03 | 0.65 | 0.13 | 0.44 | 0.49 | 0.41 | 1/12 |
| QR-GJR-GARCH-t | 0.00 | 0.93 | 0.72 | 0.38 | 0.21 | 0.34 | 0.03 | 0.65 | 0.02 | 0.51 | 0.52 | 0.56 | 3/12 |
| QR-RM | 0.00 | 0.93 | 0.77 | 0.59 | 0.28 | 0.38 | 0.03 | 0.65 | 0.01 | 0.59 | 0.20 | 0.56 | 2/12 |
| RM-CF | 0.77 | 0.03 | 0.80 | 0.51 | 0.44 | 0.00 | 0.07 | 0.00 | 0.01 | 0.00 | 0.69 | 0.21 | 5/12 |
| FHS | 0.03 | 0.29 | 0.56 | 0.08 | 0.12 | 0.23 | 0.02 | 0.90 | 0.03 | 0.84 | 0.32 | 0.13 | 3/12 |
| All above | 0.11 | 0.67 | 0.76 | 0.44 | 0.59 | 0.59 | 0.02 | 0.83 | 0.04 | 0.51 | 0.72 | 0.12 | 2/12 |
| Average1 | 0.11 | 0.67 | 0.64 | 0.27 | 0.53 | 0.43 | 0.07 | 0.90 | 0.03 | 0.67 | 0.67 | 0.33 | 1/12 |
| Average2 | 0.11 | 0.67 | 0.64 | 0.27 | 0.79 | 0.59 | 0.03 | 0.97 | 0.03 | 0.75 | 0.67 | 0.33 | 2/12 |
| Average3 | 0.11 | 0.67 | 0.68 | 0.32 | 0.67 | 0.48 | 0.02 | 0.83 | 0.03 | 0.67 | 0.69 | 0.21 | 2/12 |
| Average4 | 0.13 | 0.87 | 0.98 | 0.27 | 0.72 | 0.43 | 0.03 | 0.90 | 0.02 | 0.93 | 0.15 | 0.93 | 2/12 |

Table 28: In-sample P-values of the first 50% of the dataset for the conditional independence test(CI) and the Kupiec unconditional coverage test(UC). Red indicate that the null hypothesis of conditional independence or unconditional coverage has been rejected at the 5% significance level. The average-models are listed in table 13. The period of data is stated in table 10

| In-sample backtesting results for Value-at-Risk for Front Year Nordic power futures | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | VaR-levels | | | | | | | | | | | | Rejects |
| | Long positions | | | | | | Short positions | | | | | | at |
| | 99% | | 95% | | 90% | | 90% | | 95% | | 99% | | 5% |
| Model | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | sign. |
| GARCH-n | 0.70 | 0.00 | 0.37 | 0.68 | 0.08 | 0.01 | 0.70 | 0.41 | 0.16 | 0.86 | 0.02 | 0.00 | 4/12 |
| GARCH-t | 0.29 | 0.48 | 0.31 | 0.34 | 0.15 | 0.76 | 0.85 | 0.08 | 0.16 | 0.10 | 0.00 | 0.06 | 1/12 |
| GJR-GARCH-n | 0.14 | 0.01 | 0.09 | 0.68 | 0.03 | 0.02 | 0.85 | 0.60 | 0.39 | 0.64 | 0.01 | 0.00 | 5/12 |
| GJR-GARCH-t | 0.38 | 0.19 | 0.29 | 0.39 | 0.15 | 0.76 | 0.54 | 0.16 | 0.30 | 0.10 | 0.07 | 0.09 | 0/12 |
| QR-GARCH-n | 0.27 | 0.62 | 0.44 | 0.50 | 0.07 | 0.63 | 0.51 | 0.41 | 0.29 | 0.39 | 0.03 | 0.48 | 1/12 |
| QR-GARCH-t | 0.27 | 0.62 | 0.44 | 0.50 | 0.07 | 0.68 | 0.51 | 0.41 | 0.29 | 0.39 | 0.03 | 0.62 | 1/12 |
| QR-GJR-GARCH-n | 0.02 | 0.77 | 0.11 | 0.57 | 0.16 | 0.49 | 0.89 | 0.54 | 0.29 | 0.39 | 0.46 | 0.62 | 1/12 |
| QR-GJR-GARCH-t | 0.27 | 0.62 | 0.11 | 0.57 | 0.12 | 0.63 | 0.89 | 0.54 | 0.31 | 0.34 | 0.44 | 0.48 | 0/12 |
| QR-RM | 0.03 | 0.62 | 0.21 | 0.64 | 0.06 | 0.74 | 0.26 | 0.63 | 0.29 | 0.39 | 0.02 | 0.77 | 2/12 |
| RM-CF | 0.07 | 0.04 | 0.27 | 0.99 | 0.04 | 0.21 | 0.25 | 0.00 | 0.27 | 0.01 | 0.02 | 0.89 | 5/12 |
| FHS | 0.44 | 0.09 | 0.13 | 0.50 | 0.00 | 0.37 | 0.93 | 0.79 | 0.29 | 0.93 | 0.08 | 0.06 | 1/12 |
| All above | 0.27 | 0.62 | 0.19 | 0.71 | 0.10 | 0.81 | 0.65 | 0.63 | 0.71 | 0.50 | 0.01 | 0.19 | 1/12 |
| Average1 | 0.24 | 0.77 | 0.27 | 0.44 | 0.08 | 0.85 | 0.68 | 0.58 | 0.68 | 0.57 | 0.01 | 0.19 | 1/12 |
| Average2 | 0.29 | 0.48 | 0.31 | 0.34 | 0.08 | 0.58 | 0.72 | 0.54 | 0.39 | 0.64 | 0.00 | 0.19 | 1/12 |
| Average3 | 0.27 | 0.62 | 0.31 | 0.34 | 0.11 | 0.90 | 0.83 | 0.41 | 0.39 | 0.21 | 0.00 | 0.09 | 1/12 |
| Average4 | 0.29 | 0.48 | 0.19 | 0.71 | 0.05 | 0.41 | 0.85 | 0.85 | 0.55 | 0.86 | 0.04 | 0.26 | 1/12 |

Table 29: In-sample P-values of the first 50% of the dataset of the conditional independence test(CI) and the Kupiec unconditional coverage test(UC). Red indicate that the null hypothesis of conditional independence or unconditional coverage has been rejected at the 5% significance level. The average-models are listed in table 13. The period of data is stated in table 10

### 5.3.2 In-sample ES Results

The in-sample ES performance of the models for the first 50% of each data set is presented in this section, in table 31, 33 and 35. The VaR-backtesting results for the same models are presented in table 32, 34, 36, and summarized in table 30.

Of the individual models, GARCH-t and FHS perform very well, both with 0 ES rejections for the in-sample data. GARCH-n and RM-CF perform poor, with 13 and 18 rejections respectively. The in-sample descriptive statistics in table 11, under section 3.5, describes the fat tails of the in-sample data. From the backtesting of the ES-models, it is obvious that GARCH-n and RM-CF fail to account for the fat tails of the Nordic power futures.

To investigate the performance of simple average models, we construct an average of all the individual models, called $G_t/G_n$/FHS/RM-CF. We also construct a simple average using only GARCH-t, FHS and GARCH-n($G_t/G_n/FHS$), because of the extremely poor performance of RM-CF. FHS has the best ASL-values overall, and we therefore investigate the performance of FHS and GARCH-t ($G_t/FHS$), and FHS and GARCH-n ($G_n/FHS$).

Of the simple average models, $G_t/FHS$, $G_t/G_n/FHS$ and $G_n/FHS$ perform very well with 0 ES-rejections. $G_t/G_n$/FHS/RM-CF performs very poor with 18 rejections, probably because of the impact of RM-CF. In conclusion, especially the $G_t/FHS$, $G_t/G_n/FHS$ and $G_n/FHS$ seem to provide accurate ES-predictions. The out-of-sample performance of the models, presented in section 5.3.4, will determine whether the models perform well in a more realistic situation as well.

The VaR-backtesting of the models shows that GARCH-n and RM-CF are unable to produce the promised fraction of violations specified by the VaR-level to a large extent, with 15 and 19 rejections respectively. This further emphasize that these models are unfit in this matter. On the contrary, $G_t/G_n/FHS$ and $G_t/FHS$ perform very well with 5 and 7 rejections respectively. $G_n/FHS$, GARCH-t and FHS do not perform as well with 9, 9 and 11 rejections respectively.

| In-sample test rejections for ES | | | |
|---|---|---|---|
| Rank | Model | ES | VaR |
| 1 | $G_t/FHS$ | 0 | 6 |
| 1 | GARCH-t | 0 | 6 |
| 1 | $G_t/G_n/FHS$ | 0 | 6 |
| 1 | $G_n/FHS$ | 0 | 8 |
| 1 | FHS | 0 | 7 |
| 6 | GARCH-n | 16 | 12 |
| 7 | $G_t/G_n/FHS$/RM-CF | 18 | 7 |
| 7 | RM-CF | 18 | 15 |

Table 30: Total number of ES in-sample rejections for the first 50% of the dataset, and total VaR backtesting test rejections. The rank is based on ES-rejections. $G_t$ is GARCH-t and $G_n$ is GARCH-n

| In-sample Expected Shortfall ASL-values for Front Month Nordic power futures | | | | | | | |
|---|---|---|---|---|---|---|---|
| | ES-levels | | | | | | Rejects |
| | Long positions | | | Short positions | | | at |
| | 99% | 95% | 90% | 90% | 95% | 99% | 5% |
| GARCH-n | 0.022 | 0.002 | 0.035 | 0.001 | 0.016 | 0.004 | 6/6 |
| GARCH-t | 0.493 | 0.598 | 0.694 | 0.296 | 0.822 | 0.378 | 0/6 |
| RM-CF | 0 | 0 | 0 | 0 | 0 | 0 | 6/6 |
| FHS | 1 | 0.998 | 0.992 | 0.937 | 1 | 1 | 0/6 |
| $G_t/G_n$/FHS/RM-CF | 0 | 0 | 0 | 0 | 0 | 0 | 6/6 |
| $G_t/G_n$/FHS | 0.992 | 0.548 | 0.654 | 0.364 | 0.856 | 0.985 | 0/6 |
| $G_t$/FHS | 1 | 0.951 | 0.919 | 0.701 | 0.995 | 0.999 | 0/6 |
| $G_n$/FHS | 1 | 0.504 | 0.629 | 0.338 | 0.929 | 0.997 | 0/6 |

