Lasse Samuelsen Bøe
Jørgen Søvik Opheim
Adrian Salater Stålesen

Forecasting Value-at-Risk and Expected Shortfall in German-Nordic electricity futures spreads

June 2019

NTNU
Norwegian University of
Science and Technology

Faculty of Economics and Management
Department of Industrial Economics and Technology
Management

Master's thesis

2019

Lasse Samuelsen Bøe, Jørgen Søvik Opheim, Adrian Salater Stålesen

Master's thesis

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of
Science and Technology

NTNU

Norwegian University of
Science and Technology

Forecasting Value-at-Risk and Expected Shortfall in German-Nordic electricity futures spreads

Lasse Samuelsen Bøe

Jørgen Søvik Opheim

Adrian Salater Stålesen

Industrial Economics and Technology Management

Submission date: June 2019

Supervisor: Sjur Westgaard

Norwegian University of Science and Technology

Department of Industrial Economics and Technology Management

Preface

This thesis concludes our Master of Science in Industrial Economics and Technology Management with a specialization in Financial Engineering at the Norwegian University of Science and Technology. We would like to thank our supervisor, Professor Sjur Westgaard at the Department of Industrial Economics and Technology Management, for helpful guidance and advice. His interest in our work has been truly valuable during the completion of our master's thesis. We would also like to thank Morten Hegna in Montel for granting us access to their extensive database of energy futures data. Furthermore, we extend our gratitude to Gunnar Aronsen in Trønderenergi for valuable market insights.

Abstract

In this study, we forecast day-ahead Value-at-Risk (VaR) and Expected Shortfall (ES) in German-Nordic front-quarter and front-year electricity futures spreads. We achieve this by employing a set of univariate and bivariate generalized autoregressive conditional heteroscedasticity (GARCH) models with different distributional assumptions, along with two quantile regression models. We compare the out-of-sample performance of two alternative forecasting procedures - one approach using a fixed estimation sample and one approach which involves reestimation. We conclude that the GARCH model with an extreme value theory distributional assumption, the GJR-GARCH model with skewed Student t distribution and the constant conditional correlation (CCC) model with symmetric Student t distribution are the most appropriate specifications when forecasting VaR and ES with a fixed estimation sample. Moreover, we find that GARCH models with symmetric or skewed Student t distribution most accurately predict VaR and ES when forecasting with reestimation. The German-Nordic electricity futures spreads were subject to substantial market turmoil during September 2018, which gained international attention in the financial community. Our results can provide valuable insights for regulators, market makers or other market participants seeking to hedge or speculate on German-Nordic electricity futures spreads.

Sammendrag

I denne studien predikerer vi daglig Value-at-Risk (VaR) og Expected Shortfall (ES) i prisdifferansen mellom tyske og nordiske future-kontrakter på elektrisitet med forfall i kommende kvartal og kommende år. Til dette formålet bruker vi ulike univariate og bivariate formuleringer fra generalisert autoregressiv betinget heteroskedastisitet-modeller (GARCH-modeller). Resultatene sammenlignes basert på to alternative predikeringsmetoder - én metode som bruker konstant estimeringsutvalg og én metode som involverer re-estimering. Vi konkluderer med at GARCH-modellen med fordelingsantagelse fra ekstremverditeori, GJR-GARCH-modellen med asymmetrisk Student t-fordeling og konstant betinget korrelasjons-modellen med symmetrisk Student t-fordeling oppnår best nøyaktighet ved prediksjon av VaR og ES med konstant estimeringsutvalg. Videre finner vi at GARCH-modellen med symmetrisk og asymmetrisk Student t-fordeling gir best resultater ved prediksjon basert på re-estimering. Ekstreme prisbevegelser i de tyske og nordiske kraftmarkedene i september 2018 skapte stor internasjonal oppmerksomhet. Våre resultater kan være av interesse for ulike markedsaktører som ønsker å sikre posisjoner eller spekulere i prisdifferansen mellom tyske og nordiske future-kontrakter på elektrisitet.

Contents

List of tables

List of figures

1 Introduction

In this thesis, we investigate the financial risk involved in simultaneously holding a position in German and Nordic electricity futures. More specifically, we consider a portfolio where one futures contract is bought long, and the other one is sold short. This is equivalent to undertaking a spread position of German and Nordic electricity futures prices. Spread trading as a strategy entails identifying two financial instruments, typically commodity or interest rate derivatives, which tend to display some relationship over time and enter into long and short positions in each constituent, or leg, of the spread. This could be in two derivatives based on separate underlying, or of the identical underlying with different maturities. Thus, holding the positions which comprise the spread corresponds to speculating on the development of the spread itself, as a result of the change in the two underlying. Furthermore, the dynamics and drivers of the two underlying are not likely to be strictly analogous to the ones of the portfolio, as its components are anticipated to be highly correlated. Our analysis focuses on modelling the Value-at-Risk (VaR) and Expected Shortfall (ES) for a trader or investor positioned in the front-quarter and front-year German-Nordic electricity futures spreads. The front-quarter and front-year electricity futures refer to financial electricity derivatives with settlement at the end of the upcoming quarter or year, respectively. We accomplish this by implementing univariate and bivariate generalized autoregressive conditional heteroscedasticity (GARCH) models. Specifically, we consider the standard GARCH model, GJR-GARCH, and Markov switching GARCH models by the univariate approach. For the bivariate approach, we implement constant and dynamical conditional correlation (CCC and DCC) models. All models are estimated by employing both the normal and Student t distributions. The skewed Student t distribution is also applied for the GARCH and GJR-GARCH models. In addition to this, we include RiskMetrics and models based on quantile regression. The objective of this study is to evaluate which model that performs best in estimating VaR and ES for German-Nordic electricity futures spreads. This is achieved by backtesting and assessing the models' relative performance across forecasting procedures, quantiles and data sets.

VaR has become a widely accepted and implemented risk measure aimed at monitoring financial risk. Its popularity is due to several factors. Financial institutions and banks are obligated to use VaR for managing risk and determine required capital reserves. This is in order to comply with regulatory policies imposed on financial institutions by, e.g., Basel III. Additionally, VaR answers a question crucial to risk managers in a straightforward and easily communicated manner: with a given probability, p, what is the expected financial loss over a given time horizon? It turns out that this question also has a relatively simple statistical definition, namely that VaR corresponds to the p-quantile of a given distribution. Closely related to VaR is ES, which designates the expected loss given an occurrence which exceeds VaR. ES complements VaR as it is more sensitive to the tails of a given distribution. These risk measures are frequently used for financial securities and derivatives, and to some extent also for commodities trading.

Risk management and the application of VaR and ES for spread trading is, to our knowledge, not a well-studied area of research, especially so for commodities. However, some of the largest commodity trading losses of all time have been due to unexpected spread developments. In 1993, Metallgesellschaft, a former German industrial conglomerate, lost an estimated amount of USD 2.8 billion trying to hedge their obligations for delivery of fuel and oil products, i.e. forward contracts, with short-term futures. This effectively made them long several spreads based on the same underlying with different times to expiration (Edwards & Canter, 1995). In 2006, Amaranth Advisors, an American multi-strategy hedge-fund, lost approximately USD 5 billion trading natural gas futures of various maturities (Chincarini, 2008). As noted by (Kroner, Kneafsey, & Claessens, 1995), commodities have historically been one of the

most volatile asset classes internationally. Thus, speculating in a spread position of related commodities may represent risks for the investor that could be hard to understand, and even more so to quantify. Recently, a prominent Norwegian investor's failed spread trade of German and Nordic electricity futures caused major upheaval in the European energy markets. The event received significant attention from the international financial community ((Stafford & Sheppard, 2018), (Ewing & Schreuer, 2019)). In the week of September 3-7, 2018, the spreads of the German and Nordic electricity futures widened but did not display abnormal volatility. However, on Monday, September 10, 2018, the market experienced an extreme increase in both the front-quarter and front-year German-Nordic electricity futures spreads of 6.02 and 5.56 EUR/MWh, respectively. By considering the levels of the spreads, which were 4.93 and 9.95 EUR/MWh, respectively, it is clear that a short position in the either of the spreads on this day would lose a significant portion of its value. Subsequently, as the daily margin call was issued to the investor, he was not able to post the required collateral. This lead to the liquidation of the portfolio, and resulted in a forced auction of the positions, which rendered total estimated losses of approximately USD 140 million. The loss incidents mentioned above illustrate the need for proper risk management when engaging in spread trading, both for hedging and speculative purposes. For electricity derivatives, which may exhibit strong seasonal effects, extreme price spikes in addition to time-varying volatility and correlation, a spread position imposes significant financial risk for an investor. To efficiently manage the risk of such a portfolio, it is of great interest and importance, to model the distribution of returns, or at least the tails of the return distribution, to quantify the potential losses one could face in the event of an unexpected market event.

The modern European electricity market is the result of three decades of deregulation, which has facilitated increasingly integrated and transparent markets. The German and Nordic electricity markets are some of the most active with regards to trading in electricity derivatives, and they are connected through several transmission cables, both directly and indirectly. Moreover, Veka, Lien, Westgaard, and Higgs (2012) note that the German and Nordic electricity markets are related. They show this by using volatility and correlation models, indicating that the Nordic electricity futures prices correlate the most with its German counterparts, among a wide selection of energy commodities such as UK electricity futures, Brent crude and natural gas. As the German and Nordic electricity markets are both correlated and quite active with regards to trading of electricity derivatives, the topic and our results are of interest to market participants holding futures positions in both the German and Nordic electricity markets. These could be power producers seeking to hedge their long-term exposure to the electricity markets' volatile nature, or traders speculating on either a widening or tightening spread. To the best of our knowledge, VaR and ES forecasting for the German-Nordic futures spreads have not been treated in the literature as of June 2019. We also find that the literature is scarce on spread trading for electricity futures in general.

In the remainder of this thesis, we structure our analysis as follows. In Section 2 we provide a brief overview of the Nordic and German electricity markets, followed by a discussion on relevant literature for electricity futures with a particular emphasis on VaR and ES models in Section 3. We then proceed with a presentation of our methodology, with the univariate and bivariate models in Section 5 and 6, respectively. Backtesting procedures for both VaR and ES are provided in Section 7. Our results and a discussion of these are provided in Section 8. In Section 9, we conclude.

2 Market overview

In this section, we provide a brief overview of the Nordic and German electricity markets, with an emphasis on the two regions respective energy mix used for electricity generation and how their respective market places are organized. The Nordic and German electricity markets are currently directly linked through transmission lines and underwater power cables from Denmark and Sweden. There are also several indirect connections, e.g. through Poland. Figure 1 shows the transmission lines from the Nordics to Germany. Additional electricity cables from the Nordics to the UK and Germany are planned to be installed from 2020 to 2025, which is expected to double the total export capacity from the current levels approximately. Insight into how both the unique and shared market features affect electricity prices in the two respective regions may provide an increased qualitative understanding of their co-movements.

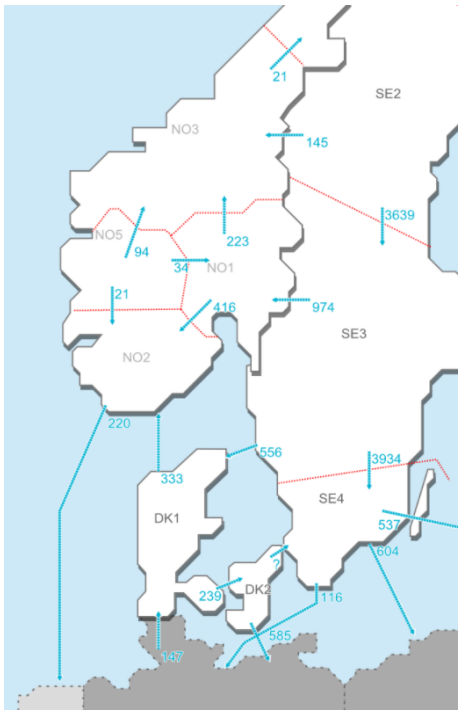


Figure 1: Typical electricity transmission (MW) in the Nordics and into Europe and Germany (blue arrows). Separate bidding regions in the Nordics are separated by red dotted lines (Statkraft, 2019)

.

2.1 Nordic electricity markets

The Nordic electricity market consists of producers, retailers, traders and other market participants in Norway, Sweden, Denmark and Finland. Through the liberalization of the Norwegian electricity market, following the Energy Act of 1990, market-based principles were introduced to broaden competition and allow for more efficient markets. Sweden later introduced similar legislation, leading to the integration of the Norwegian and Swedish electricity markets in 1996. This allowed for trading electricity across the borders of the two countries on the mutual Norwegian-Swedish power exchange, Nord Pool. Finland joined Nord Pool in 1998, and Denmark followed in 2000. The restructuring the Nordic electricity markets in the 1990s and early 2000s led to more transparent electricity trading and the introduction of derivatives for risk management such as futures, forwards and options.

Trading of electricity in the spot markets is carried out at Nord Pool. The exchange computes the daily

electricity system price for the upcoming 24 hours. The system price is a theoretical price based on the assumption of no congestion in the intra-market transmission grid and is the foundation and reference price for price setting in the Nordic financial electricity market. There are several nationally bounded spot price bidding zones within the Nordic price region, as can be seen in Figure 1. Financially settled Nordic electricity derivatives have been used since 1995, and are today primarily traded on Nasdaq. The derivatives are used by producers, retailers and end users as risk management tools, and by traders who speculate on future spot prices. They comprise futures, forwards, electricity price area differentials (EPAD) and options with forward contracts as underlying assets. The majority of Nordic futures and forward contracts use the system price as a reference price, i.e. the EUR-denominated price of 1 MWh of electricity. However, EPADs allow for speculation on the price differential between local bidding zones. Other futures contracts, including German electricity futures, can also be traded on Nasdaq.

Table 1: Electricity production in the Nordics by source (%)

Source	2010	2011	2012	2013	2014	2015	2016	2017
Hydropower	52.1	53.6	59.3	53.4	55.7	58.6	56.5	55.9
Wind	3.3	4.7	4.9	6.2	7.3	9.1	8.5	9.5
Nuclear	20.6	21.6	20.8	22.8	22.1	19.6	21.4	21.3
*Other**	24.0	20.1	15	17.6	14.9	12.7	13.6	13.3

The data is obtained and compiled from SSB (2018), Statistics Finland (2018), Danish Energy Agency (2019) and Ekonomifakta (2019). Other electricity sources refer to heat plants employing natural gas or oil, geothermal heat and other resources which are less typical to the Nordics.

In Table 1, one can see that the energy mix in the Nordics has been relatively constant since 2010. Hydropower accounts for a substantial amount of produced electricity in the Nordics. Approximately 50% of all Nordic electricity is generated from hydroelectric power plants. Norway has close to 100% hydropower while Sweden and Finland have around 50% and 20% shares, respectively. Water reservoir capacity is mainly a function of weather and precipitation. Thus, unusually cold winters and the amount of rain and snow during the year can explain much of the electricity price movements in the Nordics. Since 2010, the trend has been that fossil fuels and other carbon intensive energy carriers have been phased out in favour of wind power.

2.2 German electricity markets

As the Nordic markets were liberalized, the European Union introduced similar policies, aimed at opening cross-border electricity trading in continental Europe. Exchanges where electricity could be traded freely originated in the early 2000s. The European Energy Exchange (EEX) based in Leipzig began operations in 2002, which facilitated the integration of the European markets. The German electricity market is the largest in Europe, with a diversified energy input mix consisting of nuclear, coal, oil, gas, wind, hydropower, and solar. Due to this, several input variables could potentially impact the price formation in German electricity markets, arguably more so than in the Nordics. The energy mix in Germany has changed significantly over the last decade, with renewable energy accounting for more and more of the total produced electricity. The historical development of German electricity by source is found in Table 2. A significant factor affecting the German electricity market is the price of CO2 emission certificates. The certificates are part of a scheme regulated by the European Union Emissions Trading System (EU ETS). Producers of CO2 emissions are granted a fixed amount of emission allowances, permitting them to emit a certain amount of CO2 equivalents. Additional emission allowances must be purchased in the market. As large amounts of German electricity is produced from coal, the electricity prices are sensitive

to the cost of CO2 allowances. Also, the high share of wind power contributes significantly to the price volatility in Germany. When wind power production is high, often during weekends or at night when the demand for electricity is low, the oversupply of electricity leads to negative prices in the intra-day markets. All in all, there is a large and complex set of variables that affect German spot and futures prices.

Table 2: Electricity production in Germany by source (%)

Source	2010	2011	2012	2013	2014	2015	2016	2017
Coal	41.5	42.8	44.0	45.1	43.7	42	40.2	36.9
Nuclear	22.2	17.6	15.8	15.2	15.5	14.2	13.0	11.7
Natural Gas	14.1	14.0	12.1	10.6	9.7	9.6	12.5	13.3
Oil	1.4	1.2	1.2	1.1	0.9	1.0	0.9	0.9
Renewable energy	16.7	20.2	22.8	23.9	25.9	29.1	29.2	33.1
Wind	6.1	8.1	8.2	8.2	9.3	12.4	12.3	16.1
Solar	1.8	3.2	4.2	4.9	5.8	6.0	5.9	6.0
Biomass	4.6	5.2	6.1	6.3	6.7	6.9	6.9	6.9
Hydropower	3.3	2.9	3.5	3.6	3.1	2.9	3.2	3.1
Waste to energy	0.7	0.8	0.8	0.8	1.0	0.9	0.9	0.9
Other	4.1	4.2	4.1	4.1	4.3	4.1	4.2	4.1

The data is obtained from Energiebilanzen (2018).

Electricity trading and speculation is carried out via the EEX or EPEX SPOT. Day-ahead auctions and intra-day trading occurs on the EPEX SPOT, which operates the short-term electricity markets for Germany, France, Austria and Switzerland. The EEX derivatives market facilitates medium to long-term portfolio management and offers a broad range of electricity products for the main European markets. In addition to this, the market place provides derivatives on a wide variety of other commodities, e.g. natural gas, CO2 emissions and wind power. Derivatives with both physical delivery and financial settlement are traded on the EEX. Until recently, a system price for Germany, Austria and Luxembourg has been calculated daily by the EPEX SPOT, referred to as the Phelix Day Base or Phelix Day Peak Index. This spot price acts as the reference price for German/Austrian financial electricity derivatives (e.g. futures contracts). The German/Austrian price zone was split in the fall of 2018, as additional transmission capacity was added to the regional power grids. The effect this will have on the future development and dynamics of electricity futures remains to be studied. Henceforth, we will refer to all prices relating to the German/Austrian spot price, or the Phelix Day Base, as German.

2.3 Comparison of the German and Nordic electricity markets

The German and Nordic electricity markets are among the most active ones with regards to the trading of electricity futures. The market places are also organized similarly. The EPEX SPOT serves the same purpose in Germany as Nord Pool does for the Nordics, offering average system prices as reference for the markets. Likewise, the EEX and Nasdaq provide a more diversified offering of trading products, with financial electricity derivatives and products spanning other commodities and markets. The energy mix in the two respective markets are different in several ways. Firstly, the Nordic market relies mostly on renewable hydropower, in addition to nuclear power which is a stable and predictable source of energy with regards to production costs. Thus, the most important determinant of Nordic electricity prices is likely to be weather and the amount of precipitation during a given year. Germany, on the other hand utilize a wider range of resources as input for their electricity generation. Several of them, such as coal and natural gas are also potentially volatile commodities. A sudden price jump in either input factor

would most likely cause higher electricity prices, which again would impact the futures prices. Secondly, whereas the Nordics have maintained a large share of their produced electricity from hydropower and nuclear, Germany is undergoing a large scale energy transition. The country is currently phasing out coal and nuclear power in favour of wind, solar and other renewable sources of energy. As more transmission capacity is expected to be installed between Germany and the Nordics, it will be left to future research to determine how this will affect the two markets' impact on each other.

3 Literature review

In this section, we discuss literature which is relevant to our area of research. We emphasize that the literature on VaR and ES is quite extensive when considering other assets than electricity futures. The same applies to univariate and multivariate GARCH models. Thus we fill an interesting gap in the literature by providing a broad assessment of several GARCH specifications employed for modelling the risk of electricity futures spreads. Moreover, a significant portion of the academic literature related to energy economics and electricity price modelling has been dedicated to the forecasting of prices, returns and volatility. In addition to this, electricity spot prices are more frequently analyzed than electricity futures. The latter also extends to VaR and ES forecasting. As we in this thesis specifically consider VaR and ES forecasting for electricity futures spreads, we provide new insight into an apparently neglected field of research at the intersection of financial econometrics and energy economics.

3.1 Electricity as a commodity and derivative

Commodity prices have historically been one of the most volatile asset classes internationally (Kroner et al., 1995), a finding which also extends to electricity prices. However, electricity exhibits characteristics that differ compared to other commodities. Firstly, electricity is generally not regarded as storable. All produced electricity must be taken off the grid and utilized immediately. Secondly, there are no readily available and viable substitutes to electricity, resulting in low price elasticity of demand, see Lijesen (2007) and Labandeira, Labeaga, and López-Otero (2017). One consequence of this is high volatility in the price of electricity, leading to frequent and large price movements, causing valuation and trading of electricity products to be quite difficult. As there is almost a total absence of inventories to balance changes in both supply and demand, production and consumption of electricity must be synchronized at any given point in time. Because of this, there is limited room for arbitrage in electricity contracts over time. Electricity derivatives can serve as investment vehicles to hedge against, or speculate on developments in the underlying spot price. See Eydeland (2003) or Deng and Oren (2006) for a more elaborate overview on electricity derivatives. It is important to note that electricity futures contracts are less volatile and considerably more stable than the underlying spot prices. One explanation for this is that futures contracts represent the market's expectation of future demand, weather conditions, off-take and generation capacity during the lifetime of the derivative (Malo & Kanto, 2006). The phenomenon is in line with the Samuelson effect, originating from Samuelson (1965), who gives a general model showing that the volatility of futures contracts increases as the time to expiration decreases. He assumes that competitive forces in the futures market ensure that futures prices and spot prices converge towards maturity. See also Galloway and Kolb (1996) for an analysis of the maturity effect. The literature is largely dedicated to modelling and forecasting of electricity spot prices. Aggarwal, Saini, and Kumar (2009) provide an extensive overview of the research which has been conducted with regards to forecasting models for electricity prices and volatility. Most employ univariate time series models, assuming stationarity. See also Weron (2014) or Nowotarski and Weron (2018) for the most recent overviews of research on electricity price forecasting, which also consider volatility and VaR forecasting.

3.2 Spread trading

The notion of spread trading was introduced to the finance literature by Working (1949), who investigates the costs of storage and its effect on pricing relationships. The author demonstrated that traders could

profit from the existence of abnormal pricing relations between futures contracts with different expiration dates. The author suggests the cost-of-storage theory as an explanation of inter-temporal price relations. Spread trading has traditionally been used to speculate on the cost-of-carry between different futures contracts. However, it is important to note that spread trading also enhances liquidity in the markets, removes arbitrage and allows for hedging (Melamed, 1981).

Some studies on spread trading include Butterworth and Holmes (2002), who study the relative mispricing in the UK index futures markets, Dunis, Laws, and Evans (2006), who develop trading strategies for the gasoline crack spread, and Dunis, Laws, and Evans (2008), who assess spread trading strategies for Brent Crude, WTI Crude, gasoline and heating oil. The crack spread, i.e. the spread between heating oil futures and WTI crude oil futures prices with the same maturity, is treated by Alexander (2008c). An assessment of the relative performance of how to model changes in spreads does not appear to exist in the literature. A qualitative study of several loss incidents related to spread trading is provided by Till (2012).

For electricity futures, the literature contains few, if any, studies that specifically analyze electricity price spreads in regional and integrated markets. However, for spot markets, examples of studies which analyze the interrelations of regional electricity spot prices are found in De Vany and Walls (1999) or Hadsell and Shawky (2006) for the US markets and Worthington, Kay-Spratley, and Higgs (2005) for the Australian markets. Moreover, Spodniak, Viljainen, Makkonen, and Jantunen (2013) note that studies focusing on local electricity prices and their geographic price differences and spreads are rare for the European markets. The same authors provide empirical evidence that significant long term differences in area price spreads do exist among the bidding zones in the Nordics. See also Marckhoff and Wimschulte (2009) and Kristiansen (2004) for studies on intra-market electricity spreads in the Nordics.

3.3 Volatility modelling

Deb, Albert, Hsue, and Brown (2000) note that the most accurate way to quantify electricity price risks over any period is by modelling or simulating volatility. Extensive effort has been carried out in this respect. Considering some of the characteristics of electricity spot prices, such as time-varying mean and variance, seasonality, high levels of volatility along with a large number of price movements of high magnitude (Aggarwal et al., 2009), various modelling approaches are suggested in the literature. Large parts of the literature are concerned with the day-ahead and short-term electricity markets, see Chang and Park (2007), Garcia, Contreras, van Akkeren, and Garcia (2005), Hadsell, Marathe, and Shawky (2004), Higgs and Worthington (2008) and Knittel and Roberts (2005). Cifter (2013) analyzes the spot price formation on Nord Pool to forecast volatility and proposes the use of Markov switching models to capture time-varying volatility, as it enables more accurate forecasting compared to standard GARCH models. Other approaches for understanding volatility in electricity futures market include Haugom, Westgaard, Solibakke, and Lien (2011), who examine the link between future price volatility and current observable economic variables by using realized volatility from Nordic electricity futures. See also Solibakke (2006) for a stochastic model approach to volatility modelling for Nordic electricity prices, Haugom (2013), who consider realized variance in the same market and Solibakke (2002) who propose an ARMA-GARCH as a means to model volatility dynamics in Nordic spot prices. Volatility modelling of other commodities includes the work of, e.g. Kroner et al. (1995). The authors use a combination of implied volatility derived from option prices and GARCH models to estimate volatility in agricultural commodities futures, which outperform the individual application of the two methods. However, low trading volumes in electricity options on the futures considered in this paper cause this approach to

be unfeasible for our purposes. Wang and Wu (2012) have compiled studies modelling the volatility of crude oil futures by both univariate and multivariate GARCH methods. The authors conclude that the multivariate models perform the best overall. However, among the univariate models, GARCH models allowing for asymmetric effects achieve the greatest accuracy.

3.4 Correlation and covariance modelling

A common extension of univariate GARCH models, which can model volatility, are multivariate GARCH models. These estimate the covariance and correlation matrices and often rely on univariate GARCH models for volatility. Studies that use multivariate GARCH models for electricity futures are not very common, and there does not appear to exist a coherent framework or methodology which performs best across studies. Bauwens, Hafner, and Pierret (2013) apply the multivariate GARCH approach with DCC-models for the conditional correlations and GJR-GARCH for the volatility of German front-month, front-quarter and front-year electricity futures. Le Pen and Sévi (2010) use VAR-BEKK models to assess volatility spillover, i.e. correlation, between Dutch, German and British electricity futures markets, finding a clear connection between the three. Byström (2003) use bivariate GARCH models to assess the relationship between future and spot prices from the Nord Pool exchange. The author also investigates how time-varying volatility and correlation affects out-of-sample hedging performance for the futures and spot prices. Malo and Kanto (2006) provide a study utilizing a broader array of multivariate GARCH specifications for the Nordic electricity spot and futures prices. They consider, among others, CCC and DCC models with the normal and Student t distributions, but their findings provide no inference as to which of the models that are preferable. However, they conclude that for practitioners, the simpler models, e.g., CCC models, may suffice for hedging purposes. Sotiriadis, Tsotsos, and Kosmidou (2016) study the relationships between European energy markets but focus on spot prices rather than futures prices. They employ CCC and DCC models for conditional volatility and correlation. The authors show that the CCC and DCC models provide the same conclusions. Veka et al. (2012) also implement several multivariate GARCH models to assess the correlation between varying energy commodities such as oil, gas, coal, and electricity futures. They discover time-varying correlation and covariance matrices for all commodity classes except for oil and note that CCC models may be misleading for several assets. Also related to electricity futures markets, Balcılar, Demirer, Hammoudeh, and Nguyen (2016) employed multivariate GARCH models for EUA spot and future prices (carbon emission certificates) to assess its relationships with electricity, coal and natural gas futures prices. See also Bauwens, Laurent, and Rombouts (2006) provide a survey of the applications, model specifications and inference methods for multivariate GARCH models.

3.5 Value-at-Risk and Expected Shortfall for commodities

To the best of our knowledge, the literature and empirical studies on the modelling of VaR in European electricity markets, let alone the German and Nordic markets individually, is scarce. One of the few is Solibakke (2010) who uses EWMA and GARCH(1,1) with the normal distribution to estimate volatility, and a CCC model for the correlations. The author also conducts individual VaR forecasting for both German and Nordic front-year futures. Westgaard, Veka, Haugom, and Lien (2014) provide an examination of the risk characteristics of several energy commodities noting a lack of volatility clustering, that empirical VaR differs greatly across asset classes and that most energy commodities exhibit skew and kurtosis to a varying extent. Thus, model selection for VaR and tail risk assessment is a challenging endeavour. However, several studies estimate and model VaR for other energy commodities, e.g.

Brent crude oil and WTI crude oil. Hung, Lee, and Liu (2008) suggest using fat-tailed GARCH models to estimate VaR for energy commodities such as Brent crude oil, WTI crude oil and gas. Aloui and Mabrouk (2010) model the same commodities for both long and short positions, with innovations from the normal, Student t and skewed Student t distribution. They show that GARCH models allowing for long-range memory, fat tails and asymmetry perform the best in predicting day-ahead VaR. Zhang, Yao, and He (2015) find that linear one-regime models outperform Markov switching GJR-GARCH and eGARCH models with up to three regimes when forecasting VaR for Brent crude oil. However, a two-regime Markov switching GARCH performs best for volatility forecasting on a daily frequency. A study on more simplistic models such as RiskMetrics, historical simulation and quantile regression is found in Steen, Westgaard, and Gjølberg (2015). The authors find that quantile regression best estimates VaR across energy and agricultural commodities during 1992-2013. Dahlen, Huisman, and Westgaard (2015) employ the same models as Steen et al. (2015) for several energy commodities, including front-month and front-quarter German and Nordic electricity futures. Their results indicate that RiskMetrics and historical simulation has significant shortcomings, compared to, e.g. quantile regression. There is also evidence in the literature that the linear quantile regression models, accounting for non-linear effects of exogenous factors, outperform GARCH models with Student t in out-of-sample forecasting of price distribution quantiles (Bunn, Andresen, Chen, & Westgaard, 2016). Giot and Laurent (2003b) model long and short VaR for daily stock indices returns using parametric univariate and multivariate ARCH class models based on the skewed Student t distribution. They provide evidence that symmetric density distributions underperform compared to skewed density models when modelling both tails of the return distribution. These findings are also in line with Giot and Laurent (2003a), which propose the skewed t APARCH model to estimate day-ahead long and short VaR estimates for, among others, Brent and WTI crude oil spot prices. The study did not include electricity futures. A skewed Student t APARCH model performed best in all cases. However, a skewed Student t ARCH model delivered good results and could be preferable due to its simplicity. To cope with the extreme volatility in the electricity markets, extreme value theory (EVT), which utilizes the Generalized Pareto Distribution, has been proposed as a remedial to account for fat tails (Byström, 2005). McNeil and Frey (2000) note the advantage of a combined GARCH-EVT approach, as it enables conditional heteroscedasticity to be captured through the GARCH framework, while subsequently estimating the extreme tail distribution with EVT models. Fong Chan and Gray (2006) build on this, and conclude that combining EVT and GARCH models is a useful technique in forecasting VaR in electricity markets.

An important matter to determine before estimating VaR for a portfolio is whether to implement the model by univariate or multivariate means. Several authors conclude that univariate approaches perform the best in this respect. See e.g. Berkowitz and O' Brien (2002), Brooks and Persand (2003), Bauwens et al. (2006) and Asai, Mcaleer, and Yu (2006). These studies do not consider VaR for electricity commodities and are limited to portfolios containing relatively few assets. For portfolios with a large variety of assets, Santos, Nogales, and Ruiz (2013) provide evidence to the contrary of this. The authors suggest a DCC GARCH model with Student t as the most appropriate specification to estimate VaR for large portfolios.

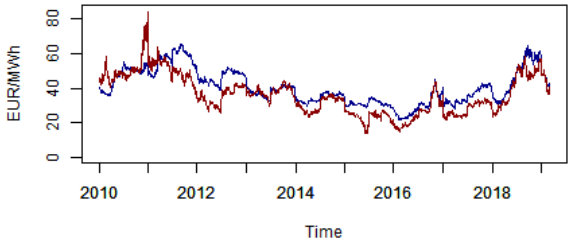
Žiković and Dizdarević (2011) evaluate the performance of a range of VaR and ES models on energy commodities such as WTI crude oil, natural gas and coal. The models are considered at confidence levels 95%, 99% and 99.5%. The obtained results, from among others EVT with GARCH, GARCH and RiskMetrics, show that the best performing VaR models are almost identical to the best performing ES models. Martins-Filho, Yao, and Torero (2018) propose the use of non-parametric models, e.g. the Generalized Pareto Distribution, to estimate ES and for five agricultural commodities. This is because few assumptions need to be made with regards to the entire distribution, and the focus is largely shifted

towards the tails. There seems to exist wide support of the idea that extreme value theory may assist in quantifying tail risk for several asset classes (McNeil & Frey, 2000), and studies from, e.g. Yamai and Yoshiba (2005), Kuester, Mittnik, and Paolella (2006), Acerbi and Tasche (2002) and Inui and Kijima (2005) suggest that EVT is the best approach when measuring and estimating the risk for the lowest and highest quantiles. However, Ghorbel and Souilmi (2014) finds that EVT does not statistically outperform standard GARCH models when estimating both VaR and ES for Brent crude oil and natural gas prices from 1998 to 2012. González-Pedraz, Moreno, and Peña (2014) consider oil, gas, coal, and power futures from 2005 to 2012 to assess the tail risk in various portfolios for both long and short positions, including the estimation of ES by using among other specifications an asymmetric DCC GARCH model.

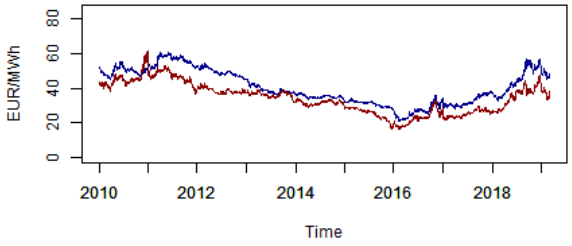
4 Data

4.1 Source of data

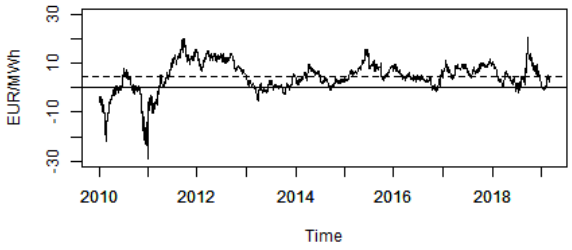
We obtain the data comprising our time series from Montel, a company which maintains a database on a wide array of commodities. In this thesis we use settlement prices from electricity deferred settlement futures contracts (DS futures) with cash settlement. The specific contract tickers for the German and Nordic futures are set forth in Appendix A. The German futures use the average of the hourly prices from the day-ahead auction for the German price zone as contract base. Correspondingly, the Nordic futures use the Nordic system price, quoted and published by Nord Pool. This yields a total of four separate time series, two for the front-quarter contracts and two for the front-year contracts. The time series are plotted in Figure 2a and 2b. The settlement prices originate from Nasdaq, and the data spans the period from 5 January 2010 to 1 March 2019.



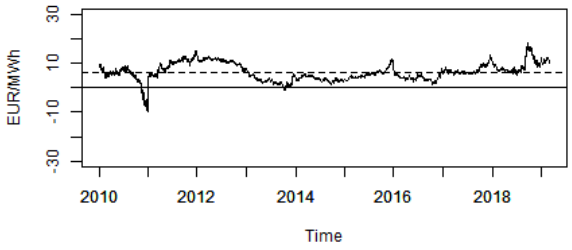
(a) Front-quarter futures prices: German (blue), Nordic (red)



(b) Front-year futures prices: German (blue), Nordic (red)



(c) Front-quarter spread (black) and sample mean (dashed)



(d) Front-year spread (black) and sample mean (dashed)

Figure 2: Top: Time series of German and Nordic futures prices. Bottom: Time series of spread of German and Nordic futures prices

4.2 Definitions

We define $P_{GER,t}$ as the electricity futures price at time t for the German contract, and $P_{NOR,t}$ as the electricity futures price at time t for the Nordic contract. The prices have units in EUR/MWh. The spread at time t, denoted by S_t, is defined as the difference between the futures price of the German and Nordic contract with corresponding expiration dates:

$$S_t = P_{GER,t} - P_{NOR,t} \tag{1}$$

This gives one time series for the front-quarter spread and another time series for the front-year spread, plotted in Figure 2c and 2d, respectively. In the rest of study, we often refer to daily profit and loss (P&L) by which we mean daily price change. In univariate analysis of VaR and ES, we model daily P&L of the spread which is defined as:

$$P\&L_t = S_t - S_{t-1} \tag{2}$$

The daily P&L of the front-quarter spread is shown in Figure 3a. In bivariate analysis of VaR and ES, rather than considering the spread as a univariate time series, we model a portfolio consisting of positions in both legs of the spread. This means that we model the daily P&L of each of the two components of the spread. We define daily P&L of the German contract and daily P&L of the Nordic contract as:

$$P\&L_{GER,t} = P_{GER,t} - P_{GER,t-1} \tag{3}$$

$$P\&L_{NOR,t} = P_{NOR,t} - P_{NOR,t-1} \tag{4}$$

The daily P&L of the front-quarter German contract and the daily P&L of the front-quarter Nordic contract is shown in Figure 3b and 3c, respectively.

4.3 Modelling considerations

How to best model the spread is an important discussion for this thesis. As emphasized by Alexander (2008c), an ordinary logarithmic return series is infeasible as the spread takes on both positive and negative values. Further, attempting to model the spread, S_t, directly is not appropriate as the front-year prices fail to reject the unit root hypothesis in the Dickey-Fuller test under the 5% significance level (Dickey & Fuller, 1981).

In the literature review, we identify two distinct approaches for modelling the dynamics of the change in commodity spreads. The first is most frequently encountered, and was proposed by Butterworth and Holmes (2002) and also used by Dunis et al. (2006) and Dunis et al. (2008). They analyze the daily change in the spread by considering the difference between the simple returns of the legs, expressed as:

$$R_t = \frac{P_t^{GER} - P_{t-1}^{GER}}{P_{t-1}^{GER}} - \frac{P_t^{NOR} - P_{t-1}^{NOR}}{P_{t-1}^{NOR}} \tag{5}$$

However, after conducting the Ljung Box test on the resulting time series, we find that both lack autocorrelation in squared returns, R_t^2. This is an important property for modelling conditional heteroscedasticity.

The second method was proposed by Alexander (2008c), where the author uses the P&L of the portfolio to capture the dynamics of the spread in a VaR setting. We will in this thesis consider this methodology. We argue that VaR and ES for a portfolio of contracts (in TWh) in this context are easily interpreted, as the two measures will be denominated in EUR/MWh.

4.4 Data processing and cleaning

As futures contracts cease to trade upon expiry, i.e. when the contracts are rolled, we remove observations from our time series when a new contract replaces the previous one as the front-quarter or front-year contract. More explicitly, the P&L of the spread on a given date where the contract is rolled is removed

from our time series, e.g. the day a Q1-contract is replaced by a Q2-contract or the 2018 contract by the 2019 contract. The P&Ls on such dates are not deemed to be caused by underlying market conditions, but rather by technicalities related to trading and clearing of the contracts. In total, the front-quarter contract involves the rolling of four contracts per year, and the front-year contract is rolled once per year. Additional processing includes the removal of observations where price data lacks in either of the legs, which primarily coincides with holidays in the Nordics and Germany.

4.5 Data characteristics

We observe from Figure 2 that the German futures contracts mostly trade higher than their Nordic counterparts for both maturities. See also Table 3 for more specific details on the futures prices and spread series. Moreover, we note that the trading ranges for the front-quarter futures and corresponding spread exceed those of the front-year equivalents. Considering this, we also observe that the standard deviations of the futures prices and spreads are higher for the front-quarter contracts compared to front-year contracts.

Table 3: Futures prices and spread characteristics

Front-quarter	T	T^-	Min	Max	Mean	Median	Std.dev
Spread	2264	371	-28.87	20.56	4.62	4.76	6.04
German	2264	-	21.82	65.38	40.65	38.35	10.06
Nordic	2264	-	13.82	84.27	36.02	34.85	11.27
Front-year	**T**	**T^-**	**Min**	**Max**	**Mean**	**Median**	**Std.dev**
Spread	2288	39	-9.99	18.10	6.46	6.09	3.56
German	2288	-	20.90	60.70	40.54	37.72	9.98
Nordic	2288	-	16.30	61.44	34.08	34.65	8.80

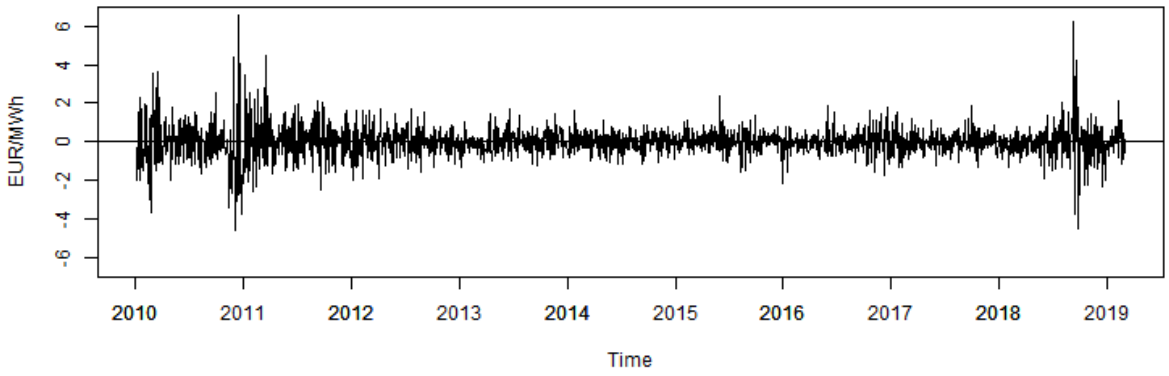
T denotes the total number of observations and T^- denotes the number of observations below zero.

The P&L of the front-quarter futures prices are plotted in Figure 3 and their characteristics are summarized in Table 4. We observe that the P&L of all contracts have a mean close to zero, which appears to be constant over time. The Nordic P&L has a higher standard deviation than the German P&L for both front-quarter and front-year contracts. Moreover, we note that the standard deviation of the spread P&L is lower than that of the Nordic P&L. As the spread is a linear combination of the German and Nordic futures (see (1)), this indicates that the futures prices are correlated. Also, from a visual inspection of the plots of the P&L in Figure 3 the series appear to exhibit time-varying volatility. The same properties are observable for the front-year contracts, given in Appendix F.1.

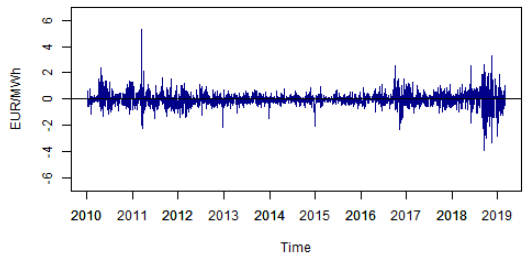
Table 4: P&L characteristics

Front-quarter	T	Min	Max	Mean	Median	Std.dev
Spread	2264	-4.55	6.60	-0.01	-0.02	0.78
German	2264	-3.89	5.27	-0.01	-0.03	0.54
Nordic	2264	-7.10	5.95	0.00	0.00	0.85
Front-year	**T**	**Min**	**Max**	**Mean**	**Median**	**Std.dev**
Spread	2288	-3.21	5.56	0.00	0.00	0.40
German	2288	-4.70	2.63	0.00	-0.01	0.44
Nordic	2288	-3.00	2.88	0.01	0.00	0.50

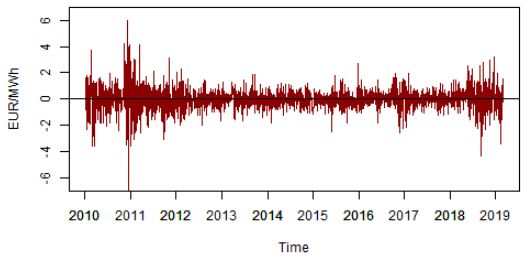
T denotes the total number of observations



(a) Daily P&L of the front-quarter spread



(b) Daily P&L of German front-quarter contracts



(c) Daily P&L of Nordic front-quarter contracts

Figure 3: Daily P&L of spread, German and Nordic front-quarter contracts

To establish an impression of the spreads' historical tail risk, we provide empirical quantiles of the front-quarter futures in Table 5. The corresponding quantiles for front-year contracts are found in Appendix F.2, and the 95% quantiles for the front-quarter and front-year contracts are given in Appendix F.3 and F.4, respectively. We also include the mean conditional on the observations exceeding the quantile. This is calculated as the arithmetic average of the daily P&Ls that exceed the 99% quantile and interpreted as the expected loss given that the P&L exceeds the given empirical quantile. The quantile and the mean of the observations beyond the quantile are analogous to historical VaR and ES. The more volatile front-quarter futures series and spreads have more extreme tails than the front-year equivalents, which

coincide with our previous observations regarding volatility. Furthermore, the historical risk appears to be have been greater for the Nordic futures than for the German futures. In Table 5 and Appendix F.2, we provide an overview of how the excess kurtosis and skew has changed over time. The measures are calculated on a per-year basis, and can be visually assessed by inspecting the histograms of the P&L per year, which we provide in Appendix F.9 and F.10.