Table 31: In-sample Expected Shortfall ASL-values for the first 50% of the dataset for the models tested for Front Month Nordic power futures. Red indicate that the null hypothesis has been rejected at the 5% significance level. $G_t$ is GARCH-t and $G_n$ is GARCH-n

| In-sample backtesting results for Value-at-Risk for Front Month Nordic power futures | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | VaR-levels | | | | | | | | | | | | Rejects |
| | Long positions | | | | | | Short positions | | | | | | at |
| | 99% | | 95% | | 90% | | 90% | | 95% | | 99% | | 5% |
| Model | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | sign. |
| GARCH-n | 0.36 | 0.03 | 0.63 | 0.90 | 0.07 | 0.55 | 0.00 | 0.03 | 0.02 | 0.66 | 0.44 | 0.21 | 4/12 |
| GARCH-t | 0.60 | 0.69 | 0.79 | 0.75 | 0.03 | 0.38 | 0.00 | 0.50 | 0.02 | 0.27 | 0.65 | 0.35 | 3/12 |
| RM-CF | 0.68 | 0.23 | 0.62 | 0.01 | 0.00 | 0.00 | 0.00 | 0.23 | 0.00 | 0.27 | 0.73 | 0.08 | 5/12 |
| FHS | 0.44 | 0.21 | 0.98 | 0.90 | 0.03 | 0.58 | 0.00 | 0.64 | 0.01 | 0.16 | 0.49 | 0.42 | 3/12 |
| $G_t/G_n$/FHS/RM-CF | 0.51 | 0.57 | 0.75 | 0.83 | 0.01 | 0.30 | 0.00 | 0.79 | 0.02 | 0.23 | 0.58 | 0.88 | 3/12 |
| $G_t/G_n$/FHS | 0.51 | 0.57 | 0.63 | 0.90 | 0.12 | 0.88 | 0.00 | 0.55 | 0.03 | 0.38 | 0.49 | 0.42 | 2/12 |
| $G_t$/FHS | 0.53 | 0.74 | 0.71 | 0.92 | 0.05 | 0.53 | 0.00 | 0.92 | 0.02 | 0.23 | 0.53 | 0.74 | 2/12 |
| $G_n$/FHS | 0.49 | 0.42 | 0.90 | 0.64 | 0.12 | 0.86 | 0.00 | 0.50 | 0.02 | 0.27 | 0.49 | 0.42 | 2/12 |

Table 32: In-sample P-values for the first 50% of the dataset for the conditional independence test(CI) and the Kupiec unconditional coverage test(UC). Red indicate that the null hypothesis of conditional independence or unconditional coverage has been rejected at the 5% significance level. $G_t$ is GARCH-t and $G_n$ is GARCH-n

| In-sample Expected Shortfall ASL-values for Front Quarter Nordic power futures | | | | | | | |
|---|---|---|---|---|---|---|---|
| | ES-levels | | | | | | Rejects |
| | Long positions | | | Short positions | | | at |
| | 99% | 95% | 90% | 90% | 95% | 99% | 5% |
| GARCH-n | 0.007 | 0.002 | 0.001 | 0.017 | 0.025 | 0.106 | 5/6 |
| GARCH-t | 0.477 | 0.451 | 0.185 | 0.794 | 0.938 | 0.175 | 0/6 |
| RM-CF | 0 | 0 | 0 | 0 | 0 | 0 | 6/6 |
| FHS | 1 | 1 | 0.979 | 0.978 | 1 | 1 | 0/6 |
| $G_t/G_n$/FHS/RM-CF | 0 | 0 | 0 | 0 | 0 | 0.001 | 6/6 |
| $G_t/G_n$/FHS | 0.976 | 0.723 | 0.208 | 0.706 | 0.914 | 0.973 | 0/6 |
| $G_t$/FHS | 1 | 0.974 | 0.634 | 0.962 | 0.994 | 0.999 | 0/6 |
| $G_n$/FHS | 1 | 0.817 | 0.171 | 0.433 | 0.878 | 0.999 | 0/6 |

Table 33: In-sample Expected Shortfall ASL-values for the models tested for Front Quarter Nordic power futures. Red indicate that the null hypothesis has been rejected at the 5% significance level. $G_t$ is GARCH-t and $G_n$ is GARCH-n

| In-sample backtesting results for Value-at-Risk for Front Quarter Nordic power futures | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | VaR-levels | | | | | | | | | | | | Rejects |
| | Long positions | | | | | | Short positions | | | | | | at |
| | 99% | | 95% | | 90% | | 90% | | 95% | | 99% | | 5% |
| Model | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | sign. |
| GARCH-n | 0.49 | 0.01 | 0.84 | 0.59 | 0.60 | 0.52 | 0.00 | 0.04 | 0.02 | 0.88 | 0.23 | 0.41 | 4/12 |
| GARCH-t | 0.26 | 0.29 | 0.64 | 0.27 | 0.59 | 0.59 | 0.03 | 0.70 | 0.04 | 0.51 | 0.75 | 0.07 | 2/12 |
| RM-CF | 0.77 | 0.03 | 0.80 | 0.51 | 0.44 | 0.00 | 0.07 | 0.00 | 0.01 | 0.00 | 0.69 | 0.21 | 5/12 |
| FHS | 0.03 | 0.29 | 0.56 | 0.08 | 0.12 | 0.23 | 0.02 | 0.90 | 0.03 | 0.84 | 0.32 | 0.13 | 3/12 |
| $G_t/G_n$/FHS/RM-CF | 0.11 | 0.67 | 0.76 | 0.44 | 0.99 | 0.59 | 0.02 | 0.77 | 0.02 | 0.44 | 0.67 | 0.33 | 2/12 |
| $G_t/G_n$/FHS | 0.25 | 0.29 | 0.98 | 0.27 | 0.77 | 0.90 | 0.01 | 0.42 | 0.02 | 0.84 | 0.64 | 0.49 | 2/12 |
| $G_t$/FHS | 0.20 | 0.56 | 0.89 | 0.15 | 0.63 | 0.53 | 0.04 | 0.83 | 0.01 | 0.67 | 0.64 | 0.49 | 2/12 |
| $G_n$/FHS | 0.04 | 0.08 | 0.90 | 0.38 | 0.68 | 0.64 | 0.00 | 0.18 | 0.02 | 0.98 | 0.18 | 0.74 | 3/12 |

Table 34: In-sample P-values for the first 50% of the dataset for the conditional independence test(CI) and the Kupiec unconditional coverage test(UC). Red indicate that the null hypothesis of conditional independence or unconditional coverage has been rejected at the 5% significance level. $G_t$ is GARCH-t and $G_n$ is GARCH-n

| In-sample Expected Shortfall ASL-values for Front Year Nordic power futures | | | | | | | |
|---|---|---|---|---|---|---|---|
| | ES-levels | | | | | | Rejects |
| | Long positions | | | Short positions | | | at |
| | 99% | 95% | 90% | 90% | 95% | 99% | 5% |
| GARCH-n | 0.001 | 0.028 | 0.071 | 0.004 | 0 | 0.004 | 5/6 |
| GARCH-t | 0.469 | 0.854 | 0.928 | 0.824 | 0.501 | 0.243 | 0/6 |
| RM-CF | 0 | 0 | 0 | 0 | 0 | 0 | 6/6 |
| FHS | 1 | 1 | 0.993 | 0.977 | 0.998 | 1 | 0/6 |
| $G_t/G_n$/FHS/RM-CF | 0 | 0 | 0 | 0 | 0 | 0 | 6/6 |
| $G_t/G_n$/FHS | 1 | 0.827 | 0.754 | 0.728 | 0.575 | 0.989 | 0/6 |
| $G_t$/FHS | 1 | 0.992 | 0.977 | 0.936 | 0.931 | 1 | 0/6 |
| $G_n$/FHS | 1 | 0.854 | 0.711 | 0.55 | 0.679 | 1 | 0/6 |

Table 35: In-sample Expected Shortfall ASL-values of the first 50% of the dataset for the models tested for Front Year Nordic power futures. Red indicate that the null hypothesis has been rejected at the 5% significance level. $G_t$ is GARCH-t and $G_n$ is GARCH-n

| In-sample backtesting results for Value-at-Risk for Front Year Nordic power futures | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | VaR-levels | | | | | | | | | | | | Rejects |
| | Long positions | | | | | | Short positions | | | | | | at |
| | 99% | | 95% | | 90% | | 90% | | 95% | | 99% | | 5% |
| Model | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | sign. |
| GARCH-n | 0.70 | 0.00 | 0.37 | 0.68 | 0.08 | 0.01 | 0.70 | 0.41 | 0.16 | 0.86 | 0.02 | 0.00 | 4/12 |
| GARCH-t | 0.29 | 0.48 | 0.31 | 0.34 | 0.15 | 0.76 | 0.85 | 0.08 | 0.16 | 0.10 | 0.00 | 0.06 | 1/12 |
| RM-CF | 0.07 | 0.04 | 0.27 | 0.99 | 0.04 | 0.21 | 0.25 | 0.00 | 0.27 | 0.01 | 0.02 | 0.89 | 5/12 |
| FHS | 0.44 | 0.09 | 0.13 | 0.50 | 0.00 | 0.37 | 0.93 | 0.79 | 0.29 | 0.93 | 0.08 | 0.06 | 1/12 |
| $G_t/G_n$/FHS/RM-CF | 0.22 | 0.94 | 0.16 | 0.86 | 0.02 | 0.87 | 0.92 | 0.19 | 0.36 | 0.25 | 0.00 | 0.09 | 2/12 |
| $G_t/G_n$/FHS | 0.47 | 0.06 | 0.21 | 0.63 | 0.10 | 0.55 | 0.96 | 0.96 | 0.23 | 0.57 | 0.00 | 0.01 | 2/12 |
| $G_t$/FHS | 0.38 | 0.19 | 0.31 | 0.34 | 0.07 | 0.68 | 0.99 | 0.41 | 0.27 | 0.44 | 0.00 | 0.02 | 2/12 |
| $G_n$/FHS | 0.47 | 0.06 | 0.27 | 0.99 | 0.04 | 0.46 | 0.84 | 0.87 | 0.18 | 0.78 | 0.00 | 0.00 | 3/12 |

Table 36: In-sample P-values for the first 50% of the dataset for the conditional independence test(CI) and the Kupiec unconditional coverage test(UC). Red indicate that the null hypothesis of conditional independence or unconditional coverage has been rejected at the 5% significance level. A$G_t$ is GARCH-t and $G_n$ is GARCH-n

### 5.3.3 Out-of-sample VaR Results

The out-of-sample VaR performance of the models is presented in this section, in table 38, 39, 40, and summarized in table 37.

The out-of-sample results are produced on the last 50% of the data. Of the individual models, the QR-GARCH-t, QR-GARCH-n, QR-GJR-GARCH-n and QR-GJR-GARCH-t perform best with a total of 6 rejections each. GARCH-t and QR-RM have a total of 8 rejections each, while FHS and GJR-GARCH-t have a total of 10 rejections each. GARCH-n and GJR-GARCH-n have 12 rejections each, while RM-CF performs very poor with a total of 16 rejections.

The individual models that utilize the quantile regression approach perform very well out-of-sample. The QR-RM performs better than the RM-CF approach. GARCH-n and GJR-GARCH-n perform inferiorly compared to GARCH-t and GJR-GARCH-t, especially for the unconditional coverage test. This is due to the fat tails of the out-of-sample Nordic power futures contracts, previously explained in table 12 under section 3.5. The 8 unconditional coverage test rejections of the RM-CF approach, reveal that the Cornish Fisher approximation, to a large degree, fails to account for the fat tails of the contracts correctly.

Apart from RM-CF, GJR-GARCH-n and GARCH-n, every individual and simple average model manage to perform well for the unconditional coverage test. This indicates that these models manage to predict the promised fraction of violations specified by the VaR-level adequately. The results of the conditional independence test for the same models are worse, indicating clustering of violations. An example illustrating this phenomenon is provided in figure 9 under section 5.2.1.

The simple average models perform well. "All above", Average1 and Average3 all have 6 rejections. Average2 and Average4 have 8 rejections each. The general tendency is similar to the in-sample test; the simple average models do not outperform the best individual models in terms of number of rejections at the 5% significance level, but perform very well in general. Only Average2 has a UC-rejection, indicating that the simple average models do produce the promised fraction of violations. The simple average models do however tend to underestimate risk when volatility is increasing fast, resulting in too low estimates of Value-at-Risk in these periods, and thus clustering of exceedances. Clustering of exceedances is illustrated in figure 9 under section 5.2.

| Total out-of-sample test rejections for VaR | | | | |
|---|---|---|---|---|
| Rank | Model | CI | UC | Sum |
| 1 | All above | 6 | 0 | 6 |
| 1 | Average1 | 6 | 0 | 6 |
| 1 | Average3 | 6 | 0 | 6 |
| 1 | QR-GARCH-t | 6 | 0 | 6 |
| 1 | QR-GARCH-n | 6 | 0 | 6 |
| 1 | QR-GJR-GARCH-n | 6 | 0 | 6 |
| 1 | QR-GJR-GARCH-t | 6 | 0 | 6 |
| 8 | Average2 | 7 | 1 | 8 |
| 8 | Average4 | 8 | 0 | 8 |
| 8 | GARCH-t | 7 | 1 | 8 |
| 8 | QR-RM | 8 | 0 | 8 |
| 12 | FHS | 7 | 3 | 10 |
| 12 | GJR-GARCH-t | 8 | 2 | 10 |
| 14 | GARCH-n | 8 | 4 | 12 |
| 14 | GJR-GARCH-n | 7 | 5 | 12 |
| 16 | RM-CF | 8 | 8 | 16 |

Table 37: Total number of out-of-sample rejections for the conditional independence test(CI) and the kupiec unconditional coverage test(UC) for VaR. The average-models are the same as listed in table 13

| Out-of-sample backtesting results for Value-at-Risk for Front Month Nordic power futures | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | VaR-levels | | | | | | | | | | | | Rejects |
| | Long positions | | | | | | Short positions | | | | | | at |
| | 99% | | 95% | | 90% | | 90% | | 95% | | 99% | | 5% |
| Model | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | sign. |
| GARCH-n | 0.56 | 0.00 | 0.04 | 0.24 | 0.01 | 0.52 | 0.50 | 0.00 | 0.06 | 0.08 | 0.06 | 0.09 | 4/12 |
| GARCH-t | 0.45 | 0.28 | 0.01 | 0.08 | 0.00 | 0.34 | 0.83 | 0.04 | 0.12 | 0.26 | 0.11 | 0.32 | 3/12 |
| GJR-GARCH-n | 0.56 | 0.00 | 0.02 | 0.20 | 0.01 | 0.52 | 0.50 | 0.00 | 0.07 | 0.09 | 0.06 | 0.09 | 4/12 |
| GJR-GARCH-t | 0.45 | 0.28 | 0.00 | 0.04 | 0.00 | 0.27 | 0.94 | 0.02 | 0.12 | 0.26 | 0.12 | 0.46 | 4/12 |
| QR-GARCH-n | 0.55 | 0.98 | 0.02 | 0.20 | 0.01 | 0.86 | 0.39 | 0.34 | 0.18 | 0.48 | 0.18 | 0.98 | 2/12 |
| QR-GARCH-t | 0.55 | 0.98 | 0.02 | 0.20 | 0.01 | 0.86 | 0.45 | 0.43 | 0.16 | 0.42 | 0.18 | 0.98 | 2/12 |
| QR-GJR-GARCH-n | 0.53 | 0.84 | 0.02 | 0.20 | 0.01 | 0.86 | 0.43 | 0.63 | 0.14 | 0.36 | 0.18 | 0.98 | 2/12 |
| QR-GJR-GARCH-t | 0.53 | 0.84 | 0.02 | 0.20 | 0.01 | 0.80 | 0.43 | 0.63 | 0.14 | 0.36 | 0.18 | 0.98 | 2/12 |
| QR-RM | 0.29 | 0.28 | 0.00 | 0.24 | 0.00 | 0.90 | 0.01 | 0.43 | 0.04 | 0.78 | 0.23 | 0.67 | 4/12 |
| RM-CF | 0.48 | 0.52 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.56 | 0.09 | 0.81 | 0.06 | 0.08 | 5/12 |
| FHS | 0.07 | 0.04 | 0.01 | 0.08 | 0.00 | 0.56 | 0.29 | 0.84 | 0.17 | 0.39 | 0.34 | 0.20 | 3/12 |
| All above | 0.45 | 0.28 | 0.02 | 0.17 | 0.00 | 0.72 | 0.88 | 0.14 | 0.11 | 0.21 | 0.18 | 0.98 | 2/12 |
| Average1 | 0.48 | 0.52 | 0.02 | 0.17 | 0.00 | 0.72 | 0.79 | 0.21 | 0.18 | 0.48 | 0.18 | 0.98 | 2/12 |
| Average2 | 0.46 | 0.39 | 0.01 | 0.04 | 0.00 | 0.56 | 0.37 | 0.31 | 0.23 | 0.70 | 0.18 | 0.98 | 3/12 |
| Average3 | 0.48 | 0.52 | 0.03 | 0.14 | 0.00 | 0.67 | 0.56 | 0.18 | 0.13 | 0.30 | 0.16 | 0.79 | 2/12 |
| Average4 | 0.46 | 0.39 | 0.01 | 0.06 | 0.00 | 0.72 | 0.17 | 0.80 | 0.25 | 0.78 | 0.18 | 0.98 | 2/12 |

Table 38: Out-of-sample P-values for the conditional independence test(CI) and the Kupiec unconditional coverage test(UC). Red indicate that the null hypothesis of conditional independence or unconditional coverage has been rejected at the 5% significance level. The average-models are listed in table 13. The period of data is stated in table 10

| Out-of-sample backtesting results for Value-at-Risk for Front Quarter Nordic power futures | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | VaR-levels | | | | | | | | | | | | Rejects |
| | Long positions | | | | | | Short positions | | | | | | at |
| | 99% | | 95% | | 90% | | 90% | | 95% | | 99% | | 5% |
| Model | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | sign. |
| GARCH-n | 0.39 | 0.08 | 0.85 | 0.34 | 0.91 | 0.12 | 0.04 | 0.01 | 0.03 | 0.24 | 0.38 | 0.05 | 3/12 |
| GARCH-t | 0.52 | 0.67 | 0.66 | 0.68 | 0.80 | 0.73 | 0.02 | 0.45 | 0.06 | 0.53 | 0.63 | 0.44 | 1/12 |
| GJR-GARCH-n | 0.36 | 0.03 | 0.48 | 0.34 | 0.98 | 0.22 | 0.03 | 0.02 | 0.06 | 0.16 | 0.47 | 0.38 | 3/12 |
| GJR-GARCH-t | 0.52 | 0.67 | 0.62 | 0.60 | 0.97 | 0.67 | 0.02 | 0.36 | 0.04 | 0.46 | 0.63 | 0.44 | 2/12 |
| QR-GARCH-n | 0.65 | 0.31 | 0.89 | 0.39 | 0.61 | 0.73 | 0.10 | 0.84 | 0.31 | 0.53 | 0.47 | 0.38 | 0/12 |
| QR-GARCH-t | 0.61 | 0.60 | 0.93 | 0.46 | 0.69 | 0.85 | 0.11 | 0.90 | 0.17 | 0.68 | 0.43 | 0.19 | 0/12 |
| QR-GJR-GARCH-n | 0.61 | 0.60 | 0.82 | 0.28 | 0.60 | 0.96 | 0.06 | 0.84 | 0.36 | 0.68 | 0.41 | 0.13 | 0/12 |
| QR-GJR-GARCH-t | 0.63 | 0.44 | 0.93 | 0.46 | 0.60 | 0.96 | 0.11 | 0.90 | 0.07 | 0.68 | 0.41 | 0.13 | 0/12 |
| QR-RM | 0.63 | 0.44 | 0.87 | 0.85 | 0.93 | 0.62 | 0.09 | 0.67 | 0.14 | 0.81 | 0.43 | 0.19 | 0/12 |
| RM-CF | 0.69 | 0.12 | 0.99 | 0.58 | 0.92 | 0.00 | 0.04 | 0.00 | 0.27 | 0.02 | 0.63 | 0.44 | 4/12 |
| FHS | 0.45 | 0.27 | 0.89 | 0.81 | 0.93 | 0.62 | 0.04 | 0.42 | 0.19 | 0.20 | 0.43 | 0.03 | 2/12 |
| All above | 0.54 | 0.84 | 0.97 | 0.53 | 0.89 | 0.85 | 0.06 | 0.60 | 0.07 | 0.68 | 0.43 | 0.19 | 0/12 |
| Average1 | 0.54 | 0.84 | 0.91 | 0.76 | 0.97 | 0.67 | 0.06 | 0.55 | 0.08 | 0.76 | 0.47 | 0.38 | 0/12 |
| Average2 | 0.54 | 0.84 | 0.66 | 0.68 | 0.84 | 0.51 | 0.03 | 0.72 | 0.10 | 0.93 | 0.45 | 0.27 | 1/12 |
| Average3 | 0.56 | 0.97 | 0.97 | 0.53 | 0.77 | 0.67 | 0.06 | 0.60 | 0.06 | 0.60 | 0.56 | 0.97 | 0/12 |
| Average4 | 0.52 | 0.67 | 0.70 | 0.76 | 0.97 | 0.67 | 0.02 | 0.96 | 0.12 | 0.98 | 0.31 | 0.19 | 1/12 |