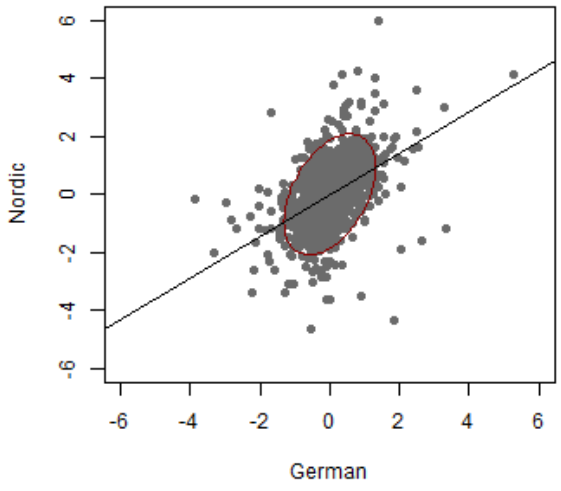
Table 5: Empirical properties of the P&L of front-quarter contracts

Contract (*Position*)	2010	2011	2012	2013	2014	2015	2016	2017	2018
99% empirical quantiles									
Spread (*Long*)	-3.54	-2.39	-1.72	-1.11	-1.11	-1.32	-1.42	-1.28	-2.54
Spread (*Short*)	3.85	2.64	1.61	1.43	0.88	1.13	1.55	1.32	3.34
German (*Long*)	-1.14	-1.39	-1.37	-0.69	-0.61	-0.55	-1.51	-0.92	-2.93
German (*Short*)	1.45	2.32	1.09	0.64	0.6	0.59	1.54	1.14	2.31
Nordic (*Long*)	-3.6	-2.86	-1.71	-1.46	-1.12	-1.15	-2.04	-1.25	-2.6
Nordic (*Short*)	4.05	3.08	1.48	1.38	1.14	1.48	1.63	1.2	2.71
Mean conditional on 99% quantile exceedance									
Spread (*Long*)	-3.99	-2.53	-1.93	-1.26	-1.14	-1.75	-1.63	-1.32	-3.66
Spread (*Short*)	5.03	3.58	1.67	1.53	1.21	1.64	1.76	1.6	4.79
German (*Long*)	-1.21	-1.78	-1.65	-1.06	-1.11	-0.63	-1.9	-1.04	-3.4
German (*Short*)	1.9	3.69	1.19	0.67	0.83	0.74	2.05	1.21	2.83
Nordic (*Long*)	-5.85	-3.24	-1.88	-1.57	-1.4	-1.83	-2.29	-1.69	-3.53
Nordic(*Short*)	4.76	3.6	1.92	1.73	1.18	1.99	1.75	1.65	3.01
Excess kurtosis									
Spread	3.96	1.49	0.59	0.56	0.18	3.74	1.56	0.82	9.76
German	2.15	16.69	2.38	1.51	12.01	1.14	3.12	0.21	3.11
Nordic	3.99	1.43	0.56	0.69	0.17	3.82	2.32	1.55	2.43
Skew									
Spread	0.65	0.36	-0.04	0.42	-0.06	-0.15	0.25	0.24	0.92
German	0.95	2.53	-0.36	-0.38	-1.4	0.31	0.22	0.26	-0.63
Nordic	-0.31	0.27	0	0.04	-0.12	0.36	-0.4	-0.29	-0.42

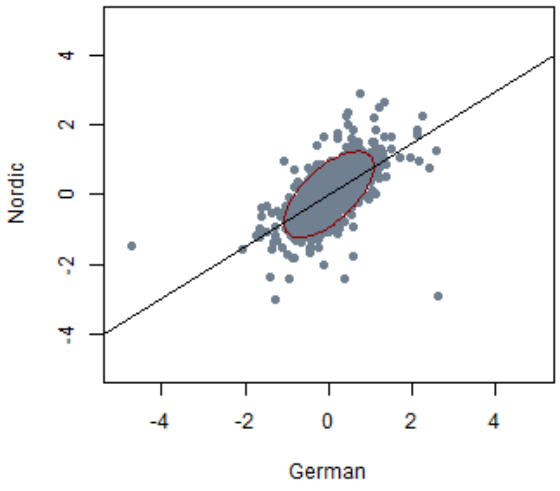
2019 not included as datasample only runs the first two months

4.6 Correlation dynamics

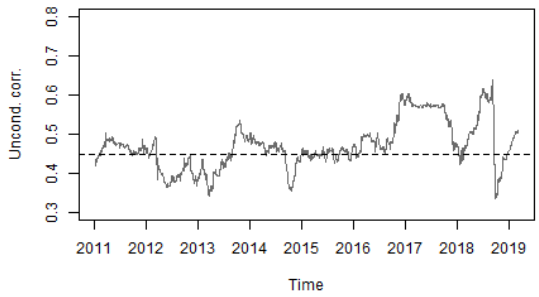
Figures 4a and 4b display scatter plots of the Nordic futures P&L and German futures P&L. By visual inspection, there seems to be a positive relationship between the two. We also include a 95% contour line from a zero-mean bivariate normal distribution, with unconditional variances and covariance estimated from the series. If the joint distribution were bivariate normal, around 5% of the data should be outside the red ellipse. Moreover, there seems to be heavy clustering around the middle, and we observe that there are more extreme observations than one would expect from a normal distribution. Figure 4c and 4d show the rolling unconditional correlation for 250 trailing trading days. These measures fluctuate across our data sample and thus motivates the use of a correlation model that captures these time-varying dynamics. Note especially the high correlation through 2017 as prices increased from the low levels in 2016. The same feature is observable throughout 2018 until the extreme movements that took place on and after 10 September 2018. Lastly, we observe that the correlation for the front-quarter series is lower than for the front-year series.



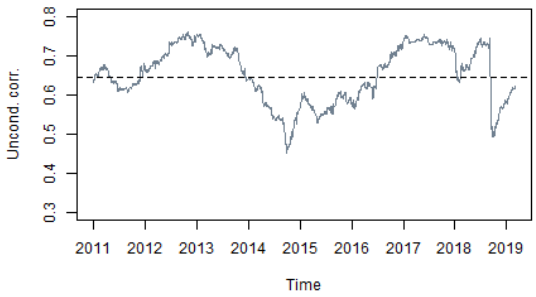
(a) Scatterplot of German and Nordic front-quarter P&L with regression line and the 95% contour line of a normal distribution



(b) Scatterplot of German and Nordic front-year P&L with regression line and the 95% contour line of a normal distribution



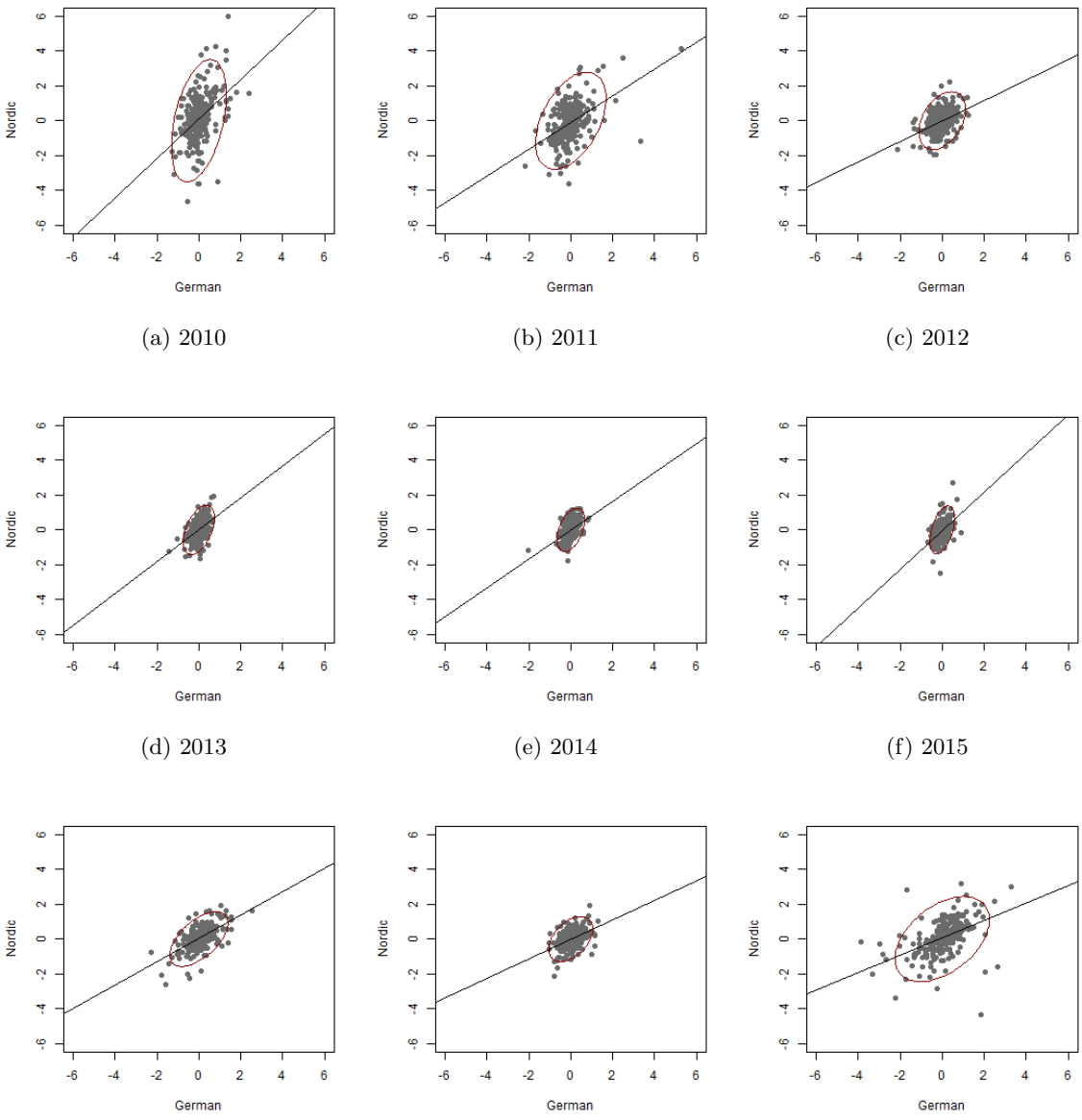
(c) 250 day trailing unconditional correlation, front-quarter contracts (gray). Unconditional correlation for sample (dashed)



(d) 250 day trailing unconditional correlation, front-year contracts (gray). Unconditional correlation for sample (dashed)

Figure 4: Summary of correlation dynamics for daily P&L of German and Nordic front-quarter and front-year contracts

In Figure 5, we give a panel of scatterplots per year in the period 2010 to 2018 the for Nordic and German front-quarter P&L. The most extreme observations stem from mainly 2010, 2011 and 2018 when the price level was higher. Whereas the data points are more clustered for the years where prices were, i.e. for the years 2012 through 2016. The corresponding panel for front-year contracts is found in Appendix F.6.



(a) 2010 (b) 2011 (c) 2012

(d) 2013 (e) 2014 (f) 2015

(g) 2016 (h) 2017 (i) 2018

Figure 5: Scatterplots of daily P&L of Nordic and German front-quarter contracts per calendar year, with a Nordic on German regression line. Red ellipse is 95% contour of a bivariate normal distribution

4.7 Descriptive statistics

We provide descriptive statistics of each individual series in Table 6. The distributions of the P&Ls of the spread, Nordic and German futures are leptokurtic, meaning that the distributions have higher peaks and heavier tails than the normal density with the same variance (Alexander, 2008a). This is visible from the histograms and quantile-quantile-plots (QQ-plots) in Figure 6, where the solid lines belong to the normal distribution. Both distributions of the spread P&L show some skewness. However, for the individual series, only the German front-quarter P&L displays skewness. Applying the method of Jarque and Bera (1980), we obtain a test statistic which follows a χ^2 distribution with 2 degrees of freedom, with a critical value of 5.99 under the 5% significance level. We reject normality according to the JB test for all series.

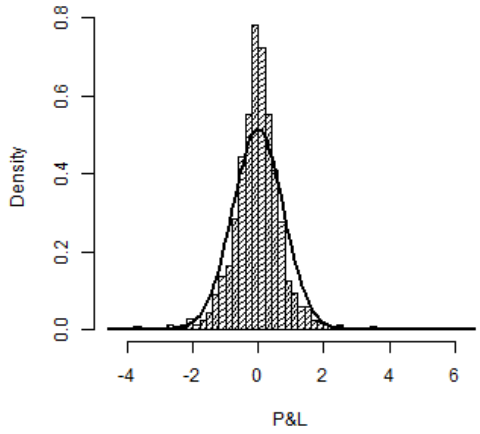
Prior to modelling the dynamics of correlation and volatility of the futures prices, stationarity must be verified with unit root tests. We conduct the Augmented Dickey Fuller tests with 1 lag (DF) and 10 lags (ADF) (Dickey & Fuller, 1981). The critical value is -2.86 under the 5% confidence level, which is exceeded for both for 1 and 10 lags. Thus, the series are stationary.

We test for autocorrelation using the Ljung-Box test (LB) in 10 lags (Ljung & Box, 1978). The test statistic follows a χ^2 distribution with degrees of freedom equal to the number of lags included. The test statistic in 10 lags displays a value larger than the critical value of 18.31 under the 5% significance level. Thus, the null hypothesis is rejected and autocorrelation is asserted. This is true for the Nordic and German futures and the spread for both front-quarter and front-year contracts. The test statistics indicate that the front-year futures have more autocorrelation in squared daily P&L than the front-quarter equivalents across all contracts.

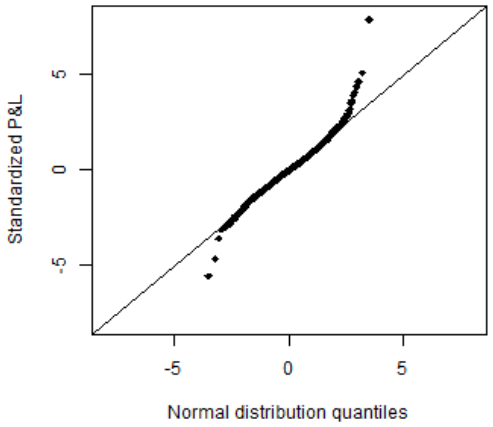
Table 6: Descriptive statistics

Front-quarter P&L series	Skewness	Excess kurtosis	JB	DF	ADF	LB
Spread	0.57	8.60	7 110*	-45.67*	-15.00*	1 091*
German	0.43	10.24	9 977*	-43.10*	-13.86*	697*
Nordic	-0.10	6.37	3 840*	-43.97*	-14.84*	1 465*
Front-year P&L	**Skewness**	**Kurtosis**	**JB**	**DF**	**ADF**	**LB**
Spread	1.00	21.22	43 391*	-46.70*	-14.85*	451*
German	-0.05	9.08	7 871*	-44.35*	-14.56*	558*
Nordic	0.04	4.07	1 582*	-44.61*	-14.34*	1 140*

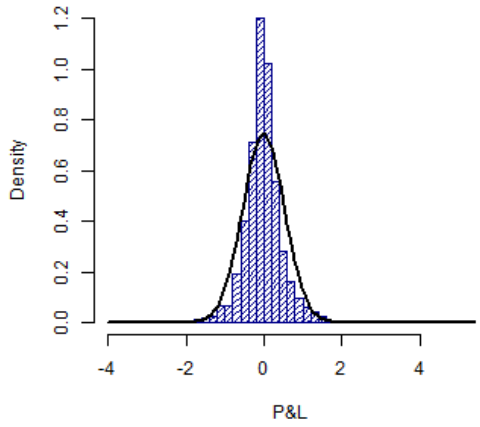
*Indicates rejection of the null hypothesis at the 5% significance level.



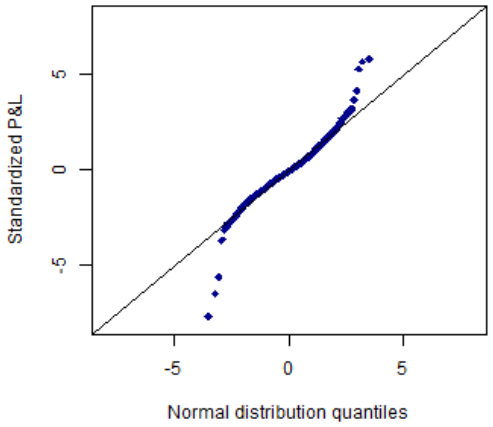
(a) Histogram of daily P&L of the spread



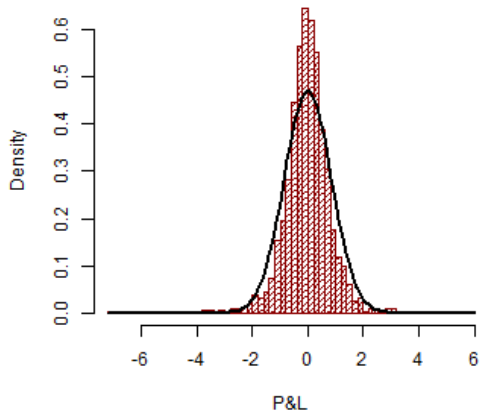
(b) QQ-plot of daily P&L of the spread against quantiles from the normal distribution



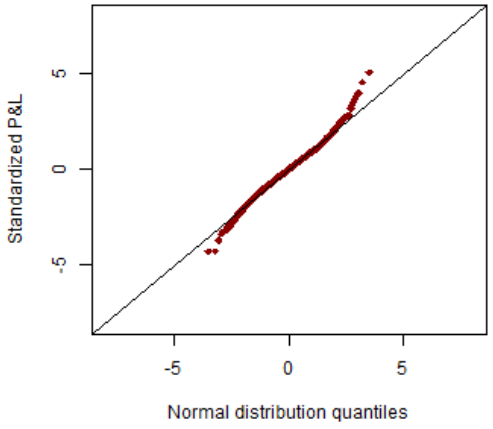
(c) Histogram of daily P&L of German contracts



(d) QQ-plot of daily P&L of the German contracts against quantiles from the normal distribution



(e) Histogram of daily P&L of Nordic contracts



(f) QQ-plot of daily P&L of the Nordic contracts against quantiles from the normal distribution

Figure 6: Histogram and QQ-plot of daily P&L, front-quarter. Spread (black), German (blue) and Nordic (red). The daily P&Ls are standardized using GARCH(1,1)-n

5 Univariate models

5.1 Conditional variance models

The daily P&L of the spread time series, as stated in (2), can be expressed in the following way:

$$P\&L_t = \mu_t + \varepsilon_t = \mu_t + \sigma_t z_t, \quad z_t \sim \text{i.i.d. } D(0,1) \tag{6}$$

where ε_t can be referred to as the market shock (Alexander, 2008a), μ_t is the conditional mean and σ_t^2 is the conditional variance. The random variable z_t is an innovation term assumed to be identically and independently distributed (i.i.d) according to some distribution D, with mean equal to zero and variance equal to one. In this study, we consider the following univariate distributions: normal distribution, Student t distribution, skewed Student t distribution and EVT.

In further analysis, we assume that the conditional mean of daily P&L, μ_t, is constant and equal to zero. The mean does not appear to be time-varying and we cannot statistically reject a zero mean. In general, on short horizons such as daily, returns or P&L, is dominated by the standard deviation (Christoffersen, 2011). Hence, an assumption of mean equal to zero is not likely to affect results on VaR and ES estimation. We aim for parsimonious models, and argue that a conditional mean of zero is a reasonable assumption for daily P&L of both front-quarter and front-year spreads. Consequently, (6) reduces to:

$$\varepsilon_t = \sigma_t z_t, \quad z_t \sim \text{i.i.d. } D(0,1) \tag{7}$$

In the next sections, we present the methods applied for calculating the conditional variance σ_t^2.

5.1.1 RiskMetrics

RiskMetrics (1996) is an exponentially weighted moving average (EWMA) model where weights on past squared returns on a financial asset decline exponentially as we move backward in time. Applied to our data, we substitute returns with daily $P\&L$ of the spread. RiskMetrics captures the empirical distribution of daily $P\&L$, but represents a random walk process for variance and thus has no mean-reversion. Estimates are changing in time merely due to the exponential weighting (Alexander, 2008a). However, an EWMA model for conditional variance can be specified as:

$$\varepsilon_t = \sigma_t z_t, \quad z_t \sim \text{i.i.d. } N(0,1) \tag{8}$$

$$\sigma_{t+1}^2 = (1-\lambda)\varepsilon_t^2 + \lambda\sigma_t^2 \tag{9}$$

where $N(0,1)$ denotes the standard normal distribution. According to the RiskMetrics specification for daily data, we set $\lambda = 0.94$.

This is a restricted version of GARCH(1,1) with the normal distribution which will be introduced shortly. The first term in (9), $(1-\lambda)\varepsilon_t^2$, is interpreted as the intensity of reaction of volatility to market events, while the second term, $\lambda\sigma_t^2$, determines persistence in volatility irrespective of market events (Alexander, 2008a). RiskMetrics has the advantage of no parameter estimation and is extensively used in practice. On the other hand, an important limitation is the fact that the model does not capture time-varying volatility.

5.1.2 GARCH(1,1)

In the data analysis, we found that the volatility in daily P&L of the spread is time-varying. For this reason, we need models that capture volatility clustering. The GARCH model proposed by Bollerslev (1986) is a generalization of the autoregressive conditional heteroscedastic (ARCH) model developed by Engle (1982). The dynamic behavior of the conditional variance is given by the following expression:

$$\sigma_{t+1}^2 = \omega + \alpha\varepsilon_t^2 + \beta\sigma_t^2 \tag{10}$$

The (1,1) specification denotes that one length of ARCH lag (α) and one length of GARCH lag (β) is used. The unconditional, or long-run average variance, is defined as follows:

$$\sigma^2 \equiv \mathbb{E}\left[\sigma_{t+1}^2\right] = \frac{\omega}{(1-\alpha-\beta)} \tag{11}$$

With $w > 0$, $\alpha, \beta \geq 0$ and $\alpha + \beta < 1$, this model represents a non-negative and stationary process. Over time, the conditional variance σ_{t+1}^2 reverts to the long-run average value. The parameters in GARCH(1,1) are estimated by maximum likelihood estimation (MLE), based on a distributional assumption. The procedure for the normal distribution, standardized Student t distribution and skewed Student t distribution is explained in Appendix C. The error parameter α measures how conditional volatility reacts to market shocks, while the lag parameter β is a measure of persistence in conditional volatility, irrespective of what happens in the market. Furthermore, $\alpha + \beta$ determines the rate of convergence of the conditional variance to the long-run average variance (Alexander, 2008a).

5.1.3 GJR-GARCH(1,1)

A GARCH(1,1) model assumes symmetric variance responses from positive and negative market shocks. However, it is often the case in financial markets that variance responds asymmetrically to positive and negative shocks, and this gives rise to the need for extended models. A model which copes with variance asymmetry is the GJR-GARCH model proposed by Glosten, Jagannathan, and Runkle (1993). The GARCH(1,1) model is extended by adding one parameter as well as an indicator variable, and augments (positively or negatively) the volatility response from only negative market shocks (Alexander, 2008a):

$$\sigma_{t+1}^2 = \omega + \alpha\varepsilon_t^2 + \gamma H_t \varepsilon_t^2 + \beta\sigma_t^2 \tag{12}$$

where

$$H_t = \begin{cases} 1, & \text{if } \varepsilon_t < 0 \\ 0, & \text{if } \varepsilon_t \geq 0 \end{cases} \tag{13}$$

and γ is the asymmetric effect parameter. Negative shocks impact variance by $(\alpha + \gamma)$ while positive shocks on the other hand impact variance by α. Depending on the sign of γ, the variance effect of a negative shock can be greater than or smaller than the effect of a positive shock. Again, the parameters in the model are found through MLE based on a distributional assumption. The unconditional, or long-run average variance is given by:

$$\sigma^2 \equiv \mathbb{E}\left[\sigma_{t+1}^2\right] = \frac{\omega}{(1-\alpha-\beta-\gamma\kappa)} \tag{14}$$

where κ is the expected value of z_t below zero. In the case of symmetric distributions, such as the normal distribution or the Student t distribution, the value of κ is simply 0.5. The rate of convergence is given by $\alpha + \beta + \gamma\kappa$, and for stationarity we require this to be less than 1. The conditions for non-negativity

will be $\omega, \alpha > 0$, $\beta \geq 0$ and $\alpha + \gamma \geq 0$. That is, the model is still valid if $\gamma < 0$, provided that $\alpha + \gamma \geq 0$ (Brooks, 2014).

There exist other extensions to GARCH(1,1) that are commonly used to model the leverage effect, such as the E-GARCH (exponential GARCH) model advocated by Nelson (1991), and the A-GARCH (asymmetric GARCH) model initially proposed by Engle (1990) and subsequently discussed by Engle and Ng (1993). We have chosen to focus on GJR-GARCH due to its simplicity.

5.2 Univariate Value-at-Risk and Expected Shortfall models

With conditional volatility models in place, we can construct univariate VaR and ES estimates. The probability of experiencing a loss greater than the VaR estimate on a given day is p. VaR can be defined mathematically as:

$$VaR^p_{t+1} = \sup_{x_{t+1}} \Big[x_{t+1} \mid P(X_{t+1} \geq x_{t+1}) \geq p \Big] \tag{15}$$

where x_{t+1} represents the $(1-p)$th quantile of the distribution of the loss function X_{t+1}, conditional on the information available at time t. The loss function is given by the following expression:

$$X_{t+1} = \begin{cases} -P\&L_{t+1}, & \text{if long position} \\ P\&L_{t+1}, & \text{if short position} \end{cases} \tag{16}$$

Throughout the rest of the study, we consider the 5%, 1%, 95% and 99% quantiles of the conditional P&L distribution. The former two and the latter two quantiles correspond to the 95% and 99% quantiles of the conditional loss distribution for long and short positions, respectively. We will focus solely on a risk horizon, meaning the number of trading days over which the VaR is measured, of 1 trading day. This is denoted by $t + 1$.

A key shortcoming of VaR is that this risk measure tells nothing about the magnitude of losses that could be incurred if the VaR estimate is exceeded on a given day. ES is a metric which addresses this limitation, and it can be expressed mathematically in the following way:

$$ES^p_{t+1} = \mathbb{E}\big[X_{t+1} \mid X_{t+1} > VaR^p_{t+1} \big] \tag{17}$$

This means that ES is the expected value of the loss, given that it exceeds the VaR estimate. In the following sections, we provide a description of the applied methods and how they are used to obtain VaR and ES estimates.

5.2.1 RiskMetrics

When the conditional distribution of market shocks is normal distributed as stated in (8), daily VaR with coverage rate p is given by the following expression:

$$VaR^p_{t+1} = \sigma_{t+1} \Phi^{-1}_{1-p} \tag{18}$$

where σ_{t+1} is calculated with the RiskMetrics model in (9) and Φ denotes the cumulative density function of the standard normal distribution. Daily ES is given by:

$$ES_{t+1}^{p} = \sigma_{t+1}\frac{\phi(\Phi_{1-p}^{-1})}{p} \tag{19}$$

where ϕ denotes the density function of the standard normal distribution. See Christoffersen (2011) for the derivation of this result. As the standard normal distribution is symmetric around zero, long and short VaR and ES estimates for the same coverage rate will be equal. We include the RiskMetrics model mainly for purposes of comparison with more sophisticated models.

5.2.2 GARCH: Normal distribution

We have employed two types of GARCH models to calculate the conditional volatility σ_{t+1} in this study. These two models are GARCH(1,1) and GJR-GARCH(1,1). In the case of the normal distribution, we use the following expression:

$$\varepsilon_t = \sigma_t z_t, \quad z_t \sim \text{i.i.d. } N(0,1) \tag{20}$$

where σ_t is given by (10) and (12) for GARCH(1,1) and GJR-GARCH(1,1), respectively. When the conditional distribution of innovations to daily P&L is standard normal distributed, the normal log likelihood function is used in MLE. See Appendix C.1.1 for the log likelihood function. Daily VaR and ES estimates with coverage rate p are obtained by (18) and (19), respectively.

5.2.3 GARCH: Student t distribution

The daily P&L of the spreads are not well described by the normal distribution, and therefore this is an inaccurate distributional assumption. The conditional distribution of daily P&L is leptokurtic, and therefore we need distributional assumptions which better capture the observed excess kurtosis. The GARCH model with Student t distribution introduced by Bollerslev (1987), assumes the conditional distribution of market shocks to be Student t distributed. We modify the expression in (20) as follows:

$$\varepsilon_t = \sigma_t z_t, \quad z_t \sim \text{i.i.d. } \tilde{t}(\upsilon) \tag{21}$$

with σ_t given in (10) and (12) for GARCH(1,1) and GJR-GARCH(1,1), respectively. $\tilde{t}(\upsilon)$ denotes the standardized Student t distribution with mean equal to zero, variance equal to one and υ degrees of freedom. υ is an additional parameter which is estimated along with the parameters in the conditional variance equation. The estimation is carried out according to MLE with the Student t log likelihood function, and from this we obtain the conditional volatility σ_{t+1}. See Appendix C.1.2 for the log likelihood function. With the GARCH models in place, we calculate daily VaR estimates in the following manner:

$$VaR_{t+1}^{p} = \sigma_{t+1}\tilde{t}_{1-p}^{-1}(\upsilon) \tag{22}$$

where $\tilde{t}_{1-p}^{-1}(\upsilon)$ denotes quantile $1-p$ of the standardized Student t distribution with υ degrees of freedom. Daily ES is given by the following expression (Alexander, 2008c):

$$ES_{t+1}^{p} = \sigma_{t+1}\frac{\upsilon - 2 + \left(\tilde{t}_{1-p}^{-1}(\upsilon)\right)^2}{p(\upsilon-1)} f_{\tilde{t}}\left(\tilde{t}_{1-p}^{-1}(\upsilon);\upsilon\right) \tag{23}$$

where $f_{\tilde{t}}(*;\upsilon)$ denotes the density function of the standardized Student t distribution with υ degrees of freedom. We require $\upsilon > 2$ for the distribution to be well defined. As the standardized Student t distribution is symmetric, VaR and ES estimates for the same coverage rate p will be equal for long and short positions.

5.2.4 GARCH: Skewed Student t distribution

The symmetric Student t distribution allows for excess kurtosis, but not for skewness. We find that the distribution of daily P&L shows some skewness. By using a skewed, or asymmetric, Student t distribution, the GARCH models can take into account both skewness and excess kurtosis. We use the extension of the Student t distribution proposed by Hansen (1994), where market shocks are described by the following expression:

$$\varepsilon_t = \sigma_t z_t, \quad z_t \sim \text{i.i.d. } t_{skew}(\upsilon, \xi) \tag{24}$$

with σ_t given in (10) and (12) for GARCH(1,1) and GJR-GARCH(1,1), respectively. $t_{skew}(\upsilon, \xi)$ is the skewed Student t distribution with mean equal to zero, variance equal to one and parameters υ and ξ. By setting $\xi = 0$, this distribution reduces to the standardized Student t distribution (Hansen, 1994).

Compared to the GARCH models with normal distribution, there are two additional parameters, υ and ξ, to be estimated along with the parameters in the conditional variance equation. To obtain parameter estimates, we use a quasi maximum likelihood estimation (QMLE) procedure as described in Appendix C.1.3. Daily VaR estimates with GARCH skewed Student t is obtained by (Christoffersen, 2011):

$$VaR^p_{t+1} = \sigma_{t+1} F^{-1}_{skew}(p; \upsilon, \xi) \tag{25}$$

where $F^{-1}_{skew}(p; \upsilon, \xi)$ is the pth quantile of the skewed Student t distribution which is given by:

$$F^{-1}_{skew}(p; \upsilon, \xi) = \begin{cases} \frac{1}{B}\left[(1-\xi)\tilde{t}^{-1}_{\frac{p}{1-\xi}}(\upsilon) - A\right], & \text{if } p < \frac{1-\xi}{2} \\ \frac{1}{B}\left[(1+\xi)\tilde{t}^{-1}_{\frac{p+\xi}{1+\xi}}(\upsilon) - A\right], & \text{if } p \geq \frac{1-\xi}{2} \end{cases} \tag{26}$$

As before, $\tilde{t}^{-1}_p(\upsilon)$ is the pth quantile of the standardized Student t distribution. The rest of the notation is defined as follows:

$$A = 4\xi C \frac{\upsilon - 2}{\upsilon - 1}, \quad B = \sqrt{1 + 3\xi^2 - A^2}, \quad C = \frac{\Gamma(\frac{\upsilon+1}{2})}{\Gamma(\frac{\upsilon}{2})\sqrt{\pi(\upsilon-2)}} \tag{27}$$

where $\Gamma(*)$ is the Gamma function, and we require $\upsilon > 2$ and $-1 < \xi < 1$. Daily ES is given by the following expression (Christoffersen, 2011):

$$ES^p_{t+1} = \sigma_{t+1} ES_{skew}(p) \tag{28}$$

$$ES_{skew}(p) = \begin{cases} \frac{C(1-\xi)^2}{Bp}\left[\left[1+\frac{1}{\upsilon-2}\left(\frac{BF^{-1}_{skew}(p;\upsilon,\xi)+A}{1-\xi}\right)^2\right]^{\frac{1-\upsilon}{2}}\frac{\upsilon-2}{1-\upsilon}\right] \\ \quad -\frac{AC(1-\xi)}{Bp}\frac{\sqrt{\pi(\upsilon-2)}\Gamma(\frac{\upsilon}{2})}{\Gamma(\frac{\upsilon+1}{2})}\tilde{t}_\upsilon\left(\frac{BF^{-1}_{skew}(p;\upsilon,\xi)+A}{1-\xi}\right), \quad \text{if } F^{-1}_{skew}(p;\upsilon,\xi) < -\frac{A}{B} \\ \quad -\frac{C(1+\xi)^2}{Bp}\left[\left[1+\frac{1}{\upsilon-2}\left(\frac{BF^{-1}_{skew}(p;\upsilon,\xi)+A}{1+\xi}\right)^2\right]^{\frac{1-\upsilon}{2}}\frac{\upsilon-2}{1-\upsilon}\right] \\ -\frac{AC(1+\xi)}{Bp}\frac{\sqrt{\pi(\upsilon-2)}\Gamma(\frac{\upsilon}{2})}{\Gamma(\frac{\upsilon+1}{2})}\left[1-\tilde{t}_\upsilon\left(\frac{BF^{-1}_{skew}(p;\upsilon,\xi)+A}{1+\xi}\right)\right], \quad \text{if } F^{-1}_{skew}(p;\upsilon,\xi) \geq -\frac{A}{B} \end{cases} \tag{29}$$

where $\tilde{t}_\upsilon(*)$ is the cumulative distribution function of the standardized Student t distribution and the rest of the notation is as defined earlier. As opposed to for the normal distribution and symmetric Student t distribution, VaR and ES estimates for the same coverage rate p can be different for long and short positions with skewed Student t distribution. We note that there exist other similar skewed Student t densities, such as the one introduced by Fernández and Steel (1998) and developed by Lambert and Laurent (2001).

5.2.5 GARCH: Extreme value theory

In this study, we aim to investigate tail risk in German-Nordic electricity futures spreads. To properly examine this topic, we need to model the extreme tails of the of the conditional distribution of daily P&L of the spread. In order to do this, we use methods from EVT, which relates to a class of distributions that are derived from considering the extreme values in a sample (Alexander, 2008b). Applications of EVT in financial risk management is discussed by among others McNeil (1999).

Before applying EVT, we must remove variance dynamics. We rewrite (7) and define standardized daily P&L as below:

$$z_t = \frac{P\&L_t}{\sigma_t} \tag{30}$$

where $z_{t+1} \sim$ i.i.d. $D(0,1)$. To calculate the conditional volatility σ_{t+1}, we use a GARCH(1,1) model with normal distributed shocks, as given in (10) together with (20). We choose this model because it captures volatility clustering but it does not accurately describe the tails of the daily P&L distribution. Hence, this is a reasonable starting point for testing the effects of an EVT-based approach.

After applying the conditional volatility model, we have a complete distribution for z_t. However, our application of EVT is concerned only with the extreme tails of the distribution. We define this to be the $T_u = 50$ most extreme observations in the distribution of z_t. Furthermore, u is defined to be the threshold value separating the extreme tail from the rest of the distribution of z_t. The choice of $T_u = 50$, and its associated u, which according to McNeil and Frey (2000) is the most critical implementation issue in EVT, is made as an attempt to balance bias and variance. By setting u too large, very few observations are left in the tail and estimates become noisy. On the contrary, if u is set too small, the data may not conform sufficiently well to the Generalized Pareto Distribution, which we introduce shortly. Consequently, this will generate biased estimates (Christoffersen, 2011).

An important result in EVT states that the extreme tails of various distributions can be approximately described by the Generalized Pareto Distribution (GPD) (Christoffersen, 2011). We employ this result

and approximate the distribution of all extreme tail observations y beyond u by:

$$GPD(y;\eta,\beta) = \begin{cases} 1-(1+\frac{\eta y}{\beta})^{-\frac{1}{\eta}} & \text{, if } \eta > 0 \\ 1-e^{-\frac{y}{\beta}} & \text{, if } \eta = 0 \end{cases} \tag{31}$$

with scaling parameter $\beta > 0$ and $y \geq u$. The tail index parameter η controls the shape of the tail distribution. The normal distribution has exponential tails and $\eta = 0$, while for heavy-tailed distributions such as Student t, the tail index parameter is positive. A more detailed discussion of the GPD and the threshold method can be found in McNeil and Frey (2000).

It is possible to use MLE to estimate the GPD. However, our data suggests that the tail index parameter η is strictly positive given that we have excess kurtosis and it is therefore possible to approximate the GPD by the Hill estimator (Hill, 1975):

$$F(y) = 1 - cy^{-\frac{1}{\eta}} \approx 1-(1+\frac{\eta y}{\beta})^{-\frac{1}{\eta}} = GPD(y;\eta,\beta) \tag{32}$$

The conditional distribution of observations y beyond the threshold u is given by:

$$f(y \mid y > u) = \frac{f(y)}{1-F(u))} \; for \; y > u \tag{33}$$

From the definition of $F(y)$ in (32), we can obtain the density function as stated below:

$$f(y) = \frac{\partial F(y)}{\partial y} = \frac{1}{\eta} c y^{-\frac{1}{\eta-1}} \tag{34}$$

From this we can construct the likelihood function for all observations y_i larger than the threshold u:

$$L = \prod_{i=1}^{T_u} \frac{f(y_i)}{1-F(u)} = \prod_{i=1}^{T_u} \frac{1}{\eta} \frac{c y_i^{-\frac{1}{\eta-1}}}{c u^{-\frac{1}{\eta}}} \text{ for } y_i > u \tag{35}$$

The log likelihood function is therefore:

$$ln(L) = \sum_{i=1}^{T_u} \Big(-ln(\eta) - (\frac{1}{\eta}+1)ln(y_i) + \frac{1}{\eta} ln(u) \Big) \tag{36}$$

By taking the derivative with respect to η and setting it equal to zero, we obtain the Hill estimator of the tail index parameter:

$$\eta = \frac{1}{T_u} \sum_{i=1}^{T_u} ln\Big(\frac{y_i}{u}\Big) \tag{37}$$

The c parameter can be estimated by ensuring that the fraction of observations beyond the threshold is accurately captured by the density as below:

$$F(u) = 1 - cu^{-\frac{1}{\eta}} = 1 - \frac{T_u}{T} \tag{38}$$

where T denotes the total number of observations in the sample. Solving for c and inserting into (32) gives closed form estimates, requiring no numerical optimization, of the cumulative distribution function for observations beyond u:

$$F(y) = 1 - cy^{-\frac{1}{\eta}} = 1 - \frac{T_u}{T}\Big(\frac{y}{u}\Big)^{-\frac{1}{\eta}} \tag{39}$$

To get quantiles, we compute the inverse cumulative distribution F^{-1}_{1-p}, which implicitly is defined by:

$$F(F^{-1}_{1-p}) = 1 - p \tag{40}$$

This means that there is only a probability of p of getting a standardized daily P&L worse than the quantile. From (39), we can solve for the quantile. Ultimately, we can construct daily VaR estimates by combining the EVT methodology with the conditional volatility model in (10) in the following manner:

$$VaR^p_{t+1} = \sigma_{t+1} F^{-1}_{1-p} = \sigma_{t+1} u \Big[\frac{p}{T_u/T}\Big]^{-\eta} \tag{41}$$

Daily ES can be computed using:

$$ES^p_{t+1} = -\sigma_{t+1} \frac{u}{\eta - 1} \Big[\frac{p}{T_u/T}\Big]^{-\eta} \tag{42}$$

See Christoffersen (2011) for the derivation of this result. As opposed to for the other methods used to obtain VaR and ES estimates, where we use confidence levels equal to 95% and 99%, we can only a confidence level equal to 99% with this EVT methodology. This is due to the choice of $T_u = 50$, which implies that we only model the 50 most extreme observations. McNeil and Frey (2000) find that, although the Hill estimator is generally the most efficient estimator of η, it does not provide the most efficient nor the most stable quantile estimator. This is a possible weakness compared to the GPD method with MLE.

5.2.6 Markov switching GARCH: Normal and Student t distributions

The Markov switching GARCH model set forth by Haas (2004) allows for modelling conditional variance by defining separate variance regimes. The process governing the regimes and the possible states of the model follows a homogeneous Markov chain. It is thus possible to estimate the dynamics of several regimes, e.g. by GARCH models, and model the unobserved transitions between them. Volatility in daily P&L of German-Nordic electricity futures spreads appears to be relatively high in periods of unusual market events, while it appears to be lower in normal market conditions. This motivates the use of Markov switching GARCH in our study. By employing this model, we can assess if using two volatility regimes with transition probabilities of moving from one regime to the next can enable more accurate VaR and ES estimation. A general Markov switching GARCH specification in the context of our study can be expressed as:

$$\varepsilon_t \mid s_t = k, I_{t-1} \sim N(0, \sigma^2_{t,k}) \tag{43}$$

$$\varepsilon_t \mid s_t = k, I_{t-1} \sim t(0, \sigma^2_{t,k}, \upsilon) \tag{44}$$

where (43) is used with a normal distribution assumption and (44) is used with a Student t distribution assumption. The information set up to $t-1$ is denoted by I_{t-1}, whereas the integer state variable s_t indicates the current regime. We use a Markov switching GARCH model with 2 regimes and so s_t is defined on $\{1, 2\}$, and evolves according to a first-order ergodic homogeneous Markov chain. The transition probability matrix of the Markov process is $\boldsymbol{P}$ with dimensions 2×2 and elements $p_{i,j} = P[s_t = j \mid s_{t-1} = i]$. Similar to Haas (2004), we assume that the conditional variance follows a GARCH(1,1) model. We rewrite (10) and define the conditional variance process for each regime k:

$$\sigma^2_{t+1,k} = \omega_k + \alpha_k \varepsilon^2_t + \beta_k \sigma^2_{t,k} \tag{45}$$

The parameters are calculated through an MLE procedure which also estimates the transition probability matrix $\boldsymbol{P}$. In the case of the Student t distribution, the degrees of freedom parameter υ is also estimated in the MLE. Like Sampid, Hasim, and Dai (2018) and Cifter (2013), we hold υ fixed across the two regimes. See Appendix C.2 for an explanation of the MLE procedure. By matrix notation we estimate the following parameters:

$$\boldsymbol{\omega} = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} \quad , \quad \boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \quad , \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{bmatrix} \quad , \quad \boldsymbol{P} = \begin{bmatrix} \pi_{1,1} & \pi_{1,2} \\ \pi_{2,1} & \pi_{2,2} \end{bmatrix} \tag{46}$$

in addition to υ in the case of the Student t distribution. The rows in the transition probability matrix $\boldsymbol{P}$ sums to 1, and the long-term probability of being in state k, denoted by Π_k, can be calculated by solving the following system of equations:

$$\begin{bmatrix} \Pi_1 & \Pi_2 \end{bmatrix} \begin{bmatrix} \pi_{1,1} & \pi_{1,2} \\ \pi_{2,1} & \pi_{2,2} \end{bmatrix} = \begin{bmatrix} \Pi_1 & \Pi_2 \end{bmatrix} \tag{47}$$

where

$$\Pi_1 + \Pi_2 = 1 \tag{48}$$

The conditional probability function of ε_{t+1} is a mixture of 2 regime-dependent distributions, and relies on the parameters obtained from the MLE, $\hat{\Psi}$. Let the conditional density of ε_{t+1} in state $s_{t+1} = k$ be denoted by $f_D(\varepsilon_{t+1} \mid s_{t+1} = k, \hat{\Psi}, I_t)$. A discrete representation of the density integral, resulting in the probability density function of ε_{t+1}, can be expressed as:

$$f(\varepsilon_{t+1} \mid \hat{\Psi}, I_t) = \sum_{i=1}^{2} \sum_{j=1}^{2} \pi_{i,j} \eta_{i,t} f_D(\varepsilon_{t+1} \mid s_{t+1} = j, \hat{\Psi}, I_t) \tag{49}$$

where $\eta_{i,t}$ refers to the filtered probability of state i at time t, and $\pi_{i,j}$ is the transition probability of moving from state i to state j. The filtered probabilities can be written as:

$$\eta_{i,t} = P[s_t = i \mid \Psi, I_t] \tag{50}$$

The filtered probabilities are obtained via the Hamilton filter. See Hamilton (1989) and Hamilton (1994) for details. The cumulative density function is obtained as follows:

$$F(\varepsilon_{t+1} \mid \hat{\Psi}, I_t) = \int_{-\infty}^{\varepsilon_{t+1}} f(\tau \mid \hat{\Psi}, I_t) d\tau \tag{51}$$

VaR and ES estimates are obtained by numerically inverting the cumulative density function in (51), see Ardia, Bluteau, Boudt, and Catania (2018) and Blasques, Koopman, Łasak, and Lucas (2016).

5.2.7 Volatility-adjusted quantile regression

The concept and theory of quantile regression was first developed by Koenker and Bassett Jr (1978). Applications of linear and non-linear regression quantile techniques in VaR estimation is covered by Taylor (1999) and Engle and Manganelli (2004). We use volatility-adjusted quantile regression based on RiskMetrics volatility, similarly as Steen et al. (2015). If $P\&L_t$ is the dependent variable and σ_t is the independent variable, then the simple linear quantile regression model can be specified as:

$$P\&L_t^p = \beta_0^p + \beta_1^p \sigma_t + \epsilon_t^p \tag{52}$$

where the distribution of the error term ϵ^p is left unspecified. The conditional pth quantile, with $p \in (0,1)$, is found by solving an optimization problem:

$$\min_{\beta_0,\beta_1} \sum_{t=1}^{T} \Big(p - \mathbf{1}_{P\&L_t \leq \beta_0+\beta_1\sigma_t}\Big)\Big(P\&L_t - (\beta_0 + \beta_1\sigma_t)\Big) \tag{53}$$

where

$$\mathbf{1}_{P\&L_t \leq \beta_0+\beta_1\sigma_t} = \begin{cases} 1, & \text{if } P\&L_t \leq \beta_0 + \beta_1\sigma_t \\ 0, & \text{if } P\&L_t > \beta_0 + \beta_1\sigma_t \end{cases} \tag{54}$$

and T is the number of observations in the sample. This means that we minimize the sum of all weighted residuals for a given quantile p to obtain the quantile regression coefficients $\hat{\beta}_0^p$ and $\hat{\beta}_1^p$. The least absolute error (the conditional mean) is a special case, but the quantile regression method explicitly allows for modelling of all quantiles of the dependent variable (Steen et al., 2015). As VaR, in our framework, is a particular conditional quantile of daily P&L of the spread, we can express the conditional quantile function as:

$$\widehat{VaR}_{t+1}^p \mid \sigma_{t+1} = \hat{\beta}_0^p + \hat{\beta}_1^p\sigma_{t+1} + \epsilon_{t+1}^p \mid \sigma_{t+1} \tag{55}$$

No distributional assumptions are required, and the technique is especially useful in stress testing and scenario analysis. However, this application of quantile regression provides only a model for quantiles, and it is not clear how to calculate the corresponding ES (Taylor, 2008a). Consequently, we only use volatility-adjusted quantile regression to estimate VaR and not to estimate ES.

6 Bivariate models

6.1 Portfolio variance and conditional correlation models

In this section, we extend the univariate analysis and present bivariate models to estimate long and short VaR and ES. In the univariate analysis, we considered a single position, either long or short, in the spread itself. We will now switch focus and consider a portfolio consisting of positions in both legs of the spread. The P&L on the portfolio on day $t+1$ is defined as:

$$P\&L_{PF,t+1} = w_{GER}P\&L_{GER,t+1} + w_{NOR}P\&L_{NOR,t+1} \tag{56}$$

where $P\&L_{GER,t+1}$ denotes the daily P&L on the German futures contract defined in (3) and $P\&L_{NOR,t+1}$ denotes the daily P&L on the Nordic futures contract defined in (4). w_{GER} and w_{NOR} refer to the portfolio weights in the German and Nordic leg of the spread, respectively. Throughout the rest of the study, we will only consider two portfolios - one long portfolio and one short portfolio. The two portfolios are defined by weights in the following way:

$$\begin{aligned} w_{GER} = 1, \quad w_{NOR} = -1, \quad &\text{if long portfolio} \\ w_{GER} = -1, \quad w_{NOR} = 1, \quad &\text{if short portfolio} \end{aligned} \tag{57}$$

That is, we consider equal weights with opposite signs in the two legs of the spread. We emphasize that the weights are not time-varying, meaning that they are considered to be constant over the entire data period. We choose the portfolio weights for a long and short portfolio in this particular way in order to enable backtesting of VaR and ES estimates from both univariate and bivariate models using the same spread time series. The reason why we can backtest using the same realized spread data, is because weights defined as in (57) lead to $P\&L_{PF,t+1} = P\&L_{t+1}$ for a long portfolio, as well as $P\&L_{PF,t+1} = -P\&L_{t+1}$ for a short portfolio. As before, $P\&L_{t+1}$ is defined as the daily P&L of the spread when considering the spread as a univariate time series. In other words, by defining these specific weights, daily P&L on the (long/short) portfolio considered in bivariate models will be equal to $(+/-)$ daily P&L of the spread considered in univariate models.

Portfolio variance is given by the following equation:

$$\sigma^2_{PF,t+1} = w^2_{GER}\sigma^2_{GER,t+1} + w^2_{NOR}\sigma^2_{NOR,t+1} + 2w_{GER}w_{NOR}\sigma_{GER-NOR,t+1} \tag{58}$$

where $\sigma_{GER-NOR,t+1}$ is the covariance of daily P&L on the German contract and daily P&L on the Nordic contract. Using matrix notation, the portfolio variance can be equivalently defined as:

$$\sigma^2_{PF,t+1} = \boldsymbol{w}^{'}\boldsymbol{\Sigma}_{t+1}\boldsymbol{w} \tag{59}$$

where $\boldsymbol{w}$ is the 2×1 vector of portfolio weights and $\boldsymbol{\Sigma}_{t+1}$ is the 2×2 conditional covariance matrix. There exists an extensive selection of multivariate GARCH models that can be used to model the dynamics of the conditional covariance matrix. See Bauwens et al. (2006) and Silvennoinen and Teräsvirta (2009) for surveys of multivariate GARCH models and their application in finance. We will assess the dynamics using models based on the decomposition of the conditional covariance into conditional volatilities and correlations. This can be written in matrix notation in the following way:

$$\boldsymbol{\Sigma}_{t+1} = \boldsymbol{D}_{t+1}\boldsymbol{\Upsilon}_{t+1}\boldsymbol{D}_{t+1} \tag{60}$$

where $\boldsymbol{D}_{t+1}$ is a 2×2 diagonal matrix of GARCH(1,1) volatilities as stated in (10) and $\boldsymbol{\Upsilon}_{t+1}$ is the 2×2 conditional correlation matrix which is symmetric positive definite with elements $\rho_{GER-NOR,t+1}$ on the off-diagonal and 1 on the diagonal. $\rho_{GER-NOR,t+1} = \frac{\sigma_{GER-NOR,t+1}}{\sigma_{GER,t+1}\sigma_{NOR,t+1}}$ is the correlation of daily P&L on the German contract and daily P&L on the Nordic contract.

In the following sections, we will introduce different specifications of the conditional correlation matrix $\boldsymbol{\Upsilon}_t$. By modelling correlation, we can assess if time-variation in covariance arises solely from time-variation in the volatilities of the two legs or if correlation has its own dynamic pattern (Christoffersen, 2011). The use of correlation models in our study is motivated by findings in the data analysis suggesting that correlation is time-varying.

6.1.1 CCC

The constant conditional correlation (CCC) model of Bollerslev (1990) assumes that the covariance matrix at time $t+1$ is described by:

$$\boldsymbol{\Sigma}_{t+1} = \boldsymbol{D}_{t+1}\boldsymbol{\Upsilon}\boldsymbol{D}_{t+1} \tag{61}$$

where $\boldsymbol{\Upsilon}$ is a correlation matrix which is not time-varying. The constant correlation can be specified in different ways in the CCC model, but we apply the unconditional correlation given by:

$$\rho_{GER-NOR} = \frac{1}{T}\sum_{t=1}^{T} z_{GER,t}z_{NOR,t} \tag{62}$$

where $z_{GER,t}$ and $z_{NOR,t}$ denote standardized daily P&L of German and Nordic futures contracts on day t, respectively, as defined in (7). As the correlation is constant, the CCC model requires no estimation beyond the estimation of the two univariate GARCH(1,1) models to obtain $z_{GER,t}$ and $z_{NOR,t}$.

6.1.2 DCC-EWMA

The assumption of constant conditional correlation can be unrealistic in many empirical applications. Engle (2002), Engle and Sheppard (2001) and Tse and Tsui (2002) extend the conditional correlation model and develop dynamic conditional correlation (DCC) models. DCC models assume that the conditional correlation matrix is time-dependent and described by:

$$\boldsymbol{\Upsilon}_{t+1} = diag(\boldsymbol{Q}_{t+1}^{-1/2})\boldsymbol{Q}_{t+1}diag(\boldsymbol{Q}_{t+1}^{-1/2}) \tag{63}$$

where $diag(\boldsymbol{Q}_{t+1})$ is the diagonal matrix that is formed from the diagonal elements of the positive definite matrix $\boldsymbol{Q}_{t+1}$. In the DCC-EWMA model described by Engle (2002), $\boldsymbol{Q}_{t+1}$ follows the process stated below:

$$\boldsymbol{Q}_{t+1} = (1-\lambda)(\boldsymbol{z}_t\boldsymbol{z}_t^{'}) + \lambda\boldsymbol{Q}_t \tag{64}$$

where λ is the exponential smoothing parameter and $\boldsymbol{z}_{t+1}$ is the 2×1 vector of standardized daily P&L. Normally, this model requires estimation of λ but we choose to set $\lambda = 0.94$ as in the RiskMetrics model for conditional variance. This simplification is done to enable a parsimonious correlation model with the advantage of no parameter estimation beyond estimation of the two univariate GARCH(1,1) models to obtain $\boldsymbol{z}_{t+1}$.

6.1.3 DCC-GARCH

Engle (2002) also proposes a DCC-GARCH model which, as opposed to DCC-EWMA, exhibits mean-reverting correlation. The conditional correlation matrix is still given by (63) whereas $\boldsymbol{Q}_{t+1}$ evolves according to:

$$\boldsymbol{Q}_{t+1} = (1 - \alpha - \beta)\bar{\boldsymbol{Q}} + \alpha \boldsymbol{z}_t \boldsymbol{z}_t^{'} + \beta \boldsymbol{Q}_t \tag{65}$$

where $\bar{\boldsymbol{Q}}$ is the 2×2 unconditional correlation matrix of $\boldsymbol{z}_t$, and α and β are non-negative scalar parameters. This model is stationary as long as $\alpha + \beta < 1$, and if the sum is equal to 1 then this reduces to the DCC-EWMA model in (64). The DCC-GARCH model captures correlation clustering, and is estimated using a two-step procedure explained in Appendix D.1 based on a multivariate distributional assumption.

6.1.4 Asymmetric DCC-GARCH

In the DCC-EWMA and DCC-GARCH models, correlation response to market shocks is symmetric. Cappiello, Engle, and Sheppard (2006) generalize the DCC model to incorporate both asymmetric correlation response and mean-reversion in correlation. In the asymmetric DCC-GARCH model, the conditional correlation matrix is given by (63) and $\boldsymbol{Q}_{t+1}$ is described by:

$$\boldsymbol{Q}_{t+1} = (1 - \alpha - \beta)\bar{\boldsymbol{Q}} + \alpha(\boldsymbol{z}_t \boldsymbol{z}_t^{'}) + \beta \boldsymbol{Q}_t + \gamma(\boldsymbol{\eta}_t \boldsymbol{\eta}_t^{'} - \bar{\boldsymbol{\Gamma}}) \tag{66}$$

where $\eta_{i,t}$ is defined by the negative part of $z_{i,t}$ in the following way for $i \in \{GER, NOR\}$:

$$\eta_{i,t} = \begin{cases} z_{i,t}, & \text{if } z_{i,t} < 0 \\ 0, & \text{if } z_{i,t} \geq 0 \end{cases} \tag{67}$$

Also, $\bar{\boldsymbol{\Gamma}} = \mathbb{E}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t^{'})$. We note that γ is the asymmetric effect parameter, and if γ is positive then the correlation will increase more when $z_{GER,t}$ and $z_{NOR,t}$ are negative than in any other case. On the other hand, if γ is negative, correlation will decrease more when $z_{GER,t}$ and $z_{NOR,t}$ are negative. The model is motivated by the phenomenon often observed for risky assets, where correlation may increase more in down markets compared to up markets (Christoffersen, 2011). The model is stationary if $\alpha + \beta + \kappa\gamma < 1$, where κ is the maximum eigenvalue of $\bar{\boldsymbol{Q}}^{-1/2} \bar{\boldsymbol{\Gamma}} \bar{\boldsymbol{Q}}^{-1/2}$. The asymmetric DCC-GARCH model is, like its symmetric equivalent, estimated using a two-step procedure explained in Appendix D.1.

An important feature of the DCC models described here is that $\boldsymbol{Q}_{t+1}$ is positive semidefinite as it is a weighted average of positive semidefinite and positive definite matrices. This will consequently ensure that the correlation matrix $\boldsymbol{\Upsilon}_{t+1}$ and covariance matrix $\boldsymbol{\Sigma}_{t+1}$ is positive semidefinite, meaning that $\sigma^2_{PF,t+1} = \boldsymbol{w}^{'} \boldsymbol{\Sigma}_{t+1} \boldsymbol{w} \geq 0$ regardless of portfolio weights.

6.2 Portfolio Value-at-Risk and Expected Shortfall models

6.2.1 CCC and DCC: Bivariate normal distribution

If we assume that the daily P&L of German and Nordic futures contracts are bivariate normal distributed, the model can be specified as follows:

$$\boldsymbol{\varepsilon}_t \mid I_{t-1} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_t = \boldsymbol{D}_t \boldsymbol{\Upsilon}_t \boldsymbol{D}_t) \tag{68}$$

where $\boldsymbol{\varepsilon}_t$ is the 2×1 vector of market shocks and as before I_{t-1} denotes the information set containing all relevant information up to time $t-1$. $\boldsymbol{D}_t$ consists of GARCH(1,1) volatilities from (10) and $\boldsymbol{\Upsilon}_t$ is described by one of the correlation models already introduced. We report results from CCC, DCC-EWMA, DCC-GARCH and asymmetric DCC-GARCH assuming bivariate normal distribution.

The multivariate normal distribution has the property that a linear combination of multivariate normal variables is also normally distributed. The long and short portfolio we consider are linear combinations of daily P&L of German and Nordic futures contracts, and thus the day-ahead VaR and ES can be estimated by the following expressions:

$$VaR^p_{t+1} = \sigma_{PF,t+1} \Phi^{-1}_{1-p} \tag{69}$$

$$ES^p_{t+1} = \sigma_{PF,t+1} \frac{\phi(\Phi^{-1}_{1-p})}{p} \tag{70}$$

where the portfolio volatility is defined by (59). See Appendix D.1.1 for the two-step estimation procedure with bivariate normal distribution.

6.2.2 CCC and DCC: Bivariate Student t distribution

In many applications, the multivariate normal distribution fails to adequately capture the true multivariate risk. This motivates the use of the multivariate Student t distribution. A bivariate Student t distribution model with υ degrees of freedom can in our case be specified in the following manner:

$$\boldsymbol{\varepsilon}_t \mid I_{t-1} \sim t(\mathbf{0}, \boldsymbol{\Sigma}_t = \boldsymbol{D}_t \boldsymbol{\Upsilon}_t \boldsymbol{D}_t, \upsilon) \tag{71}$$

where $\boldsymbol{D}_t$ still consists of GARCH(1,1) volatilities from (10), and $\boldsymbol{\Upsilon}_t$ is modelled by a conditional correlation model. We report results from CCC, DCC-GARCH and asymmetric DCC-GARCH with a distributional assumption of multivariate Student t. See Appendix D.1.2 for explanation of the two-step estimation procedure with bivariate Student t distribution.

As a portfolio of multivariate Student t distributed variables does not itself follow the Student t distribution in general, we need to use Monte Carlo simulation to estimate portfolio day-ahead VaR and ES. The following representation can be applied to simulate standardized multivariate Student t variables (Demarta & McNeil, 2005):

$$\boldsymbol{z} = \sqrt{\frac{\upsilon - 2}{\upsilon}} \sqrt{W} \boldsymbol{U} \tag{72}$$

where W is a univariate inverse gamma random variable defined by $W \sim IG(\frac{\upsilon}{2}, \frac{\upsilon}{2})$, $\boldsymbol{U}$ is a vector of multivariate standard normal variables described by $\boldsymbol{U} \sim N(\mathbf{0}, \boldsymbol{\Upsilon}_t)$, and $\boldsymbol{U}$ and W are independent. The simulated vector $\boldsymbol{z}$ is standardized multivariate Student t distributed with zero mean, variance equal to

one and correlation matrix $\boldsymbol{\Upsilon}_t$.

In the next step, we simulate day-ahead P&L on the portfolio as $P\&L^j_{PF,t+1} = \boldsymbol{z}^j \sigma_{PF,t+1}$ for $j = 1, \ldots, j^*$ simulations. From this, VaR^p_{t+1} is computed as the pth quantile of $P\&L^j_{PF,t+1}$ over the j^* simulations. Finally, we calculate $ES^p_{t+1} = \frac{1}{pj^*} \sum_{j=1}^{j^*} P\&L^j_{PF,t+1} \cdot \mathbf{1}(P\&L^j_{PF,t+1} < VaR^p_{t+1})$ for a long portfolio and $ES^p_{t+1} = \frac{1}{pj^*} \sum_{j=1}^{j^*} P\&L^j_{PF,t+1} \cdot \mathbf{1}(P\&L^j_{PF,t+1} > VaR^p_{t+1})$ for a short portfolio, where $\mathbf{1}(*)$ takes the value 1 if the argument is true and 0 otherwise. Like Bauwens and Laurent (2005), we use $j^* = 10,000$ Monte Carlo simulations.

6.2.3 Volatility and correlation-adjusted quantile regression

In this section, we present a model which, to the best of our knowledge, has not been proposed in the existing literature with this specific formulation. We will refer to the model as volatility and correlation-adjusted quantile regression. As the name suggests, the concept is similar the volatility-adjusted quantile regression, which was introduced in Section 5.2.7. We note that this is a univariate model, as opposed to the other models in Section 6. Nevertheless, we choose to include the model in this part of the study as one of the independent variables is based on a conditional correlation model.

In volatility and correlation-adjusted quantile regression, we assume that $P\&L_t$ is the dependent variable and that $\sigma_{GER,t}$, $\sigma_{NOR,t}$ and $\rho_{GER-NOR,t}$ are the three explanatory variables. $\sigma_{GER,t}$ and $\sigma_{NOR,t}$ are based on RiskMetrics volatility as stated in (9), whereas $\rho_{GER-NOR,t}$ comes from the DCC-EWMA model summarized by (63) and (64). This leads to the following linear quantile regression model:

$$P\&L^p_t = \beta^p_0 + \beta^p_1 \sigma_{GER,t} + \beta^p_2 \sigma_{NOR,t} + \beta^p_3 \rho_{GER-NOR,t} + \epsilon^p_t \tag{73}$$

where the distribution of the error term ϵ^p is left unspecified. The conditional pth quantile, with $p \in (0,1)$, is found by solving an optimization problem:

$$\min_{\beta_0,\beta_1,\beta_2,\beta_3} \sum_{t=1}^{T} \Big(p - \mathbf{1}_{P\&L_t \leq \beta_0+\beta_1\sigma_{GER,t}+\beta_2\sigma_{NOR,t}+\beta_3\rho_{GER-NOR,t}}\Big)\Big(P\&L_t - (\beta_0 + \beta_1\sigma_{GER,t} + \beta_2\sigma_{NOR,t} + \beta_3\rho_{GER-NOR,t})\Big) \tag{74}$$

where

$$\mathbf{1}_{P\&L_t \leq \beta_0+\beta_1\sigma_{GER,t}+\beta_2\sigma_{NOR,t}+\beta_3\rho_{GER-NOR,t}} = \begin{cases} 1, & \text{if } P\&L_t \leq \beta_0 + \beta_1\sigma_{GER,t} + \beta_2\sigma_{NOR,t} + \beta_3\rho_{GER-NOR,t} \\ 0, & \text{if } P\&L_t > \beta_0 + \beta_1\sigma_{GER,t} + \beta_2\sigma_{NOR,t} + \beta_3\rho_{GER-NOR,t} \end{cases} \tag{75}$$

and T is the number of observations in the sample. That is, we obtain the quantile regression coefficients $\hat{\beta}^p_0$, $\hat{\beta}^p_1$, $\hat{\beta}^p_2$ and $\hat{\beta}^p_3$ by minimizing the sum of all weighted residuals for a given quantile p. Similar as in (55), VaR can be expressed as the following conditional quantile function:

$$\widehat{VaR}^p_{t+1} \mid \sigma_{GER,t+1}, \sigma_{NOR,t+1}, \rho_{GER-NOR,t+1} = \hat{\beta}^p_0 + \hat{\beta}^p_1\sigma_{GER,t+1} + \hat{\beta}^p_2\sigma_{NOR,t+1} + \hat{\beta}^p_3\rho_{GER-NOR,t+1} + \epsilon^p_{t+1} \mid \sigma_{GER,t+1}, \sigma_{NOR,t+1}, \rho_{GER-NOR,t+1} \tag{76}$$

We only use volatility and correlation-adjusted quantile regression to estimate VaR and not to estimate

ES.

The model has an intuitive interpretation in the sense that it assumes that daily P&L of the spread is described by the volatility of both legs of the spread in addition to the correlation. Furthermore, we argue that the simplicity of quantile regression is preserved by using relatively simple volatility and correlation models. However, we note that the model implicitly assumes that there exists a linear relationship between $P\&L_t$ and $\rho_{GER-NOR,t}$. Even if we accept that there may exist a linear relationship between daily P&L and volatility, it is arguably more difficult to justify the same for correlation. A possible extension of the model could include a non-linear relationship between $P\&L_t$ and $\rho_{GER-NOR,t}$. Nevertheless, we have chosen the formulation described above for purposes of simplicity.

7 Backtesting Value-at-Risk and Expected Shortfall

Backtesting refers to testing the accuracy of a VaR or ES model over a historical period when the true outcome is known. We use the standard approach for backtesting, which in our study involves recording the number of occasions over the period under consideration where the actual daily P&L of the spread exceeds the model VaR or ES, and compare this number to the prespecified levels.

According to Christoffersen (2011), a proper VaR model should satisfy two conditions: (I) the number of exceedances should be as close as possible to the number implied by the VaR quantile and (II) the exceedances should be randomly distributed across the sample, meaning that we observe no clustering of exceedances. In other words, we want to avoid that the model overestimates or underestimates the number of exceedances in certain periods. To test the first condition, we use the unconditional test of Kupiec (1995), and to test both conditions, we employ the conditional coverage test of Christoffersen (1998). In addition to this, Engle and Manganelli (2004) argue that in each period, the probability of exceeding the VaR should be independent of all past information, including the VaR estimate as well as previous exceedances. To test this, we employ the dynamic quantile test. In summary, this gives three different tests that we use to assess the accuracy and appropriateness of the VaR models.

7.1 Unconditional coverage test

The Kupiec (1995) test is a likelihood ratio test designed to reveal whether a VaR model provides the desirable unconditional coverage. An indicator sequence is defined as follows:

$$H_t^p = \begin{cases} 1, & \text{if } X_t > VaR_t^p \\ 0, & \text{if } X_t \leq VaR_t^p \end{cases} \tag{77}$$

where X_t is the loss function defined in (16). The indicator sequence implies that if the loss is greater than the VaR estimate on a given day, we count an exceedance. Under the null hypothesis that the number of exceedances is equal to the prespecified VaR, the test statistic is (Kupiec, 1995):

$$-2ln(LR_{UC}) = -2\big[n_0 ln(1-\pi_{exp}) + n_1 ln(\pi_{exp}) - n_0 ln(1-\pi_{obs}) - n_1 ln(\pi_{obs})\big] \tag{78}$$

where n_1 and n_0 is the number of exceedances and non-exceedances, respectively. $\pi_{exp} = p$ is the expected proportion of exceedances, while $\pi_{obs} = \frac{n_1}{n_0+n_1}$ represents the observed fraction of exceedances. The asymptotic distribution of $-2ln(LR_{UC})$ is χ^2 distributed with one degree of freedom.

7.2 Conditional coverage test

The unconditional test does not take into account whether several violations occur in rapid succession or if they tend to be isolated. Christoffersen (1998) extended the Kupiec test and proposed a joint test for correct coverage and for detection of whether a quantile exceedance today has implications for the probability of a quantile exceedance tomorrow. Under the null hypothesis that the number of violations is equal to the prespecified VaR and the violations are randomly distributed, the test statistic is:

$$-2ln(LR_{CC}) = -2\big[n_0 ln(1-\pi_{exp}) + n_1 ln(\pi_{exp}) - n_{00} ln(1-\pi_{01}) - n_{01} ln(\pi_{01}) - n_{10} ln(1-\pi_{11}) - n_{11} ln(\pi_{11})\big] \tag{79}$$

where n_{ij} denotes the number of times an indicator variable with value i is immediately followed by an indicator variable with value j. Further, $\pi_{01} = \frac{n_{01}}{n_{00}+n_{01}}$, $\pi_{11} = \frac{n_{11}}{n_{10}+n_{11}}$ and the rest of the notation is as described for the Kupiec test. The test statistic follows a χ^2 distribution with two degrees of freedom. As a practical matter, one may incur samples where $\pi_{11} = 0$. In this case, the test statistic is stated as (Christoffersen, 2011):

$$-2ln(LR_{CC}) = -2\big[n_0 ln(1-\pi_{exp}) + n_1 ln(\pi_{exp}) - n_{00} ln(1-\pi_{01}) - n_{01} ln(\pi_{01})\big] \tag{80}$$

A drawback of the conditional coverage test of Christoffersen is that the test only takes into account one violation immediately followed by another, ignoring all other patterns of clustering.

7.3 Dynamic quantile test

The dynamic quantile test by Engle and Manganelli (2004) complements the backtesting framework proposed by Christoffersen (1998). The test defines a hit variable similar to the indicator sequence H_t^p in the unconditional coverage test as follows:

$$Hit_t^p = \begin{cases} 1-p, & \text{if } X_t > VaR_t^p \\ -p, & \text{if } X_t \leq VaR_t^p \end{cases} \tag{81}$$

where X_t again is the loss function defined in (16). Engle and Manganelli (2004) argue that Hit_t^p should be uncorrelated with its own lagged values, any other lagged values of the past information set and have an expected value of zero. This can be tested by regressing Hit_t^p on its own lagged values and other variables which the hit variable should be uncorrelated with in a good VaR model. As Engle and Manganelli (2004), we choose a lag order of 4 when conducting the dynamic quantile test. We also include the VaR estimate on day t as well as the squared loss on day $t-1$, as these two variables also should be uncorrelated with the hit variable. This leads to the following regression on day t:

$$Hit_t^p = \phi_0 + \phi_1 Hit_{t-1}^p + \phi_2 Hit_{t-2}^p + \phi_3 Hit_{t-3}^p + \phi_4 Hit_{t-4}^p + \phi_5 VaR_t^p + \phi_6 X_{t-1}^2 \tag{82}$$

In matrix notation notation, this can be expressed as:

$$\boldsymbol{Hit}^p = \boldsymbol{Y}\boldsymbol{\phi} \tag{83}$$

where $\boldsymbol{Hit}^p$ consists of Hit_t^p for all days $t = 1, \dots T$, $\boldsymbol{\phi}$ is the set of coefficients and $\boldsymbol{Y}$ is the set of regressors from (82). The null hypothesis of no influence from the regressors is expressed as:

$$H_0 : \boldsymbol{\phi} = \mathbf{0} \tag{84}$$

Engle and Manganelli (2004) show that the ordinary least squares solution to this regression is given by:

$$\hat{\boldsymbol{\phi}} = (\boldsymbol{Y}'\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{Hit}^p \sim N(\mathbf{0}, p(1-p)\boldsymbol{Y}'\boldsymbol{Y}) \tag{85}$$

They also give the following test statistic:

$$DQ_{CC} = \frac{\hat{\boldsymbol{\phi}}\boldsymbol{Y}'\boldsymbol{Y}\hat{\boldsymbol{\phi}}}{p(1-p)} \tag{86}$$

The asymptotic distribution of DQ_{CC} is χ^2 distributed with q degrees of freedom. The degrees of freedom is equal to the number of coefficients that are estimated during the regression, which in our application gives $q = 7$.

7.4 Backtesting Expected Shortfall

McNeil and Frey (2000) propose a test for ES which considers the difference between realized losses and ES estimates given that the loss is greater than the VaR estimate, called standardized exceedance residuals. The test is based on the assumption that, if the estimated process dynamics are correct and ES is an unbiased estimate of the expectation in the tail beyond the VaR, the standardized exceedance residuals should behave as a sample from an i.i.d. zero mean process. By bootstrapping, one can assess whether the mean of the residuals is statistically different from zero. The bootstrapping procedure is defined and described in Appendix E. As McNeil and Frey (2000), we obtain the standardized exceedance residuals $\hat{z}_t$ by:

$$\hat{z}_{t+1} = \left\{ \frac{X_{t+1} - ES^p_{t+1}}{\sigma_{t+1}} \mid X_{t+1} > VaR^p_{t+1} \right\} \tag{87}$$

where X_{t+1} is the loss function as defined in (16). The test results in p-values for a one-sided bootstrap test of the null hypothesis that the standardized exceedance residuals have mean equal to zero against the alternative hypothesis that the mean is greater than zero. As McNeil and Frey (2000), the null hypothesis and alternative hypothesis are expressed as follows:

$$H_0 : \ \mu_{\hat{z}} = 0, \quad H_1 : \ \mu_{\hat{z}} > 0 \tag{88}$$

where $\mu_{\hat{z}}$ is the mean of the standardized exceedance residuals. The argument for using a one-sided test is that for risk management purposes, it is more critical to underestimate risk rather than to overestimate risk. If $\mu_{\hat{z}}$ is positive, then the ES estimates are too low on average. The test could be altered by considering $H_1 : \ \mu_{\hat{z}} \neq 0$, but it would then not be possible to infer whether risk is overestimated or underestimated. Another approach is suggested by Taylor (2008b), who propose standardizing the exceedance residuals by the VaR estimate instead of the conditional volatility. However, this method increases the dependence on proper VaR models and may give false indications on the suitability and performance of the ES model (Embrechts, Kaufmann, & Patie, 2005).

An apparent drawback of the framework offered by McNeil and Frey (2000), is that it relies explicitly on the VaR estimates. If the conditional volatility model is incorrectly specified, or the underlying distributional assumption is wrong, the results from the ES test could be misleading. Another matter of concern is the size of the sample, as we only use observations corresponding to exceedances. If the sample size is small, we would expect few exceedances. This will in turn result in few data points to use during the bootstrap, which could potentially lead to less reliable results.

We choose to only use the one-sided test of McNeil and Frey (2000) to backtest ES. We are primarily concerned with underestimation of risk and consider this to be a more appropriate ES backtesting methodology for the purposes of our study compared to the two-sided version of the test and the approach suggested by Taylor (2008b). Furthermore, we regard three test for VaR and one for ES to be a sufficient number of criteria to assess the accuracy of our selection of models.

8 Results

In order to evaluate the predictive ability of our selection of univariate and bivariate models, we create day-ahead forecasts of VaR and ES. We define the out-of-sample window to be 1000 data points, which implies an in-sample window, or estimation sample, of 1264 observations for the front-quarter spread and 1288 observations for the front-year spread. This can be illustrated as:

$$\text{Front-quarter:} \quad \underbrace{P\&L_1,\ P\&L_2,\ P\&L_3,\ ...,\ P\&L_{1264=T^*}}_{\text{In-sample window}},\ \underbrace{P\&L_{1265=T^*+1},\ P\&L_{1266},...,\ P\&L_{2264=T}}_{\text{Out-of-sample window}}$$

$$\text{Front-year:} \quad \underbrace{P\&L_1,\ P\&L_2,\ P\&L_3,\ ...,\ P\&L_{1288=T^*}}_{\text{In-sample window}},\ \underbrace{P\&L_{1289=T^*+1},\ P\&L_{1266},...,\ P\&L_{2288=T}}_{\text{Out-of-sample window}}$$

The final dates of the in-sample window corresponds to February 2015. We set the out-of-sample window equal to 1000 data points to balance two considerations. Firstly, we need an estimation sample sufficiently large to enable good parameter estimates which in turn can facilitate predictive ability. Secondly, we need a sufficiently large number of expected VaR exceedances at the 95% and 99% confidence levels so that forecasting performance can be investigated through backtesting of VaR and ES estimates. Our selected out-of-sample window implies that we expect 50 VaR violations at the 95% confidence level and 10 VaR violations at the 99% confidence level.

We predict VaR and ES using two different approaches. (I) In the first approach, we estimate model parameters from the in-sample window and create 1000 rolling forecasts without any reestimation of model parameters. That is, we fix the model parameters from the initial estimation, and compute day-ahead forecast $VaR^p_{T^*+1}$ as well as $ES^p_{T^*+1}$. Then the models observe $P\&L_{T^*+1}$, and compute new day-ahead forecast $VaR^p_{T^*+2}$ as well as $ES^p_{T^*+2}$. This procedure is repeated, without any reestimation, until we have 1000 day-ahead VaR and ES predictions. (II) In the second approach, we also start by estimating model parameters based on the in-sample window. The remaining procedure is however different from the first approach, as we include reestimation every 20th day using a moving window of T^* data points. This implies that we first create 20 rolling forecasts from the initial estimation, before we reestimate all model parameters based on a moving estimation sample, including the most recent 20 realized daily P&L and excluding the 20 first data points in the in-sample window. This is repeated until we have 1000 day-ahead VaR and ES predictions. In summary, this gives 49 refits beyond the initial estimation. An alternative procedure could include reestimation every day, but this would be very computationally intensive. Bauwens et al. (2006) find that the results in VaR estimation are qualitatively the same when updating model parameters every 5, 10 and 50 observations. Through testing with different refit horizons on our data, we observe that we achieve better backtesting performance by refitting every 20th day compared to every 50th day, but backtesting results are not significantly improved by reestimating even more frequently. We also observe better backtesting performance by using a moving window compared to an expanding window including all past data points. The only exceptions are the Markov switching GARCH models. Therefore, in the reported results, Markov switching GARCH models are reestimated with expanding window including all past observations, whereas the other models are reestimated using a moving estimation window.

The main advantage of approach (I) is that it entails less computational effort, given that all models are estimated only once. However, if the characteristics of the out-of-sample data differ from those of the in-sample data, we would expect this procedure to yield inadequate predictive accuracy. From the data analysis, we observe significant variation in excess kurtosis, skewness, empirical quantiles and mean

conditional on quantile exceedances in different periods across the total data sample. Consequently, it is reasonable to assume that VaR and ES forecasting without any estimation beyond the initial fit will be challenging for the models. On the contrary, approach (II) involves much greater computational effort but the resulting out-of-sample forecasts can gain accuracy.

We will proceed by first analyzing estimation results based on the in-sample window in Section 8.1. Then we present backtesting results in Section 8.2, before we summarize and discuss results in Section 8.3. Finally, we provide additional details on how to interpret the VaR and ES forecasts in Section 8.4.

8.1 Estimation results

8.1.1 Univariate GARCH models

Table 7: Estimation results for univariate GARCH models, front-quarter

Model	ω	α	β	υ	ξ	γ
GARCH(1,1)-n	0.006 (2.34)	0.088 (5.00)	0.903 (48.47)	-	-	-
GARCH(1,1)-t	0.007 (2.14)	0.084 (4.30)	0.905 (43.16)	10.999 (3.65)	-	-
GARCH(1,1)-skew-t*	0.006 (2.34)	0.088 (5.00)	0.903 (48.47)	10.992 (3.85)	0.043 (1.08)	-
GJR(1,1)-n	0.005 (1.92)	0.034 (1.96)	0.922 (47.62)	-	-	0.073 (3.72)
GJR(1,1)-t	0.005 (1.74)	0.035 (1.78)	0.920 (40.30)	12.055 (3.34)	-	0.071 (3.09)
GJR(1,1)-skew-t*	0.005 (1.92)	0.034 (1.96)	0.922 (47.62)	12.020 (3.51)	0.043 (1.07)	0.073 (3.72)
MSGARCH(1,1)-n						
Regime 1	0.000 (0.39)	0.026 (0.02)	0.974 (16537.51)	-	-	-
Regime 2	0.038 (2.18)	0.135 (1.59)	0.839 (46.77)	-	-	-
MSGARCH(1,1)-t						
Regime 1	0.001 (1.83)	0.006 (1.57)	0.989 (191.10)	13.158 (3.00)	-	-
Regime 2	0.039 (1.92)	0.128 (3.20)	0.845 (21.20)	13.158 (3.00)	-	-

t-statistics are shown in parentheses. The critical value is approximately 1.96 at the 5% significance level. * denotes that the model is estimated using QMLE.

Table 8: Estimation results for univariate GARCH models, front-year

Model	ω	α	β	υ	ξ	γ
GARCH(1,1)n	0.002 (2.00)	0.097 (3.82)	0.890 (30.35)	-	-	-
GARCH(1,1)-t	0.002 (2.16)	0.093 (4.12)	0.892 (34.83)	10.931 (3.84)	-	-
GARCH(1,1)-skew-t*	0.002 (2.00)	0.097 (3.82)	0.890 (30.35)	10.888 (4.03)	-0.036 (0.88)	-
GJR(1,1)-n	0.002 (1.58)	0.081 (2.43)	0.897 (25.46)	-	-	0.018 (0.85)
GJR(1,1)-t	0.002 (2.09)	0.091 (3.13)	0.893 (33.82)	10.945 (3.83)	-	0.002 (0.09)
GJR(1,1)-skew-t*	0.002 (1.58)	0.081 (2.43)	0.897 (25.46)	10.981 (3.97)	-0.036 (0.88)	0.018 (0.85)
MSGARCH(1,1)-n						
Regime 1	0.004 (1.11)	0.018 (0.51)	0.910 (14.16)	-	-	-
Regime 2	0.030 (2.80)	0.159 (1.96)	0.729 (16.27)	-	-	-
MSGARCH(1,1)-t						
Regime 1	0.004 (1.71)	0.048 (2.19)	0.891 (18.10)	17.98 (2.23)	-	-
Regime 2	0.031 (2.48)	0.159 (3.23)	0.728 (9.50)	17.98 (2.23)	-	-

t-statistics are shown in parentheses. The critical value is approximately 1.96 at the 5% significance level. * denotes that the model is estimated using QMLE.

Table 7 shows front-quarter estimation results for GARCH and GJR-GARCH models assuming normal, Student t and skewed Student t distributions of market shocks, as well as estimation results for Markov switching GARCH with normal and Student t distributions. Table 8 presents the corresponding results

for front-year daily P&L. In the following discussion of the significance of parameters, we base our conclusions on a significance level of 5%. This means that if a parameter has a t-statistic which, in absolute value, is greater than the critical value from the Student t distribution with $(T^* - 1)$ degrees of freedom at the 5% significance level in a two-tailed test, we argue that this parameter is significant.

The estimation results show that ω , α and β are significant for all GARCH(1,1) models. The fact that the latter two parameters are significant, suggests that there exists conditional heteroscedasticity effects, meaning that volatility in daily P&L is time-varying. Moreover, the degrees of freedom parameter υ is significant, indicating excess kurtosis. This suggests that the normal distribution is an inaccurate distributional assumption, and that the Student t distribution offers a better fit for the data. On the other hand, the skewness parameter ξ is insignificant, implying that we do not obtain a statistically significant better fit with the skewed Student t distribution compared to the symmetric version. These conclusions are supported by the QQ-plots, which are given in Appendix F.7 and F.8 due to space considerations. We use standardized daily P&L from the in-sample period with conditional volatility as defined in (10). The subfigures (a), (e) and (g) show that the data is not well-described by the normal distribution. On the contrary, subfigures (b), (f) and (h) show that Student t better captures the data characteristics. It is not clear that the skewed Student t in (c) offers improved fit compared to the Student t distribution in (b). These observations are consistent across front-quarter and front-year daily P&L.

The front-quarter estimation results show that, for the GJR-GARCH(1,1) models, ω and α parameters are insignificant with the exception of ω with Student t distribution. Furthermore, β is significant, and the same conclusions that were made concerning υ and ξ for GARCH(1,1) models also hold for GJR-GARCH(1,1). The asymmetry parameter γ is statistically significant for front-quarter data, implying that variance responds asymmetrically to positive and negative market shocks of the same magnitude. We also perform LR tests (see Appendix B.1) from which we conclude that the added parameter in GJR-GARCH(1,1) relative to GARCH(1,1) is significant. The opposite is the case for front-year data, where we cannot reject the null hypothesis of an insignificant added γ parameter.

Comparing estimation results for front-quarter and front-year daily P&L, we see that α is higher for the latter and β is higher for the former. This signals that volatility is slightly more sensitive to market shocks in the front-year time series, while persistence irrespective of market events is somewhat higher for the front-quarter equivalent. These effects balance out in a fairly similar rate of convergence of volatility, $\alpha + \beta$. Naturally, as front-quarter contracts are more closely related to the spot price, unconditional volatility is higher for front-quarter daily P&L according to (11) and (14).

The Markov switching GARCH models estimate parameters for two separate volatility regimes. Both for front-quarter and front-year data, regime 2 have the highest long-run average variance according to (11) and can be interpreted as the high-volatility regime. This is largely due to higher ω parameters in regime 2 compared to regime 1. Moreover, we observe higher α and lower β parameters relative to regime 1, indicating that volatility is more sensitive to market shocks and less persistent irrespective of market movements in regime 2. For Markov switching GARCH with Student t distribution, we conclude that υ is significant, indicating that this is a more precise distributional assumption.

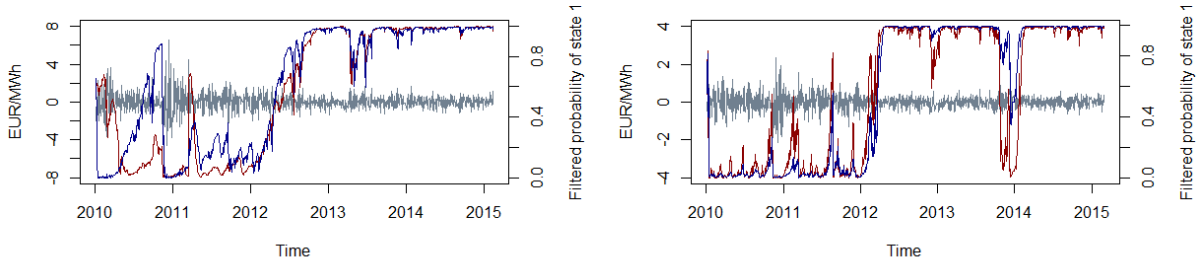
In Table 9, we present the estimated transition probabilities in the Markov switching GARCH models. The transition probabilities, π_{11} and π_{21}, indicate very low probabilities of switching regimes on two consecutive days. We observe that the long-term probability of being in state 1, Π_1, is slightly higher both for front-quarter and front-year data. We note that β and π_{11} have very high t-statistics as a consequence of very low standard errors in the estimation. This is consistent with Marcucci (2005) and

Bauwens, Preminger, and Rombouts (2010), who report comparable standard errors. In Figure 7, we show the filtered probabilities of state 1 for Markov switching GARCH with normal distribution and Student t distribution. The low probability of moving from one state to the next on two consecutive days is clearly visible from the filtered probabilities.

Table 9: Transition probabilities and long-term probabilities for MSGARCH(1,1)

	Front-quarter				Front-year			
Model	π_{11}	π_{21}	Π_1	Π_2	π_{11}	π_{21}	Π_1	Π_2
MSGARCH(1,1)-n	0.9991 (676.92)	0.0012 (0.90)	0.5726	0.4274	0.9976 (428.40)	0.003 (1.50)	0.5492	0.4508
MSGARCH(1,1)-t	0.9988 (522.02)	0.0017 (0.49)	0.5879	0.4121	0.9992 (1040.10)	0.001 (0.76)	0.5567	0.4433

t-statistics are shown in parentheses. The critical value is approximately 1.96 at the 5% significance level.