Table 39: Out-of-sample P-values for the conditional independence test(CI) and the Kupiec unconditional coverage test(UC). Red indicate that the null hypothesis of conditional independence or unconditional coverage has been rejected at the 5% significance level. The average-models are listed in table 13. The period of data is stated in table 10

| Out-of-sample backtesting results for Value-at-Risk for Front Year Nordic power futures | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | VaR-levels | | | | | | | | | | | | Rejects |
| | Long positions | | | | | | Short positions | | | | | | at |
| | 99% | | 95% | | 90% | | 90% | | 95% | | 99% | | 5% |
| Model | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | sign. |
| GARCH-n | 0.33 | 0.44 | 0.01 | 0.23 | 0.07 | 0.16 | 0.00 | 0.11 | 0.00 | 0.59 | 0.00 | 0.01 | 5/12 |
| GARCH-t | 0.54 | 0.51 | 0.00 | 0.30 | 0.26 | 0.39 | 0.00 | 0.39 | 0.00 | 0.44 | 0.00 | 0.56 | 4/12 |
| GJR-GARCH-n | 0.42 | 0.18 | 0.00 | 0.59 | 0.06 | 0.26 | 0.00 | 0.12 | 0.00 | 0.90 | 0.00 | 0.00 | 5/12 |
| GJR-GARCH-t | 0.24 | 0.98 | 0.00 | 0.22 | 0.23 | 0.29 | 0.00 | 0.36 | 0.00 | 0.39 | 0.00 | 0.71 | 4/12 |
| QR-GARCH-n | 0.16 | 0.38 | 0.00 | 0.80 | 0.28 | 0.07 | 0.00 | 0.43 | 0.00 | 0.80 | 0.00 | 0.98 | 4/12 |
| QR-GARCH-t | 0.58 | 0.28 | 0.00 | 0.88 | 0.26 | 0.09 | 0.00 | 0.51 | 0.00 | 0.80 | 0.02 | 0.98 | 4/12 |
| QR-GJR-GARCH-n | 0.16 | 0.38 | 0.00 | 0.83 | 0.14 | 0.07 | 0.00 | 0.36 | 0.00 | 0.66 | 0.02 | 0.82 | 4/12 |
| QR-GJR-GARCH-t | 0.16 | 0.38 | 0.00 | 0.83 | 0.22 | 0.06 | 0.00 | 0.56 | 0.00 | 0.80 | 0.02 | 0.82 | 4/12 |
| QR-RM | 0.12 | 0.19 | 0.02 | 0.88 | 0.19 | 0.09 | 0.00 | 0.47 | 0.01 | 0.80 | 0.02 | 0.86 | 4/12 |
| RM-CF | 0.66 | 0.45 | 0.01 | 0.44 | 0.24 | 0.00 | 0.01 | 0.00 | 0.00 | 0.04 | 0.02 | 0.86 | 7/12 |
| FHS | 0.45 | 0.71 | 0.00 | 0.62 | 0.07 | 0.65 | 0.00 | 0.65 | 0.00 | 0.08 | 0.00 | 0.01 | 5/12 |
| All above | 0.18 | 0.51 | 0.00 | 0.83 | 0.20 | 0.36 | 0.01 | 0.60 | 0.00 | 0.90 | 0.02 | 0.98 | 4/12 |
| Average1 | 0.16 | 0.38 | 0.01 | 0.69 | 0.21 | 0.21 | 0.01 | 0.56 | 0.00 | 0.90 | 0.02 | 0.82 | 4/12 |
| Average2 | 0.56 | 0.38 | 0.01 | 0.69 | 0.05 | 0.26 | 0.00 | 0.65 | 0.00 | 0.50 | 0.00 | 0.71 | 4/12 |
| Average3 | 0.58 | 0.28 | 0.00 | 0.69 | 0.29 | 0.21 | 0.00 | 0.56 | 0.00 | 0.90 | 0.00 | 0.98 | 4/12 |
| Average4 | 0.54 | 0.51 | 0.00 | 0.62 | 0.03 | 0.19 | 0.00 | 0.75 | 0.00 | 0.50 | 0.00 | 0.86 | 5/12 |

Table 40: Out-of-sample P-values for the conditional independence test(CI) and the Kupiec unconditional coverage test(UC). Red indicate that the null hypothesis of conditional independence or unconditional coverage has been rejected at the 5% significance level. The average-models are listed in table 13. The period of data is stated in table 10

### 5.3.4 Out-of-sample ES Results

The out-of-sample ES performance of the models is presented in this section, in table 42, 44 and 46. The VaR-backtesting results for the same models is presented in table 43, 45, 47, and summarized in table 41.

$G_t/FHS$, GARCH-t, $G_t/G_n/FHS$, $G_n/FHS$ and FHS perform very well with 0 ES rejections. GARCH-n, $G_t/G_n/FHS$/RM-CF and RM-CF in contrast, perform very poor. They have a total of 13, 13 and 18 Expected Shortfall rejections respectively. This result is in line with previous results; RM-CF and GARCH-n fail to account for the fat tails of Nordic power futures. The out-of-sample descriptive statistics under section 3.5, in table 12, describe the fat tails of the out-of-sample window.

A closer look at the ASL-values of the models, divulges ASL-values of RM-CF of 0 for all quantiles in all contracts. This indicates that RM-CF consistently underestimates risk, and this model is thus next to useless. This is illustrated in the t-statistics of the bootstrap in figure 10. FHS has ASL-values close to 1 for all quantiles in all contracts, which indicates that FHS consistently overestimates risk since the ES-backtest procedure is a one-sided test, testing underestimation of risk. This is illustrated in figure 12. Although preventing underestimation of risk is of much higher importance than preventing overestimation of risk, an ES-model consistently overestimating risk is not appealing. An optimal risk model aims to predict the correct level of risk, and FHS does not seem trustworthy in this regard, given ASL-values consistently very close to 1. An example of two more trustworthy models in this regard is illustrated in figure 11 and 13. GARCH-t, $G_t/G_n$/FHS, $G_t$/FHS and $G_n$/FHS produce results that seem to neither overestimate- nor underestimate the risk, and are in our opinion preferred to the FHS-model.

RM-CF performs terrible, and seems to be the factor that makes the average model including all the individual models, $G_t/G_n$/FHS/RM-CF, useless. The other simple average models perform very well, and are the preferred models in addition to the GARCH-t model. In conclusion, the simple average approach shows very promising results for ES-averaging. However, GARCH-t performs very well, and it raises the question whether the performance of the simple average models is worth the additional work required to calculate simple averages of several models.

| Out-of-sample test rejections for ES | | | |
|---|---|---|---|
| Rank | Model | ES | VaR |
| 1 | $G_t/FHS$ | 0 | 7 |
| 1 | GARCH-t | 0 | 8 |
| 1 | $G_t/G_n/FHS$ | 0 | 8 |
| 1 | $G_n/FHS$ | 0 | 10 |
| 1 | FHS | 0 | 10 |
| 6 | GARCH-n | 15 | 12 |
| 7 | $G_t/G_n/FHS$/RM-CF | 18 | 8 |
| 7 | RM-CF | 18 | 16 |

Table 41: Total number of out-of-sample rejections for ES, and total VaR backtesting test rejections. The rank is based on ES-rejections. $G_t$ is GARCH-t and $G_n$ is GARCH-n

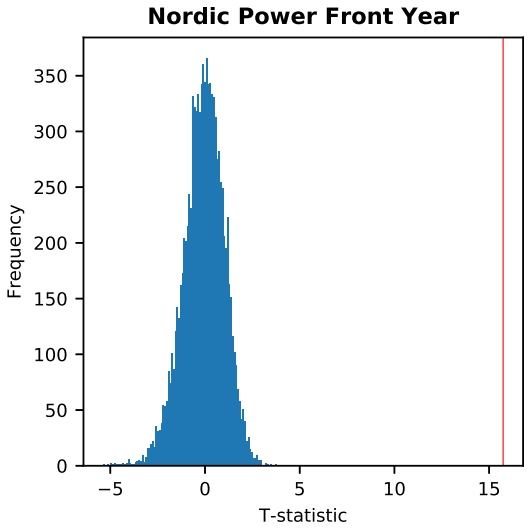



Figure 10: Histogram of bootstrap T-statistic for RM-CF 0.99 long. This bootstrap yields an ASL of 0. The red line show the T-statistic of the original sample of exceedences