(a) Front-quarter (b) Front-year

Figure 7: Filtered probabilities of $s_t = 1$, MSGARCH(1,1)-n in red and MSGARCH(1,1)-t in blue

8.1.2 GARCH: Extreme value theory

Table 10: Results for GPD using the Hill estimator

	Front-quarter				Front-year			
Position	T	T_u	u	η	T	T_u	u	η
Long	1264	50	-1.738	0.249	1288	50	-1.817	0.227
Short	1264	50	1.771	0.242	1288	50	1.723	0.236

Table 10 presents results for the GARCH(1,1)-EVT model for front-quarter and front-year time series. In the reported results, u denote the 51st most negative observation of $P\&L_t$ for a long position and the 51st most positive observation for a short position in the in-sample window, as we consider the $T_u = 50$ most extreme tail observations. For the front-quarter P&L, the tail index parameter η is larger for a long position compared to a short position. For the front-year P&L, we observe the opposite. Different η for long and short positions indicate asymmetry in the tail distributions. QQ-plots for standardized daily P&L against quantiles from the EVT distribution show that EVT provides more extreme estimates than the other distributional assumptions. The plots are given in subfigures (d) of Figure 19 and 20 in Appendix F.7 and F.8, respectively.

Table 11: Estimation results for volatility-adjusted quantile regression

Data - quantile (position)	$\hat{\beta}_0^p$	$\hat{\beta}_1^p$
Front-quarter - 95% (long)	-0.201 (-1.67)	-1.344 (-7.41)
Front-quarter - 95% (short)	-0.030 (-0.23)	1.736 (7.82)
Front-quarter - 99% (long)	-0.396 (-1.38)	-1.949 (-3.52)
Front-quarter - 99% (short)	-0.044 (-0.10)	2.599 (3.03)
Front-year - 95% (long)	-0.070 (-1.06)	-1.055 (-6.95)
Front-year - 95% (short)	-0.013 (-0.19)	1.618 (6.52)
Front-year - 99% (long)	-0.119 (-1.04)	-2.199 (-6.14)
Front-year - 99% (short)	0.020 (0.13)	2.547 (4.43)

t-statistics are shown in parentheses. The critical value is approximately 1.96 at the 5% significance level.

8.1.3 Volatility adjusted quantile regression

Table 11 presents the output from volatility-adjusted quantile regression for front-quarter and front-year daily P&L. The estimation results show that $\hat{\beta}_1^p$ is significant for all quantiles. The parameter $\hat{\beta}_1^p$ is an indicator of how VaR reacts with respect to changes in the volatility obtained by RiskMetrics. Naturally, this parameter differs in sign for long and short positions, as VaR is a negative number for a long position and a positive number for a short position. As expected, the absolute value of $\hat{\beta}_1^p$ is higher as we investigate the more extreme 99% quantile compared to the 95% quantile. The intercept parameter $\hat{\beta}_0^p$ is insignificant across all quantiles.

8.1.4 Bivariate models

Table 12 presents estimation results for bivariate models based on front-quarter data, whereas Table 13 shows the same for front-year data. For each of the models, a GARCH(1,1) model is estimated for both daily P&L on the German contract and daily P&L on the Nordic contract. In addition to this, a conditional correlation model is estimated. From Table 12, we see that all GARCH(1,1) parameters are significant. The DCC-GARCH parameters are also significant, with the exception of α for aDCC-GARCH-t as well as the correlation asymmetry parameter γ in the aDCC models. Statistically significant α and β parameters in the DCC models indicate that conditional correlation is time-varying. Maximized log likelihood values are almost identical for symmetric and asymmetric DCC-GARCH models with the same distributional assumption (see Table 15), and according to the LR test (see Appendix B.1), we cannot reject the null hypothesis of an insignificant added γ parameter. Thus, we can infer that we do not obtain a statistically significant better fit by allowing for asymmetric correlation response to market shocks. The degrees of freedom parameter υ is significant in all of the conditional correlation models using Student t. This suggests that bivariate Student t is a more accurate distributional assumption compared to the bivariate normal distribution. From Table 13 with front-year data, we can draw similar conclusions. The asymmetric correlation parameter γ is insignificant and Student t offers a better fit. However, we observe that ω is insignificant for most of the GARCH(1,1) models.

Table 12: Estimation results for bivariate models, front-quarter

	GARCH(1,1)				GARCH(1,1)				CCC/DCC			
Model	ω_{GER}	α_{GER}	β_{GER}	υ_{GER}	ω_{NOR}	α_{NOR}	β_{NOR}	υ_{NOR}	α	β	γ	υ
CCC-n	0.011(4.42)	0.144(6.12)	0.807(29.46)	-	0.008(2.46)	0.094(5.57)	0.895(50.15)	-	-	-	-	-
CCC-t	0.005(2.46)	0.110(3.78)	0.873(29.11)	4.562(7.92)	0.008(2.18)	0.083(4.53)	0.906(46.49)	11.764(3.48)	-	-	-	6.865(9.94)
DCC-EWMA-n	0.011(4.42)	0.144(6.12)	0.807(29.46)	-	0.008(2.46)	0.094(5.57)	0.895(50.15)	-	-	-	-	-
DCC-GARCH-n	0.011(2.17)	0.144(2.91)	0.807(15.47)	-	0.008(1.99)	0.094(3.61)	0.895(33.59)	-	0.023(2.37)	0.936(42.36)	-	-
DCC-GARCH-t	0.005(2.05)	0.110(2.88)	0.873(21.96)	4.562(8.10)	0.008(2.06)	0.083(3.71)	0.906(38.79)	11.764(3.72)	0.022(2.01)	0.931(40.14)	-	6.907(8.83)
aDCC-GARCH-n	0.011(2.17)	0.144(2.92)	0.807(15.48)	-	0.008(1.99)	0.094(3.61)	0.895(33.60)	-	0.022(2.10)	0.937(41.41)	0.002(0.19)	-
aDCC-GARCH-t	0.005(2.05)	0.110(2.88)	0.873(21.95)	4.562(7.99)	0.008(2.06)	0.083(3.71)	0.906(38.79)	11.764(3.72)	0.022(1.89)	0.931(39.89)	0.000(0.01)	6.907(8.78)

t-statistics are shown in parentheses. The critical value is approximately 1.96 at the 5% significance level. t statistics for GARCH(1,1) parameters are based on MLE with univariate log likelihood functions for CCC-n, CCC-t and DCC-EWMA-n, and MLE with bivariate log likelihood functions for the other models.

Table 13: Estimation results for bivariate models, front-year

	GARCH(1,1)				GARCH(1,1)				CCC/DCC			
Model	ω_{GER}	α_{GER}	β_{GER}	υ_{GER}	ω_{NOR}	α_{NOR}	β_{NOR}	υ_{NOR}	α	β	γ	υ
CCC-n	0.001(2.53)	0.112(4.95)	0.883(40.23)	-	0.002(1.81)	0.084(4.63)	0.910(47.10)	-	-	-	-	-
CCC-t	0.001(1.70)	0.092(4.52)	0.907(47.43)	7.409(5.34)	0.002(2.04)	0.088(4.41)	0.901(41.91)	12.597(3.68)	-	-	-	9.460(7.15)
DCC-EWMA-n	0.001(2.53)	0.112(4.95)	0.883(40.23)	-	0.002(1.81)	0.084(4.63)	0.910(47.10)	-	-	-	-	-
DCC-GARCH-n	0.001(1.73)	0.112(3.00)	0.883(24.68)	-	0.002(1.04)	0.084(2.25)	0.910(21.73)	-	0.026(2.95)	0.954(61.26)	-	-
DCC-GARCH-t	0.001(1.49)	0.092(3.66)	0.907(38.69)	7.409(5.55)	0.002(1.75)	0.088(3.31)	0.901(30.66)	12.597(3.29)	0.034(2.79)	0.939(36.66)	-	9.425(6.94)
aDCC-GARCH-n	0.001(1.74)	0.112(3.01)	0.883(24.72)	-	0.002(1.04)	0.084(2.25)	0.910(21.75)	-	0.023(2.56)	0.949(48.15)	0.013(1.00)	-
aDCC-GARCH-t	0.001(1.49)	0.092(3.67)	0.907(38.69)	7.409(5.48)	0.002(1.75)	0.088(3.31)	0.901(30.67)	12.597(3.28)	0.032(2.63)	0.933(29.49)	0.012(0.69)	9.625(6.92)

t-statistics are shown in parentheses. The critical value is approximately 1.96 at the 5% significance level. t statistics for GARCH(1,1) parameters are based on MLE with univariate log likelihood functions for CCC-n, CCC-t and DCC-EWMA-n, and MLE with bivariate log likelihood functions for the other models.

8.1.5 Volatility and correlation-adjusted quantile regression

Table 14: Estimation results for volatility and correlation-adjusted quantile regression

Data - quantile (position)	$\hat{\beta}_0^p$	$\hat{\beta}_1^p$	$\hat{\beta}_2^p$	$\hat{\beta}_3^p$
Front-quarter - 95‘% (long)	-0.492 (-2.91)	0.212 (0.74)	-1.239 (-6.11)	0.526 (1.65)
Front quarter - 95% (short)	0.243 (1.27)	0.084 (0.19)	1.564 (7.11)	-0.654 (-2.28)
Front-quarter - 99% (long)	-0.422 (-1.84)	-0.337 (-0.37)	-1.751 (-2.92)	0.469 (1.96)
Front quarter - 99% (short)	-0.079 (-0.18)	-0.346 (-0.27)	2.850 (3.10)	-0.211 (-0.33)
Front-year - 95% (long)	-0.265 (-2.15)	-0.092 (-0.20)	-1.181 (-3.31)	0.321 (1.66)
Front-year - 95% (short)	0.168 (2.07)	0.197 (0.67)	1.219 (4.36)	-0.344 (-2.64)
Front-year - 99% (long)	-0.338 (-1.86)	0.289 (0.45)	-1.898 (-2.78)	0.240 (0.80)
Front-year - 99% (short)	0.589 (4.68)	1.013 (1.66)	1.330 (1.93)	-1.022 (-5.12)

t-statistics are shown in parentheses. The critical value is approximately 1.96 at the 5% significance level.

In Table 14, we show the output of the volatility and correlation-adjusted quantile regression model. In this model, $\hat{\beta}_1^p$ is a measure of how VaR reacts with respect to changes in RiskMetrics volatility of daily P&L on the German contract, $\hat{\beta}_2^p$ is the same for daily P&L on the Nordic contract, and $\hat{\beta}_3^p$ shows how VaR is affected by changes in conditional correlation obtained by DCC-EWMA. We observe that all $\hat{\beta}_1^p$ parameters are insignificant, whereas all $\hat{\beta}_2^p$ parameters are significant with the exception of the short 99% quantile for front-year data. Furthermore, $\hat{\beta}_2^p$ parameters are a lot higher than $\hat{\beta}_1^p$ parameters, which may imply that volatility of the Nordic contract is a greater determinant of VaR than volatility of the German contract. This coincides well with the observation that standard deviation in the daily P&L of the Nordic contract is higher than standard deviation in the daily P&L of the German contract, see Table 4. The correlation parameter, $\hat{\beta}_3^p$, is significant for all short quantiles excluding 99% front-quarter, but is insignificant for all long quantiles. We also note that short VaR, which is a positive number, appears to be negatively related to conditional correlation based on this output.

8.1.6 Information criteria

In Table 15, we present AIC and BIC values as well as maximum log likelihood values of the models which are estimated using MLE. We omit the CCC models and DCC-EWMA as conditional correlation is not estimated by MLE, even though the models include GARCH(1,1) for each leg of the spread which indeed is estimated by MLE. The improvements offered by the Student t distribution over the normal distribution is evident. Furthermore, as already verified by LR tests, the asymmetric DCC-GARCH models do not offer a significantly better fit compared to their symmetric equivalents. GJR models show lower AIC and BIC values compared to GARCH models for front-quarter data, and the opposite holds for front-year data. This was previously verified by examining the significance of the γ parameters. According to AIC, MSGARCH models and DCC-GARCH-t give the best fit among the univariate models and bivariate models, respectively. On the other hand, among the univariate models, BIC tells us that GJR(1,1) offer the best fit on front-quarter data while GARCH(1,1)-t offers the best fit on front-year data. Among the bivariate models, DCC-GARCH-t gives the lowest statistic. See Appendix B.2 for an explanation of the AIC and BIC criteria.

Table 15: Information criteria and maximum log likelihood values

	Front-quarter			Front-year		
Model	AIC	BIC	LL	AIC	BIC	LL
GARCH(1,1)-n	2.106	2.118	-1327.7	0.625	0.637	-399.8
GARCH(1,1)-t	2.092	2.109	-1318.3	0.609	0.625	-388.4
GARCH(1,1)-skew-t*	-	-	-	-	-	-
GJR(1,1)-n	2.096	2.113	-1320.8	0.626	0.642	-399.4
GJR(1,1)-t	2.086	2.106	-1313.3	0.611	0.631	-388.4
GJR(1,1)-skew-t*	-	-	-	-	-	-
MSGARCH(1,1)-n	2.099	2.132	-1318.9	0.603	0.635	-380.3
MSGARCH(1,1)-t	2.084	2.121	-1308.0	0.599	0.635	-376.7
DCC-GARCH-n	3.139	3.176	-1975.0	1.181	1.217	-751.6
DCC-GARCH-t	3.021	3.070	-1897.3	1.130	1.178	-715.9
aDCC-GARCH-n	3.141	3.182	-1975.0	1.181	1.221	-750.8
aDCC-GARCH-t	3.023	3.076	-1897.3	1.131	1.183	-715.6

Note that log likelihood functions have different distributions. * denotes that the model is estimated using QMLE. Hence, AIC, BIC and LL values are not directly comparable to the other models and are therefore omitted.

8.2 Backtesting results

In this section, we assess the predictive ability of our selection of models by analyzing VaR and ES backtesting results. We only report out-of-sample results, and not in-sample results, as we want to evaluate the models on the basis of how well they forecast VaR and ES. We use three criteria to evaluate the VaR forecasts - the unconditional coverage test, the conditional coverage test and the dynamic quantile test. ES forecasts are evaluated by the ES test of McNeil and Frey (2000). A proper tail risk model should not be rejected according to any of the four tests. However, we emphasize that correct unconditional and conditional coverage is regarded as most important, meaning that the number of exceedances should be as close as possible to the number implied by the VaR quantile and that we observe no clustering of exceedances. If a model fails on either of these two criteria, the performance on the dynamic quantile test and the ES test is of less relevance. For both front-quarter and front-year data, we report results for confidence levels 95% and 99% for long and short positions. Moreover, we show results using both approach (I) with a fixed estimation window and approach (II) with reestimation. The tables reporting the results show test statistics along with p-values for the three VaR criteria, as well as p-values and bootstrapped p-values for the ES test. Bold numbers indicate the highest p-values according to each criteria, meaning the highest performance on the specific test, for each quantile. Rejection at the 10%, 5% and 1% significance level is denoted by *, ** and ***, respectively. We stress that several models have very similar p-values on the different criteria. Hence, there can be marginal differences in performance between a model which is highlighted in bold and other models. We are interested in the best overall models, and we will therefore base our analysis on overall backtesting results rather than specific test statistics at specific quantiles.

We note that GARCH(1,1)-EVT is not included in the 95% confidence level tables as we only model the 50 most extreme observations. Furthermore, the two quantile regression models are only used to estimate VaR and not ES, and RiskMetrics is omitted from the forecasting with reestimation tables as the model cannot be reestimated when λ is fixed to 0.94.

8.2.1 Front-quarter data

Table 16 shows backtesting results for 95% VaR and ES forecasts obtained using a fixed estimation window for front-quarter data, whereas Table 17 shows the same for 99% VaR and ES forecasts. At the 95% confidence level, we observe that the conditional correlation models fail to predict the VaR quantile which results in too few exceedances. The only exception is CCC-t, which interestingly passes all four backtests both at the long and short quantile. From the actual number of VaR exceedances, it is evident that the short quantile is more difficult to model than the long quantile. This is an expected result, as the most extreme daily P&L occurrences in the out-of-sample window are positive values. A visual inspection of Figure 3a verifies this observation. Moreover, we observe that the GJR models outperform GARCH models with the same distributional assumption at the 95% confidence level when forecasting with fixed estimation window. This is consistent with the significance of the asymmetry parameter, γ, established in the analysis of estimation results for front-quarter data.

From the results reported for the 99% confidence level in Table 17, we see that the models in general perform better on the three VaR backtests compared to the 95% confidence level. This result may seem surprising, and indicates that when using forecasting without reestimating, predictive ability is better on the more extreme 99% long and short quantiles. However, different performance at the two confidence levels can be explained by studying the empirical quantiles of the daily P&L of the front-quarter spread. In Table 5, we observe that, for the out-of-sample period, the absolute value of the short 99% quantile exceeds the long 99% quantile. The opposite is the case for the 95% front-quarter empirical quantiles in 2018 (see Table 31), where we observe that the absolute value of the long 95% quantile exceeds the short 95% quantile. In summary, this implies that it is difficult to achieve simultaneous forecasting accuracy at the 95% and 99% confidence levels, and can consequently serve as a possible explanation for why the backtesting results differ. Moreover, from the backtesting results in Table 17, we see that RiskMetrics does not provide the correct unconditional and conditional coverage. This is consistent with the academic literature, which commonly reports that RiskMetrics fails to predict extreme tail risk. At the 99% confidence level, we also observe that the DCC models show better performance on the VaR backtests. They are however largely rejected on the ES test, implying that the mean of the excess VaR violations is greater than zero. Furthermore, the improvements offered by a Student t distributional assumption relative to the normal distribution, is evident from the results. GARCH(1,1)-EVT predicts the exact number of expected exceedances both on the long and short quantile, and interestingly performs very well on the backtests. Also, as for the 95% confidence level, CCC-t shows good overall performance. Figure 8 and 9 show the 99% VaR and ES forecasts across the out-of-sample window.

To address the limitations imposed by forecasting approach (I), we consider the alternative estimation procedure (II) which includes reestimation every 20th day. Table 18 presents backtesting results for 95% VaR and ES forecasts, while Table 19 shows the same for 99% VaR and ES forecasts. By comparing the two forecasting approaches, quite mixed results emerge. At the 95% confidence level, we see significant improvements in forecasting performance across all models when introducing reestimation. On the contrary, at the 99% confidence level, we observe that forecasting performance is worsened when including reestimation. These mixed results illustrate the challenges in forecasting the out-of-sample data which exhibit different data characteristics compared to the in-sample data.

At the 95% confidence level (Table 18), we observe that the Markov switching GARCH models and DCC models show the worst performance in the model set as they produce too high VaR forecasts leading to fewer exceedances than expected. Moreover, models assuming Student t or skewed Student t outperform their normal distribution equivalents. High-performing models include GARCH and GJR

with symmetric and skewed Student t, as well as volatility-adjusted quantile regression and CCC-t.

At the 99% confidence level (Table 19), MSGARCH models perform well on the VaR backtests. This indicates that, at this specific confidence level, allowing for two volatility regimes can be beneficial for forecasting VaR. Nevertheless, these models are rejected on the ES test, implying that the ES forecasts are not unbiased estimates of the expectation in the tail beyond the VaR. We also observe that the conditional correlation models with Student t distribution show good backtesting performance. As an interesting note, we see that the proposed volatility and correlation-adjusted quantile regression has an extremely high statistic on the dynamic quantile test. This is a consequence of the model severely failing to model the extreme market movement in September 2018. Rejection on the dynamic quantile test means that the probability of a VaR exceedance is not independent of all past information. Figure 12 and 13 show the 99% VaR and ES forecasts using reestimation across the out-of-sample window.

8.2.2 Front-year data

Table 20 shows backtesting results for 95% VaR and ES forecasts obtained using a fixed estimation window for front-year data, whereas Table 21 shows the same for 99% VaR and ES forecasts. At the 95% confidence level, we note better overall model performance for the short quantile compared to the long quantile. This can be attributed to the positive skewness observed for front-year data. The conditional correlation models accurately predict the short VaR quantile, but are consistently rejected according to the unconditional and conditional coverage tests for the long quantile. Overall, the univariate models offer better performance on the backtests.

At the 99% confidence level (Table 21), all models, excluding RiskMetrics and DCC-EWMA, pass the VaR backtests on both the long and short quantiles. The observation that these two models fail to produce accurate 99% VaR forecasts is an expected result, as they fail to capture volatility clustering.

Forecasting approach (II) with reestimation every 20th day produce mixed results compared to the fixed estimation window approach. Table 22 presents backtesting results for 95% VaR and ES forecasts with reestimation, while Table 23 shows the same for 99% VaR and ES forecasts. At the 95% confidence level, we observe improved predictive ability for univariate models for the long quantile when including reestimation. This is especially evident for volatility-adjusted quantile regression. The effect of changing forecasting approach is however ambiguous for the conditional correlation models. DCC models are accurate with both forecasting procedures at this confidence level, while the CCC models benefits from using a fixed estimation sample. At the 99% confidence level, using forecasting approach (II) leads to worsened performance for the short quantile and comparable performance for the long quantile. Again, the mixed results from changing forecasting approach illustrate the challenges in forecasting the out-of-sample data which exhibit different data characteristics compared to the in-sample data.

At the 95% confidence level (Table 22), several of the models do not pass the ES test. The GARCH(1,1) models perform well on the VaR backtests on both the long and short quantiles. For the latter quantile, DCC models also display good results.

At the 99% confidence level (Table 23), we note better backtesting results for the long quantile compared to the short quantile for univariate GARCH models. On the other hand, volatility-adjusted quantile regression shows good performance on both quantiles.

8.3 Summary and discussion of results

It follows from the analysis of the backtesting results that no single model specification is clearly preferable. However, there are several important findings which may aid in model selection. As noted initially in Section 8.2, correct unconditional and conditional coverage are the two most important criteria when evaluating the predictive ability of the models. In Table 24 and 25, we show aggregated results for the unconditional and conditional coverage tests with forecasting approach (I) and (II), respectively. In the tables, X denotes rejection at the 5% significance level. By assessing these two tables, we are able to conclude on which models perform the best overall in forecasting VaR in German-Nordic electricity futures spreads, according to the unconditional and condtional coverage tests. The most successful models when forecasting with fixed estimation window are GARCH(1,1)-EVT, GJR models with normal distribution, Student t distribution or skewed Student distribution, as well as CCC-t. We argue that these models show the highest performance as they are not rejected on neither the Kupiec test nor the Christoffersen test for any of the quantiles for both front-quarter and front-year data. For these four highest-performing models, we show aggregated results on the dynamic quantile test and ES test in Table 26. By assessing these results, we can further refine our conclusions, and we find that GARCH(1,1)-EVT, GJR(1,1)-skew-t and CCC-t are the three most successful models when forecasting VaR and ES with fixed estimation window.

By the same reasoning, GARCH(1,1)-t, GARCH(1,1)-skew-t and volatility-adjusted quantile regression are the most successful models when forecasting VaR with reestimation, according to the unconditional and conditional coverage tests. By assessing Table 27, with aggregated dynamic quantile and ES test results for these three models, we arrive upon our final conclusion that GARCH(1,1)-t and GARCH(1,1)-skew-t are the two highest-performing models when forecasting VaR and ES with reestimation.

Apart from the overall conclusions on model performance, several interesting findings have emerged from our results. One of most important findings is that the univariate models show overall better performance compared to the bivariate models. This implies that when forecasting VaR and ES in German-Nordic electricity futures spreads, modelling the spreads as univariate time series accommodate improved predictive ability compared to modelling the two legs of the spreads as well as the correlation. A notable exception among the bivariate models is CCC-t, which consistently show good backtesting results. In the analysis of in-sample estimation results, we established statistically significant α and β parameters for the DCC models, implying that correlation is time-varying. However, from the backtesting results, it is evident that the DCC models fail to accurately forecast VaR and ES. At the 95% confidence level, the DCC models overestimate the risk of a short position in the front-quarter spread and overestimate the risk of a long position in the front-year spread. At these quantiles, the DCC models predict excessively high portfolio variance which in turn leads to overestimation of VaR. We also observe that the asymmetric DCC-GARCH models do not outperform the symmetric equivalents.

From the overall results, we also see that non-normal distributional assumptions produce more accurate VaR and ES forecasts. This is an expected result, as the conditional distribution of the daily P&L of the spreads exhibit excess kurtosis and skewness. Among the univariate GARCH models, we find that GARCH(1,1) and GJR-GARCH(1,1) with Student t and skewed Student t show good predictive ability of VaR and ES. However, we observe only minor differences in backtesting performance from using a skewed version of Student t distribution compared to the symmetric version. The asymmetric parameter in GJR-GARCH, γ, is statistically significant when estimating the models based on front-quarter data, but not significant when estimating based on front-year data. Consequently, the GJR-GARCH model specification outperforms the symmetric variance response GARCH model specification for front-quarter

data, whereas backtesting results are similar for front-year data. Moreover, GARCH(1,1)-EVT, which in our formulation requires no numerical optimization as we use the Hill estimator, shows good predictive ability. This model also shows consistently good performance on the ES test.

The Markov switching GARCH models show considerably better backtesting performance at the 99% quantiles compared to the 95% quantiles. We find that specifying two volatility regimes does not improve forecasting accuracy compared to more simplistic GARCH models. As for the other univariate GARCH models, we find that Markov switching GARCH performs better with a Student t distributional assumption compared to the normal distribution.

An important result that has emerged from the present study is that volatility-adjusted quantile regression serves well the purpose of forecasting VaR in German-Nordic electricity futures spreads. Among the models showing the highest performance, volatility-adjusted quantile regression is arguably both simplest and most widely used in practice. Furthermore, we find that our proposed volatility and correlation-adjusted quantile regression model does not outperform the simpler specification with only one explanatory variable. Although showing desirable performance on the unconditional and conditional coverage tests, we find that the quantile regression models perform worse on the dynamic quantile test compared to other high-performing models.

Finally, we get mixed results by comparing the two forecasting approaches employed in this study. At the 95% confidence level, many of the models show improved performance when forecasting with reestimation compared to forecasting with fixed estimation window. A similar improvement is however not observed at the 99% confidence level.

Table 16: Backtesting of 95% VaR and ES forecasts with fixed estimation window, front-quarter

	50 expected	Kupiec	Christoffersen	Engle & Manganelli	ES test
Model	Exceedances	$-2ln(LR_{UC})$(p-val)	$-2ln(LR_{CC})$(p-val)	DQ(p-val)	b.p-val(p-val)
Long:					
RiskMetrics-n	49	0.021(0.884)	**0.174(0.917)**	2.181(0.949)	0.059*(0.020)**
Vol-adj qreg	42	1.421(0.233)	2.211(0.331)	4.377(0.735)	-
GARCH(1,1)-n	40	2.253(0.133)	2.354(0.308)	3.794(0.803)	0.150(0.078)*
GARCH(1,1)-t	42	1.421(0.233)	1.454(0.483)	3.293(0.857)	0.480(0.439)
GARCH(1,1)-skew t	43	1.081(0.299)	1.093(0.579)	3.292(0.857)	0.368(0.294)
GARCH(1,1)-EVT	-	-	-	-	-
MSGARCH(1,1)-n	41	1.812(0.178)	2.161(0.339)	4.511(0.719)	0.166(0.092)*
MSGARCH(1,1)-t	45	0.544(0.461)	0.544(0.762)	3.699(0.814)	0.352(0.272)
GJR(1,1)-n	42	1.421(0.233)	1.845(0.397)	2.395(0.935)	0.056*(0.020)**
GJR(1,1)-t	44	0.788(0.375)	1.384(0.501)	4.689(0.698)	0.284(0.198)
GJR(1,1)-skew t	47	0.193(0.660)	0.216(0.898)	**0.396(1.000)**	0.225(0.142)
Vol-corr-adj qreg	50	**0.000(1.000)**	0.855(0.652)	4.554(0.714)	-
CCC-n	39	2.747(0.097)*	2.896(0.235)	7.452(0.383)	0.165(0.091)*
CCC-t	41	1.812(0.178)	1.874(0.392)	5.369(0.615)	**0.682(0.726)**
DCC-EWMA-n	43	1.081(0.299)	1.093(0.579)	3.791(0.803)	0.112(0.051)*
DCC-GARCH-n	32	7.777(0.005)***	7.777(0.020)**	10.723(0.151)	0.144(0.074)*
DCC-GARCH-t	34	6.043(0.014)**	6.067(0.048)**	8.998(0.253)	0.673(0.717)
aDCC-GARCH-n	33	6.878(0.009)***	6.887(0.032)**	10.667(0.154)	0.157(0.081)*
aDCC-GARCH-t	34	6.043(0.014)**	6.067(0.048)**	8.998(0.253)	0.673(0.717)
Short:					
RiskMetrics-n	44	**0.788(0.375)**	4.844(0.089)*	5.908(0.551)	0.016**(0.012)**
Vol-adj qreg	43	1.081(0.299)	4.950(0.084)*	5.708(0.574)	-
GARCH(1,1)-n	38	3.294(0.070)*	6.300(0.043)**	7.405(0.388)	0.060*(0.040)**
GARCH(1,1)-t	38	3.294(0.070)*	6.300(0.043)**	7.399(0.389)	0.140(0.086)*
GARCH(1,1)-skew t	38	3.294(0.070)*	6.300(0.043)**	7.405(0.388)	0.207(0.129)
GARCH(1,1)-EVT	-	-	-	-	-
MSGARCH(1,1)-n	33	6.878(0.009)***	9.133(0.010)**	8.782(0.269)	0.079*(0.049)**
MSGARCH(1,1)-t	37	3.895(0.048)**	6.742(0.034)**	8.670(0.277)	0.234(0.144)
GJR(1,1)-n	40	2.253(0.133)	2.534(0.282)	5.104(0.647)	0.019**(0.016)**
GJR(1,1)-t	41	1.812(0.178)	**2.161(0.339)**	**4.689(0.698)**	0.075*(0.047)**
GJR(1,1)-skew t	40	2.253(0.133)	2.534(0.282)	5.104(0.647)	0.087*(0.057)*
Vol-corr-adj qreg	41	1.812(0.178)	**2.161(0.339)**	7.891(0.342)	-
CCC-n	37	3.895(0.048)**	6.742(0.034)**	6.288(0.507)	0.035**(0.019)**
CCC-t	40	2.253(0.133)	2.534(0.282)	6.599(0.472)	0.228(0.142)
DCC-EWMA-n	37	3.895(0.048)**	6.742(0.034)**	7.788(0.352)	0.023**(0.019)**
DCC-GARCH-n	31	8.739(0.003)***	10.725(0.005)***	9.806(0.200)	0.051*(0.034)**
DCC-GARCH-t	34	6.043(0.014)**	8.439(0.015)**	8.678(0.277)	**0.278(0.178)**
aDCC-GARCH-n	32	7.777(0.005)***	9.895(0.007)***	9.938(0.192)	0.056*(0.036)**
aDCC-GARCH-t	34	6.043(0.014)**	8.439(0.015)**	8.678(0.277)	**0.278(0.178)**

*, **, *** indicates rejection at the 10%, 5% and 1% significance level, respectively. The decision is based on critical values from the χ^2 distribution with 1 degree of freedom for $-2ln(LR_{UC})$, 2 degrees of freedom for $-2ln(LR_{CC})$ and 7 degrees of freedom for the DQ, respectively. For the ES test, the bootstrapped p-value is given. p-values are shown in parentheses. Bold numbers indicate highest p-value.

Table 17: Backtesting of 99% VaR and ES forecasts with fixed estimation window, front-quarter

	10 expected	Kupiec	Christoffersen	Engle & Manganelli	ES test
Model	Exceedances	$-2ln(LR_{UC})$(p-val)	$-2ln(LR_{CC})$(p-val)	DQ(p-val)	b.p-val(p-val)
Long:					
RiskMetrics-n	17	4.091(0.043)**	4.680(0.096)*	10.973(0.140)	0.120(0.069)*
Vol-adj qreg	11	0.098(0.754)	0.343(0.842)	3.429(0.843)	-
GARCH(1,1)-n	13	0.831(0.362)	1.173(0.556)	3.016(0.883)	0.182(0.114)
GARCH(1,1)-t	10	**0.000(1.000)**	**0.202(0.904)**	1.520(0.982)	0.425(0.328)
GARCH(1,1)-skew t	11	0.098(0.754)	0.343(0.842)	2.025(0.958)	0.368(0.271)
GARCH(1,1)-EVT	10	**0.000(1.000)**	**0.202(0.904)**	1.521(0.982)	0.721(0.744)
MSGARCH(1,1)-n	10	**0.000(1.000)**	**0.202(0.904)**	3.149(0.871)	0.154(0.091)*
MSGARCH(1,1)-t	11	0.098(0.754)	0.343(0.842)	3.232(0.863)	0.395(0.320)
GJR(1,1)-n	15	2.189(0.139)	2.647(0.266)	4.691(0.698)	0.095*(0.056)*
GJR(1,1)-t	12	0.380(0.538)	0.672(0.715)	2.486(0.928)	0.482(0.384)
GJR(1,1)-skew t	13	0.831(0.362)	1.173(0.556)	2.446(0.931)	0.347(0.247)
Vol-corr-adj qreg	11	0.098(0.754)	0.343(0.842)	1.487(0.983)	-
CCC-n	11	0.098(0.754)	0.343(0.842)	2.179(0.949)	0.093*(0.048)**
CCC-t	10	**0.000(1.000)**	**0.202(0.904)**	0.788(0.998)	0.804(0.865)
DCC-EWMA-n	12	0.380(0.538)	0.672(0.715)	4.507(0.720)	0.188(0.120)
DCC-GARCH-n	11	0.098(0.754)	0.343(0.842)	3.060(0.879)	0.365(0.256)
DCC-GARCH-t	8	0.434(0.510)	0.563(0.755)	**0.786(0.998)**	0.805(0.864)
aDCC-GARCH-n	11	0.098(0.754)	0.343(0.842)	1.245(0.990)	0.354(0.239)
aDCC-GARCH-t	8	0.434(0.510)	0.563(0.755)	1.860(0.967)	**0.860(0.928)**
Short:					
RiskMetrics-n	18	5.225(0.022)**	5.886(0.053)*	14.957(0.037)**	0.034**(0.032)**
Vol-adj qreg	13	0.831(0.362)	1.173(0.556)	5.326(0.620)	-
GARCH(1,1)-n	12	0.380(0.538)	0.672(0.715)	2.137(0.952)	0.041**(0.038)**
GARCH(1,1)-t	10	**0.000(1.000)**	**0.202(0.904)**	2.047(0.957)	0.069*(0.059)*
GARCH(1,1)-skew t	10	**0.000(1.000)**	**0.202(0.904)**	2.063(0.956)	0.084*(0.067)*
GARCH(1,1)-EVT	10	**0.000(1.000)**	**0.202(0.904)**	2.063(0.956)	0.156(0.107)
MSGARCH(1,1)-n	10	**0.000(1.000)**	**0.202(0.904)**	1.697(0.975)	0.045**(0.039)**
MSGARCH(1,1)-t	10	**0.000(1.000)**	**0.202(0.904)**	1.190(0.991)	0.082*(0.066)*
GJR(1,1)-n	15	2.189(0.139)	2.647(0.266)	5.872(0.555)	0.097*(0.066)*
GJR(1,1)-t	10	**0.000(1.000)**	**0.202(0.904)**	2.486(0.928)	0.069*(0.059)*
GJR(1,1)-skew t	9	0.105(0.746)	0.268(0.875)	1.933(0.963)	0.067*(0.058)*
Vol-corr-adj qreg	15	2.189(0.139)	2.647(0.266)	8.963(0.255)	-
CCC-n	15	2.189(0.139)	2.647(0.266)	9.582(0.214)	0.068*(0.049)**
CCC-t	10	**0.000(1.000)**	**0.202(0.904)**	1.459(0.984)	**0.284(0.199)**
DCC-EWMA-n	13	0.831(0.362)	1.173(0.556)	3.040(0.881)	0.054*(0.046)**
DCC-GARCH-n	10	**0.000(1.000)**	**0.202(0.904)**	3.061(0.879)	0.043**(0.043)**
DCC-GARCH-t	9	0.105(0.746)	0.268(0.875)	**0.846(0.997)**	0.232(0.149)
aDCC-GARCH-n	10	**0.000(1.000)**	**0.202(0.904)**	1.254(0.990)	0.041**(0.042)**
aDCC-GARCH-t	7	1.016(0.314)	1.114(0.573)	1.283(0.989)	0.151(0.102)

*, **, *** indicates rejection at the 10%, 5% and 1% significance level, respectively. The decision is based on critical values from the χ^2 distribution with 1 degree of freedom for $-2ln(LR_{UC})$, 2 degrees of freedom for $-2ln(LR_{CC})$ and 7 degrees of freedom for the DQ, respectively. For the ES test, the bootstrapped p-value is given. p-values are shown in parentheses. Bold numbers indicate highest p-value.

Table 18: Backtesting of 95% VaR and ES forecasts with reestimation, front-quarter

	50 expected	Kupiec	Christoffersen	Engle & Manganelli	ES test
Model	Exceedances	$-2ln(LR_{UC})$(p-val)	$-2ln(LR_{CC})$(p-val)	DQ(p-val)	b.p-val(p-val)
Long:					
RiskMetrics-n	-	-	-	-	-
Vol-adj qreg	52	0.083(0.773)	0.691(0.708)	3.407(0.845)	-
GARCH(1,1)-n	46	0.346(0.557)	1.836(0.399)	4.043(0.775)	0.055*(0.018)**
GARCH(1,1)-t	51	0.021(0.885)	0.747(0.688)	4.454(0.726)	0.578(0.588)
GARCH(1,1)-skew t	50	**0.000(1.000)**	0.855(0.652)	2.158(0.951)	0.414(0.362)
GARCH(1,1)-EVT	-	-	-	-	-
MSGARCH(1,1)-n	41	1.812(0.178)	2.161(0.339)	4.511(0.719)	0.166(0.092)*
MSGARCH(1,1)-t	45	0.544(0.461)	0.544(0.762)	3.699(0.814)	0.352(0.272)
GJR(1,1)-n	49	0.021(0.884)	0.174(0.917)	**1.821(0.969)**	0.065*(0.022)**
GJR(1,1)-t	51	0.021(0.885)	0.189(0.910)	5.285(0.625)	0.558(0.558)
GJR(1,1)-skew t	51	0.021(0.885)	**0.085(0.958)**	2.002(0.960)	0.378(0.308)
Vol-corr-adj qreg	43	1.081(0.299)	1.093(0.579)	6.408(0.493)	-
CCC-n	48	0.085(0.770)	0.297(0.862)	3.606(0.824)	0.057*(0.021)**
CCC-t	51	0.021(0.885)	**0.085(0.958)**	2.932(0.891)	0.551(0.540)
DCC-EWMA-n	51	0.021(0.885)	0.747(0.688)	2.080(0.955)	0.054*(0.017)**
DCC-GARCH-n	38	3.294(0.070)*	3.459(0.177)	9.011(0.252)	0.170(0.096)*
DCC-GARCH-t	44	0.788(0.375)	1.384(0.501)	5.992(0.541)	**0.815(0.890)**
aDCC-GARCH-n	38	3.294(0.070)*	3.459(0.177)	7.772(0.353)	0.158(0.083)*
aDCC-GARCH-t	43	1.081(0.299)	1.587(0.452)	6.003(0.539)	0.779(0.851)
Short:					
RiskMetrics-n	-	-	-	-	-
Vol-adj qreg	50	**0.000(1.000)**	0.104(0.949)	7.159(0.413)	-
GARCH(1,1)-n	44	0.788(0.375)	0.791(0.673)	3.900(0.791)	0.024**(0.016)**
GARCH(1,1)-t	46	0.346(0.557)	0.353(0.838)	3.763(0.807)	0.174(0.105)
GARCH(1,1)-skew t	47	0.193(0.660)	0.475(0.789)	**3.292(0.857)**	0.169(0.101)
GARCH(1,1)-EVT	-	-	-	-	-
MSGARCH(1,1)-n	33	6.878(0.009)***	9.133(0.010)**	8.782(0.269)	0.079*(0.049)**
MSGARCH(1,1)-t	37	3.895(0.048)**	6.742(0.034)**	8.670(0.277)	0.234(0.144)
GJR(1,1)-n	46	0.346(0.557)	0.707(0.702)	4.887(0.674)	0.020**(0.015)**
GJR(1,1)-t	49	0.021(0.884)	**0.100(0.951)**	5.285(0.625)	0.221(0.132)
GJR(1,1)-skew t	48	0.085(0.770)	1.233(0.540)	6.132(0.524)	0.166(0.099)*
Vol-corr-adj qreg	47	0.193(0.660)	0.216(0.898)	7.510(0.378)	-
CCC-n	47	0.193(0.660)	0.216(0.898)	4.902(0.672)	0.023**(0.012)**
CCC-t	50	**0.000(1.000)**	0.119(0.942)	5.018(0.658)	0.212(0.128)
DCC-EWMA-n	42	1.421(0.233)	5.109(0.078)*	4.870(0.676)	0.015**(0.012)**
DCC-GARCH-n	37	3.895(0.048)**	4.014(0.134)	8.938(0.257)	0.055*(0.031)**
DCC-GARCH-t	38	3.294(0.070)*	6.300(0.043)**	5.975(0.543)	0.197(0.119)
aDCC-GARCH-n	39	2.747(0.097)*	2.966(0.227)	8.689(0.276)	0.065*(0.036)**
aDCC-GARCH-t	39	2.747(0.097)*	2.966(0.227)	6.668(0.464)	**0.259(0.168)**

*, **, *** indicates rejection at the 10%, 5% and 1% significance level, respectively. The decision is based on critical values from the χ^2 distribution with 1 degree of freedom for $-2ln(LR_{UC})$, 2 degrees of freedom for $-2ln(LR_{CC})$ and 7 degrees of freedom for the DQ, respectively. For the ES test, the bootstrapped p-value is given. p-values are shown in parentheses. Bold numbers indicate highest p-value.

Table 19: Backtesting of 99% VaR and ES forecasts with reestimation, front-quarter

	10 expected	Kupiec	Christoffersen	Engle & Manganelli	ES test
Model	Exceedances	$-2ln(LR_{UC})$(p-val)	$-2ln(LR_{CC})$(p-val)	DQ(p-val)	b.p-val(p-val)
Long:					
RiskMetrics-n	-	-	-	-	-
Vol-adj qreg	14	1.437(0.231)	1.835(0.399)	3.356(0.850)	-
GARCH(1,1)-n	18	5.225(0.022)**	5.886(0.053)*	13.991(0.051)*	0.262(0.172)
GARCH(1,1)-t	10	**0.000(1.000)**	**0.202(0.904)**	1.512(0.982)	0.471(0.378)
GARCH(1,1)-skew t	13	0.831(0.362)	1.173(0.556)	6.093(0.529)	0.815(0.880)
GARCH(1,1)-EVT	14	1.437(0.231)	1.835(0.399)	6.050(0.534)	**0.937(0.982)**
MSGARCH(1,1)-n	10	**0.000(1.000)**	**0.202(0.904)**	3.149(0.871)	0.154(0.091)*
MSGARCH(1,1)-t	11	0.098(0.754)	0.343(0.842)	3.232(0.863)	0.395(0.320)
GJR(1,1)-n	17	4.091(0.043)**	4.680(0.096)*	13.408(0.063)*	0.255(0.178)
GJR(1,1)-t	12	0.380(0.538)	0.672(0.715)	2.947(0.890)	0.556(0.499)
GJR(1,1)-skew t	11	0.098(0.754)	0.343(0.842)	1.510(0.982)	0.462(0.370)
Vol-corr-adj qreg	12	0.380(0.538)	0.672(0.715)	8.283(0.308)	-
CCC-n	14	1.437(0.231)	1.835(0.399)	3.433(0.842)	0.037**(0.014)**
CCC-t	10	**0.000(1.000)**	**0.202(0.904)**	0.713(0.998)	0.577(0.517)
DCC-EWMA-n	18	5.225(0.022)**	5.886(0.053)*	10.305(0.172)	0.147(0.087)*
DCC-GARCH-n	12	0.380(0.538)	0.672(0.715)	10.359(0.169)	0.132(0.078)*
DCC-GARCH-t	9	0.105(0.746)	0.268(0.875)	**0.639(0.999)**	0.803(0.863)
aDCC-GARCH-n	12	0.380(0.538)	0.672(0.715)	2.451(0.931)	0.128(0.073)*
aDCC-GARCH-t	9	0.105(0.746)	0.268(0.875)	0.718(0.998)	0.828(0.891)
Short:					
RiskMetrics-n	-	-	-	-	-
Vol-adj qreg	15	2.189(0.139)	2.647(0.266)	9.059(0.248)	-
GARCH(1,1)-n	16	3.077(0.079)*	3.597(0.166)	6.901(0.439)	0.050*(0.047)**
GARCH(1,1)-t	12	0.380(0.538)	0.672(0.715)	2.267(0.944)	0.149(0.101)
GARCH(1,1)-skew t	12	0.380(0.538)	0.672(0.715)	2.597(0.920)	0.275(0.182)
GARCH(1,1)-EVT	16	3.077(0.079)*	3.597(0.166)	6.866(0.443)	**0.616(0.588)**
MSGARCH(1,1)-n	10	**0.000(1.000)**	**0.202(0.904)**	1.697(0.975)	0.045**(0.039)**
MSGARCH(1,1)-t	10	**0.000(1.000)**	**0.202(0.904)**	**1.190(0.991)**	0.083*(0.066)*
GJR(1,1)-n	18	5.225(0.022)**	5.886(0.053)*	10.301(0.172)	0.074*(0.056)*
GJR(1,1)-t	13	0.831(0.362)	1.173(0.556)	2.947(0.890)	0.184(0.117)
GJR(1,1)-skew t	14	1.437(0.231)	1.835(0.399)	3.319(0.854)	0.268(0.172)
Vol-corr-adj qreg	15	2.189(0.139)	3.704(0.157)	36.343(0.000)***	-
CCC-n	18	5.225(0.022)**	5.886(0.053)*	12.141(0.096)*	0.023**(0.023)**
CCC-t	14	1.437(0.231)	1.835(0.399)	9.389(0.226)	0.337(0.227)
DCC-EWMA-n	18	5.225(0.022)**	5.886(0.053)*	10.359(0.169)	0.084*(0.061)*
DCC-GARCH-n	14	1.437(0.231)	1.835(0.399)	10.503(0.162)	0.089*(0.064)*
DCC-GARCH-t	10	**0.000(1.000)**	**0.202(0.904)**	1.389(0.986)	0.325(0.224)
aDCC-GARCH-n	14	1.437(0.231)	1.835(0.399)	10.489(0.163)	0.082*(0.061)*
aDCC-GARCH-t	10	**0.000(1.000)**	**0.202(0.904)**	10.767(0.149)	0.395(0.315)

*, **, *** indicates rejection at the 10%, 5% and 1% significance level, respectively. The decision is based on critical values from the χ^2 distribution with 1 degree of freedom for $-2ln(LR_{UC})$, 2 degrees of freedom for $-2ln(LR_{CC})$ and 7 degrees of freedom for the DQ, respectively. For the ES test, the bootstrapped p-value is given. p-values are shown in parentheses. Bold numbers indicate highest p-value.

Table 20: Backtesting of 95% VaR and ES forecasts with fixed estimation window, front-year

	50 expected	Kupiec	Christoffersen	Engle & Manganelli	ES test
Model	Exceedances	$-2ln(LR_{UC})$(p-val)	$-2ln(LR_{CC})$(p-val)	DQ(p-val)	b.p-val(p-val)
Long:					
RiskMetrics-n	40	2.253(0.133)	2.354(0.308)	4.572(0.712)	0.090*(0.039)**
Vol-adj qreg	29	10.867(0.001)***	12.601(0.002)***	10.412(0.166)	-
GARCH(1,1)-n	41	1.812(0.178)	1.874(0.392)	3.121(0.874)	0.371(0.283)
GARCH(1,1)-t	43	1.081(0.299)	**1.093(0.579)**	**1.927(0.964)**	0.672(0.697)
GARCH(1,1)-skew t	41	1.812(0.178)	1.874(0.392)	3.121(0.874)	**0.723(0.772)**
GARCH(1,1)-EVT	-	-	-	-	-
MSGARCH(1,1)-n	41	1.812(0.178)	5.322(0.070)*	6.852(0.444)	0.192(0.113)
MSGARCH(1,1)-t	52	**0.083(0.773)**	1.625(0.444)	5.230(0.632)	0.410(0.336)
GJR(1,1)-n	39	2.747(0.097)*	5.917(0.052)*	5.577(0.590)	0.191(0.113)
GJR(1,1)-t	42	1.421(0.233)	1.454(0.483)	6.483(0.485)	0.626(0.625)
GJR(1,1)-skew t	39	2.747(0.097)*	5.917(0.052)*	5.577(0.590)	0.556(0.526)
Vol-corr-adj qreg	33	6.878(0.009)***	9.133(0.010)**	8.589(0.284)	-
CCC-n	37	3.895(0.048)**	5.492(0.064)*	6.548(0.477)	0.206(0.122)
CCC-t	38	3.294(0.070)*	4.703(0.095)*	5.117(0.646)	0.442(0.357)
DCC-EWMA-n	37	3.895(0.048)**	4.014(0.134)	7.260(0.402)	0.179(0.101)
DCC-GARCH-n	31	8.739(0.003)***	8.741(0.013)**	8.857(0.263)	0.154(0.081)*
DCC-GARCH-t	32	7.777(0.005)***	7.777(0.020)**	7.862(0.345)	0.393(0.311)
aDCC-GARCH-n	34	6.043(0.014)**	6.067(0.048)**	4.948(0.666)	0.208(0.123)
aDCC-GARCH-t	33	6.878(0.009)***	6.887(0.032)**	7.714(0.358)	0.390(0.307)
Short:					
RiskMetrics-n	52	0.083(0.773)	0.691(0.708)	8.763(0.270)	0.037**(0.051)*
Vol-adj qreg	59	1.616(0.204)	2.282(0.319)	12.709(0.080)*	-
GARCH(1,1)-n	46	0.346(0.557)	0.353(0.838)	5.046(0.654)	0.047**(0.063)*
GARCH(1,1)-t	48	0.085(0.770)	0.297(0.862)	5.132(0.644)	0.150(0.114)
GARCH(1,1)-skew t	51	0.021(0.885)	0.747(0.688)	6.417(0.492)	0.136(0.108)
GARCH(1,1)-EVT	-	-	-	-	-
MSGARCH(1,1)-n	44	0.788(0.375)	1.384(0.501)	3.869(0.795)	0.039**(0.059)*
MSGARCH(1,1)-t	47	0.193(0.660)	1.095(0.578)	3.961(0.784)	0.067*(0.075)*
GJR(1,1)-n	48	0.085(0.770)	1.233(0.540)	6.638(0.467)	0.042**(0.064)*
GJR(1,1)-t	49	0.021(0.884)	1.017(0.601)	6.483(0.485)	0.152(0.117)
GJR(1,1)-skew t	48	0.085(0.770)	1.233(0.540)	6.638(0.467)	0.078*(0.083)*
Vol-corr-adj qreg	64	3.805(0.051)*	4.018(0.134)	11.381(0.123)	-
CCC-n	51	0.021(0.885)	**0.085(0.958)**	4.042(0.775)	0.181(0.125)
CCC-t	52	0.083(0.773)	0.117(0.943)	3.927(0.788)	0.396(0.232)
DCC-EWMA-n	49	0.021(0.884)	0.174(0.917)	5.464(0.603)	0.057*(0.072)*
DCC-GARCH-n	47	0.193(0.660)	0.216(0.898)	7.090(0.420)	0.284(0.178)
DCC-GARCH-t	49	0.021(0.884)	0.174(0.917)	4.899(0.672)	**0.447(0.286)**
aDCC-GARCH-n	50	**0.000(1.000)**	0.119(0.942)	3.712(0.812)	0.252(0.160)
aDCC-GARCH-t	49	0.021(0.884)	0.100(0.951)	**3.602(0.824)**	0.420(0.262)

*, **, *** indicates rejection at the 10%, 5% and 1% significance level, respectively. The decision is based on critical values from the χ^2 distribution with 1 degree of freedom for $-2ln(LR_{UC})$, 2 degrees of freedom for $-2ln(LR_{CC})$ and 7 degrees of freedom for the DQ, respectively. For the ES test, the bootstrapped p-value is given. p-values are shown in parentheses. Bold numbers indicate highest p-value.

Table 21: Backtesting of 99% VaR and ES forecasts with fixed estimation window, front-year

	10 expected	Kupiec	Christoffersen	Engle & Manganelli	ES test
Model	Exceedances	$-2ln(LR_{UC})$(p-val)	$-2ln(LR_{CC})$(p-val)	DQ(p-val)	b.p-val(p-val)
Long:					
RiskMetrics-n	13	0.831(0.362)	1.173(0.556)	5.979(0.542)	0.047**(0.031)**
Vol-adj qreg	9	0.105(0.746)	0.268(0.875)	1.069(0.994)	-
GARCH(1,1)-n	8	0.434(0.510)	0.563(0.755)	0.999(0.995)	0.014**(0.027)**
GARCH(1,1)-t	8	0.434(0.510)	0.563(0.755)	1.012(0.995)	0.308(0.193)
GARCH(1,1)-skew t	8	0.434(0.510)	0.563(0.755)	0.999(0.995)	0.423(0.285)
GARCH(1,1)-EVT	8	0.434(0.510)	0.563(0.755)	0.999(0.995)	0.574(0.478)
MSGARCH(1,1)-n	10	**0.000(1.000)**	**0.202(0.904)**	2.089(0.955)	0.118(0.065)*
MSGARCH(1,1)-t	9	0.105(0.746)	0.268(0.875)	1.163(0.992)	0.062*(0.040)**
GJR(1,1)-n	9	0.105(0.746)	0.268(0.875)	1.169(0.992)	0.025**(0.031)**
GJR(1,1)-t	8	0.434(0.510)	0.563(0.755)	1.938(0.963)	0.304(0.189)
GJR(1,1)-skew t	8	0.434(0.510)	0.563(0.755)	1.055(0.994)	0.352(0.233)
Vol-corr-adj qreg	7	1.016(0.314)	1.114(0.573)	3.290(0.857)	-
CCC-n	10	**0.000(1.000)**	**0.202(0.904)**	8.712(0.274)	0.139(0.075)*
CCC-t	6	1.886(0.170)	1.959(0.376)	1.746(0.973)	0.146(0.077)*
DCC-EWMA-n	10	**0.000(1.000)**	**0.202(0.904)**	**0.998(0.995)**	0.225(0.163)
DCC-GARCH-n	7	1.016(0.314)	1.114(0.573)	2.010(0.959)	0.103(0.081)*
DCC-GARCH-t	7	1.016(0.314)	1.114(0.573)	2.052(0.957)	0.577(0.526)
aDCC-GARCH-n	8	0.434(0.510)	0.563(0.755)	8.678(0.277)	0.127(0.089)*
aDCC-GARCH-t	7	1.016(0.314)	1.114(0.573)	2.159(0.951)	**0.564(0.531)**
Short:					
RiskMetrics-n	17	4.091(0.043)**	4.680(0.096)*	9.549(0.216)	0.081*(0.099)*
Vol-adj qreg	10	**0.000(1.000)**	**0.202(0.904)**	1.937(0.963)	-
GARCH(1,1)-n	15	2.189(0.139)	2.647(0.266)	4.093(0.769)	0.167(0.132)
GARCH(1,1)-t	12	0.380(0.538)	0.672(0.715)	1.924(0.964)	0.286(0.166)
GARCH(1,1)-skew t	14	1.437(0.231)	1.835(0.399)	2.983(0.887)	0.327(0.184)
GARCH(1,1)-EVT	14	1.437(0.231)	1.835(0.399)	2.983(0.887)	**0.434(0.248)**
MSGARCH(1,1)-n	13	0.831(0.362)	1.173(0.556)	2.307(0.941)	0.144(0.124)
MSGARCH(1,1)-t	13	0.831(0.362)	1.173(0.556)	2.214(0.947)	0.174(0.134)
GJR(1,1)-n	15	2.189(0.139)	2.647(0.266)	4.110(0.767)	0.160(0.130)
GJR(1,1)-t	12	0.380(0.538)	0.672(0.715)	1.938(0.963)	0.292(0.165)
GJR(1,1)-skew t	12	0.380(0.538)	0.672(0.715)	2.044(0.957)	0.211(0.146)
Vol-corr-adj qreg	11	0.098(0.754)	0.343(0.842)	4.053(0.774)	-
CCC-n	14	1.437(0.231)	1.835(0.399)	10.622(0.156)	0.229(0.143)
CCC-t	9	0.105(0.746)	0.268(0.875)	11.351(0.124)	0.303(0.171)
DCC-EWMA-n	16	3.077(0.079)*	3.597(0.166)	5.831(0.560)	0.195(0.141)
DCC-GARCH-n	11	0.098(0.754)	0.343(0.842)	1.225(0.990)	0.210(0.140)
DCC-GARCH-t	8	0.434(0.510)	0.563(0.755)	**0.861(0.997)**	0.316(0.175)
aDCC-GARCH-n	11	0.098(0.754)	0.343(0.842)	8.653(0.279)	0.172(0.133)
aDCC-GARCH-t	10	**0.000(1.000)**	**0.202(0.904)**	0.916(0.996)	0.368(0.193)

*, **, *** indicates rejection at the 10%, 5% and 1% significance level, respectively. The decision is based on critical values from the χ^2 distribution with 1 degree of freedom for $-2ln(LR_{UC})$, 2 degrees of freedom for $-2ln(LR_{CC})$ and 7 degrees of freedom for the DQ, respectively. For the ES test, the bootstrapped p-value is given. p-values are shown in parentheses. Bold numbers indicate highest p-value.

Table 22: Backtesting of 95% VaR and ES forecasts with reestimation, front-year

	50 expected	Kupiec	Christoffersen	Engle & Manganelli	ES test
Model	Exceedances	$-2ln(LR_{UC})$(p-val)	$-2ln(LR_{CC})$(p-val)	DQ(p-val)	b.p-val(p-val)
Long:					
RiskMetrics-n	-	-	-	-	-
Vol-adj qreg	54	0.329(0.566)	0.688(0.709)	**4.552(0.715)**	-
GARCH(1,1)-n	47	0.193(0.660)	1.095(0.578)	5.346(0.618)	0.294(0.208)
GARCH(1,1)-t	54	0.329(0.566)	0.688(0.709)	5.054(0.653)	0.793(0.862)
GARCH(1,1)-skew t	50	**0.000(1.000)**	1.267(0.531)	4.843(0.679)	0.787(0.849)
GARCH(1,1)-EVT	-	-	-	-	-
MSGARCH(1,1)-n	41	1.812(0.178)	5.322(0.070)*	6.852(0.444)	0.195(0.113)
MSGARCH(1,1)-t	52	0.083(0.773)	1.625(0.444)	5.230(0.632)	0.413(0.336)
GJR(1,1)-n	48	0.085(0.770)	**0.132(0.936)**	6.691(0.462)	0.280(0.196)
GJR(1,1)-t	55	0.510(0.475)	0.949(0.622)	7.060(0.423)	**0.810(0.883)**
GJR(1,1)-skew t	51	0.021(0.885)	0.189(0.910)	5.379(0.614)	0.753(0.806)
Vol-corr-adj qreg	49	0.021(0.884)	0.174(0.917)	10.124(0.182)	-
CCC-n	54	0.329(0.566)	0.735(0.693)	6.501(0.483)	0.222(0.135)
CCC-t	52	0.083(0.773)	0.691(0.708)	7.430(0.386)	0.478(0.406)
DCC-EWMA-n	40	2.253(0.133)	2.534(0.282)	5.092(0.649)	0.159(0.086)*
DCC-GARCH-n	32	7.777(0.005)***	7.777(0.020)**	5.065(0.652)	0.158(0.087)*
DCC-GARCH-t	33	6.878(0.009)***	6.887(0.032)**	7.626(0.367)	0.499(0.451)
aDCC-GARCH-n	33	6.878(0.009)***	9.370(0.009)***	10.630(0.156)	0.146(0.081)*
aDCC-GARCH-t	34	6.043(0.014)**	6.067(0.048)**	7.026(0.426)	0.479(0.418)
Short:					
RiskMetrics-n	-	-	-	-	-
Vol-adj qreg	60	1.984(0.159)	3.522(0.172)	14.147(0.049)**	-
GARCH(1,1)-n	49	0.021(0.884)	0.174(0.917)	6.666(0.464)	0.022**(0.038)**
GARCH(1,1)-t	51	**0.021(0.885)**	**0.085(0.958)**	8.800(0.267)	0.123(0.096)*
GARCH(1,1)-skew t	52	0.083(0.773)	0.117(0.943)	7.001(0.429)	0.056*(0.069)*
GARCH(1,1)-EVT	-	-	-	-	-
MSGARCH(1,1)-n	44	0.788(0.375)	1.384(0.501)	3.869(0.795)	0.040**(0.059)*
MSGARCH(1,1)-t	47	0.193(0.660)	1.095(0.578)	3.961(0.784)	0.067*(0.075)*
GJR(1,1)-n	54	0.329(0.566)	0.331(0.847)	7.331(0.395)	0.025**(0.040)**
GJR(1,1)-t	55	0.510(0.475)	0.511(0.775)	7.060(0.423)	0.159(0.116)
GJR(1,1)-skew t	57	0.989(0.320)	1.011(0.603)	7.812(0.349)	0.088*(0.081)*
Vol-corr-adj qreg	62	2.826(0.093)*	4.019(0.134)	13.852(0.054)*	-
CCC-n	63	3.299(0.069)*	3.581(0.167)	10.211(0.177)	0.046**(0.059)*
CCC-t	63	3.299(0.069)*	3.299(0.192)	10.307(0.172)	0.203(0.134)
DCC-EWMA-n	51	**0.021(0.885)**	0.747(0.688)	5.065(0.652)	0.040**(0.057)*
DCC-GARCH-n	46	0.346(0.557)	1.139(0.566)	4.826(0.681)	0.167(0.123)
DCC-GARCH-t	49	0.021(0.884)	0.100(0.951)	**2.782(0.904)**	**0.437(0.260)**
aDCC-GARCH-n	51	**0.021(0.885)**	0.189(0.910)	4.948(0.666)	0.172(0.123)
aDCC-GARCH-t	49	0.021(0.884)	1.160(0.560)	4.281(0.747)	0.353(0.200)

*, **, *** indicates rejection at the 10%, 5% and 1% significance level, respectively. The decision is based on critical values from the χ^2 distribution with 1 degree of freedom for $-2ln(LR_{UC})$, 2 degrees of freedom for $-2ln(LR_{CC})$ and 7 degrees of freedom for the DQ, respectively. For the ES test, the bootstrapped p-value is given. p-values are shown in parentheses. Bold numbers indicate highest p-value.

Table 23: Backtesting of 99% VaR and ES forecasts with reestimation, front-year

	10 expected	Kupiec	Christoffersen	Engle & Manganelli	ES test
Model	Exceedances	$-2ln(LR_{UC})$(p-val)	$-2ln(LR_{CC})$(p-val)	DQ(p-val)	b.p-val(p-val)
Long:					
RiskMetrics-n	-	-	-	-	-
Vol-adj qreg	14	1.437(0.231)	1.835(0.399)	3.785(0.804)	-
GARCH(1,1)-n	10	**0.000(1.000)**	**0.202(0.904)**	**0.772(0.998)**	0.060*(0.048)**
GARCH(1,1)-t	8	0.434(0.510)	0.563(0.755)	1.014(0.995)	0.145(0.122)
GARCH(1,1)-skew t	8	0.434(0.510)	0.563(0.755)	0.906(0.996)	0.433(0.317)
GARCH(1,1)-EVT	10	**0.000(1.000)**	**0.202(0.904)**	0.803(0.997)	**0.606(0.557)**
MSGARCH(1,1)-n	10	**0.000(1.000)**	**0.202(0.904)**	2.089(0.955)	0.119(0.065)*
MSGARCH(1,1)-t	9	0.105(0.746)	0.268(0.875)	1.163(0.992)	0.061*(0.040)**
GJR(1,1)-n	10	**0.000(1.000)**	**0.202(0.904)**	0.818(0.997)	0.038**(0.030)**
GJR(1,1)-t	9	0.105(0.746)	0.268(0.875)	9.231(0.236)	0.231(0.166)
GJR(1,1)-skew t	8	0.434(0.510)	0.563(0.755)	0.963(0.995)	0.319(0.194)
Vol-corr-adj qreg	15	2.189(0.139)	2.647(0.266)	11.423(0.121)	-
CCC-n	16	3.077(0.079)*	4.384(0.112)	16.846(0.018)**	0.108(0.059)*
CCC-t	11	0.098(0.754)	0.343(0.842)	7.780(0.352)	0.266(0.170)
DCC-EWMA-n	11	0.098(0.754)	0.343(0.842)	1.461(0.984)	0.263(0.201)
DCC-GARCH-n	8	0.434(0.510)	0.563(0.755)	1.912(0.965)	0.443(0.363)
DCC-GARCH-t	5	3.094(0.079)*	3.144(0.208)	2.714(0.910)	0.421(0.372)
aDCC-GARCH-n	9	0.105(0.746)	0.268(0.875)	8.468(0.293)	0.436(0.345)
aDCC-GARCH-t	5	3.094(0.079)*	3.144(0.208)	3.183(0.868)	0.474(0.478)
Short:					
RiskMetrics-n	-	-	-	-	-
Vol-adj qreg	14	1.437(0.231)	1.835(0.399)	4.833(0.680)	-
GARCH(1,1)-n	20	7.827(0.005)***	8.644(0.013)**	13.815(0.055)*	0.132(0.122)
GARCH(1,1)-t	16	3.077(0.079)*	3.597(0.166)	5.726(0.572)	0.328(0.184)
GARCH(1,1)-skew t	16	3.077(0.079)*	3.597(0.166)	6.097(0.528)	0.304(0.177)
GARCH(1,1)-EVT	17	4.091(0.043)**	4.680(0.096)*	7.692(0.360)	**0.481(0.296)**
MSGARCH(1,1)-n	13	0.831(0.362)	1.173(0.556)	2.307(0.941)	0.147(0.124)
MSGARCH(1,1)-t	13	0.831(0.362)	1.173(0.556)	2.214(0.947)	0.173(0.134)
GJR(1,1)-n	21	9.284(0.002)***	10.186(0.006)***	16.494(0.021)**	0.110(0.111)
GJR(1,1)-t	18	5.225(0.022)**	5.886(0.053)*	9.231(0.236)	0.352(0.199)
GJR(1,1)-skew t	18	5.225(0.022)**	5.886(0.053)*	9.243(0.236)	0.336(0.196)
Vol-corr-adj qreg	17	4.091(0.043)**	4.680(0.096)*	32.447(0.000)***	-
CCC-n	22	10.838(0.001)***	11.829(0.003)***	20.345(0.005)***	0.222(0.142)
CCC-t	18	5.225(0.022)**	5.886(0.053)*	13.705(0.057)*	0.425(0.240)
DCC-EWMA-n	18	5.225(0.022)**	5.886(0.053)*	11.469(0.119)	0.183(0.135)
DCC-GARCH-n	11	**0.098(0.754)**	0.343(0.842)	1.252(0.990)	0.232(0.144)
DCC-GARCH-t	8	0.434(0.510)	0.563(0.755)	0.793(0.998)	0.350(0.188)
aDCC-GARCH-n	13	0.831(0.362)	1.173(0.556)	8.453(0.294)	0.244(0.148)
aDCC-GARCH-t	9	0.105(0.746)	**0.268(0.875)**	**0.644(0.999)**	0.375(0.206)

*, **, *** indicates rejection at the 10%, 5% and 1% significance level, respectively. The decision is based on critical values from the χ^2 distribution with 1 degree of freedom for $-2ln(LR_{UC})$, 2 degrees of freedom for $-2ln(LR_{CC})$ and 7 degrees of freedom for the DQ, respectively. For the ES test, the bootstrapped p-value is given. p-values are shown in parentheses. Bold numbers indicate highest p-value.

Table 24: Aggregated Kupiec and Christoffersen backtesting results for VaR forecasts with fixed estimation window at the 5% significance level

	Kupiec									Christoffersen								
	Front-quarter				Front-year					Front-quarter				Front-year				
	95 Long	95 Short	99 Long	99 Short	95 Long	95 Short	99 Long	99 Short	**Sum Kupiec**	95 Long	95 Short	99 Long	99 Short	95 Long	95 Short	99 Long	99 Short	**Sum Christoffersen**
RiskMetrics-n	-	-	X	X	-	-	-	X	3	-	-	-	-	-	-	-	-	**0**
Vol-adj qreg	-	-	-	-	X	-	-	-	1	-	-	-	-	X	-	-	-	1
GARCH(1,1)-n	-	-	-	-	-	-	-	-	**0**	-	X	-	-	-	-	-	-	1
GARCH(1,1)-t	-	-	-	-	-	-	-	-	**0**	-	X	-	-	-	-	-	-	1
GARCH(1,1)-skew t	-	-	-	-	-	-	-	-	**0**	-	X	-	-	-	-	-	-	1
GARCH(1,1)-EVT	*	*	-	-	*	*	-	-	**0**	*	*	-	-	*	*	-	-	**0**
MSGARCH(1,1)-n	-	X	-	-	-	-	-	-	1	-	X	-	-	-	-	-	-	1
MSGARCH(1,1)-t	-	X	-	-	-	-	-	-	1	-	X	-	-	-	-	-	-	1
GJR(1,1)-n	-	-	-	-	-	-	-	-	**0**	-	-	-	-	-	-	-	-	**0**
GJR(1,1)-t	-	-	-	-	-	-	-	-	**0**	-	-	-	-	-	-	-	-	**0**
GJR(1,1)-skew t	-	-	-	-	-	-	-	-	**0**	-	-	-	-	-	-	-	-	**0**
Vol-corr-adj qreg	-	-	-	-	X	-	-	-	1	-	-	-	-	X	-	-	-	1
CCC-n	-	X	-	-	X	-	-	-	2	-	X	-	-	-	-	-	-	1
CCC-t	-	-	-	-	-	-	-	-	**0**	-	-	-	-	-	-	-	-	**0**
DCC-EWMA-n	-	X	-	-	X	-	-	-	2	-	X	-	-	-	-	-	-	1
DCC-GARCH-n	X	X	-	-	X	-	-	-	3	X	X	-	-	X	-	-	-	3
DCC-GARCH-t	X	X	-	-	X	-	-	-	3	X	X	-	-	X	-	-	-	3
aDCC-GARCH-n	X	X	-	-	X	-	-	-	3	X	X	-	-	X	-	-	-	3
aDCC-GARCH-t	X	X	-	-	X	-	-	-	3	X	X	-	-	X	-	-	-	3

X denotes rejection at the 5% significance level for the given test. Sum Kupiec and Sum Christoffersen give the aggregated number of rejections over the tests. * denotes that GARCH(1,1)-EVT is not included at the 95% quantiles.

Table 25: Aggregated Kupiec and Christoffersen backtesting results for VaR forecasts with reestimation at the 5% significance level

	Kupiec									Christoffersen								
	Front-quarter				Front-year					Front-quarter				Front-year				
	95 Long	95 Short	99 Long	99 Short	95 Long	95 Short	99 Long	99 Short	**Sum Kupiec**	95 Long	95 Short	99 Long	99 Short	95 Long	95 Short	99 Long	99 Short	**Sum Christoffersen**
Vol-adj qreg	-	-	-	-	-	-	-	-	**0**	-	-	-	-	-	-	-	-	**0**
GARCH(1,1)-n	-	-	X	-	-	-	-	X	2	-	-	-	-	-	-	-	X	1
GARCH(1,1)-t	-	-	-	-	-	-	-	-	**0**	-	-	-	-	-	-	-	-	**0**
GARCH(1,1)-skew t	-	-	-	-	-	-	-	-	**0**	-	-	-	-	-	-	-	-	**0**
GARCH(1,1)-EVT	*	*	-	-	*	*	-	X	1	*	*	-	-	*	*	-	-	**0**
MSGARCH(1,1)-n	-	X	-	-	-	-	-	-	1	-	X	-	-	-	-	-	-	1
MSGARCH(1,1)-t	-	X	-	-	-	-	-	-	1	-	X	-	-	-	-	-	-	1
GJR(1,1)-n	-	-	X	X	-	-	-	X	3	-	-	-	-	-	-	-	X	1
GJR(1,1)-t	-	-	-	-	-	-	-	X	1	-	-	-	-	-	-	-	-	**0**
GJR(1,1)-skew t	-	-	-	-	-	-	-	X	1	-	-	-	-	-	-	-	-	**0**
Vol-corr-adj qreg	-	-	-	-	-	-	-	X	1	-	-	-	-	-	-	-	-	**0**
CCC-n	-	-	-	X	-	-	-	X	2	-	-	-	-	-	-	-	X	1
CCC-t	-	-	-	-	-	-	-	X	1	-	-	-	-	-	-	-	-	**0**
DCC-EWMA-n	-	-	X	X	-	-	-	X	3	-	-	-	-	-	-	-	-	**0**
DCC-GARCH-n	-	X	-	-	X	-	-	-	2	-	-	-	-	X	-	-	-	1
DCC-GARCH-t	-	-	-	-	X	-	-	-	1	-	X	-	-	X	-	-	-	2
aDCC-GARCH-n	-	-	-	-	X	-	-	-	1	-	-	-	-	X	-	-	-	1
aDCC-GARCH-t	-	-	-	-	X	-	-	-	1	-	-	-	-	X	-	-	-	1

X denotes rejection at the 5% significance level for the given test. Sum Kupiec and Sum Christoffersen give the aggregated number of rejections over the tests. * denotes that GARCH(1,1)-EVT is not included at the 95% quantiles.

Table 26: Aggregated Engle & Manganelli and ES-test backtesting results for VaR forecasts with fixed estimation window at the 5% significance level

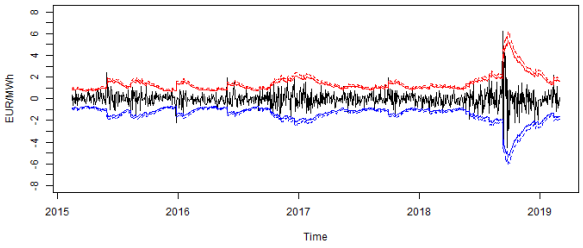
	Engle & Manganelli									ES-test								
	Front-quarter				Front-year					Front-quarter				Front-year				
	95 Long	95 Short	99 Long	99 Short	95 Long	95 Short	99 Long	99 Short	**Sum Engle & Manganelli**	95 Long	95 Short	99 Long	99 Short	95 Long	95 Short	99 Long	99 Short	**Sum ES-test**
GARCH(1,1)-EVT	*	*	-	-	*	*	-	-	**0**	*	*	-	-	*	*	-	-	**0**
GJR(1,1)-n	-	-	-	-	-	-	-	-	**0**	X	X	-	-	-	-	X	-	3
GJR(1,1)-t	-	-	-	-	-	-	-	-	**0**	-	X	-	-	-	-	-	-	1
GJR(1,1)-skew t	-	-	-	-	-	-	-	-	**0**	-	-	-	-	-	-	-	-	**0**
CCC-t	-	-	-	-	-	-	-	-	**0**	-	-	-	-	-	-	-	-	**0**

X denotes rejection at the 5% significance level for the given test. Sum Kupiec and Sum Christoffersen give the aggregated number of rejections over the tests. * denotes that GARCH(1,1)-EVT is not included at the 95% quantiles.

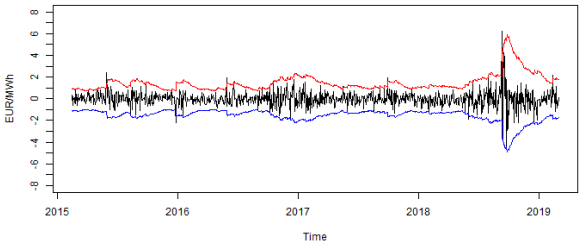
Table 27: Aggregated Engle & Manganelli and ES-test backtesting results for VaR forecasts with reestimation at the 5% significance level

	Engle & Manganelli									ES-test								
	Front-quarter				Front-year					Front-quarter				Front-year				
	95 Long	95 Short	99 Long	99 Short	95 Long	95 Short	99 Long	99 Short	**Sum Engle & Manganelli**	95 Long	95 Short	99 Long	99 Short	95 Long	95 Short	99 Long	99 Short	**Sum ES-test**
Vol-adj qreg	-	-	-	-	-	X	-	-	1	*	*	*	*	*	*	*	*	*
GARCH(1,1)-t	-	-	-	-	-	-	-	-	**0**	-	-	-	-	-	-	-	-	**0**
GARCH(1,1)-skew t	-	-	-	-	-	-	-	-	**0**	-	-	-	-	-	-	-	-	**0**

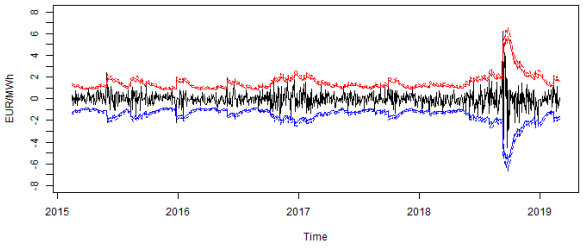
X denotes rejection at the 5% significance level for the given test. Sum Engle & Manganellig and Sum ES-test give the aggregated number of rejections over the tests. * denotes that the model is not included in the test.



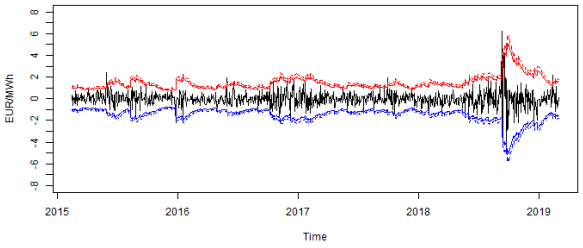
(a) RiskMetrics-n



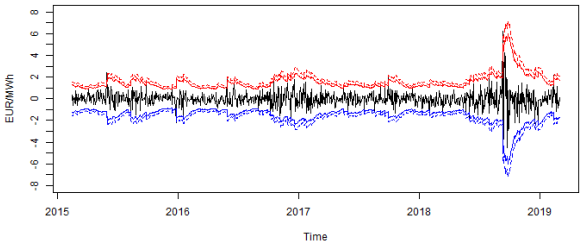
(b) Volatility-adjusted quantile regression



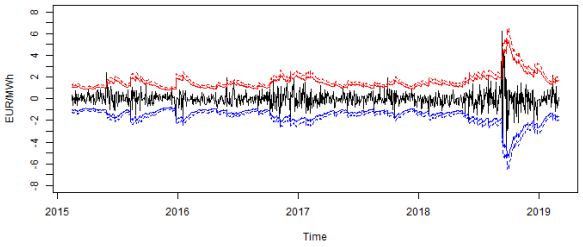
(c) GARCH(1,1)-n



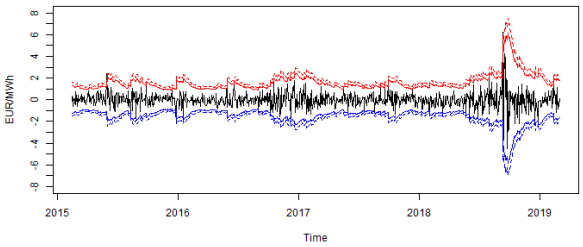
(d) GJR-GARCH(1,1)-n



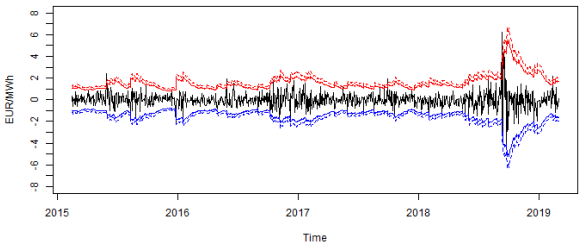
(e) GARCH(1,1)-t



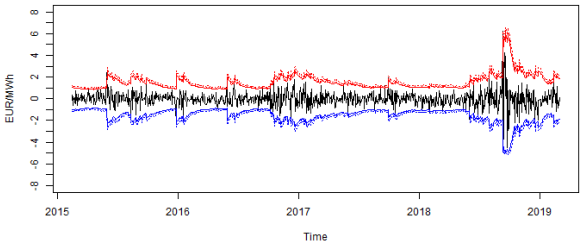
(f) GJR-GARCH(1,1)-t



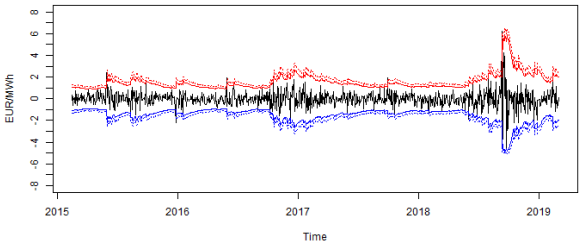
(g) GARCH(1,1)-skew t



(h) GJR-GARCH(1,1)-skew t

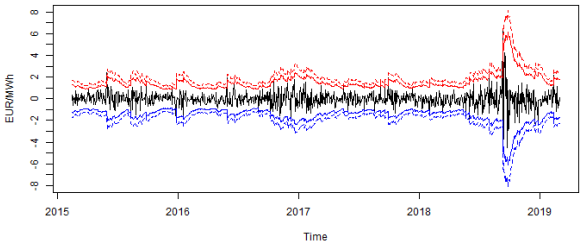


(i) MSGARCH(1,1)-n

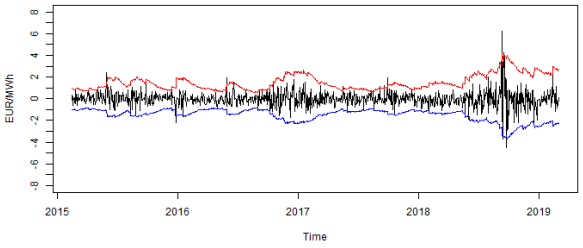


(j) MSGARCH(1,1)-t

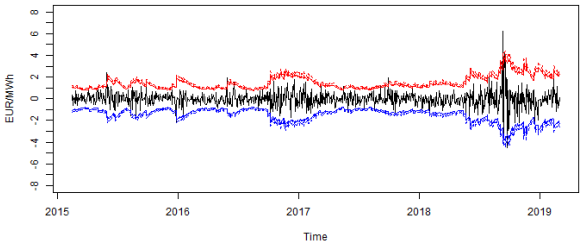
Figure 8: 99% VaR and ES forecasts with fixed estimation window for a long position (solid blue and dotted blue) and short position (solid red and dotted red), front-quarter



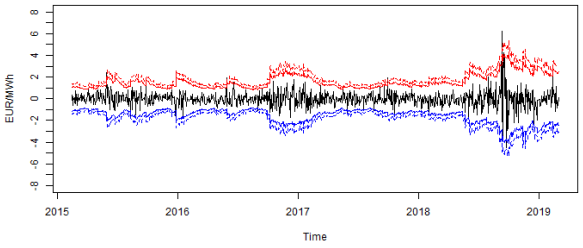
(a) GARCH(1,1)-EVT



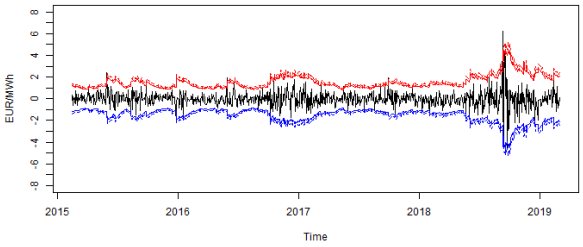
(b) Vol-corr-adj qreg



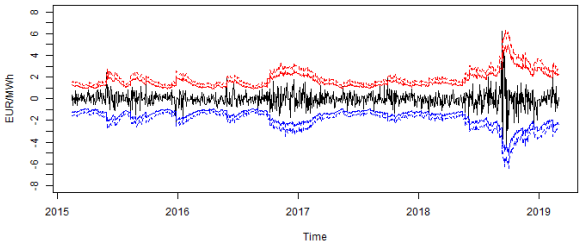
(c) CCC-n



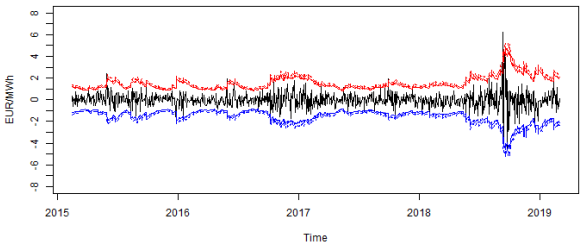
(d) CCC-t



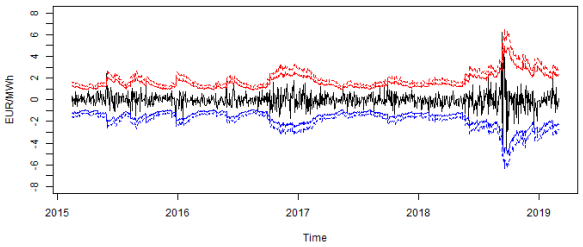
(e) DCC-n



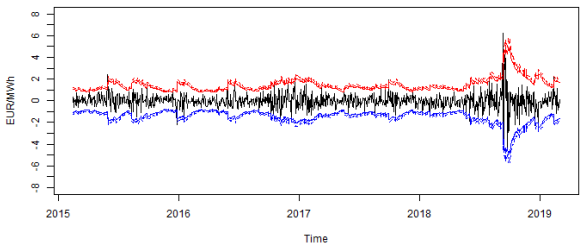
(f) DCC-t



(g) aDCC-n

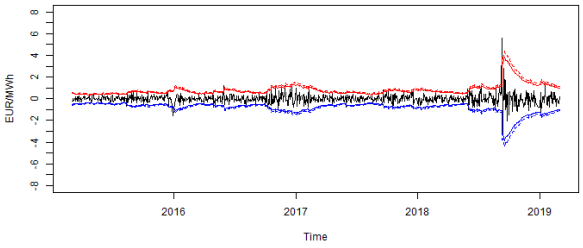


(h) aDCC-t

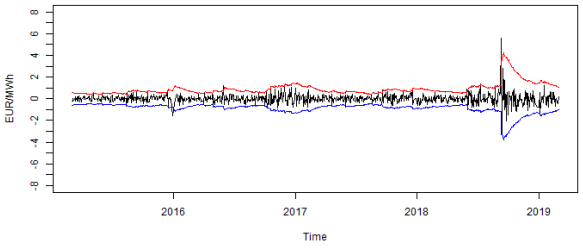


(i) DCC-EWMA

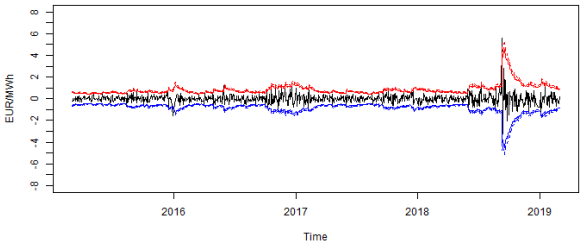
Figure 9: 99% VaR and ES forecasts with fixed estimation window for a long position (solid blue and dotted blue) and short position (solid red and dotted red), front-quarter



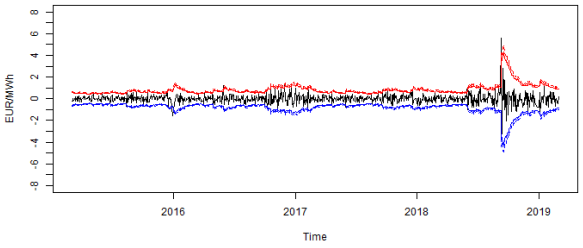
(a) RiskMetrics-n



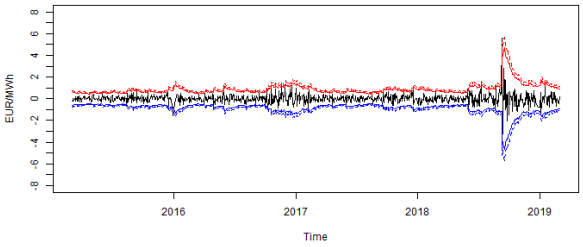
(b) Volatility-adjusted quantile regression



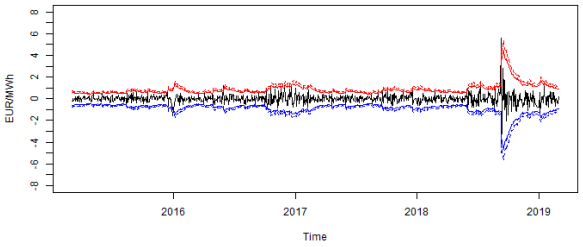
(c) GARCH(1,1)-n



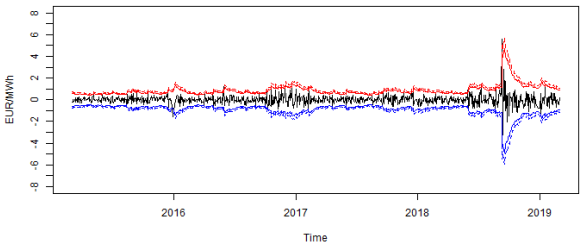
(d) GJR-GARCH(1,1)-n



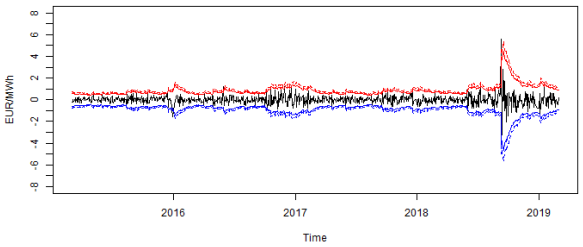
(e) GARCH(1,1)-t



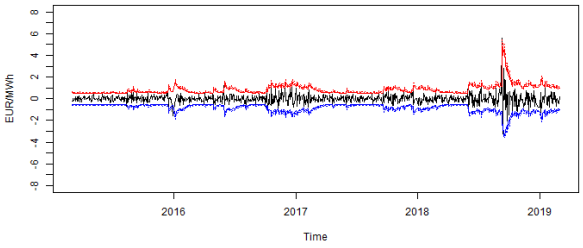
(f) GJR-GARCH(1,1)-t



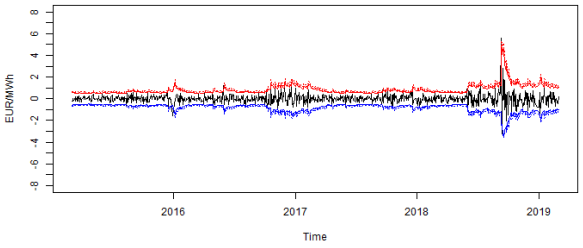
(g) GARCH(1,1)-skew t



(h) GJR-GARCH(1,1)-skew t

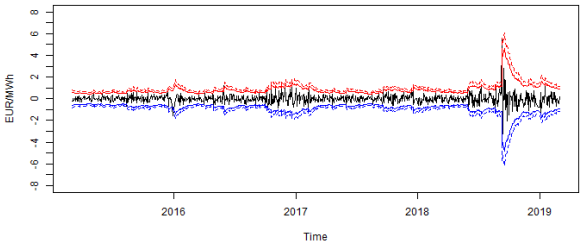


(i) MSGARCH(1,1)-n

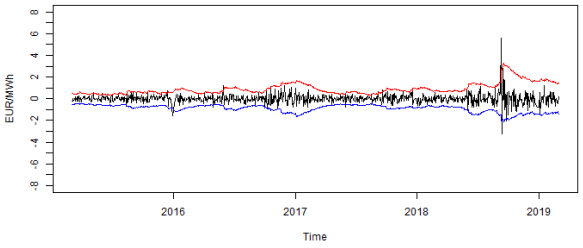


(j) MSGARCH(1,1)-t

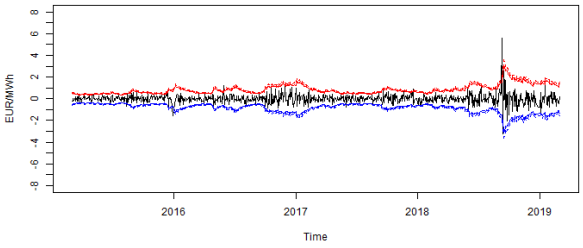
Figure 10: 99% VaR and ES forecasts with fixed estimation window for a long position (solid blue and dotted blue) and short position (solid red and dotted red), front-year