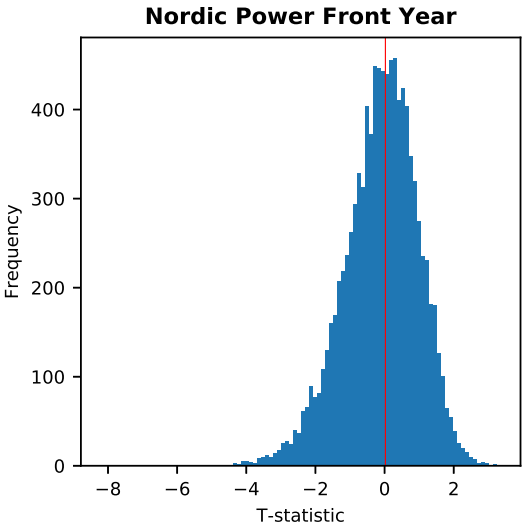



Figure 11: Histogram of bootstrap T-statistic for GARCH-t 0.99 short. This bootstrap yields an ASL of 0.471. The red line show the T-statistic of the original sample of exceedences

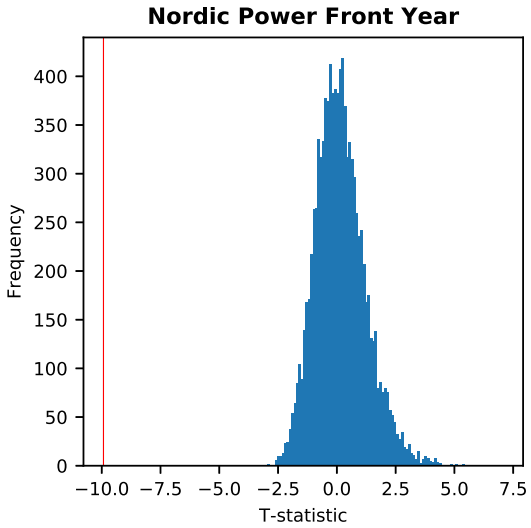



Figure 12: Histogram of bootstrap T-statistic for FHS 0.99 long. This bootstrap yields an ASL of 1. The red line show the T-statistic of the original sample of exceedences

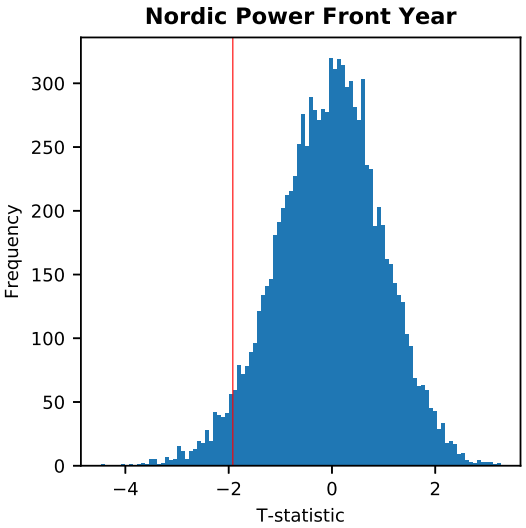



Figure 13: Histogram of bootstrap T-statistic for $G_t$/FHS 0.9 long. This bootstrap yields an ASL of 0.96. The red line show the T-statistic of the original sample of exceedences

| Out-of-sample Expected Shortfall ASL-values for Front Month Nordic power futures | | | | | | | |
|---|---|---|---|---|---|---|---|
| | ES-levels | | | | | | Rejects |
| | Long positions | | | Short positions | | | at |
| | 99% | 95% | 90% | 90% | 95% | 99% | 5% |
| GARCH-n | 0 | 0.001 | 0.002 | 0.004 | 0 | 0.006 | 6/6 |
| GARCH-t | 0.634 | 0.436 | 0.316 | 0.372 | 0.29 | 0.163 | 0/6 |
| RM-CF | 0 | 0 | 0 | 0 | 0 | 0 | 6/6 |
| FHS | 1 | 0.998 | 0.54 | 0.963 | 0.99 | 1 | 0/6 |
| $G_t/G_n$/FHS/RM-CF | 0 | 0 | 0 | 0 | 0 | 0.001 | 6/6 |
| $G_t/G_n$/FHS | 0.996 | 0.413 | 0.217 | 0.316 | 0.46 | 0.734 | 0/6 |
| $G_t$/FHS | 1 | 0.893 | 0.511 | 0.775 | 0.796 | 0.9 | 0/6 |
| $G_n$/FHS | 1 | 0.457 | 0.184 | 0.297 | 0.531 | 0.887 | 0/6 |

Table 42: Out-of-sample Expected Shortfall ASL-values for the models tested for Front Month Nordic power futures. Red indicate that the null hypothesis has been rejected at the 5% significance level. $G_t$ is GARCH-t and $G_n$ is GARCH-n

| Out-of-sample backtesting results for Value-at-Risk for Front Month Nordic power futures | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | VaR-levels | | | | | | | | | | | | Rejects |
| | Long positions | | | | | | Short positions | | | | | | at |
| | 99% | | 95% | | 90% | | 90% | | 95% | | 99% | | 5% |
| Model | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | sign. |
| GARCH-n | 0.56 | 0.00 | 0.04 | 0.24 | 0.01 | 0.52 | 0.50 | 0.00 | 0.06 | 0.08 | 0.06 | 0.09 | 4/12 |
| GARCH-t | 0.45 | 0.28 | 0.01 | 0.08 | 0.00 | 0.34 | 0.83 | 0.04 | 0.12 | 0.26 | 0.11 | 0.32 | 3/12 |
| RM-CF | 0.48 | 0.52 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.56 | 0.09 | 0.81 | 0.06 | 0.08 | 5/12 |
| FHS | 0.07 | 0.04 | 0.01 | 0.08 | 0.00 | 0.56 | 0.29 | 0.84 | 0.17 | 0.39 | 0.34 | 0.20 | 3/12 |
| $G_t/G_n$/FHS/RM-CF | 0.43 | 0.20 | 0.00 | 0.02 | 0.00 | 0.24 | 0.88 | 0.14 | 0.14 | 0.36 | 0.12 | 0.46 | 3/12 |
| $G_t/G_n$/FHS | 0.40 | 0.09 | 0.02 | 0.06 | 0.00 | 0.61 | 0.71 | 0.02 | 0.12 | 0.26 | 0.20 | 0.84 | 3/12 |
| $G_t$/FHS | 0.43 | 0.20 | 0.01 | 0.06 | 0.00 | 0.34 | 0.63 | 0.24 | 0.23 | 0.70 | 0.18 | 0.98 | 2/12 |
| $G_n$/FHS | 0.43 | 0.04 | 0.01 | 0.12 | 0.00 | 0.92 | 0.75 | 0.03 | 0.16 | 0.42 | 0.28 | 0.39 | 4/12 |

Table 43: Out-of-sample P-values for the last 50% of the dataset for the conditional independence test(CI) and the Kupiec unconditional coverage test(UC). Red indicate that the null hypothesis of conditional independence or unconditional coverage has been rejected at the 5% significance level. $G_t$ is GARCH-t and $G_n$ is GARCH-n

| Out-of-sample Expected Shortfall ASL-values for Front Quarter Nordic power futures | | | | | | | |
|---|---|---|---|---|---|---|---|
| | ES-levels | | | | | | Rejects |
| | Long positions | | | Short positions | | | at |
| | 99% | 95% | 90% | 90% | 95% | 99% | 5% |
| GARCH-n | 0.002 | 0 | 0.003 | 0.054 | 0.034 | 0.117 | 4/6 |
| GARCH-t | 0.443 | 0.225 | 0.356 | 0.684 | 0.703 | 0.509 | 0/6 |
| RM-CF | 0 | 0 | 0 | 0 | 0 | 0 | 6/6 |
| FHS | 1 | 0.985 | 0.894 | 0.932 | 0.997 | 1 | 0/6 |
| $G_t/G_n$/FHS/RM-CF | 0 | 0 | 0 | 0 | 0 | 0.006 | 6/6 |
| $G_t/G_n$/FHS | 0.985 | 0.308 | 0.263 | 0.423 | 0.776 | 0.839 | 0/6 |
| $G_t$/FHS | 1 | 0.815 | 0.675 | 0.778 | 0.945 | 0.984 | 0/6 |
| $G_n$/FHS | 1 | 0.426 | 0.264 | 0.375 | 0.753 | 0.998 | 0/6 |

Table 44: Out-of-sample Expected Shortfall ASL-values for the models tested for Front Quarter Nordic power futures. Red indicate that the null hypothesis has been rejected at the 5% significance level. $G_t$ is GARCH-t and $G_n$ is GARCH-n

| Out-of-sample backtesting results for Value-at-Risk for Front Quarter Nordic power futures | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | VaR-levels | | | | | | | | | | | | Rejects |
| | Long positions | | | | | | Short positions | | | | | | at |
| | 99% | | 95% | | 90% | | 90% | | 95% | | 99% | | 5% |
| Model | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | sign. |
| GARCH-n | 0.39 | 0.08 | 0.85 | 0.34 | 0.91 | 0.12 | 0.04 | 0.01 | 0.03 | 0.24 | 0.38 | 0.05 | 3/12 |
| GARCH-t | 0.52 | 0.67 | 0.66 | 0.68 | 0.80 | 0.73 | 0.02 | 0.45 | 0.06 | 0.53 | 0.63 | 0.44 | 1/12 |
| RM-CF | 0.69 | 0.12 | 0.99 | 0.58 | 0.92 | 0.00 | 0.04 | 0.00 | 0.27 | 0.02 | 0.63 | 0.44 | 4/12 |
| FHS | 0.45 | 0.27 | 0.89 | 0.81 | 0.93 | 0.62 | 0.04 | 0.42 | 0.19 | 0.20 | 0.43 | 0.03 | 2/12 |
| $G_t/G_n$/FHS/RM-CF | 0.56 | 0.97 | 0.73 | 0.85 | 0.80 | 0.47 | 0.03 | 0.78 | 0.09 | 0.85 | 0.49 | 0.51 | 1/12 |
| $G_t/G_n$/FHS | 0.47 | 0.38 | 0.66 | 0.68 | 0.49 | 0.84 | 0.03 | 0.22 | 0.10 | 0.93 | 0.47 | 0.38 | 1/12 |
| $G_t$/FHS | 0.52 | 0.67 | 0.89 | 0.81 | 0.84 | 0.51 | 0.03 | 0.72 | 0.10 | 0.93 | 0.52 | 0.67 | 1/12 |
| $G_n$/FHS | 0.39 | 0.08 | 0.70 | 0.76 | 0.70 | 0.50 | 0.04 | 0.28 | 0.09 | 0.85 | 0.31 | 0.19 | 1/12 |