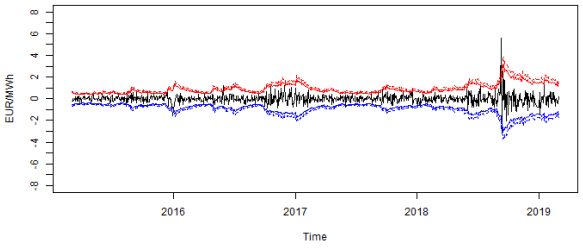
(a) GARCH(1,1)-EVT



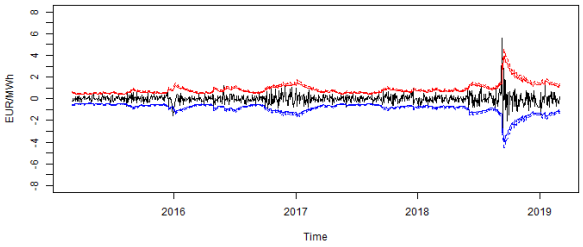
(b) Vol-corr-adj qreg



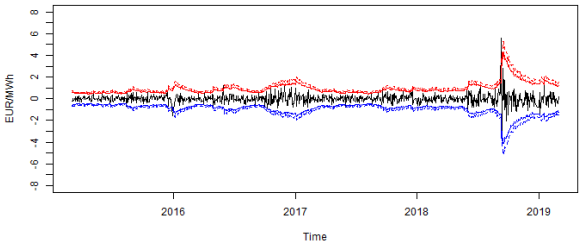
(c) CCC-n



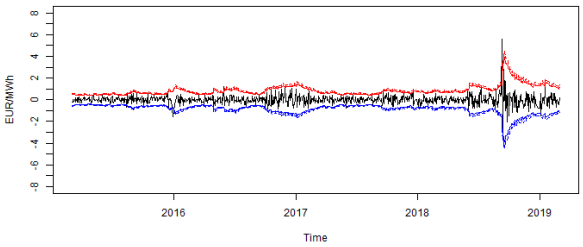
(d) CCC-t



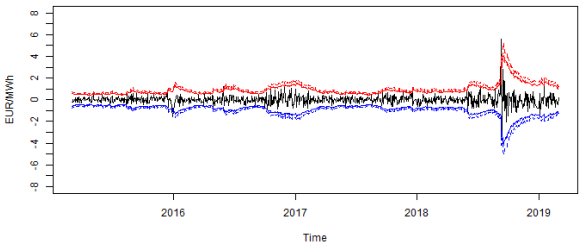
(e) DCC-n



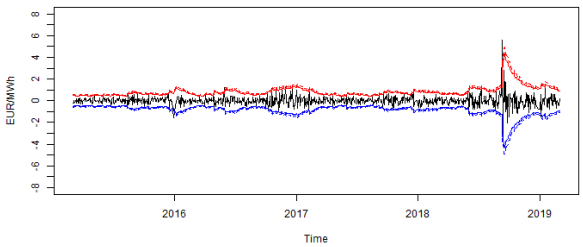
(f) DCC-t



(g) aDCC-n

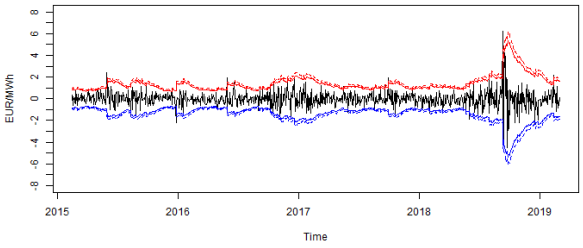


(h) aDCC-t

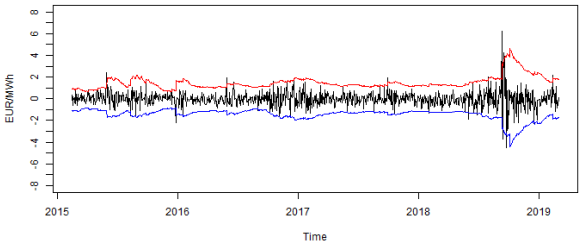


(i) DCC-EWMA

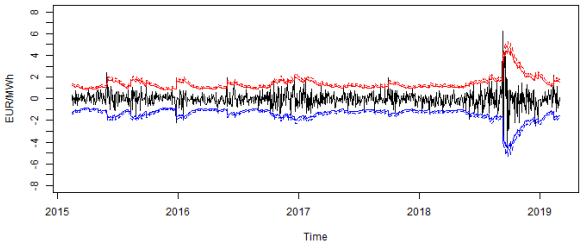
Figure 11: 99% VaR and ES forecasts with fixed estimation window for a long position (solid blue and dotted blue) and short position (solid red and dotted red), front-year



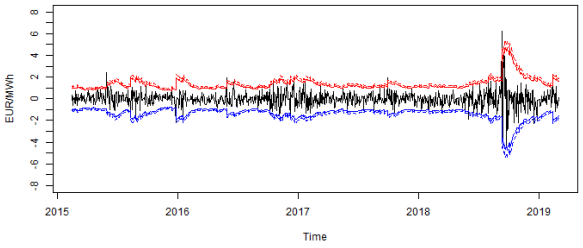
(a) RiskMetrics-n



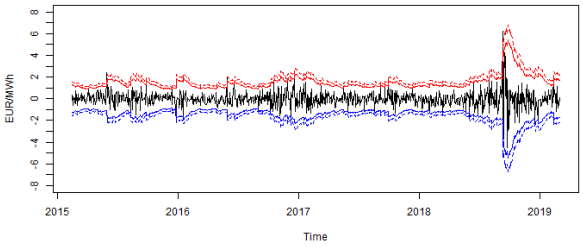
(b) Volatility-adjusted quantile regression



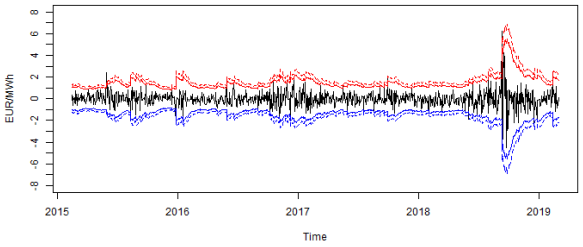
(c) GARCH(1,1)-n



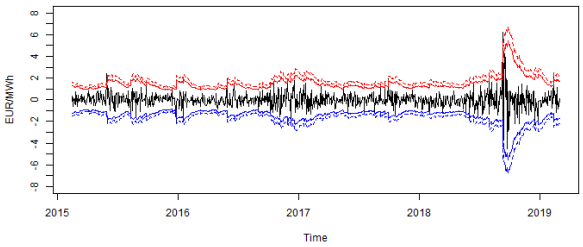
(d) GJR-GARCH(1,1)-n



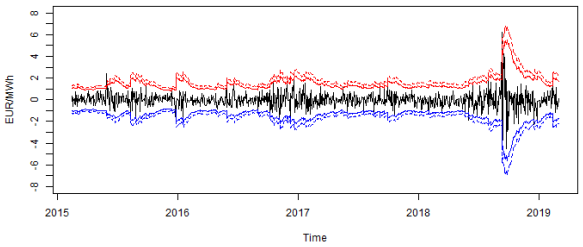
(e) GARCH(1,1)-t



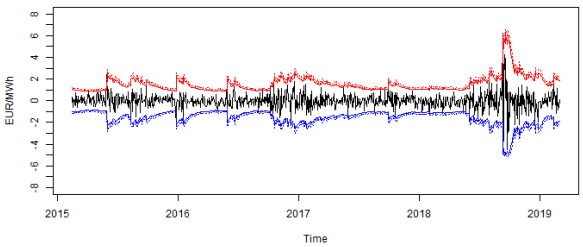
(f) GJR-GARCH(1,1)-t



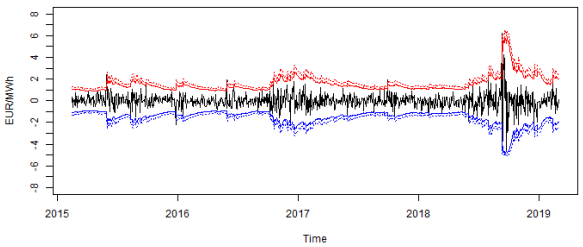
(g) GARCH(1,1)-skew t



(h) GJR-GARCH(1,1)-skew t

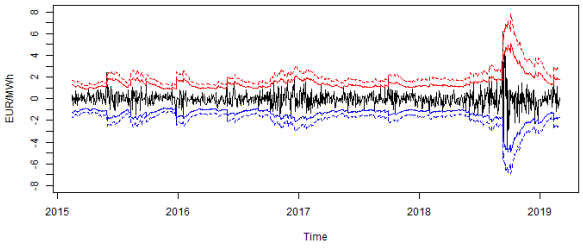


(i) MSGARCH(1,1)-n

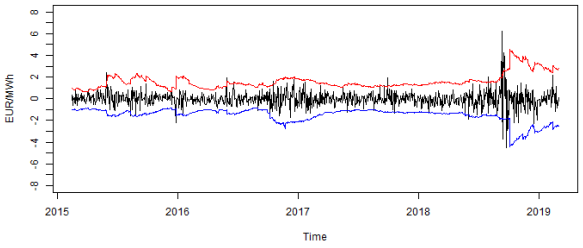


(j) MSGARCH(1,1)-t

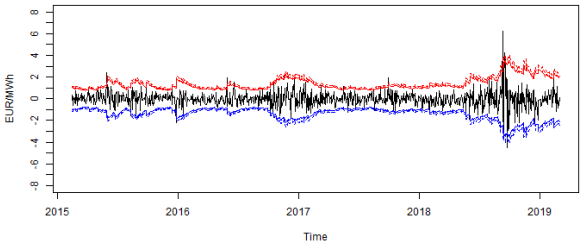
Figure 12: 99% VaR and ES forecasts with reestimation for a long position (solid blue and dotted blue), short position (solid red and dotted red), front-quarter



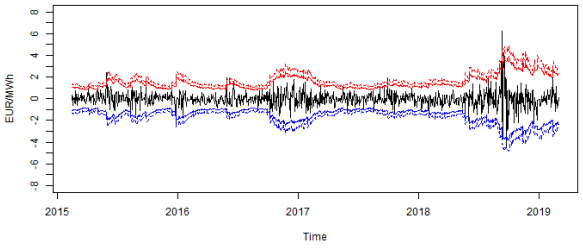
(a) GARCH(1,1)-EVT



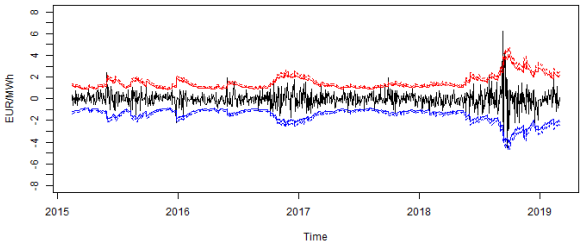
(b) Vol-corr-adj qreg



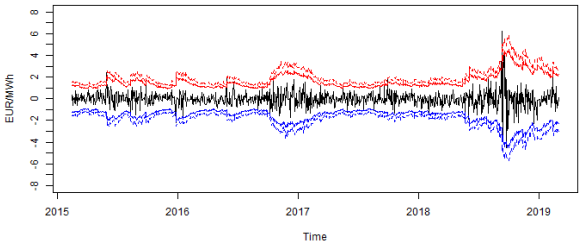
(c) CCC-n



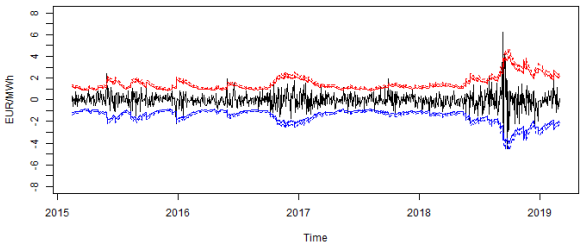
(d) CCC-t



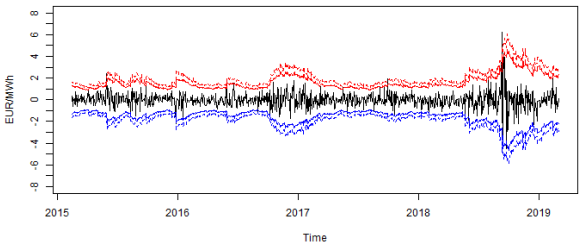
(e) DCC-n



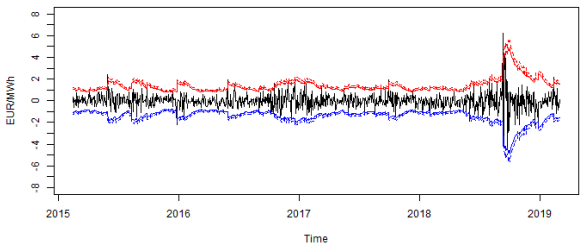
(f) DCC-t



(g) aDCC-n

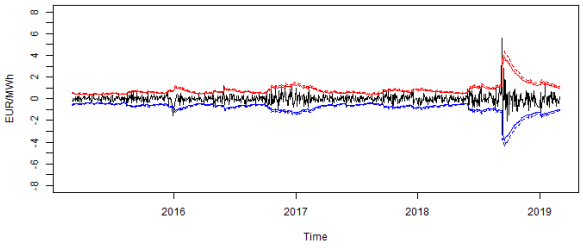


(h) aDCC-t

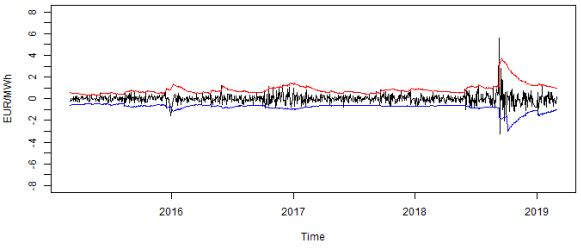


(i) DCC-EWMA

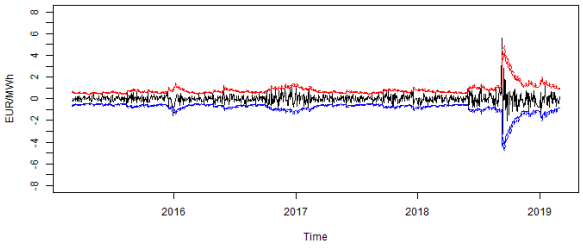
Figure 13: 99% VaR and ES forecasts with reestimation for a long position (solid blue and dotted blue), short position (solid red and dotted red), front-quarter



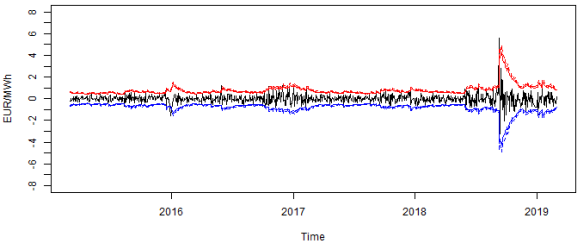
(a) RiskMetrics-n



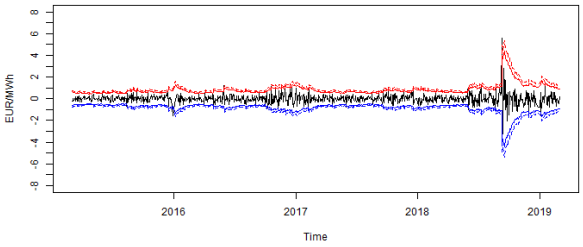
(b) Volatility-adjusted quantile regression



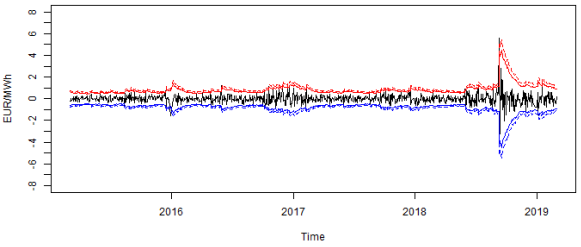
(c) GARCH(1,1)-n



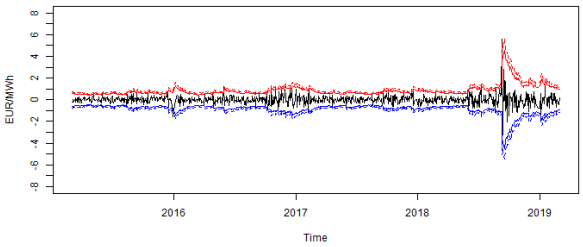
(d) GJR-GARCH(1,1)-n



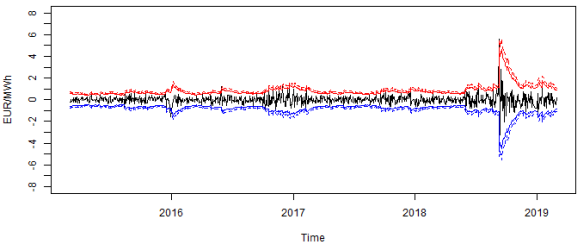
(e) GARCH(1,1)-t



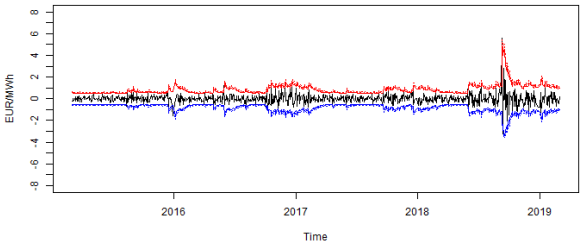
(f) GJR-GARCH(1,1)-t



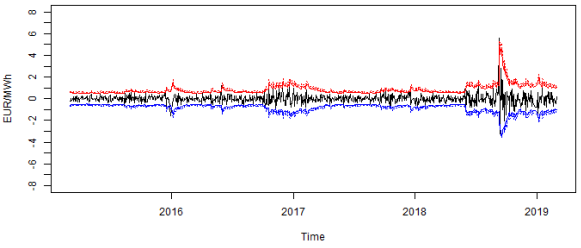
(g) GARCH(1,1)-skew t



(h) GJR-GARCH(1,1)-skew t

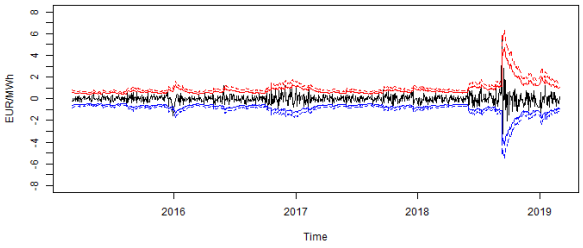


(i) MSGARCH(1,1)-n

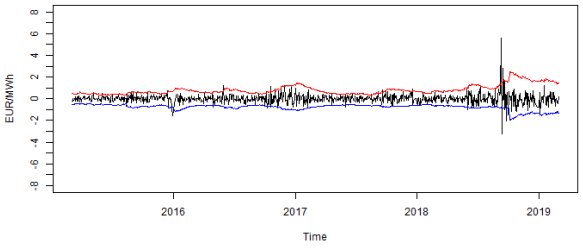


(j) MSGARCH(1,1)-t

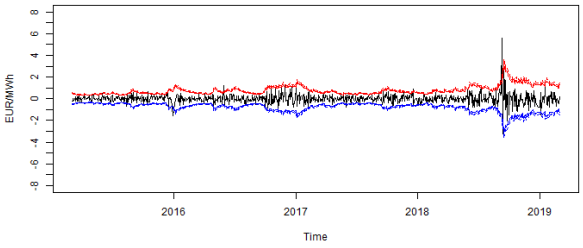
Figure 14: 99% VaR and ES forecasts with reestimation for a long position (solid blue and dotted blue), short position (solid red and dotted red), front-year



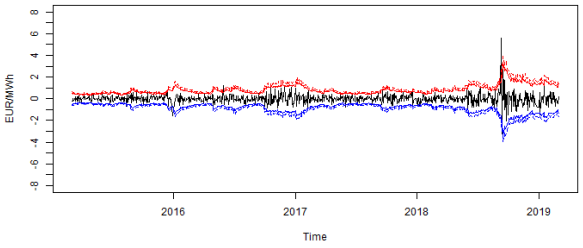
(a) GARCH(1,1)-EVT



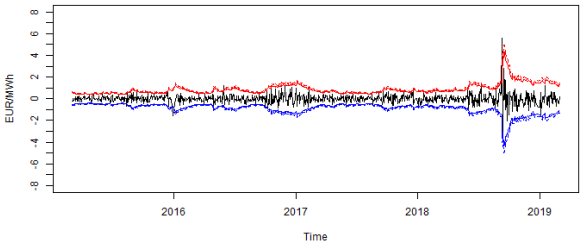
(b) Vol-corr-adj qreg



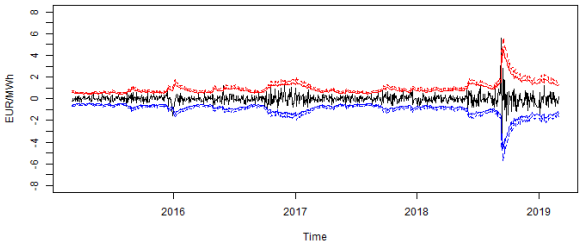
(c) CCC-n



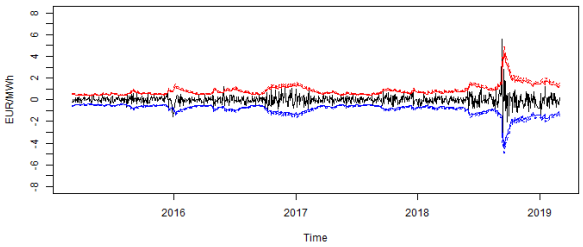
(d) CCC-t



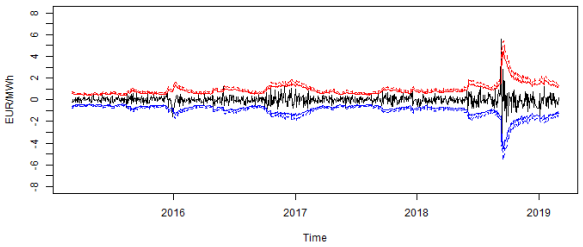
(e) DCC-n



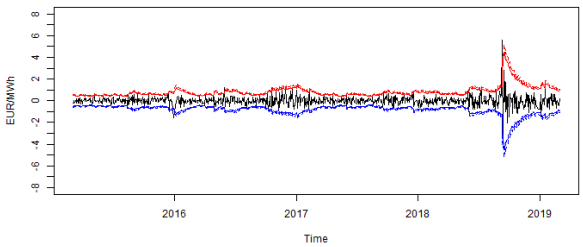
(f) DCC-t



(g) aDCC-n



(h) aDCC-t



(i) DCC-EWMA

Figure 15: 99% VaR and ES forecasts with reestimation for a long position (solid blue and dotted blue), short position (solid red and dotted red), front-year

8.4 Interpretation of VaR and ES forecasts

The most volatile period of our out-of-sample forecast window takes place during September 2018. Table 28 exhibit VaR and ES forecasts for a short position in the spread, estimated by procedure (II), for the days which coincide with the most extreme out-of-sample observations for both the front-quarter and front-year electricity futures spreads. We note that on 10, September, all models forecast VaR and ES that is materially lower than the realized P&L. However, we emphasize that at the 99% level, our models are not expected to capture these extreme dynamics. Naturally, the market shock of 10 September, causes the 99% VaR and ES forecasts for the subsequent days to inflate notably.

Table 28: Forecast of 99% VaR and ES on extreme events

	10 September 2018				13 September 2019			
	Front-quarter 6.02 EUR/MWh		Front-year 5.56 EUR/MWh		Front-quarter -3.74 EUR/MWh		Front-year -3.21 EUR/MWh	
Model	VaR	ES	VaR	ES	VaR	ES	VaR	ES
RiskMetrics-n	1.88	2.16	1.10	1.26	-3.76	-4.31	-3.21	-3.68
Vol-adj qreg	1.80	-	1.16	-	-2.64	-	-1.65	-
GARCH(1,1)-n	1.61	1.84	1.04	1.19	-3.40	-3.89	-3.58	-4.10
GARCH(1,1)-t	1.84	2.29	1.11	1.35	-3.95	-4.92	-3.69	-4.51
GARCH(1,1)-skew t	1.74	2.18	1.12	1.37	-3.71	-4.66	-3.78	-4.61
GARCH(1,1)-EVT	1.72	2.52	1.16	1.56	-3.62	-5.06	-3.61	-4.63
MSGARCH(1,1)-n	1.94	2.24	1.16	1.32	-4.46	-4.87	-2.40	-2.39
MSGARCH(1,1)-t	2.09	2.49	1.20	1.43	-4.43	-4.80	-2.39	-2.41
GJR(1,1)-n	1.65	1.89	1.07	1.22	-3.66	-4.20	-3.73	-4.27
GJR(1,1)-t	1.88	2.33	1.14	1.39	-4.19	-5.22	-3.90	-4.76
GJR(1,1)-skew t	1.78	2.21	1.15	1.40	-3.98	-4.96	-3.92	-4.77
Vol-corr-adj qreg	1.91	-	1.28	-	-1.83	-	-0.72	-
CCC-n	2.45	2.81	1.45	1.66	-2.90	-3.32	-1.91	-2.19
CCC-t	2.57	3.22	1.56	1.82	-3.13	-3.86	-1.99	-2.44
DCC-EWMA	1.85	2.12	1.28	1.46	-3.32	-3.81	-3.18	-3.64
DCC-n	2.40	2.75	1.48	1.69	-3.09	-3.54	-3.08	-3.53
DCC-t	2.65	3.32	1.54	1.86	-3.24	-4.18	-3.25	-4.00
aDCC-n	2.40	2.75	1.46	1.69	-3.08	-3.54	-3.04	-3.53
aDCC-t	2.62	3.28	1.53	1.94	-3.35	-4.42	-3.13	-3.81

6.02, 5.56, -3.74 and -3.21 EUR/MWh denote the daily P&L on the respective days as seen above.

To interpret the VaR and ES forecasts, it is necessary to consider the level of the spreads. On 10 September, the spreads widened from 4.93 to 11.13 EUR/MWh and 9.95 to 15.51 EUR/MWh for the front-quarter and front-year spreads, respectively. On 13 September, the spreads narrowed from 10.04 to 6.30 EUR/MWh and 16.20 to 12.99 EUR/MWh, for the front-quarter and front-year spreads, respectively. By entering a short position in the front-quarter spread on the preceding day, and forecast VaR and ES for September 10, one could infer that VaR corresponds to nearly 50% of the position's value. Comparing this to a scenario where the front-quarter spread lies around its peak of 20.56 EUR/MWh (see Table 3), where GARCH(1,1)-t produce short VaR and ES forecasts of 4.94 and 6.14 EUR/MWh, respectively, the VaR and ES forecasts indicate losses of approximately 25% of the portfolio for the 99% quantile. It is thus clear that our proposed framework requires a two-step evaluation to quantify market risk, by first forecasting VaR and ES, and then by evaluating these in relation to the level of the spread.

9 Conclusion

In this thesis, we have studied tail risk in the spreads of German and Nordic electricity futures prices for front-quarter and front-year contracts. We have done this by forecasting VaR and ES for a portfolio consisting of a long position in either of the aforementioned futures contracts, and a short position in the other contract of the corresponding maturity. The spreads have been modelled from both a univariate and bivariate perspective. By the univariate approach, we have considered the daily P&L of the spread itself, calculated as the daily change in the difference of the German and Nordic electricity futures prices. The bivariate approach involves the daily P&L of both the German and Nordic electricity futures contracts. The former procedure lends itself directly to the computation of the spreads' conditional volatility, and subsequently to VaR and ES estimation. The latter requires the estimation of the conditional volatility for both series, in addition to the conditional correlation between the German and Nordic electricity futures prices, to arrive upon portfolio VaR and ES. Conditional volatility has been modelled by RiskMetrics and GARCH, GJR-GARCH and Markov switching GARCH models, while CCC, DCC and aDCC models have been utilized to establish conditional correlation. Moreover, VaR has been modelled directly by volatility-adjusted quantile regression, and by a novel approach which we call volatility and correlation-adjusted quantile regression. The former has been specified to only include the volatility of the spread itself and an intercept, whereas the latter incorporates an intercept, the conditional volatility of both futures P&L, in addition to the correlation between the German and Nordic electricity futures prices. The volatilities and correlation employed for quantile regression have been estimated by RiskMetrics and DCC-EWMA. We have considered market shocks from the normal, Student t and skewed Student t distributions and EVT. All models and their performance for the 95% and 99% quantiles for long and short positions have been compared by backtesting of unconditional and conditional coverage, in addition to the dynamical quantile test and an expected shortfall test. This thesis is, to the authors' knowledge, the first attempt at modelling VaR and ES for electricity futures spreads. Furthermore, this extends to our approach of modelling the daily P&L of the spread directly, inspired by Alexander (2008c). Additionally, we have yet to identify an academic treatment of the front-quarter and front-year German-Nordic electricity futures spreads. Thus, we provide new insight into an apparently neglected field of research at the intersection of financial econometrics and energy economics.

Financial literature on European energy markets provides evidence of high volatility in electricity futures prices, leading to frequent and large price movements, which further complicates hedging and trading. The historical development of the futures spreads in this study conforms with these findings. The spreads are characterized by substantial price changes and high volatility, which can be attributed to a combination of market drivers. Moreover, our results support other contributions in suggesting that the German and Nordic markets are closely related. We find that stationary conditional volatility and correlations models have explanatory power, indicating that volatility in the P&L of the spread is mean-reverting over longer periods. This is an important feature which justifies trading and speculation in the spreads. This thesis addresses the question of whether univariate or multivariate portfolio modelling is most appropriate for forecasting VaR and ES for electricity futures spreads. The findings of this thesis have several important implications from both academic and practitioner points of view. As noted by several researchers, multivariate modelling of tail risk does not increase predictive power compared to univariate modelling for portfolios of few assets, see e.g., Berkowitz and O' Brien (2002), Brooks and Persand (2003), Bauwens et al. (2006) and Asai et al. (2006). Our results provide similar inference, as we find that allowing for time-varying correlation is detrimental to the predictive power of VaR and ES forecasts. A notable exception of this is the CCC-t model implemented with moving reestimation, which predicts VaR and ES well for the front-quarter contracts, across all quantiles assessed

in Section 8. For the univariate GARCH models, we identify a small set of models which consistently outperform other specifications. Specifically, we find that that GARCH(1,1)-EVT, GJR(1,1)-skew-t and CCC-t are the three most successful models when forecasting VaR and ES with fixed estimation window. Furthermore, GARCH(1,1)-t and GARCH(1,1)-skew-t are the two highest-performing models when forecasting VaR and ES with reestimation. As noted by Steen et al. (2015), volatility-adjusted quantile regression performs well in estimating in-sample VaR for a wide range of commodities, which did not include electricity futures. Our research extends on this approach by forecasting VaR out-of-sample for electricity futures. We show that volatility-adjusted quantile regression gives appropriate forecasts of VaR for German-Nordic electricity futures spreads. However, we find that our proposed volatility and correlation-adjusted quantile regression model does not outperform the simpler specification with only one explanatory variable. For practitioners, which could be traders, market-makers such as Nasdaq and regulators who impose risk management policies on market-makers, these results are promising. Our results indicate that the simpler univariate approach, which is considerably less computationally expensive than the bivariate, provide adequate predictive power when forecasting tail risk. Furthermore, the parsimonious GARCH with symmetric and skewed Student t serves the purpose of forecasting VaR and ES well. In addition to this, the volatility-adjusted quantile regression, which is easily implemented in spreadsheet-like environments, can be employed for VaR forecasting. Finally, GARCH with EVT shows strong performance with regards to ES forecasting. Considering that we in this thesis employ the easily implemented Hill estimator for EVT, this could complement volatility-adjusted quantile regression and aid practitioners in establishing a more holistic impression of market risk.

With regards to the potential weaknesses of this study, we highlight that a longer data window is desirable. This would allow for a larger amount of observations to estimate are models upon, which could result in more granular backtesting results. The predictive power of all models could then be asserted to a greater extent, and consequently, ease model selection. We also acknowledge that implementing additional ES backtests could help differentiate between which models that best estimates ES, e.g. those of Taylor (2008a) and Taylor (2008b), or by altering the test of McNeil and Frey (2000) to include a two-sided alternative hypothesis of zero-mean standardized exceedances. There are also other GARCH models which could be estimated on the dataset used in this thesis. An interesting model would be the APARCH model, which is found by Giot and Laurent (2003a) to exhibit strong performance in forecasting VaR for both long and short positions in Brent and WTI crude oil spot prices.

Further studies should aim at facing the challenges described above. As our models achieve predictive power when estimating the dynamics of the univariate series, i.e. the spreads P&L, this approach could be evaluated for other spread series. Moreover, an analysis of the fundamental drivers of the German-Nordic electricity futures spreads could be a fruitful direction for further research. Both regression analysis and principal component analysis could be possible tools in this respect. The volatility and correlation-adjusted quantile regression model proposed in this thesis show adequate performance in quantifying tail risk for German-Nordic electricity futures spreads. This particular specification of quantile regression could be tested for other data series. Furthermore, an assessment of how volatility and correlation should be included for this model deserves further research, e.g., by examining whether a non-linear term for the correlation improves predictive performance. Also, by including additional risk factors as explanatory variables for quantile regression, such as, e.g., Dahlen et al. (2015), the model is likely to capture the true risk in these markets better. Another approach for incorporating additional explanatory variables could be a DCCx model, which allow exogenous factors, such as, e.g., daily import and export of electricity, to impact correlation. Finally, a study of possible spread trading or hedging strategies, based on the approach we lay out in this thesis, in European electricity markets is an interesting topic left to future research.

References

Acerbi, C., & Tasche, D. (2002). Expected Shortfall: A natural coherent alternative to Value-at-Risk. *Economic Notes*, *31*(2), 379–388.

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

Alexander, C. (2008a). *Market risk analysis: Practical financial econometrics* (Vol. 2). John Wiley & Sons.

Alexander, C. (2008b). *Market risk analysis: Quantitative methods in finance* (Vol. 1). John Wiley & Sons.

Alexander, C. (2008c). *Market risk analysis: Value-at-risk models* (Vol. 4). John Wiley & Sons.

Aloui, C., & Mabrouk, S. (2010). Value-at-Risk estimations of energy commodities via long-memory, asymmetry and fat-tailed GARCH models. *Energy Policy*, *38*(5), 2326–2339.

Ardia, D., Bluteau, K., Boudt, K., & Catania, L. (2018). Forecasting risk with Markov-switching GARCH models: A large-scale performance study. *International Journal of Forecasting*, *34*(4), 733–747.

Asai, M., Mcaleer, M., & Yu, J. (2006). Multivariate stochastic volatility: A review. *Econometric Reviews*, *25*(2-3), 145–175.

Balcılar, M., Demirer, R., Hammoudeh, S., & Nguyen, D. K. (2016). Risk spillovers across the energy and carbon markets and hedging strategies for carbon risk. *Energy Economics*, *54*(C), 159–172.

Bauwens, L., Hafner, C. M., & Pierret, D. (2013). Multivariate volatility modeling of electricity futures. *Journal of Applied Econometrics*, *28*(5), 743–761.

Bauwens, L., & Laurent, S. (2005). A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. *Journal of Business & Economic Statistics*, *23*(3), 346–354.

Bauwens, L., Laurent, S., & Rombouts, J. V. (2006). Multivariate GARCH models: A survey. *Journal of Applied Econometrics*, *21*(1), 79–109.

Bauwens, L., Preminger, A., & Rombouts, J. V. K. (2010). Theory and inference for a Markov switching GARCH model. *Econometrics Journal*, *13*(2), 218–244.

Berkowitz, J., & O' Brien, J. (2002). How accurate are Value-at-Risk models at commercial banks? *Journal of Finance*, *57*(3), 1093–1111.

Blasques, F., Koopman, S. J., Łasak, K., & Lucas, A. (2016). In-sample confidence bands and out-of-sample forecast bands for time-varying parameters in observation-driven models. *International Journal of Forecasting*, *32*(3), 875–887.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 542–547.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *Review of Economics and Statistics*, *72*(3), 498–505.

Bollerslev, T., & Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, *11*(2), 143–172.

Brooks, C. (2014). *Introductory econometrics for finance*. Cambridge University Press.

Brooks, C., & Persand, G. (2003). Volatility forecasting for risk management. *Journal of Forecasting*, *22*(1), 1–22.

Bunn, D., Andresen, A., Chen, D., & Westgaard, S. (2016). Analysis and forecasting of electricty price risks with quantile factor models. *The Energy Journal*, *37*(1).

Butterworth, D., & Holmes, P. (2002). Inter-market spread trading: Evidence from UK index futures markets. *Applied Financial Economics*, *12*(11), 783–790.

Byström, H. (2003). The hedging performance of electricity futures on the Nordic power exchange. *Applied Economics*, *35*(1), 1–11.

Byström, H. (2005). Extreme value theory and extremely large electricity price changes. *International Review of Economics & Finance*, *14*(1), 41–55.

Cappiello, L., Engle, R. F., & Sheppard, K. (2006). Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial econometrics*, *4*(4), 537–572.

Chang, Y., & Park, C. (2007). Electricity market structure, electricity price, and its volatility. *Economics Letters*, *95*(2), 192–197.

Chincarini, L. (2008). A case study on risk management: Lessons from the collapse of Amaranth Advisors L.L.C. *Journal of Applied Finance*, *18*(1), 152–174.

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

Christoffersen, P. F. (2011). *Elements of financial risk management* (2nd ed.). Academic Press.

Cifter, A. (2013). Forecasting electricity price volatility with the Markov-switching GARCH model: Evidence from the Nordic electric power market. *Electric Power Systems Research*, *102*, 61–67.

Dahlen, K. E., Huisman, R., & Westgaard, S. (2015). Risk modelling of energy futures: A comparison of RiskMetrics, historical simulation, filtered historical simulation, and quantile regression. In *Springer proceedings in mathematics and statistics* (Vol. 122, pp. 283–291). Springer New York LLC.

Danish Energy Agency. (2019, March). Retrieved from https://ens.dk/en/our-services/statistics-data-key-figures-and-energy-maps/annual-and-monthly-statistics (Accessed: 05.06.2019)

Deb, R., Albert, R., Hsue, L.-L., & Brown, N. (2000). How to incorporate volatility and risk in electricity price forecasting. *The Electricity Journal*, *13*(4), 65–75.

Demarta, S., & McNeil, A. J. (2005). The t copula and related copulas. *International statistical review*, *73*(1), 111–129.

Deng, S. J., & Oren, S. S. (2006). Electricity derivatives and risk management. *Energy*, *31*(6-7), 940–953.

De Vany, A., & Walls, W. (1999). Cointegration analysis of spot electricity prices: Insights on transmission efficiency in the Western US. *Energy Economics*, *21*(5), 435–448.

Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: Journal of the Econometric Society*, 1057–1072.

Dunis, C. L., Laws, J., & Evans, B. (2006). Modelling and trading the gasoline crack spread: A non-linear story. *Derivatives Use, Trading & Regulation*, *12*(1-2), 126–145.

Dunis, C. L., Laws, J., & Evans, B. (2008). Trading futures spread portfolios: Applications of higher order and recurrent networks. *The European Journal of Finance*, *14*(6), 503–521.

Edwards, R. F., & Canter, M. S. (1995). The collapse of metallgesellschaft: Unhedgeable risks, poor hedging strategy, or just bad luck? *The Journal of Futures Markets*, *15*(3), 211–264.

Efron, B. (1993). *An introduction to the bootstrap* (Vol. 57). Chapman Hall.

Ekonomifakta. (2019, March). Retrieved from https://www.ekonomifakta.se/fakta/energi/energibalans-i-sverige/elproduktion/ (Accessed: 05.06.2019)

Embrechts, P., Kaufmann, R., & Patie, P. (2005). Strategic long-term financial risks: Single risk factors. *Computational Optimization and Applications*, *32*(1-2), 61–90.

Energiebilanzen, A. (2018, December). *Stromerzeugung nach energieträgern (strommix) von 1990*

bis 2018 (in twh) deutschland insgesamt (Tech. Rep.). Retrieved from https://www.ag-energiebilanzen.de/28-0-Zusatzinformationen. (Accessed: 05.06.2019)

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987–1007.

Engle, R. F. (1990). Stock volatility and the crash of'87: Discussion. *The Review of Financial Studies*, *3*(1), 103–106.

Engle, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, *20*(3), 339–350.

Engle, R. F., & Manganelli, S. (2004). CAViaR: Conditional autoregressive Value-at-Risk by regression quantiles. *Journal of Business & Economic Statistics*, *22*(4), 367–381.

Engle, R. F., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance*, *48*(5), 1749–1778.

Engle, R. F., & Sheppard, K. (2001). *Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH* (Tech. Rep.). National Bureau of Economic Research.

Ewing, J., & Schreuer, M. (2019, May). How a lone norwegian trader shook the world's financial system. *New York Times*. Retrieved from https://www.nytimes.com/2019/05/03/business/central-counterparties-financial-meltdown.html (Accessed: 05.06.2019)

Eydeland, A. (2003). *Energy and power risk management: new developments in modeling, pricing, and hedging*. John Wiley & Sons.

Fernández, C., & Steel, M. F. (1998). On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, *93*(441), 359–371.

Fong Chan, K., & 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.

Galloway, T., & Kolb, R. (1996). Futures prices and the maturity effect. *The Journal of Futures Markets (1986-1998)*, *16*(7).

Garcia, R. C., Contreras, J., van Akkeren, M., & Garcia, J. B. C. (2005). A GARCH forecasting model to predict day-ahead electricity prices. *IEEE Transactions on Power Systems*, *20*(2), 867–874.

Ghorbel, A., & Souilmi, S. (2014). Risk measurement in commodities markets using conditional extreme value theory. *International Journal of Econometrics and Financial Management*, *2*(5), 188–205.

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

Giot, P., & Laurent, S. (2003b). Value-at-Risk for long and short trading positions. *Journal of Applied Econometrics*, *18*(6), 641–663.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, *48*(5), 1779–1801.

González-Pedraz, C., Moreno, M., & Peña, J. I. (2014). Tail risk in energy portfolios. *Energy Economics*, *46*, 422–434.

Haas, M. (2004). A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics*, *2*(4), 493–530.

Hadsell, L., Marathe, A., & Shawky, H. A. (2004). Estimating the volatility of wholesale electricity spot prices in the US. *The Energy Journal*, *25*(4).

Hadsell, L., & Shawky, H. (2006). Electricity price volatility and the marginal cost of congestion: An empirical study of peak hours on the NYISO market, 2001-2004. *Energy Journal*, *27*(2), 157–179.

Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, *57*(2).

Hamilton, J. (1994). *Time series analysis*. Princeton, N.J: Princeton University Press.

Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*, 705–730.

Haugom, E. (2013). Predicting realized volatility for Nord Pool forward prices by including volatility spillover and covariance effects. *The Journal of Energy Markets*, *6*(3), 3–27.

Haugom, E., Westgaard, S., Solibakke, P. B., & Lien, G. (2011). Realized volatility and the influence of market measures on predictability: Analysis of Nord Pool forward electricity data. *Energy Economics*, *33*(6), 1206–1215.

Higgs, H., & Worthington, A. (2008). Stochastic price modeling of high volatility, mean-reverting, spike-prone commodities: The Australian wholesale spot electricity market. *Energy Economics*, *30*(6), 3172–3185.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, *3*(5), 1163–1174.

Hung, J.-C., Lee, M.-C., & Liu, H.-C. (2008). Estimation of Value-at-Risk for energy commodities via fat-tailed GARCH models. *Energy Economics*, *30*(3), 1173–1191.

Inui, K., & Kijima, M. (2005). On the significance of Expected Shortfall as a coherent risk measure. *Journal of Banking and Finance*, *29*(4), 853–864.

Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, *6*(3), 255–259.