Table 45: Out-of-sample P-values for the last 50% of the dataset for the conditional independence test(CI) and the Kupiec unconditional coverage test(UC). Red indicate that the null hypothesis of conditional independence or unconditional coverage has been rejected at the 5% significance level. $G_t$ is GARCH-t and $G_n$ is GARCH-n

| Out-of-sample Expected Shortfall ASL-values for Front Year Nordic power futures | | | | | | | |
|---|---|---|---|---|---|---|---|
| | ES-levels | | | | | | Rejects |
| | Long positions | | | Short positions | | | at |
| | 99% | 95% | 90% | 90% | 95% | 99% | 5% |
| GARCH-n | 0.001 | 0.028 | 0.071 | 0.004 | 0 | 0.004 | 5/6 |
| GARCH-t | 0.469 | 0.854 | 0.928 | 0.824 | 0.501 | 0.243 | 0/6 |
| RM-CF | 0 | 0 | 0 | 0 | 0 | 0 | 6/6 |
| FHS | 1 | 1 | 0.993 | 0.977 | 0.998 | 1 | 0/6 |
| $G_t/G_n$/FHS/RM-CF | 0 | 0 | 0 | 0 | 0 | 0 | 6/6 |
| $G_t/G_n$/FHS | 1 | 0.827 | 0.754 | 0.728 | 0.575 | 0.989 | 0/6 |
| $G_t$/FHS | 1 | 0.992 | 0.977 | 0.936 | 0.931 | 1 | 0/6 |
| $G_n$/FHS | 1 | 0.854 | 0.711 | 0.55 | 0.679 | 1 | 0/6 |

Table 46: Out-of-sample Expected Shortfall ASL-values for the models tested for Front Year Nordic power futures. Red indicate that the null hypothesis has been rejected at the 5% significance level. $G_t$ is GARCH-t and $G_n$ is GARCH-n

| Out-of-sample backtesting results for Value-at-Risk for Front Year Nordic power futures | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | VaR-levels | | | | | | | | | | | | Rejects |
| | Long positions | | | | | | Short positions | | | | | | at |
| | 99% | | 95% | | 90% | | 90% | | 95% | | 99% | | 5% |
| Model | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | CI | UC | sign. |
| GARCH-n | 0.33 | 0.44 | 0.01 | 0.23 | 0.07 | 0.16 | 0.00 | 0.11 | 0.00 | 0.59 | 0.00 | 0.01 | 5/12 |
| GARCH-t | 0.54 | 0.51 | 0.00 | 0.30 | 0.26 | 0.39 | 0.00 | 0.39 | 0.00 | 0.44 | 0.00 | 0.56 | 4/12 |
| RM-CF | 0.66 | 0.45 | 0.01 | 0.44 | 0.24 | 0.00 | 0.01 | 0.00 | 0.00 | 0.04 | 0.02 | 0.86 | 7/12 |
| FHS | 0.45 | 0.71 | 0.00 | 0.62 | 0.07 | 0.65 | 0.00 | 0.65 | 0.00 | 0.08 | 0.00 | 0.01 | 5/12 |
| $G_t/G_n$/FHS/RM-CF | 0.56 | 0.38 | 0.00 | 0.76 | 0.19 | 0.56 | 0.00 | 0.43 | 0.00 | 0.50 | 0.04 | 0.44 | 4/12 |
| $G_t/G_n$/FHS | 0.51 | 0.82 | 0.01 | 0.62 | 0.22 | 0.91 | 0.00 | 0.93 | 0.00 | 0.69 | 0.00 | 0.25 | 4/12 |
| $G_t$/FHS | 0.52 | 0.66 | 0.00 | 0.39 | 0.09 | 0.51 | 0.01 | 0.43 | 0.00 | 16 | 0.00 | 0.56 | 4/12 |
| $G_n$/FHS | 0.47 | 0.86 | 0.00 | 0.76 | 0.15 | 0.62 | 0.00 | 0.48 | 0.00 | 0.69 | 0.00 | 0.01 | 5/12 |

Table 47: Out-of-sample P-values for the last 50% of the dataset for the conditional independence test(CI) and the Kupiec unconditional coverage test(UC). Red indicate that the null hypothesis of conditional independence or unconditional coverage has been rejected at the 5% significance level. $G_t$ is GARCH-t and $G_n$ is GARCH-n

## 5.4 Summary of results

The in-sample fit VaR backtesting of the entire dataset reveals that:

- The simple average models generally perform very well for Value-at-Risk, with a total of 6, 6, 7, 7 and 8 test rejections, while the individual models have 5 to 19 test rejections
- Of the individual models, the ones exploiting the quantile regression approach perform very well. This is in line with the findings of Steen et al. (2015) and Dahlen et al. (2015)
- The GARCH-t approach seems adequate as well, contrary to the GARCH-n approach, indicating that the latter fails to account for the fat tails previously described in section 3.4.
- FHS performs poorly with a total of 11 rejections, while the inferior performance of RM-CF indicates that the Cornish-Fisher approximation fails to correctly account for the tails of the distributions

For the out-of-sample VaR backtesting, we can conclude that:

- The simple average models perform well in general, with 6, 6, 6, 8 and 8 rejections in total for the five models, while the individual models have 6 to 16 test rejections
- The quantile regression approaches performs very well, as the findings of Steen et al. (2015) and Dahlen et al. (2015) also indicated. The quantile regression approaches again seem most adequate among stand alone models
- The GARCH-t models outperform the GARCH-n models because of the fat tails of the Nordic power futures logreturn distributions, previously elaborated in section 3.5
- FHS performs quite poor, while RM-CF performs terrible, indicating that the Cornish Fisher approach fails to correctly account for the tails of the distribution

For the in-sample fit backtest of ES on the entire dataset, we conclude that:

- $G_t/G_n$/FHS, $G_t$/FHS and $G_n$/FHS, and FHS produce superior results with 0 ES-rejections in total for all 4 models
- GARCH-t performs promising as well, with a total of 1 ES-rejection. As Zhu and Galbraith (2011) emphasized in their study, the additional parameters of the GARCH-t approach provides discernible improvements compared to GARCH-n in this regard
- GARCH-n, $G_t/G_n$/FHS/RM-CF and RM-CF perform very poor with 17, 18 and 18 ES-rejections respectively, making these models virtually useless for risk management purposes
- Analyzing the VaR-rejections for the models, the general tendency is that the simple average models have the least amount of VaR-rejections as well, making the average models; $G_t/G_n$/FHS, $G_t$/FHS and $G_n$/FHS very promising in this regard

The out-of-sample backtesting of ES reveals that:

- The results are very similar to those seen in-sample fit on the entire dataset; $G_t/G_n$/FHS, $G_t$/FHS and $G_n$/FHS perform very well with 0 ES-rejections, and few VaR-rejections
- GARCH-t and FHS are the best individual models with 0 ES-rejections, and few VaR-rejections. The additional parameters of GARCH-t compared to GARCH-n improved the model, for the out-of-sample backtesting as well, as Zhu and Galbraith (2011) emphasized in their study
- GARCH-n, $G_t/G_n$/FHS/RM-CF and RM-CF have a total of 15, 18 and 18 ES-rejections respectively, making these models virtually useless for risk management purposes
- The t-statistics of the bootstraps, presented in figure 10 and 12 respectively, reveals that RM-CF consistently underestimates risk, while FHS consistently overestimates risk, making GARCH-t the only individual model performing adequately, while the simple average approach of $G_t/G_n$/FHS, $G_t$/FHS and $G_n$/FHS perform superior among the simple average models

# 6 Conclusion

The highly volatile nature of the Nordic power futures strongly entails a requirement for risk management for actors affected by fluctuations in electricity prices and investors in the power futures market.

The aim of this paper is threefold:

1. First, to compare the in-sample fit of well known univariate risk models, for both Value-at-Risk and Expected Shortfall, in the Nordic power futures market.
2. Second, to study the out-of-sample performance of the models.
3. Third, to investigate both the in-sample fit and the out-of-sample performance of equally weighted averages of the same risk models, to find out whether these simple average models are more adequate than the individual models alone for risk management in the Nordic power futures market.

The paper written by Nowotarski et al. (2014) was the inspiration behind the idea of combining individual models in forecast averaging. Their quantile regression of the average of several point forecasts, proved to be superior under normal market conditions, but failed in a more volatile environment. Contrary to Nowotarski et al. (2014), we calculate Value-at-Risk for several models first, and make a simple average of the quantiles afterwards. The same procedure is carried out for Expected Shortfall. In our study, we aim to study if this simple average approach outperforms the individual models for the highly volatile Nordic power futures, and we regard our approach to model-averaging as the main contribution of this study.

The risk models are used to obtain Value-at-Risk and Expected Shortfall estimates for the 90%-, 95%-, and 99%-quantiles of the loss distribution for both long and short positions for the Nordic Power Front Month futures data, Front Quarter futures data, and Front Year futures data. The in-sample fit was conducted using the entire dataset of Front Month, Front Quarter and Front Year Nordic power futures. The out-of-sample testing was conducted using an in-sample window of the first 50% of the total data to decide simple average model constellations, and starting parameter values. The simple average models chosen were based on the in-sample results of the individual models. The constellations can be displayed in table 13 and 14 for VaR and ES respectively. The last 50% of the total data was used for the out-of-sample performance testing, using an expanding window.

The general tendency of both VaR in-sample fit and VaR out-of-sample backtesting, is that the quantile regression approach seems to perform well for the highly volatile Nordic power futures contracts. This is in line with the results of Dahlen et al. (2015) and Steen et al. (2015). The GARCH-n approaches fail to account for the fat tails of the Nordic power futures, elaborated in section 3.4, resulting in more rejections than for the GARCH-t approaches. RM-CF performs very poor as well, entailing that the Cornish Fisher approach fails to produce the promised fraction of violations specified by the VaR-level. FHS performs acceptable, while the simple average models perform very well in general.

For the ES in-sample fit and out-of-sample backtesting, the conclusion is quite clear cut; GARCH-t are best by far of the individual models, while $G_t/G_n$/FHS, $G_t$/FHS and $G_n$/FHS perform superior among the simple average models. RM-CF, GARCH-n and $G_t/G_n$/FHS/

RM-CF perform extremely poor. As Zhu and Galbraith (2011) emphasized in their study, the additional parameters of the GARCH-t approach provides discernible improvements compared to GARCH-n in this regard as well. A close look at the test values of the models, shows ASL-values of RM-CF of 0 and ASL-values of 1 for FHS, for all quantiles in all contracts. This indicates that RM-CF consistently underestimates risk, and FHS consistently overestimates risk, making these models next to useless for risk management purposes. This is further illustrated in figure 10 and 12. GARCH-t, $G_t/G_n$/FHS, $G_t$/FHS and $G_n$/FHS have reliable results, and are preferred in this regard.

## 6.1 Further work

The results show that the simple average models chosen perform well, for both Value-at-Risk and Expected Shortfall, and for both in-sample fit and out-of-sample use. Based on the results, the model-averaging approach is an area of research worth looking more into, both in terms of including other models and apply it for other commodities.

To further improve the average models presented in this paper, weighted averages is an approach that might yield better results than our simple average approach. This can be researched further in combination with machine learning to find the optimal weighted average. The quantile regression approach can be further researched for the Nordic power futures using multiple variables to enhance the predictability of the model.

An area that needs more research, is backtesting of Expected Shortfall. There were very few studies on this matter, and the backtesting procedure of McNeil and Frey (2000) used in this study, is an one-sided test only testing if the model is systematically underestimating the risk. A model that systematically overestimates risk will thus yield superior test results, falsely indicating that the model fits the purpose. The filtered historical simulations approach in this study is one example of this occurrence, and we had to study the bootstrap of the model closely in order to disclose the overestimation of the risk. Expected Shortfall as a risk metric is probably getting more used for risk management in general in the future, following the shortcomings of VaR, elaborated on in section 1.4. To be able to fairly assess which models are best suited for ES predictions, a better backtesting procedure to account for both overestimation of risk and underestimation of risk would be beneficial.

# A Appendix

## A.1 GARCH parameter values

| In-sample GARCH-n parameters | | | |
|---|---|---|---|
| Parameter | Front Month | Front Quarter | Front Year |
| Omega | 0.2129981448640945 | 0.0873320331816989 | 0.014956354246876058 |
| Beta | 0.8642084636289018 | 0.8904995367760722 | 0.8995710494915841 |
| Alpha | 0.11375143488790014 | 0.09876149591137398 | 0.10042895157223047 |

Table 48: In-sample parameters for GARCH-n for every contract. These estimates are based on the whole range of data for the front month contracts

| In-sample GARCH-t parameters | | | |
|---|---|---|---|
| Parameter | Front Month | Front Quarter | Front Year |
| Omega | 0.2210979205142998 | 0.06737420457556025 | 0.013220455682573459 |
| Beta | 0.8631438599342088 | 0.9030639147149002 | 0.9032380952002583 |
| Alpha | 0.1135780874521573 | 0.0896500119592205 | 0.09676190271344341 |
| DoF | 6.48229334493245 | 7.447429564727474 | 7.56674631292566 |

Table 49: In-sample parameters for GARCH-n for every contract. These estimates are based on the whole range of data for the front month contracts

| In-sample GJR-GARCH-n parameters | | | |
|---|---|---|---|
| Parameter | Front Month | Front Quarter | Front Year |
| Omega | 0.21349163956718195 | 0.08597027012273213 | 0.01372091740850213 |
| Beta | 0.8639293611334875 | 0.8901112398788718 | 0.9024149238730055 |
| Alpha | 0.11566769416441489 | 0.11029589043357409 | 0.11435794194932992 |
| Leverage | -0.0031817960110339947 | -0.020754742886720544 | -0.03354573164747952 |

Table 50: In-sample parameters for GJR-GARCH-n for every contract. These estimates are based on the whole range of data for the front month contracts

| In-sample GJR-GARCH-t parameters | | | |
|---|---|---|---|
| Parameter | Front Month | Front Quarter | Front Year |
| Omega | 0.22211314543459307 | 0.06722338887182545 | 0.012472199022702647 |
| Beta | 0.8627651949930552 | 0.902439846152967 | 0.9048236440607549 |
| Alpha | 0.11711259091660381 | 0.09650373010571808 | 0.10747752663591338 |
| Leverage | -0.006241046281465197 | -0.011854172166651763 | -0.024602341390929115 |
| DoF | 6.4884563355405565 | 7.490566987640095 | 7.674257800387197 |

Table 51: In-sample parameters for GJR-GARCH-t for every contract. These estimates are based on the whole range of data for the front month contracts

## A.2 QR parameter values

| In-sample QR parameters for QR-GARCH-n | | | | | | |
|---|---|---|---|---|---|---|
| Position | Long position | | | | | |
| Quantile | 0.99 | | 0.95 | | 0.90 | |
| Parameter | Alpha | Beta | Alpha | Beta | Alpha | Beta |
| Front Month | -0.786742 | -2.326134 | -0.549740 | -1.409688 | -0.213773 | -1.149712 |
| Front Quarter | -1.576704 | -1.926729 | -0.301259 | -1.484594 | -0.221654 | -1.094221 |
| Front Year | -0.542196 | -2.068166 | -0.225204 | -1.401407 | -0.053128 | -1.140745 |
| Position | Short position | | | | | |
| Quantile | 0.99 | | 0.95 | | 0.90 | |
| Parameter | Alpha | Beta | Alpha | Beta | Alpha | Beta |
| Front Month | 0.427976 | 2.297199 | -0.071603 | 1.607855 | -0.333709 | 1.270289 |
| Front Quarter | 0.550053 | 2.120663 | 0.190017 | 1.505147 | 0.073541 | 1.140661 |
| Front Year | 0.199840 | 2.356109 | 0.304410 | 1.362925 | 0.240114 | 1.022617 |

Table 52: In-sample fit QR parameters of the entire dataset for QR-GARCH-n

| In-sample QR parameters for QR-GARCH-t | | | | | | |
|---|---|---|---|---|---|---|
| Position | Long position | | | | | |
| Quantile | 0.99 | | 0.95 | | 0.90 | |
| Parameter | Alpha | Beta | Alpha | Beta | Alpha | Beta |
| Front Month | -0.744664 | -2.341312 | -0.587613 | -1.385844 | -0.183658 | -1.160294 |
| Front Quarter | -1.430947 | -1.952515 | -0.302047 | -1.479424 | -0.234736 | -1.087631 |
| Front Year | -0.516106 | -2.095031 | -0.221170 | -1.409187 | -0.063342 | -1.135275 |
| Position | Short position | | | | | |
| Quantile | 0.99 | | 0.95 | | 0.90 | |
| Parameter | Alpha | Beta | Alpha | Beta | Alpha | Beta |
| Front Month | 0.405775 | 2.304686 | -0.124303 | 1.627745 | -0.355065 | 1.278722 |
| Front Quarter | 0.528049 | 2.121223 | 0.270301 | 1.453424 | 0.075400 | 1.139276 |
| Front Year | 0.224564 | 2.347242 | 0.307562 | 1.361179 | 0.240831 | 1.023948 |

Table 53: In-sample fit QR parameters of the entire dataset for QR-GARCH-t

| In-sample QR parameters for QR-GJR-GARCH-n | | | | | | |
|---|---|---|---|---|---|---|
| Position | Long position | | | | | |
| Quantile | 0.99 | | 0.95 | | 0.90 | |
| Parameter | Alpha | Beta | Alpha | Beta | Alpha | Beta |
| Front Month | -0.766524 | -2.336229 | -0.602153 | -1.383585 | -0.217732 | -1.148474 |
| Front Quarter | -1.492660 | -1.967282 | -0.295165 | -1.485819 | -0.303799 | -1.061299 |
| Front Year | -0.505766 | -2.121990 | -0.234650 | -1.400292 | -0.071410 | -1.128968 |
| Position | Short position | | | | | |
| Quantile | 0.99 | | 0.95 | | 0.90 | |
| Parameter | Alpha | Beta | Alpha | Beta | Alpha | Beta |
| Front Month | 0.462666 | 2.280966 | -0.108587 | 1.615526 | -0.320866 | 1.265522 |
| Front Quarter | 0.382109 | 2.152703 | 0.214016 | 1.493033 | -0.016522 | 1.177301 |
| Front Year | 0.012895 | 2.522448 | 0.294463 | 1.369123 | 0.228261 | 1.034271 |

Table 54: In-sample fit QR parameters of the entire dataset for QR-GJR-GARCH-n

| In-sample QR parameters for QR-GJR-GARCH-t | | | | | | |
|---|---|---|---|---|---|---|
| Position | Long position | | | | | |
| Quantile | 0.99 | | 0.95 | | 0.90 | |
| Parameter | Alpha | Beta | Alpha | Beta | Alpha | Beta |
| Front Month | -0.700215 | -2.363192 | -0.566253 | -1.400964 | -0.192816 | -1.161437 |
| Front Quarter | -1.501962 | -1.957387 | -0.314191 | -1.474914 | -0.321847 | -1.053783 |
| Front Year | -0.467810 | -2.141092 | -0.242716 | -1.399015 | -0.083816 | -1.123699 |
| Position | Short position | | | | | |
| Quantile | 0.99 | | 0.95 | | 0.90 | |
| Parameter | Alpha | Beta | Alpha | Beta | Alpha | Beta |
| Front Month | 0.466478 | 2.275551 | -0.138070 | 1.629615 | -0.321630 | 1.264767 |
| Front Quarter | 0.507823 | 2.106378 | 0.158843 | 1.502409 | 0.019820 | 1.155522 |
| Front Year | 0.017699 | 2.510345 | 0.310330 | 1.362877 | 0.218432 | 1.042126 |

Table 55: In-sample fit QR parameters of the entire dataset for QR-GJR-GARCH-t

| In-sample QR parameters for QR-RiskMetrics | | | | | | |
|---|---|---|---|---|---|---|
| Position | Long position | | | | | |
| Quantile | 0.99 | | 0.95 | | 0.90 | |
| Parameter | Alpha | Beta | Alpha | Beta | Alpha | Beta |
| Front Month | -1.428824 | -2.138305 | -1.117127 | -1.261537 | -0.631650 | -1.028819 |
| Front Quarter | -1.921611 | -1.721105 | -0.572989 | -1.397532 | -0.458375 | -1.036499 |
| Front Year | -0.202489 | -2.405227 | -0.284481 | -1.404229 | -0.153937 | -1.107609 |
| Position | Short position | | | | | |
| Quantile | 0.99 | | 0.95 | | 0.90 | |
| Parameter | Alpha | Beta | Alpha | Beta | Alpha | Beta |
| Front Month | 1.189093 | 2.045513 | 0.724088 | 1.386227 | 0.309443 | 1.029717 |
| Front Quarter | 0.898112 | 2.017267 | 0.466015 | 1.445773 | 0.267870 | 1.080846 |
| Front Year | 0.535105 | 2.255488 | 0.464664 | 1.290773 | 0.281256 | 1.034079 |