Knittel, C. R., & Roberts, M. R. (2005). An empirical examination of restructured electricity prices. *Energy Economics*, *27*(5), 791–817.

Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, 33–50.

Kristiansen, T. (2004). Congestion management, transmission pricing and area price hedging in the Nordic region. *International Journal of Electrical Power and Energy Systems*, *26*(9), 685–695.

Kroner, K. F., Kneafsey, K. P., & Claessens, S. (1995). Forecasting volatility in commodity markets. *Journal of Forecasting*, *14*(2), 77–95.

Kuester, K., Mittnik, S., & Paolella, M. S. (2006). Value-at-Risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics*, *4*(1), 53–89.

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

Labandeira, X., Labeaga, J. M., & López-Otero, X. (2017). A meta-analysis on the price elasticity of energy demand. *Energy Policy*, *102*, 549–568.

Lambert, P., & Laurent, S. (2001). *Modelling financial time series using GARCH-type models with a skewed student distribution for the innovations* (Tech. Rep.). Louvain-la-Neuve, Belgium: Institut de Statistique, Université Catholique de Louvain.

Le Pen, Y., & Sévi, B. (2010). Volatility transmission and volatility impulse response functions in European electricity forward markets. *Energy Economics*, *32*(4), 758–770.

Lijesen, M. G. (2007). The real-time price elasticity of electricity. *Energy Economics*, *29*(2), 249–258.

Ljung, G. M., & Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*(2), 297–303.

Malo, P., & Kanto, A. (2006). Evaluating multivariate GARCH models in the Nordic electricity markets. *Communications in Statistics - Simulation and Computation*, *35*(1), 117–148.

Marckhoff, J., & Wimschulte, J. (2009). Locational price spreads and the pricing of contracts for difference: Evidence from the Nordic market. *Energy Economics*, *31*(2), 257–268.

Marcucci, J. (2005). Forecasting stock market volatility with regime-switching GARCH models. *Studies in Nonlinear Dynamics Econometrics*, *9*(4).

Martins-Filho, C., Yao, F., & Torero, M. (2018). Nonparametric estimation of conditional Value-at-Risk

and Expected Shortfall based on extreme value theory. , *34*(1), 23–67.

McNeil, A. J. (1999). Extreme value theory for risk managers. *Departement Mathematik ETH Zentrum*.

McNeil, A. J., & Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance*, *7*(3), 271–300.

Melamed, L. (1981). The futures market: Liquidity and the technique of spreading. *The Journal of Futures Markets*(3), 405–411.

Nasdaq. (2018). *Trading appendix 2 - contract specifications for commodity derivatives.* https://business.nasdaq.com/trade/commodities/regulation/legal-framework.html. (Accessed: 09.06.2018)

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

Nowotarski, J., & Weron, R. (2018). Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews*, *81*, 1548–1568.

RiskMetrics. (1996). Technical document. *JP Morgan, New York*.

Sampid, M. G., Hasim, H. M., & Dai, H. (2018). Refining Value-at-Risk estimates using a bayesian Markov-switching GJR-GARCH copula-EVT model. *PLoS ONE*, *13*(6).

Samuelson, P. (1965). Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review (pre-1986)*, *6*(2).

Santos, A. A. P., Nogales, F. J., & Ruiz, E. (2013). Comparing univariate and multivariate models to forecast portfolio Value-at-Risk. *Journal of Financial Econometrics*, *11*(2), 400–441.

Silvennoinen, A., & Teräsvirta, T. (2009). Multivariate GARCH models. In *Handbook of financial time series* (pp. 201–229). Springer.

Solibakke, P. (2002, 01). Efficient estimated mean and volatility characteristics for the Nordic spot electricity power market. *International Journal of Business*, *7*.

Solibakke, P. (2006). Describing the Nordic forward electric-power market: A stochastic model approach. *International Journal of Business*, *11*(4).

Solibakke, P. (2010). Corporate risk management in European energy markets. *The Journal of Energy Markets*, *3*(1), 93–131.

Sotiriadis, M., Tsotsos, R., & Kosmidou, K. (2016). Price and volatility interrelationships in the wholesale spot electricity markets of the Central-Western European and Nordic region: a multivariate GARCH approach. *Energy Systems*, *7*(1), 5–32.

Spodniak, P., Viljainen, S., Makkonen, M., & Jantunen, A. (2013). Area price spreads in the Nordic electricity market: The role of transmission lines and electricity import dependency. *10th International Conference on the European Energy Market (EEM)*, 1-8.

SSB. (2018, November). Retrieved from https://www.ssb.no/energi-og-industri/statistikker/elektrisitet/aar (Accessed: 05.06.2019)

Stafford, P., & Sheppard, D. (2018, September). Trader blows 100m hole in nasdaq's nordic power market. *Financial Times*. Retrieved from https://www.ft.com/content/43c74e02-b749-11e8-bbc3-ccd7de085ffe (Accessed: 05.06.2019)

Statistics Finland. (2018, November). Retrieved from http://pxnet2.stat.fi/PXWeb/pxweb/en/StatFin/StatFin__ene__salatuo/statfin_salatuo_pxt_001.px/table/tableViewLayout1/ (Accessed: 05.06.2019)

Statkraft. (2019, June). Retrieved from https://driftsdata.statnett.no/Web/Map/?language=en (Accessed: 05.06.2019)

Steen, M., Westgaard, S., & Gjølberg, O. (2015). Commodity Value-at-Risk modeling: Comparing RiskMetrics, historic simulation and quantile regression. *Journal of Risk Model Validation*, *9*(2), 49–78.

Taylor, J. W. (1999). A quantile regression approach to estimating the distribution of multiperiod returns. *The Journal of Derivatives*, *7*(1), 64–78.

Taylor, J. W. (2008a). Estimating Value-at-Risk and Expected Shortfall using expectiles. *Journal of Financial Econometrics*, *6*(2), 231–252.

Taylor, J. W. (2008b). Using exponentially weighted quantile regression to estimate Value-at-Risk and Expected Shortfall. *Journal of Financial Econometrics*, *6*(3), 382–406.

Till, H. (2012). Case studies and risk management in commodity derivatives trading. In *Risk management in commodity markets: From shipping to agriculturals and energy* (pp. 255–291). John Wiley and Sons.

Tse, Y. K., & Tsui, A. K. C. (2002). A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business & Economic Statistics*, *20*(3), 351–362.

Veka, S., Lien, G., Westgaard, S., & Higgs, H. (2012). Time-varying dependency in European energy markets: An analysis of Nord Pool, European Energy Exchange and Intercontinental Exchange energy commodities. *Journal of Energy Markets*, *5*(2), 3–32.

Wang, Y., & Wu, C. (2012). Forecasting energy market volatility using GARCH models: Can multivariate models beat univariate models? *Energy Economics*, *34*(6), 2167–2181.

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., & Lien, G. (2014). A note on the risk characteristics of European energy futures markets. *Beta*(01), 6–19.

Working, H. (1949). The theory of price of storage. *The American Economic Review*, *39*(6), 1254–1262.

Worthington, A., Kay-Spratley, A., & Higgs, H. (2005). Transmission of prices and price volatility in Australian electricity spot markets: A multivariate GARCH analysis. *Energy Economics*, *27*(2), 337–350.

Yamai, Y., & Yoshiba, T. (2005). Value-at-risk versus expected shortfall: A practical perspective. *Journal of Banking and Finance*, *29*(4), 997–1015.

Zhang, Y.-J., Yao, T., & He, L.-Y. (2015). Forecasting crude oil market volatility: can the Regime switching GARCH model beat the single-regime GARCH models?

Žiković, S., & Dizdarević, N. V. (2011). Similarities between Expected Shotrfall and Value-at-Risk: Application to energy markts. *International Journal of Management Cases*, *13*(3), 386–399.

Appendices

A Contract tickers

The contract tickers for Nordic and German DS futures are stated in Table 29.

Table 29: Contract ticker specification

Spread series	Nordic DS futures	German DS futures
Front-quarter	ENOQ[Q][YY]	EDEBLQ[Q][YY]
Front-year	ENOYR[YY]	EDEBLYR[YY]

YY and Q refer to the relevant year and quarter, respectively.

For a more extensive overview of the specific terms of the contracts, see Nasdaq (2018).

B GARCH model evaluation

The GARCH models in this article are estimated using maximum likelihood. The maximized value of the likelihood based on the estimation sample can be used as an indication of the goodness of fit of different models (Alexander, 2008a). In the following sections, we describe the likelihood-based criteria used in this study.

B.1 Likelihood ratio (LR)

When extending a GARCH model by adding parameters, the LR test provides a way to judge if the added parameters are statistically significant (Christoffersen, 2011). If two different models have maximum likelihood values L_0 and L_1, and model 0 is a special case of model 1, the likelihood ratio statistic is:

$$-2ln(LR) = 2\Big[ln(L_1) - ln(L_0)\Big] \tag{89}$$

Given that model 1 contains model 0, the former will always fit the data better. The test statistic follows a χ^2 distribution with degrees of freedom equal to the number of added parameters in model 1 compared to model 0. The null hypothesis states that the additional parameters in model 1 are insignificant (Christoffersen, 2011). We have used the LR test to judge if the additional parameter γ in the GJR-GARCH(1,1) model compared to GARCH(1,1), and DCC-GARCH compared to asymmetric DCC-GARCH, is statistically significant.

B.2 Information criteria

The Akaike information criterion (AIC) and the Bayesian information criterion (BIC) penalize models for additional parameters, and can be used as indicators of goodness of fit of different models (Alexander, 2008a). AIC is defined as:

$$AIC = \frac{1}{T}\Big[2k - 2ln(L)\Big] \tag{90}$$

where $ln(L)$ is the maximum log likelihood value, k is the number of parameters to be estimated and T is number of observations in the sample. BIC is defined as:

$$BIC = \frac{1}{T}\Big[kln(T) - 2ln(L)\Big] \tag{91}$$

The model that offers the lowest value of AIC or BIC is considered to be the best fit, according to these criteria.

C Estimation of univariate models

C.1 Maximum likelihood estimation with univariate distributions

The parameters in GARCH(1,1) as given in (10) or GJR-GARCH(1,1) as given in (12) are obtained by MLE based on a distributional assumption. In the following sections, we show the log likelihood functions of the normal distribution, the standardized Student t distribution and the skewed Student t distribution.

C.1.1 Normal distribution

We restate (20):

$$\varepsilon_t = \sigma_t z_t, \quad z_t \sim \text{i.i.d. } N(0,1) \tag{92}$$

The density function of the normal distribution is given by the following expression:

$$f(\varepsilon_t) = \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{(\varepsilon_t)^2}{2\sigma_t^2}} \tag{93}$$

The joint likelihood of the entire sample is specified as:

$$L = \prod_{t=1}^{T} f(\varepsilon_t) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{(\varepsilon_t)^2}{2\sigma_t^2}} \tag{94}$$

From this, the log likelihood of all observations is given by:

$$\ln(L) = \sum_{t=1}^{T} \Big[-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma_t^2) - \frac{1}{2}\frac{\varepsilon_t^2}{\sigma_t^2} \Big] \tag{95}$$

σ_t is the conditional volatility which is modelled by either a GARCH(1,1) model or a GJR-GARCH(1,1) model. We maximize the expression above and from this we obtain the optimal parameters ω, α, β and γ.

C.1.2 Standardized Student t distribution

We restate (21):

$$\varepsilon_t = \sigma_t z_t, \quad z_t \sim \text{i.i.d. } \tilde{t}(\upsilon) \tag{96}$$

The standardized Student t distribution with $\upsilon > 2$ has the following density function :

$$f_{\tilde{t}}(\varepsilon_t;\upsilon) = \Big((\upsilon-2)\pi\Big)^{-1/2}\Gamma\Big(\frac{\upsilon}{2}\Big)^{-1}\Gamma\Big(\frac{\upsilon+1}{2}\Big)\Big(1+\frac{\varepsilon_t^2}{\upsilon-2}\Big)^{-\frac{\upsilon+1}{2}} \tag{97}$$

The joint log likelihood of the entire sample is given by:

$$\begin{aligned} ln(L) = -\sum_{t=1}^{T}\Bigg[\ln(\sigma_t) + \Big(\frac{\upsilon+1}{2}\Big)\ln\Big(1+(\upsilon-2)^{-1}\Big(\frac{\varepsilon_t}{\sigma_t}\Big)^2\Big)\Bigg] \\ + T\ln\Bigg[\Big((\upsilon-2)\pi\Big)^{-\frac{1}{2}}\Gamma\Big(\frac{\upsilon}{2}\Big)^{-1}\Gamma\Big(\frac{\upsilon+1}{2}\Big)\Bigg] \end{aligned} \tag{98}$$

where the conditional volatility σ_t is included to estimate all parameters simultaneously. Again, σ_t is modelled by either a GARCH(1,1) model or a GJR-GARCH(1,1) model. We maximize the expression above and from this we obtain the optimal parameters ω, α, β and γ, along with the υ parameter of the standardized Student t distribution.

C.1.3 Skewed Student t distribution

We restate (24):

$$\varepsilon_t = \sigma_t z_t, \quad z_t \sim \text{i.i.d. } F_{skew}(\upsilon,\xi) \tag{99}$$

The skewed Student t distribution proposed by Hansen (1994) with $\upsilon > 2$ and $-1 < \xi < 1$ has the following density function:

$$f_{skew}(\varepsilon_t;\upsilon,\xi) = \begin{cases} BC\Big[1+\frac{(B\varepsilon_t+A)^2}{\big((1-\xi)^2(\upsilon-2)\big)}\Big]^{-\frac{1+\upsilon}{2}}, & \text{if } \varepsilon_t < -\frac{A}{B} \\ BC\Big[1+\frac{(B\varepsilon_t+A)^2}{\big((1+\xi)^2(\upsilon-2)\big)}\Big]^{-\frac{1+\upsilon}{2}}, & \text{if } \varepsilon_t \geq -\frac{A}{B} \end{cases} \tag{100}$$

We repeat the following definitions:

$$A = 4\xi C\frac{\upsilon-2}{\upsilon-1}, \quad B = \sqrt{1+3\xi^2-A^2}, \quad C = \frac{\Gamma(\frac{\upsilon+1}{2})}{\Gamma(\frac{\upsilon}{2})\sqrt{\pi(\upsilon-2)}} \tag{101}$$

To estimate the GARCH(1,1) and GJR-GARCH(1,1) models with skewed Student t, we use QMLE as developed in Bollerslev and Wooldridge (1992). First, we estimate the parameters included in the GARCH models (ω, α, β and γ) by maximizing the log likelihood function assuming normal distribution (95), and from this we obtain standardized daily P&L, z_t. Then we treat z_t as a regular random variable, and thus implicitly assume that conditional variance is estimated without error. Then we estimate the υ and ξ parameters in the skewed Student t distribution by maximizing the log likelihood function stated below:

$$ln\big(f_{skew}(\varepsilon_t;\upsilon,\xi)\big) = \begin{cases} ln(BC) - (\frac{1+\upsilon}{2})ln\Big(1+\frac{(Bz_t+A)^2}{\big((1-\xi)^2(\upsilon-2)\big)}\Big), & \text{if } z_t < -\frac{A}{B} \\ ln(BC) - (\frac{1+\upsilon}{2})ln\Big(1+\frac{(Bz_t+A)^2}{\big((1+\xi)^2(\upsilon-2)\big)}\Big), & \text{if } z_t \geq -\frac{A}{B} \end{cases} \tag{102}$$

$$ln(L) = \sum_{t=1}^{T} ln\big(f_{skew}(\varepsilon_t;\upsilon,\xi)\big) \tag{103}$$

In general, the QMLE method will give consistent but inefficient estimates. We choose to use QMLE to

make the numerical optimization easier, and thus we trade trade theoretical asymptotic parameter efficiency for practicality (Christoffersen, 2011). Nevertheless, using this approach compared to estimating all parameters simultaneously is not likely to represent an issue in our application of the results.

C.2 Estimation of Markov switching GARCH

We estimate the Markov switching GARCH models by maximizing the log likelihood function of the density, which we obtain by integrating over the state variable s_t. Let the vector Ψ encompass all parameters across the two regimes $\{1, 2\}$. Also, let the conditional density of ε_{t+1} in state $s_t = k$ be denoted by $f_D(\varepsilon_{t+1} \mid s_{t+1} = k, \hat{\Psi}, I_t)$. The discrete representation of the density integral from (49) is repeated:

$$f(\varepsilon_{t+1} \mid \hat{\Psi}, I_t) = \sum_{i=1}^{2} \sum_{j=1}^{2} \pi_{i,j} \eta_{i,t} f_D(\varepsilon_{t+1} \mid s_{t+1} = j, \hat{\Psi}, I_t) \tag{104}$$

where $\eta_{i,t-1}$ refers to the filtered probability of state i at time $t-1$, and $p_{i,j}$ is the transition probability of moving from state i to state j. The filtered probabilities can be written in the following way as stated in (50):

$$\eta_{i,t} = P[s_t = i \mid \Psi, I_t] \tag{105}$$

The filtered probabilities are obtained via the Hamilton filter. See Hamilton (1989) and Hamilton (1994) for details. The likelihood function is then expressed as:

$$\ln(L) = \ln\Big(\prod_{t=1}^{T} f(\varepsilon_t \mid \hat{\Psi}, I_{t-1}) \Big) \tag{106}$$

This expression can be combined with the log likelihood of the normal distribution in (95) or the log likelihood of the Student t distribution in (98) to estimate all required parameters simultaneously.

D Estimation of bivariate models

D.1 Two-step estimation of DCC-GARCH and asymmetric DCC-GARCH

The parameters in the DCC-GARCH and asymmetric DCC-GARCH models are estimated using a two-step approach. In step one, we estimate two univariate GARCH(1,1) models - one for daily P&L on the German futures contract and one for daily P&L on the Nordic futures contract. From this, we obtain two sets of GARCH parameters - ω_{GER}, α_{GER} and β_{GER}, as well as ω_{NOR}, α_{NOR} and β_{NOR}. In the case of the Student t distribution, we also obtain υ_{GER} and υ_{NOR} in the first step of the estimation procedure.

From the GARCH(1,1) models, we calculate $\boldsymbol{z}_t$, which is the 2×1 vector of standardized daily P&L. Then we move on to the second step of the estimation procedure, where we maximize the log likelihood function of a bivariate distribution while keeping the two sets of GARCH parameters fixed. From this, we obtain the DCC parameters α, β and γ. In the case of the bivariate Student t distribution, we also obtain υ in the second step.

We note that when estimating a DCC model with bivariate normal distribution in step two, we use univariate normal distribution in each GARCH(1,1) model in step one. Similarly, when estimating a

DCC model with bivariate Student t distribution in step two, we use univariate Student t distribution in each GARCH(1,1) model in step one.

Engle (2002) showed that the two-step approach gives consistent estimates under the assumption of multivariate normal distribution. The disadvantage of this method is, as for QMLE, that estimates are not fully efficient. However, Bauwens and Laurent (2005) argue that this is not relevant for forecasting VaR. They also show that the two-step procedure gives very similar estimates compared to one-step estimation with multivariate Student t.

The first step is accomplished by maximizing the log likelihood functions in (95) or (98). In the following sections, we show the log likelihood functions which are maximized in step two for bivariate standard normal distribution and bivariate standardized Student t distribution.

D.1.1 Bivariate standard normal distribution

We repeat the following assumption from (68):

$$\boldsymbol{\varepsilon}_t \mid I_{t-1} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_t = \boldsymbol{D}_t \boldsymbol{\Upsilon}_t \boldsymbol{D}_t) \tag{107}$$

The bivariate standard normal density is given by:

$$f(\boldsymbol{z}_t; \Upsilon_t) = \frac{1}{2\pi \mid \boldsymbol{\Upsilon}_t \mid^{1/2}} e^{-\frac{1}{2}\boldsymbol{z}_t' \boldsymbol{\Upsilon}_t^{-1} \boldsymbol{z}_t} \tag{108}$$

where $\mid \boldsymbol{\Upsilon}_t \mid$ denotes the determinant of the correlation matrix $\boldsymbol{\Upsilon}_t$ and $\boldsymbol{\Upsilon}_t$ is modelled by a DCC model. The second step of the estimation procedure is performed by maximizing the following expression:

$$ln(L) = \sum_{t=1}^{T} \Big[- ln(2\pi) - \frac{1}{2} ln\big(\mid \boldsymbol{\Upsilon}_t \mid \big) - \frac{1}{2} \boldsymbol{z}_t' \boldsymbol{\Upsilon}_t^{-1} \boldsymbol{z}_t \Big] \tag{109}$$

From this we obtain the DCC parameters α and β, as well as γ in the case of an asymmetric DCC model.

D.1.2 Bivariate standardized Student t distribution

We repeat the following assumption from (71):

$$\boldsymbol{\varepsilon}_t \mid I_{t-1} \sim t(\mathbf{0}, \boldsymbol{\Sigma}_t = \boldsymbol{D}_t \boldsymbol{\Upsilon}_t \boldsymbol{D}_t, \upsilon) \tag{110}$$

The density function of the bivariate standardized Student t distribution with $\upsilon > 2$ is given by:

$$f_{\tilde{t}}(\boldsymbol{z}_t; \boldsymbol{\Upsilon}_t, \upsilon) = C(\boldsymbol{\Upsilon}_t, \upsilon) \left(1 + \frac{\boldsymbol{z}_t' \boldsymbol{\Upsilon}_t^{-1} \boldsymbol{z}_t}{\upsilon - 2} \right)^{-\frac{\upsilon+2}{2}} \tag{111}$$

where

$$C(\boldsymbol{\Upsilon}_t, \upsilon) = \frac{\Gamma\big(\frac{\upsilon+2}{2}\big)}{\Gamma\big(\frac{\upsilon}{2}\big)\big((\upsilon-2)\pi\big) \mid \boldsymbol{\Upsilon}_t \mid^{1/2}} \tag{112}$$

The second step of the estimation procedure is performed by maximizing the following expression:

$$ln(L) = \sum_{t=1}^{T} \left[ln\big(C(\boldsymbol{\Upsilon}_t, \upsilon)\big) - \frac{\upsilon+2}{2} \left(1 + \frac{\boldsymbol{z}_t' \boldsymbol{\Upsilon}_t^{-1} \boldsymbol{z}_t}{\upsilon - 2} \right) \right] \tag{113}$$

From this, we obtain the DCC and asymmetric DCC parameters α and β and γ, as well as the υ parameter of the bivariate standardized Student t distribution.

We note that when estimating the CCC model with bivariate Student t distribution, we still need to maximize (113) to obtain υ. However, the conditional correlation is constant so we set $\boldsymbol{\Upsilon}_t = \boldsymbol{\Upsilon}$.

E Bootstrapping in the Expected Shortfall test

The bootstrapping procedure used in the ES test is explained in detail in Efron (1993). The method is based on empirically finding an appropriate null distribution of a test statistic, rather than assuming one. We then compare the test statistic with the bootstrapped test statistic to assess the hypothesis of zero mean standardized exceedance residuals. We consider the following test statistic:

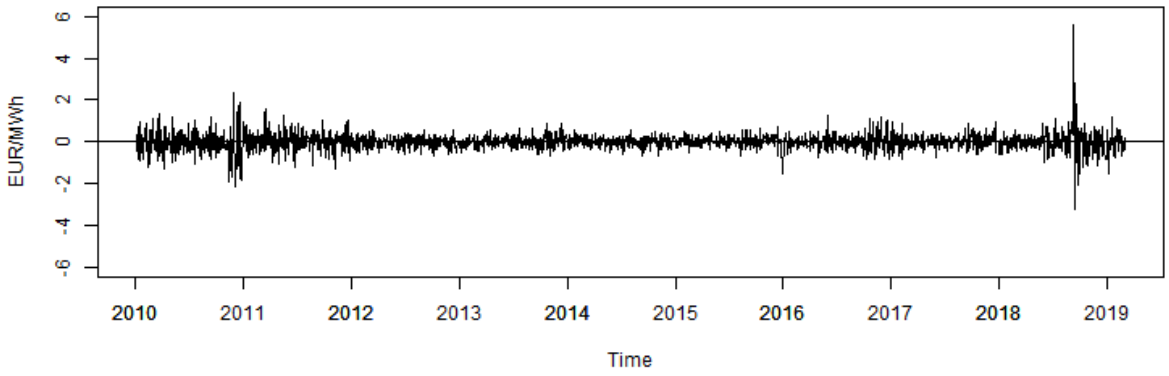
$$T(\hat{\boldsymbol{z}}) = \frac{\hat{\mu}_{\hat{z}} - \mu_{\hat{z}}}{\frac{\hat{\sigma}_{\hat{z}}}{\sqrt{n_{\hat{z}}}}} = \frac{\hat{\mu}_{\hat{z}}}{\frac{\hat{\sigma}_{\hat{z}}}{\sqrt{n_{\hat{z}}}}} \tag{114}$$

where $\hat{\mu}_{\hat{z}}$ and $\hat{\sigma}_{\hat{z}}$ are the mean and standard deviation of the variable we use to test $\hat{z}_t$ by bootstrapping, and $n_{\hat{z}}$ is the number of VaR exceedances. The appropriate null distribution should follow the null hypothesis. In our case, this is for $\hat{z}_t$ to have a zero mean. We define the zero mean variable $\tilde{z}_t = \hat{z}_t - \hat{\mu}_{\hat{z}}$. Random bootstrap samples $\tilde{\boldsymbol{z}}^*$ of size $n_{\hat{z}}$ is then drawn from $\tilde{\boldsymbol{z}}$ with replacement. An empirical null distribution of T is obtained by calculating corresponding test statistics for each sample:

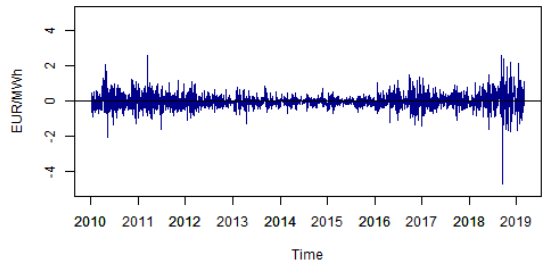
$$T(\tilde{\boldsymbol{z}}^*) = \frac{\hat{\mu}_{\tilde{z}^*}}{\frac{\hat{\sigma}_{\tilde{z}^*}}{\sqrt{n_{\tilde{z}^*}}}} = \frac{\hat{\mu}_{\tilde{z}^*}}{\frac{\hat{\sigma}_{\tilde{z}^*}}{\sqrt{n_{\hat{z}}}}} \tag{115}$$

F Data

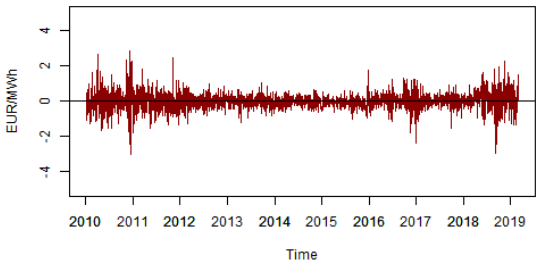
F.1 P&L of front-year contracts



(a) Daily P&L of the front-year spread



(b) Daily P&L of German front-year contracts



(c) Daily P&L of Nordic front-year contracts

Figure 16: P&L of spread, German and Nordic front-year contracts

F.2 Empirical properties of the daily P&L of front-year contracts

Table 30: Empirical properties of the daily P&L of front-year contracts

Contract (*Position*)	2010	2011	2012	2013	2014	2015	2016	2017	2018
99% empirical quantiles									
Spread (*Long*)	-1.8	-1.09	-0.61	-0.57	-0.55	-0.61	-0.78	-0.69	-1.36
Spread (*Short*)	1.59	1.15	0.53	0.66	0.46	0.58	1.09	0.85	1.42
German (*Long*)	-1.12	-0.96	-0.79	-0.70	-0.46	-0.52	-1.05	-0.85	-1.64
German (*Short*)	1.29	1.12	0.84	0.70	0.45	0.43	1.23	0.93	2.10
Nordic (*Long*)	-1.89	-1.43	-0.80	-0.90	-0.57	-0.7	-1.22	-1.08	-1.65
Nordic (*Short*)	2.30	1.26	0.85	0.80	0.60	0.59	1.26	0.90	1.76
Mean conditional on 99% quantile exceedance									
Spread (*Long*)	-1.96	-1.23	-0.63	-0.62	-0.65	-0.97	-0.82	-0.75	-2.26
Spread (*Short*)	2.00	1.37	0.56	0.82	0.53	0.64	1.19	0.94	3.35
German (*Long*)	-1.56	-1.19	-0.96	-0.94	-0.58	-0.62	-1.29	-1.25	-2.7
German (*Short*)	1.71	1.96	1.00	0.78	0.58	0.49	1.38	1.05	2.43
Nordic (*Long*)	-2.48	-1.67	-0.99	-1.03	-0.60	-0.88	-1.46	-1.67	-2.39
Nordic(*Short*)	2.63	1.87	1.01	0.91	0.62	1.08	1.27	1.03	2.03
Excess kurtosis									
Spread	1.74	0.78	-0.29	0.53	-0.09	6.94	1.82	0.82	25.77
German	1.84	3.54	1.17	1.95	1.70	0.76	0.56	0.92	7.57
Nordic	1.82	1.32	-0.06	0.67	-0.33	5.69	1.43	6.07	2.23
Skew									
Spread	0.13	-0.03	-0.23	0.38	-0.23	-1.00	0.64	0.15	2.36
German	0.36	0.91	0.32	-0.08	-0.08	-0.26	0.12	-0.08	-0.87
Nordic	0.08	0.19	0.03	-0.07	0.01	0.70	-0.12	-1.05	-0.4

2019 not included as datasample only runs the first two months

F.3 Empirical properties of the 95% quantile of the P&L of all contracts, front-quarter

Table 31: Empirical properties of the 95% quantile of the P&L of all contracts, front-quarter

Contract (*Position*)	2010	2011	2012	2013	2014	2015	2016	2017	2018
Empirical quantiles									
Spread (*Long*)	-2.14	-1.49	-1.12	-0.76	-0.81	-0.68	-0.89	-0.81	-1.40
Spread (*Short*)	1.85	1.80	1.02	0.80	0.68	0.77	0.95	0.84	1.25
German (*Long*)	-0.74	-0.86	-0.72	-0.50	-0.42	-0.39	-0.82	-0.63	-1.43
German (*Short*)	1.23	0.88	0.69	0.45	0.35	0.38	1.05	0.77	1.37
Nordic (*Long*)	-1.98	-1.83	-1.19	-0.99	-0.89	-0.89	-0.93	-0.94	-1.57
Nordic (*Short*)	2.15	1.51	1.07	0.85	0.80	0.85	1.20	0.81	1.44
Mean conditional on 95% quantile exceedance									
Spread (*Long*)	-2.86	-1.94	-1.43	-0.97	-0.98	-1.11	-1.21	-1.05	-2.23
Spread (*Short*)	3.10	2.46	1.44	1.17	0.86	1.10	1.32	1.16	2.36
German (*Long*)	-0.97	-1.20	-1.13	-0.72	-0.65	-0.52	-1.25	-0.78	-2.33
German (*Short*)	1.45	1.85	0.91	0.57	0.55	0.53	1.42	0.98	2.00
Nordic (*Long*)	-3.24	-2.57	-1.49	-1.26	-1.08	-1.22	-1.52	-1.24	-2.27
Nordic (*Short*)	3.41	2.55	1.41	1.24	1.00	1.31	1.49	1.09	2.22

2019 not included as datasample only runs the first two months

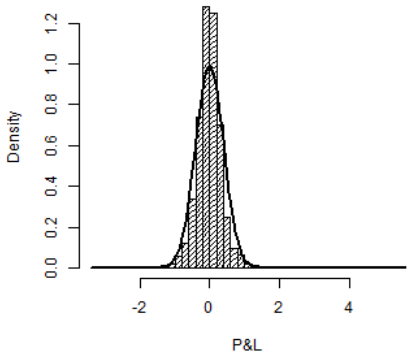
F.4 Empirical properties of the 95% quantile of the P&L of all contracts, front-year

Table 32: Empirical properties of the 95% quantile of the P&L of all contracts, front-year

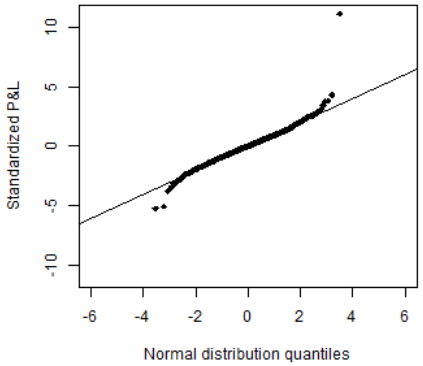
95% empirical quantiles									
Spread (*Long*)	-1.00	-0.68	-0.49	-0.41	-0.38	-0.30	-0.54	-0.45	-0.78
Spread (*Short*)	0.93	0.76	0.35	0.46	0.34	0.37	0.52	0.45	0.71
German (*Long*)	-0.70	-0.70	-0.50	-0.47	-0.33	-0.33	-0.74	-0.57	-1.01
German (*Short*)	0.86	0.80	0.50	0.49	0.27	0.27	0.81	0.65	1.15
Nordic (*Long*)	-1.18	-0.95	-0.60	-0.58	-0.47	-0.50	-0.79	-0.55	-1.02
Nordic (*Short*)	1.39	0.85	0.65	0.54	0.40	0.38	0.78	0.51	1.06
Mean conditional on 95% quantile exceedance									
Spread (*Long*)	-1.44	-0.93	-0.56	-0.51	-0.47	-0.52	-0.67	-0.58	-1.27
Spread (*Short*)	1.33	1.01	0.45	0.61	0.42	0.51	0.88	0.67	1.46
German (*Long*)	-1.00	-0.93	-0.69	-0.65	-0.44	-0.45	-0.96	-0.80	-1.63
German (*Short*)	1.19	1.16	0.72	0.62	0.42	0.37	1.07	0.84	1.61
Nordic (*Long*)	-1.67	-1.31	-0.75	-0.78	-0.53	-0.67	-1.06	-1.01	-1.46
Nordic (*Short*)	1.93	1.27	0.77	0.69	0.55	0.64	1.10	0.75	1.47

2019 not included as datasample only runs the first two months

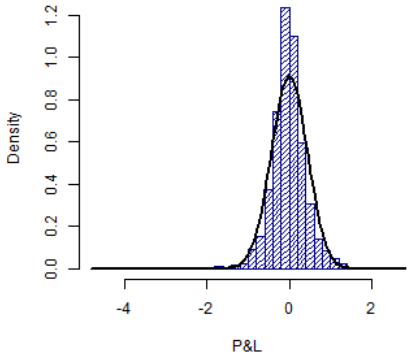
F.5 Histogram and QQ-plot of front-year contracts



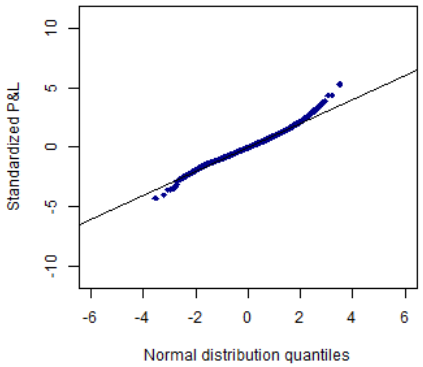
(a) Histogram of daily P&L of the spread



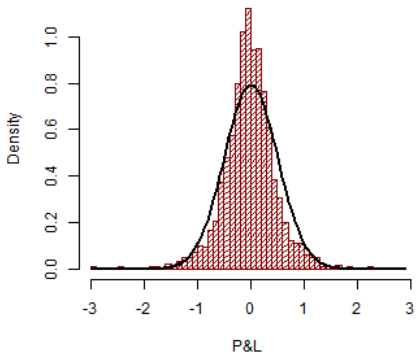
(b) QQ-plot of daily P&L of the spread against quantiles from the normal distribution



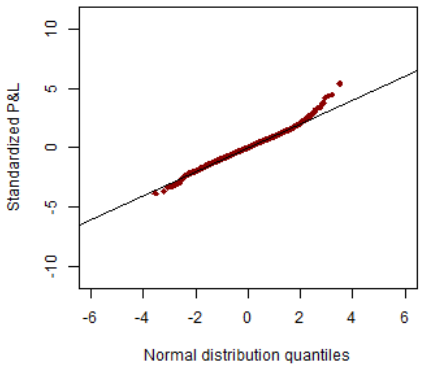
(c) Histogram of daily P&L of German contracts



(d) QQ-plot of daily P&L of German contracts against quantiles from the normal distribution



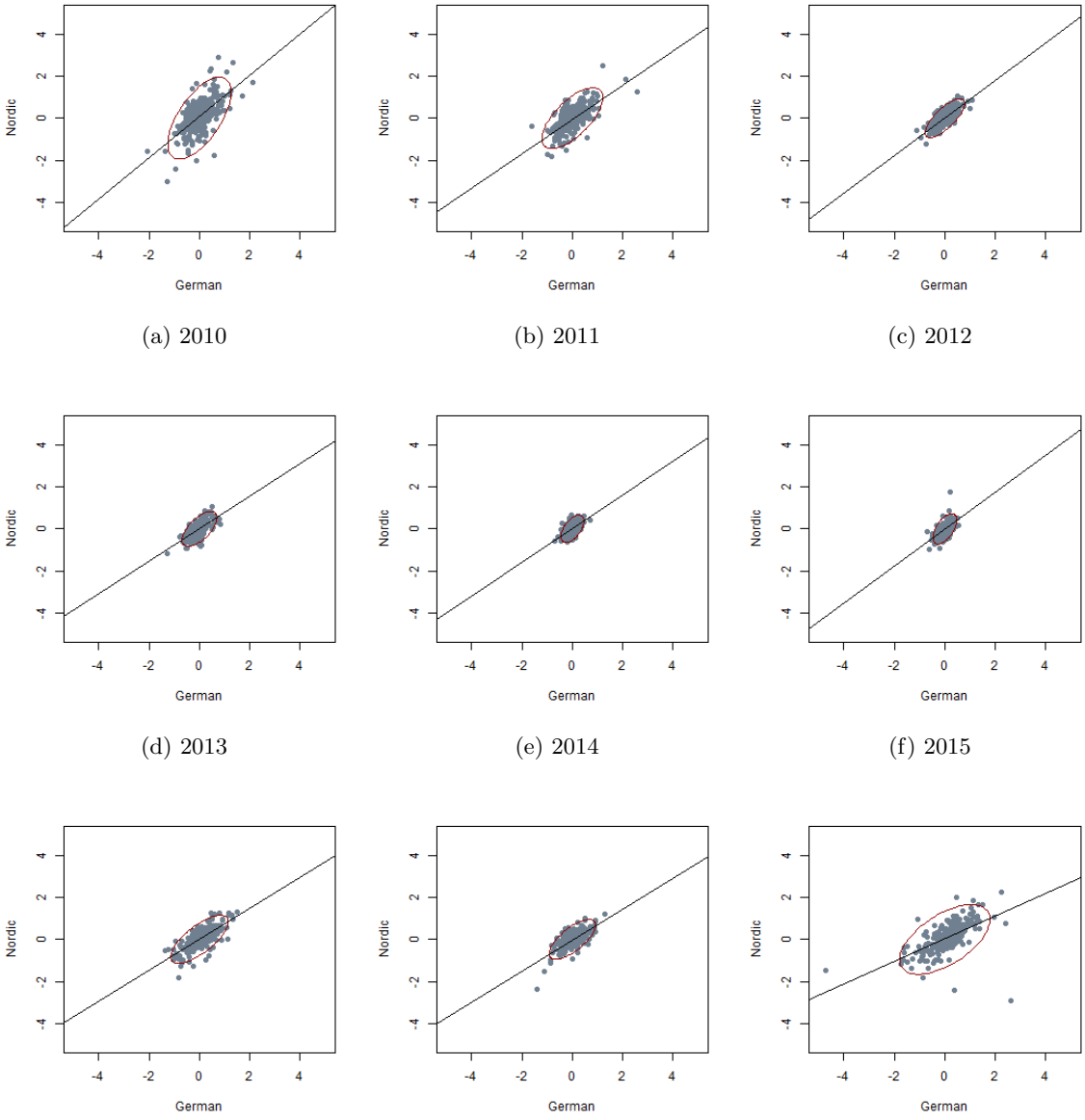
(e) Histogram of P&L of Nordic contracts



(f) QQ-plot of daily P&L of Nordic contracts against quantiles from the normal distribution

Figure 17: Histogram and QQ-plot of P&L of front-year contracts. Spread (black), German (blue) and Nordic (red). The P&Ls are standardized using GARCH(1,1)

F.6 Scatterplots per year of P&L of German and Nordic contracts, front-quarter



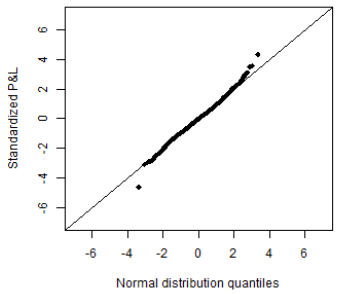
(a) 2010 (b) 2011 (c) 2012

(d) 2013 (e) 2014 (f) 2015

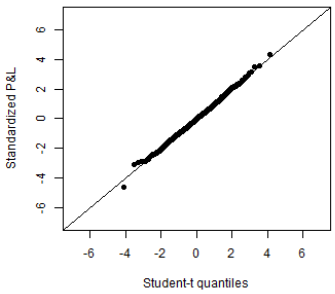
(g) 2016 (h) 2017 (i) 2018

Figure 18: Scatterplots of daily P&L of Nordic and German front-year contracts per calendar year, with a Nordic on German regression line. Red ellipse is 95% contour of a bivariate normal distribution

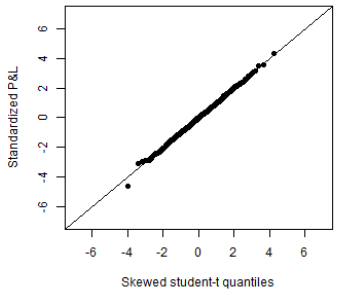
F.7 QQ-plots of in-sample P&L, front-quarter



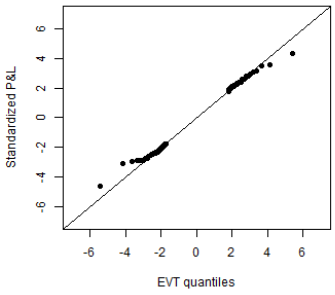
(a) Spread daily P&L vs Normal quantiles



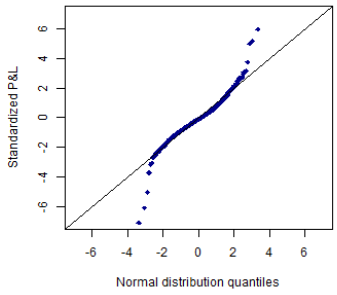
(b) Spread daily P&L vs Student t quantiles



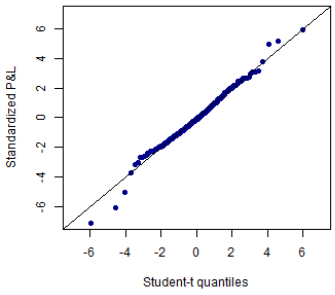
(c) Spread daily P&L vs skewed Student t quantiles



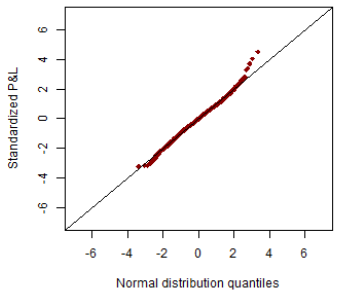
(d) Spread daily P&L of spread vs EVT quantiles



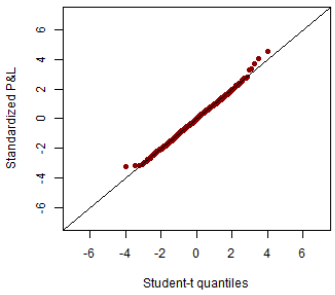
(e) German daily P&L vs normal quantiles



(f) German daily P&L vs Student t quantiles



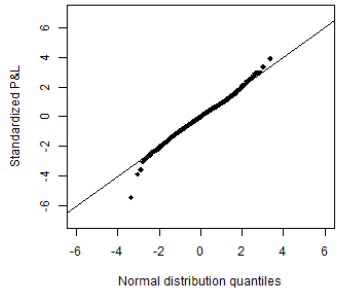
(g) Nordic daily P&L vs normal quantiles



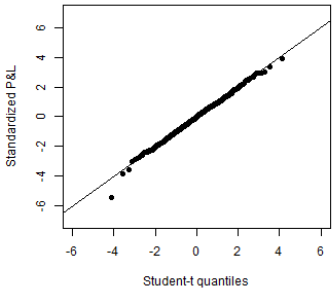
(h) Nordic daily P&L vs Student t quantiles

Figure 19: QQ-plots - P&L of front-quarter Spread, and German and Nordic futures prices by using a GARCH(1,1)-n and plotted vs selected distribution quantiles

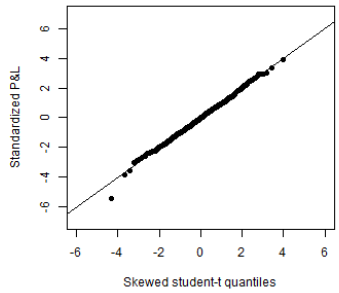
F.8 QQ-plots of in-sample P&L, front-year