Table 56: In-sample fit QR parameters of the entire dataset for QR-RiskMetrics

## A.3 Histograms of logreturns

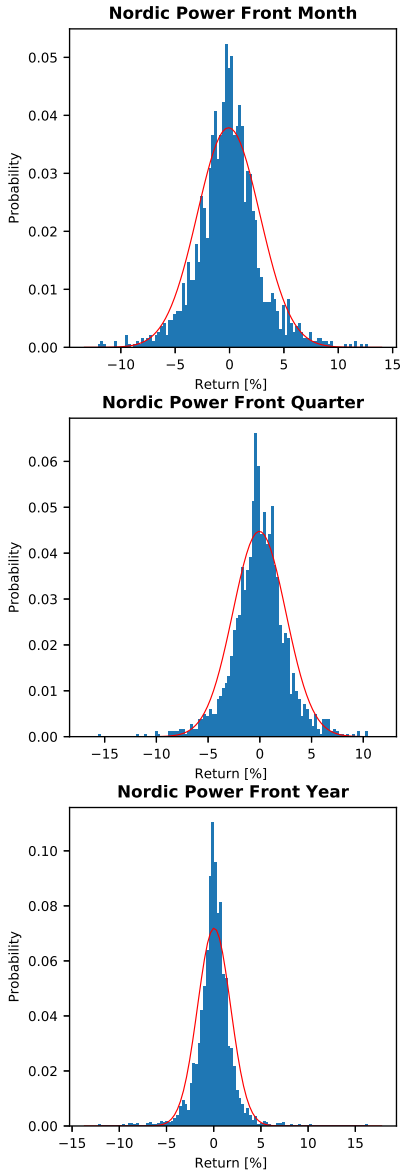



Figure 14: Histogram of in-sample returns with the corresponding normal probability density function. Data far out in the tails, and high density close to zero is evident, demonstrating the leptokurtic shape.

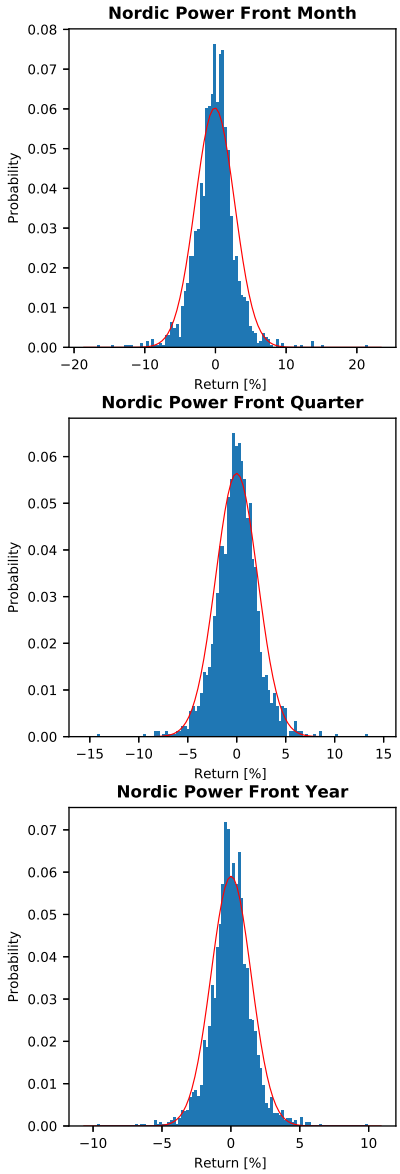



Figure 15: Histogram of out-of-sample returns with the corresponding normal probability density function. Data far out in the tails, and high density close to zero is evident, demonstrating the leptokurtic shape.

# References


Aggarwal, S. K., Saini, L. M. and Kumar, A. (2009), 'Electricity price forecasting in deregulated markets: A review and evaluation', *International Journal of Electrical Power Energy Systems* **31**(1), 13 – 22.

Alexander, C. (2008*a*), *Practical Financial Econometrics*, John Wiley & Sons Ltd.

Alexander, C. (2008*b*), *Value-at-Risk models*, John Wiley & Sons Ltd.

Ardia, D. and Hoogerheide, L. F. (2014), 'Garch models for daily stock returns: Impact of estimation frequency on value-at-risk and expected shortfall forecasts', *Economics Letters* **123**(2), 187–190.

BIS (2018*a*), 'Baseliii', `https://www.bis.org/bcbs/publ/d424.pdf`. Accessed: 2018-12-07.

BIS (2018*b*), 'The market risk framework: 25 years in the making', `https://www.bis.org/speeches/sp180425.pdf`. Accessed: 2018-12-07.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of econometrics* **31**(3), 307–327.

Botterud, A., Kristiansen, T. and Ilic, M. D. (2010), 'The relationship between spot and futures prices in the nord pool electricity market', *Energy Economics* **32**(5), 967–978.

Bye, T. and Hope, E. (2005), 'Deregulation of electricity markets: the norwegian experience', *Economic and Political Weekly* pp. 5269–5278.

Chan, K. F. and Gray, P. (2006), 'Using extreme value theory to measure value-at-risk for daily electricity spot prices', *International Journal of forecasting* **22**(2), 283–300.

Christoffersen, P. (2012), *Elements of Financial Risk Management*, Elsevier.

Christoffersen, P. F. (1998), 'Evaluating interval forecasts', *International Economic Review* **39**(4), 841–862.

Dahlen, K. E., Huisman, R. and Westgaard, S. (2015), Risk modelling of energy futures: A comparison of riskmetrics, historical simulation, filtered historical simulation, and quantile regression, *in* 'Stochastic Models, Statistics and Their Applications', Springer, pp. 283–291.

E24 (2018), 'Milliardtap felte krafttrader einar aas', `https://e24.no/privat/investeringer/krafttrader-einar-aas-er-trolig-personlig-konkurs-etter-stortap-i-kraftmarkedet/24439978`. Accessed: 2018-12-15.

Efron, B. and Tibshirani, R. J. (1993), *An introduction to the bootstrap*, Chapman  Hall.

Füss, R., Kaiser, D. G. and Adams, Z. (2016), *Value at Risk, GARCH Modelling and the Forecasting of Hedge Fund Return Volatility*, Palgrave Macmillan UK, London, pp. 91–117.

Füss, R., Adams, Z. and Kaiser, D. (2008), 'The predictive power of value-at-risk models in commodity futures markets', *Journal of Asset Management* **11**.

Gabrielsen, A., Kirchner, A., Liu, Z. and Zagaglia, P. (2015), 'Forecasting value-at-risk with time-varying variance, skewness and kurtosis in an exponential weighted moving average framework', *Annals of Financial Economics* **10**(01), 1550005.

Garcia, R. C., Contreras, J., Van Akkeren, M. and Garcia, J. B. C. (2005), 'A garch forecasting model to predict day-ahead electricity prices', *IEEE transactions on power systems* **20**(2), 867–874.

Giot, P. and Laurent, S. (2003), 'Market risk in commodity markets: a var approach', *Energy Economics* **25**(5), 435 – 457.

Harmantzis, F. C., Miao, L. and Chien, Y. (2006), 'Empirical study of value-at-risk and expected shortfall models with heavy tails', *The journal of risk finance* **7**(2), 117–135.

Kupiec, P. (1995), 'Techniques for verifying the accuracy of risk measurement models', *The journal of Derivatives* **3**(2), 73–84.

McNeil, A. J. and Frey, R. (2000), 'Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach', *Journal of empirical finance* **7**(3-4), 271–300.

Mina, J., Xiao, J. Y. et al. (2001), 'Return to riskmetrics: the evolution of a standard', *RiskMetrics Group* **1**, 1–11.

Nasdaq (2018), 'Who we are & our history', `https://business.nasdaq.com/trade/commodities/who-we-are/index.html`. Accessed: 2018-12-06.

NasdaqCommodities (2018), 'Power futures', `https://business.nasdaq.com/trade/commodities/products/power-derivatives/power-futures.html`. Accessed: 2018-12-06.

Nowotarski, J., Raviv, E., Trück, S. and Weron, R. (2014), 'An empirical comparison of alternative schemes for combining electricity spot price forecasts', *Energy Economics* **46**, 395–412.

NRK (2018), 'Dette er einar aas-saken', `https://www.nrk.no/norge/dette-er-einar-aas-saken-1.14211611`. Accessed: 2018-12-15.

PwC (2016), 'Basel iv: Revised standardised approach for market risk', `https://www.pwc.com/gx/en/advisory-services/basel-iv/basel-iv-revised-standardised-.pdf`. Accessed: 2019-05-07.

Sheedy, E. A. (2008), 'Why var models fail and what can be done'.

Steen, M., Westgaard, S. and Gjølberg, O. (2015), 'Commodity value-at-risk modeling: comparing riskmetrics, historic simulation and quantile regression'.

Vehviläinen, I. and Keppo, J. (2003), 'Managing electricity market price risk', *European Journal of Operational Research* **145**(1), 136–147.

Weron, R. (2014), 'Electricity price forecasting: A review of the state-of-the-art with a look into the future', *International Journal of Forecasting* **30**(4), 1030 – 1081.

Westgaard, S., Veka, S., Haugom, E. and Lien, G. (2014), 'A note on the risk characteristics of european energy futures markets', *Beta* **28**(01), 6–19.

Zhu, D. and Galbraith, J. W. (2011), 'Modeling and forecasting expected shortfall with the generalized asymmetric student-t and asymmetric exponential power distributions', *Journal of Empirical Finance* **18**(4), 765–778.

Master's thesis 2019

Maurits Mogenssøn Aaløkken, Jørgen Andersen Sveinsson

NTNU
Norwegian University of
Science and Technology