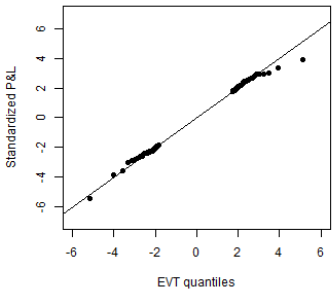
(a) Spread daily P&L vs Normal quantiles



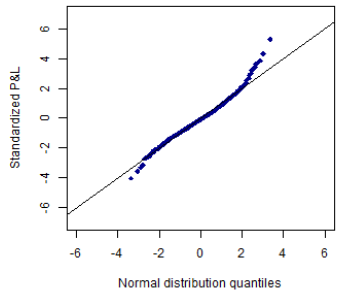
(b) Spread daily P&L vs Student t quantiles



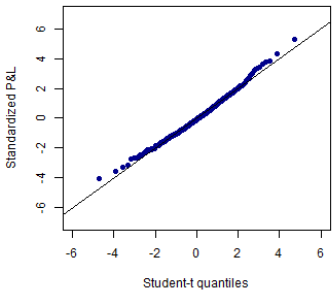
(c) Spread daily P&L vs skewed Student t quantiles



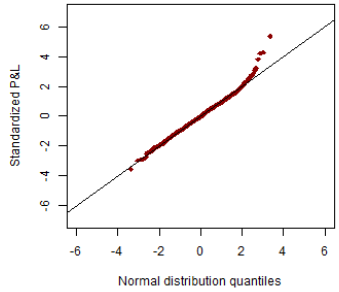
(d) Spread daily P&L of spread vs EVT quantiles



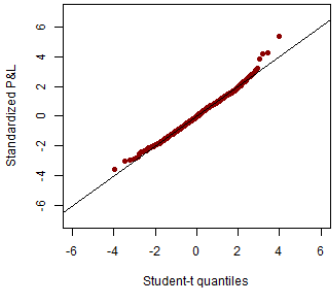
(e) German daily P&L vs normal quantiles



(f) German daily P&L vs Student t quantiles



(g) Nordic daily P&L vs normal quantiles

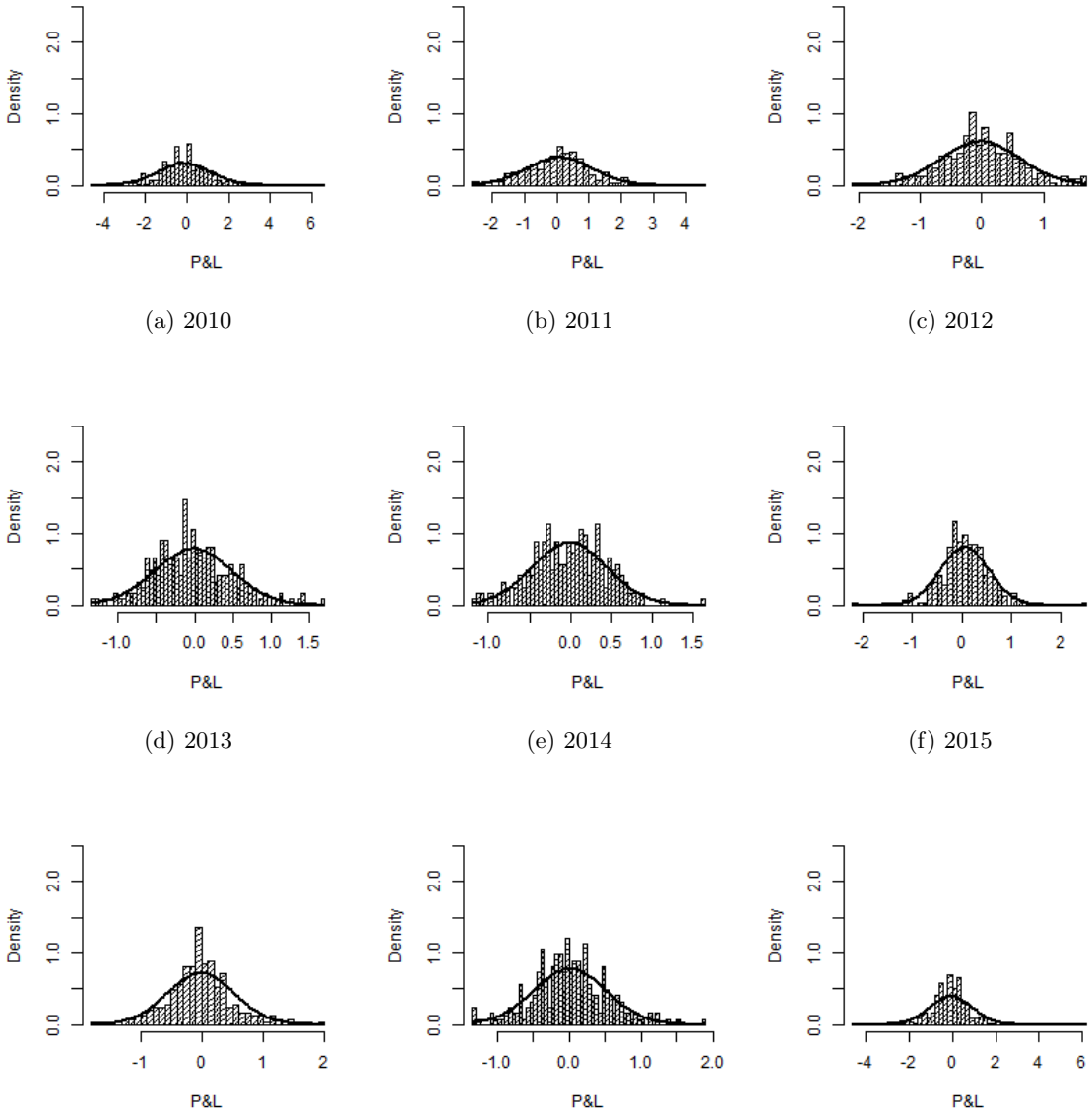


(h) Nordic daily P&L vs Student t quantiles

Figure 20: QQ-plots - Standardized P&L of front-year Spread, and German and Nordic futures prices by using a GARCH(1,1)-n and plotted vs selected distribution quantiles

F.9 Histogram of P&L, front-quarter

F.9.1 Histogram of spread P&L, front-quarter



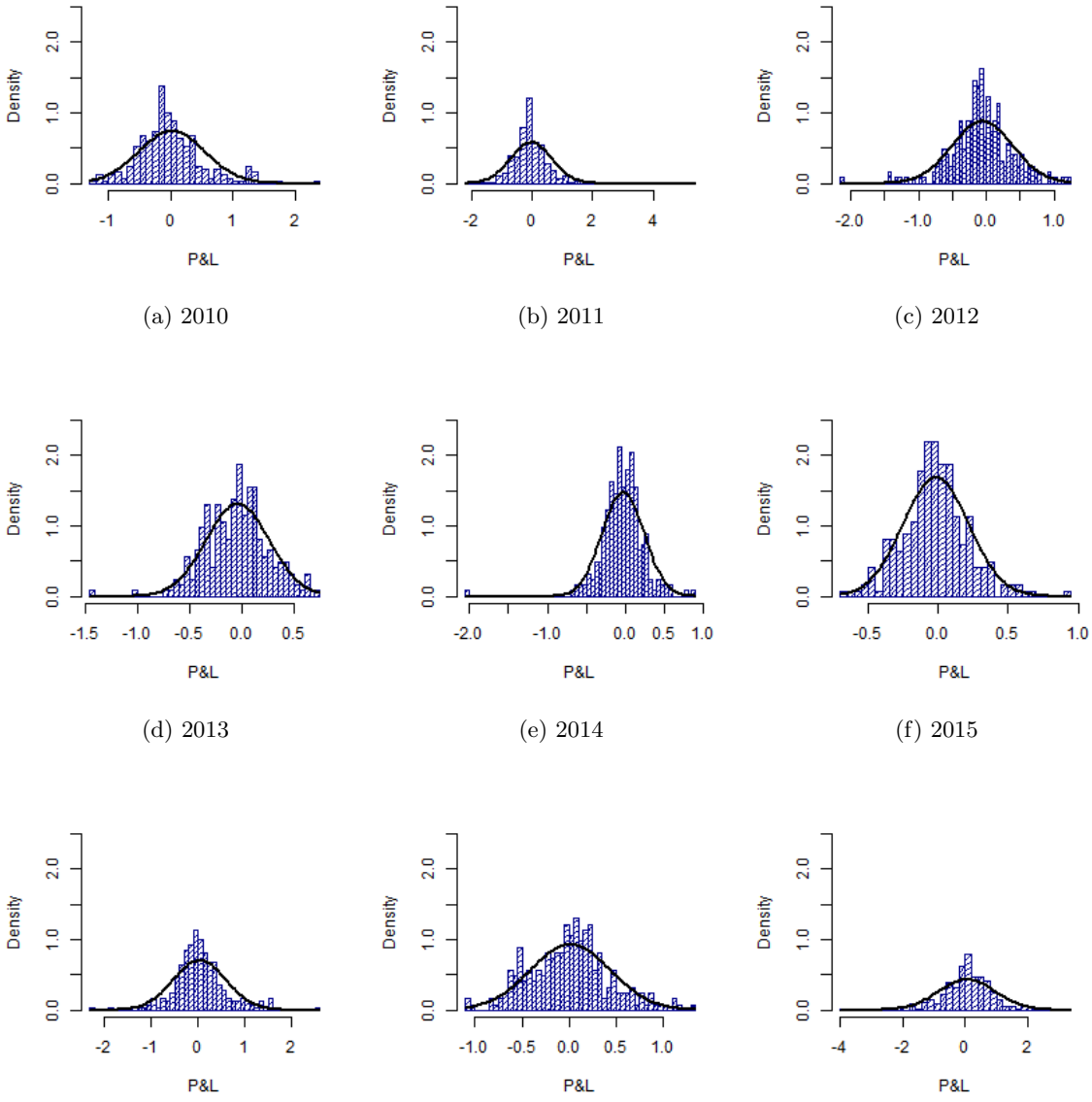
(a) 2010 (b) 2011 (c) 2012

(d) 2013 (e) 2014 (f) 2015

(g) 2016 (h) 2017 (i) 2018

Figure 21: Histogram of front-quarter spread P&L per year, with an overlay from the normal distribution (solid black)

F.9.2 Histogram of German P&L, front-quarter



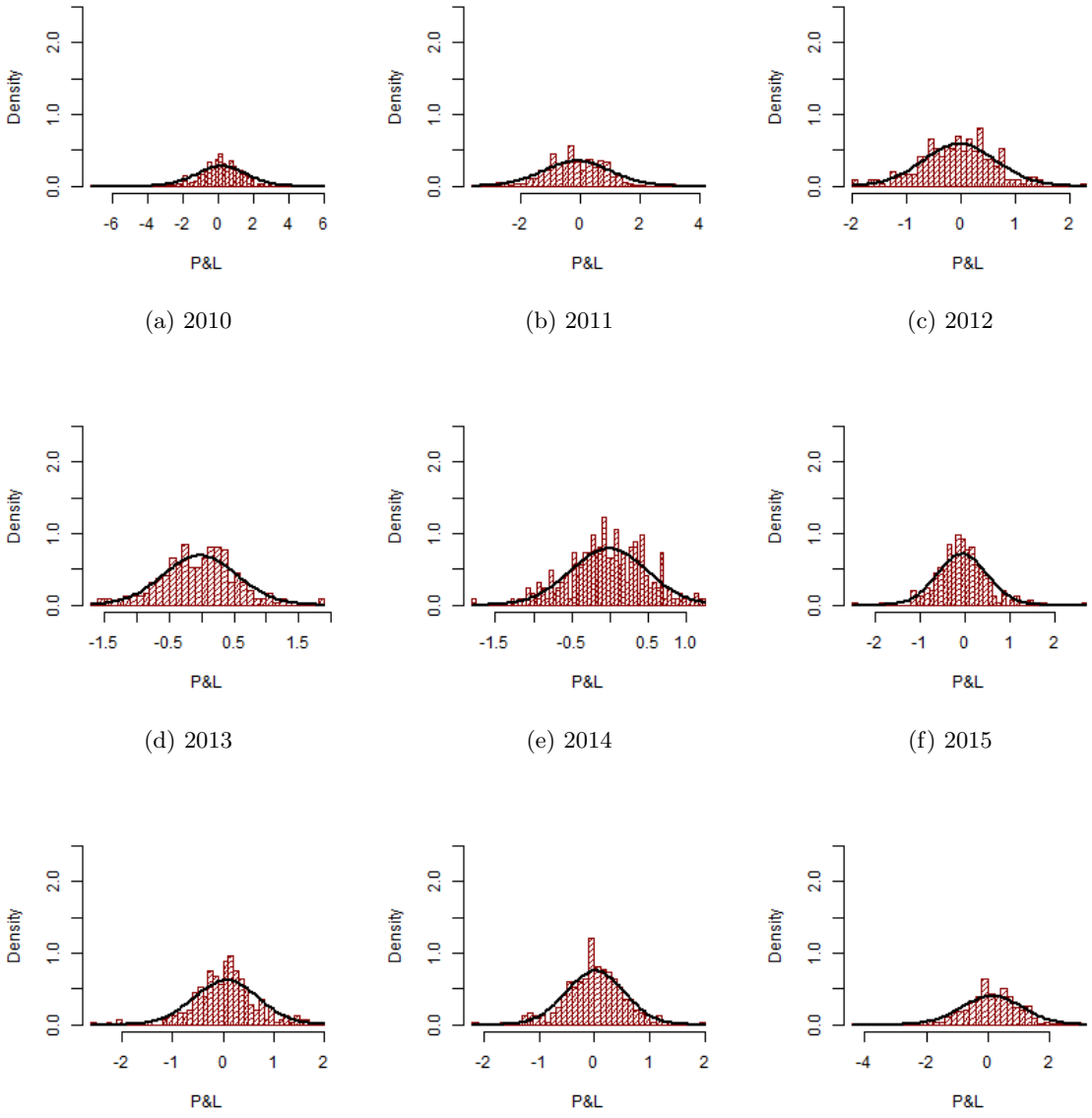
(a) 2010 (b) 2011 (c) 2012

(d) 2013 (e) 2014 (f) 2015

(g) 2016 (h) 2017 (i) 2018

Figure 22: Histogram of front-quarter German P&L per year, with an overlay from the normal distribution (solid black)

F.9.3 Histogram of Nordic P&L, front-quarter



(a) 2010 (b) 2011 (c) 2012

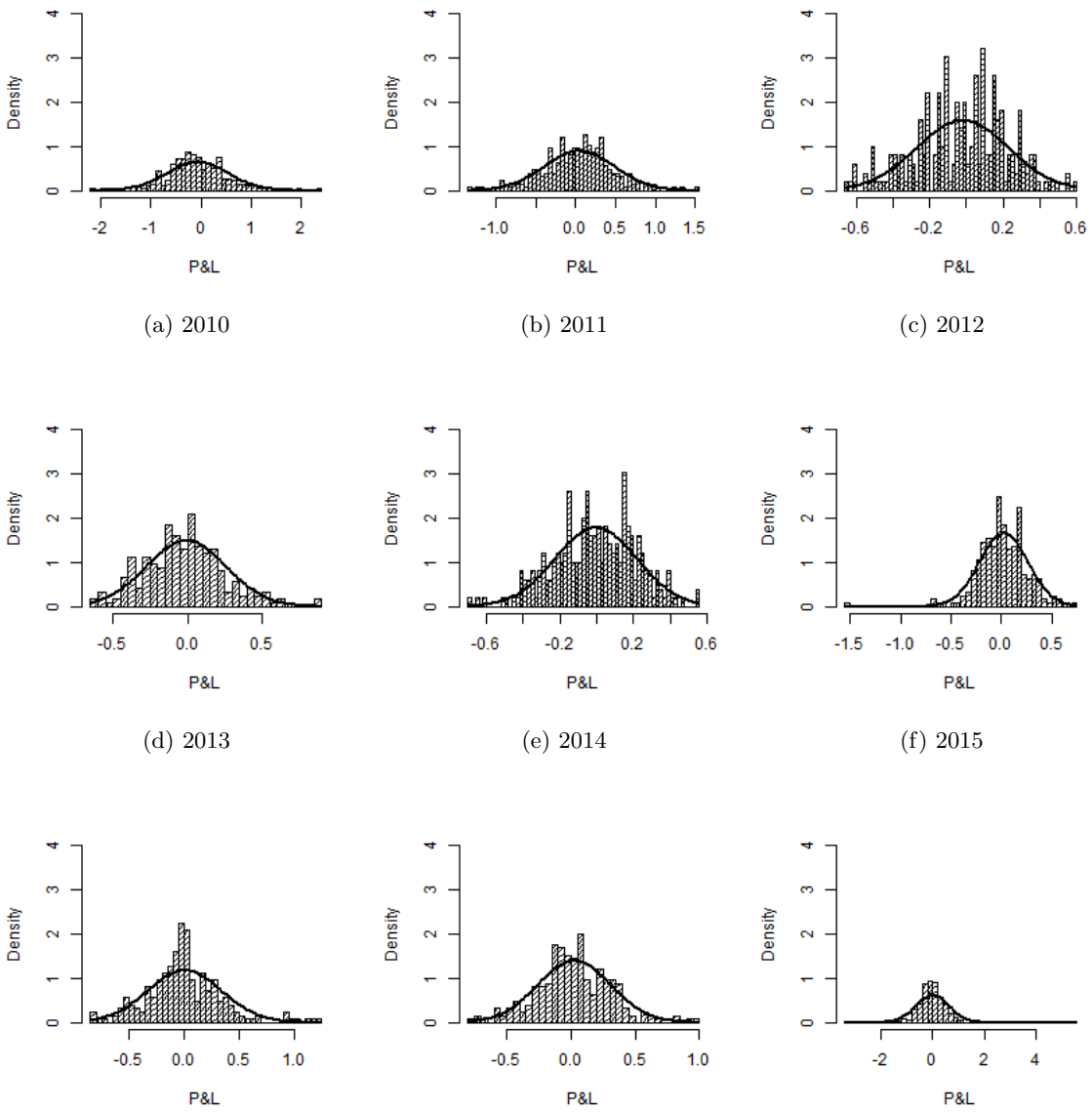
(d) 2013 (e) 2014 (f) 2015

(g) 2016 (h) 2017 (i) 2018

Figure 23: Histogram of front-year Nordic P&L per year, with an overlay from the normal distribution (solid black)

F.10 Histogram of P&L, front-year

F.10.1 Histogram of spread P&L, front-year



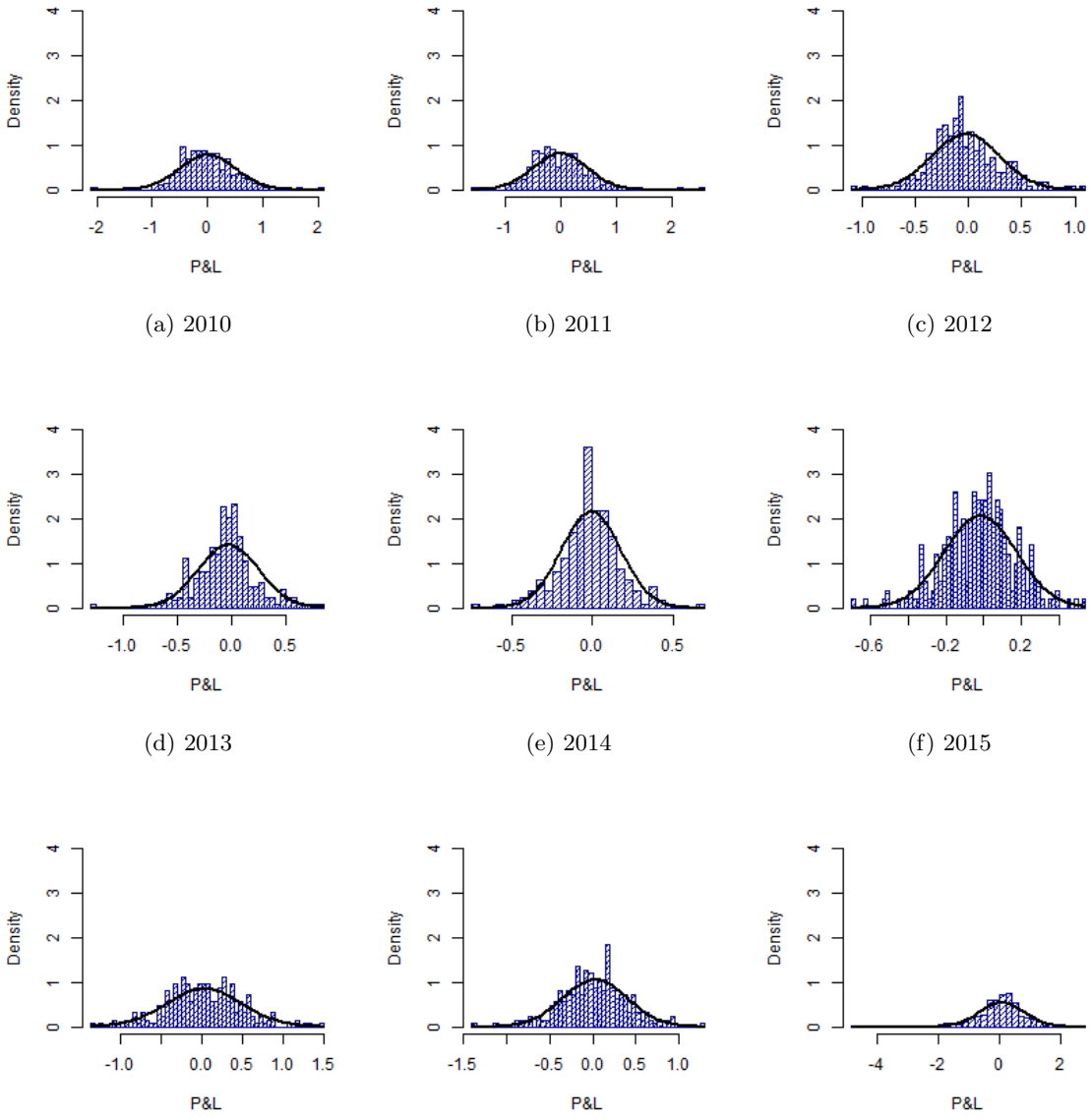
(a) 2010 (b) 2011 (c) 2012

(d) 2013 (e) 2014 (f) 2015

(g) 2016 (h) 2017 (i) 2018

Figure 24: Histogram of front-year spread P&L per year, with an overlay from the normal distribution (solid black)

F.10.2 Histogram of German P&L, front-year



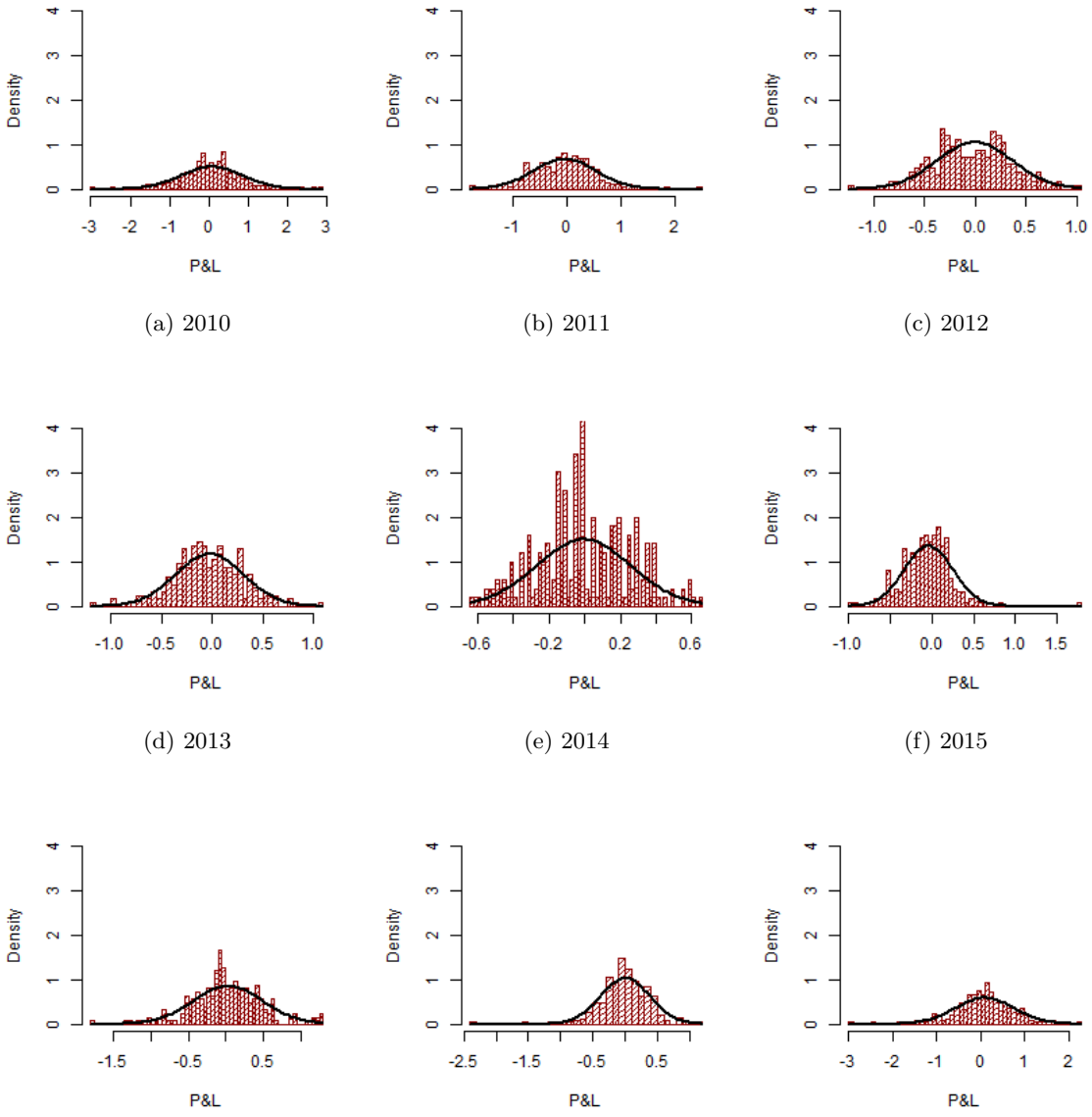
(a) 2010 (b) 2011 (c) 2012

(d) 2013 (e) 2014 (f) 2015

(g) 2016 (h) 2017 (i) 2018

Figure 25: Histogram of front-year German P&L per year, with an overlay from the normal distribution (solid black)

F.10.3 Histogram of Nordic P&L, front-year



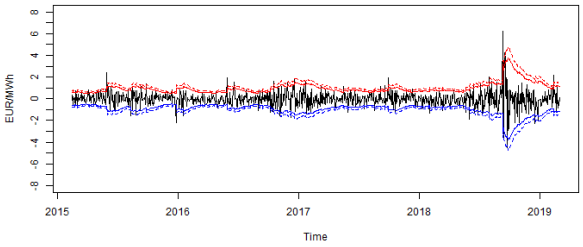
(a) 2010 (b) 2011 (c) 2012

(d) 2013 (e) 2014 (f) 2015

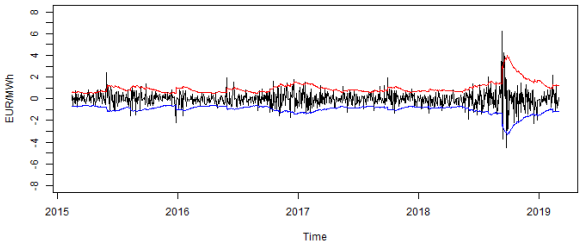
(g) 2016 (h) 2017 (i) 2018

Figure 26: Histogram of front-year Nordic P&L per year, with an overlay from the normal distribution (solid black)

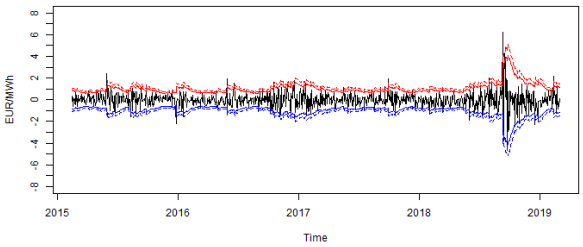
F.11 95% VaR and ES forecasts with fixed estimation window



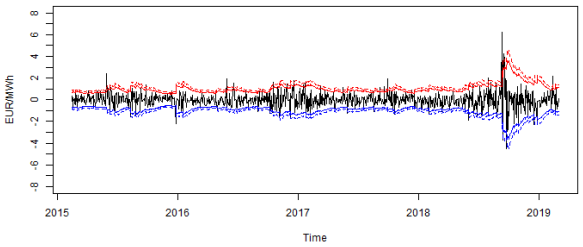
(a) RiskMetrics-n



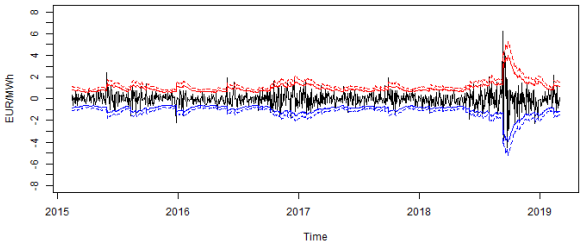
(b) Volatility-adjusted quantile regression



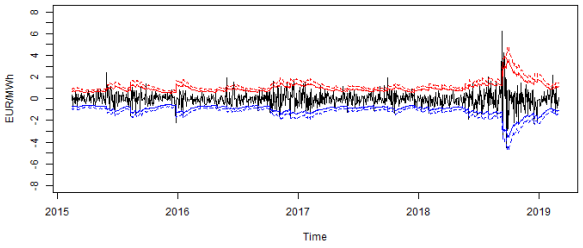
(c) GARCH(1,1)-n



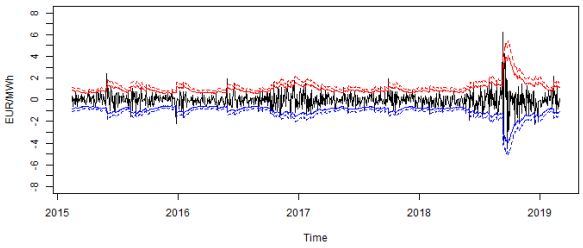
(d) GJR-GARCH(1,1)-n



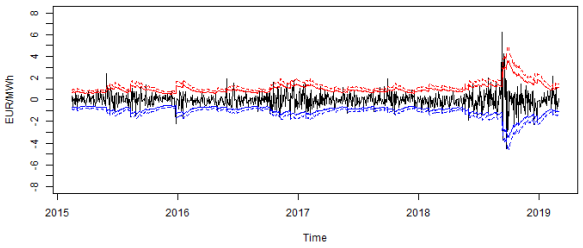
(e) GARCH(1,1)-t



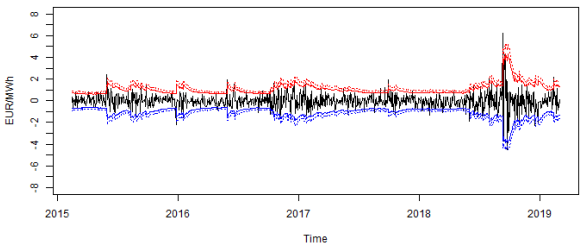
(f) GJR-GARCH(1,1)-t



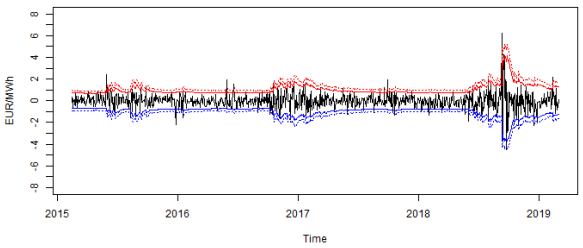
(g) GARCH(1,1)-skew t



(h) GJR-GARCH(1,1)-skew t

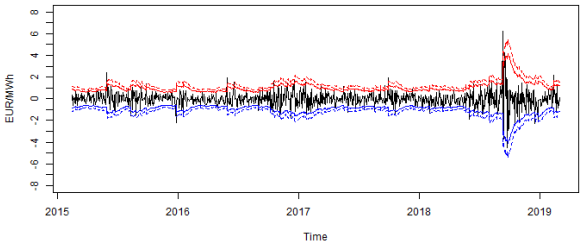


(i) MSGARCH(1,1)-n

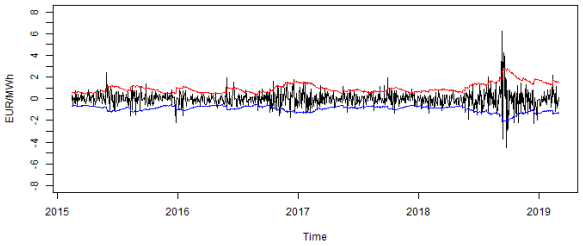


(j) MSGARCH(1,1)-t

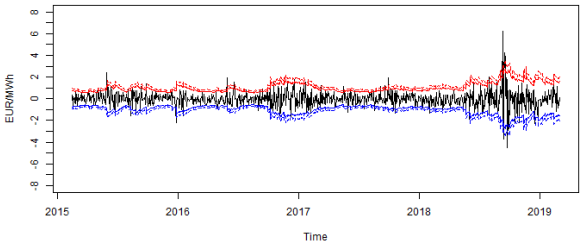
Figure 27: 95% VaR and ES forecasts with fixed estimation window for a long position (solid blue and dotted blue) and short position (solid red and dotted red), front-quarter



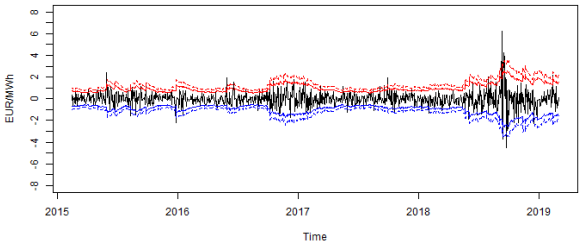
(a) GARCH(1,1)-EVT



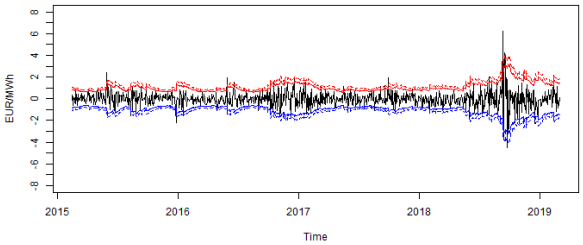
(b) Vol-corr-adj qreg



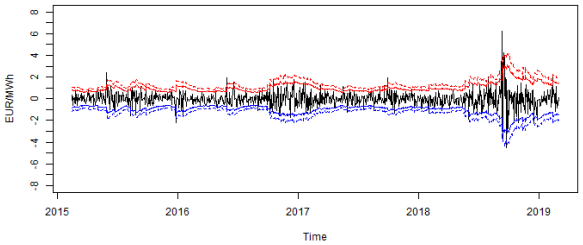
(c) CCC-n



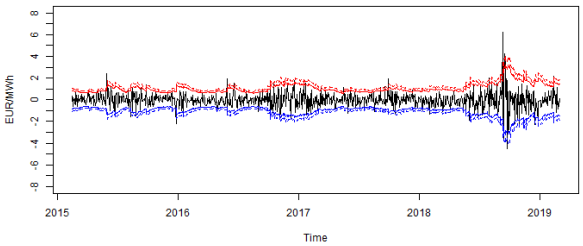
(d) CCC-t



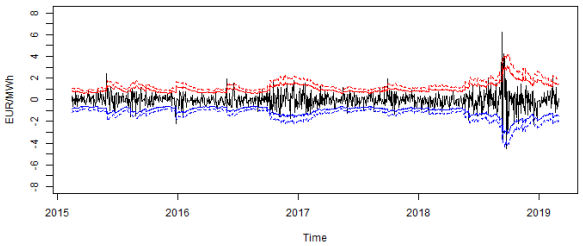
(e) DCC-n



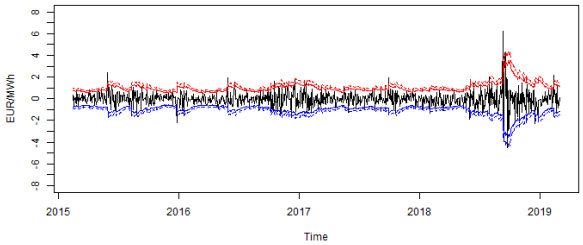
(f) DCC-t



(g) aDCC-n

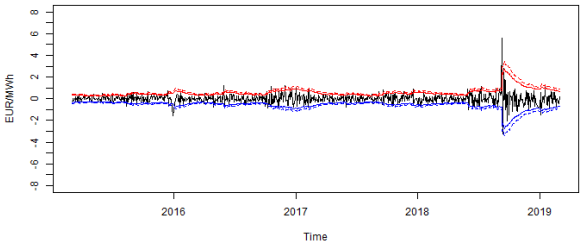


(h) aDCC-t

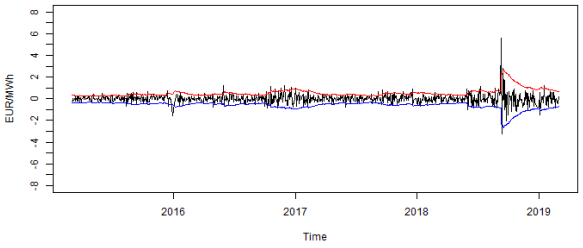


(i) DCC-EWMA

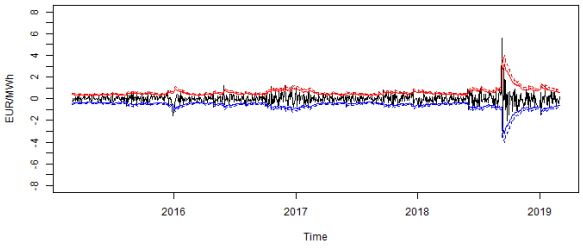
Figure 28: 95% VaR and ES forecasts with fixed estimation window for a long position (solid blue and dotted blue) and short position (solid red and dotted red), front-quarter



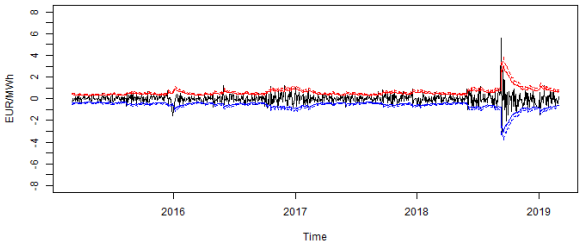
(a) RiskMetrics-n



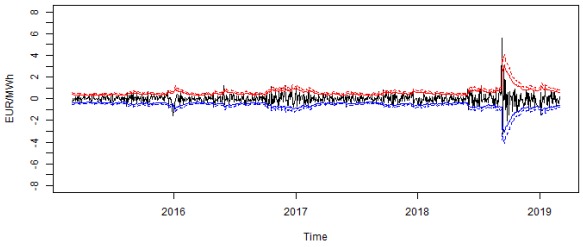
(b) Volatility-adjusted quantile regression



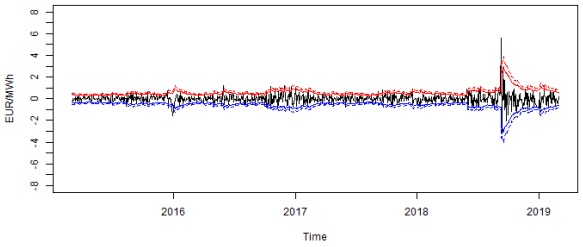
(c) GARCH(1,1)-n



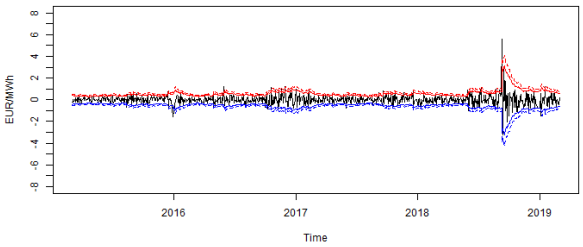
(d) GJR-GARCH(1,1)-n



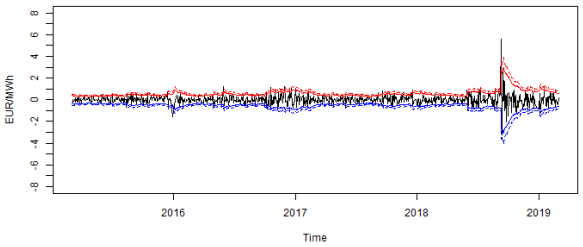
(e) GARCH(1,1)-t



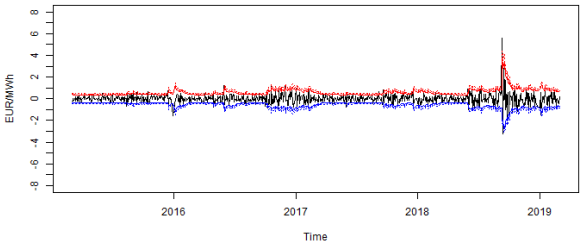
(f) GJR-GARCH(1,1)-t



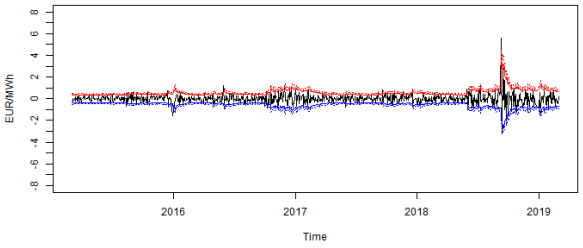
(g) GARCH(1,1)-skew t



(h) GJR-GARCH(1,1)-skew t

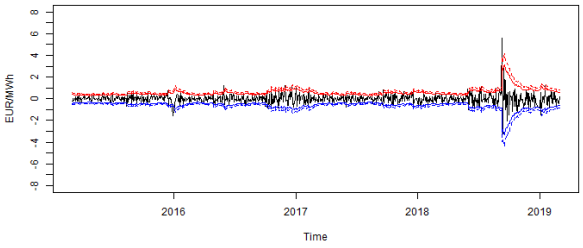


(i) MSGARCH(1,1)-n

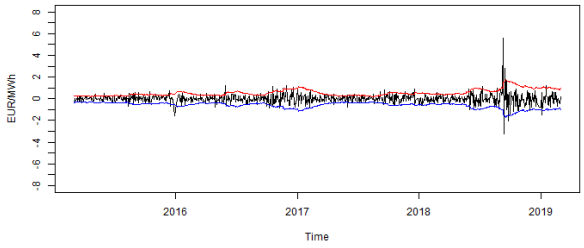


(j) MSGARCH(1,1)-t

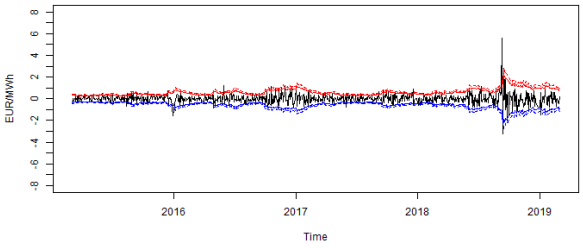
Figure 29: 95% VaR and ES forecasts with fixed estimation window for a long position (solid blue and dotted blue) and short position (solid red and dotted red), front-year



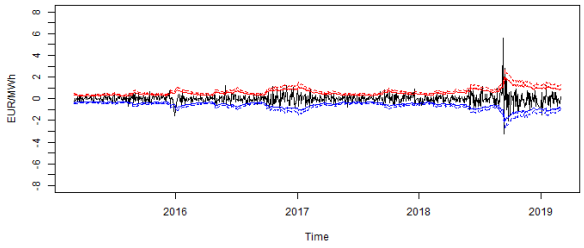
(a) GARCH(1,1)-EVT



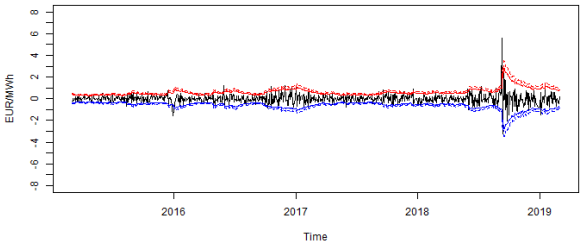
(b) Vol-corr-adj qreg



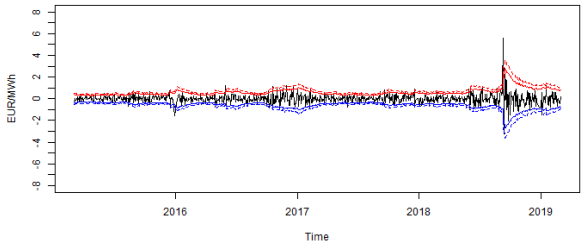
(c) CCC-n



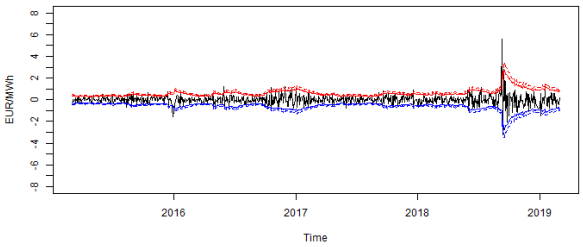
(d) CCC-t



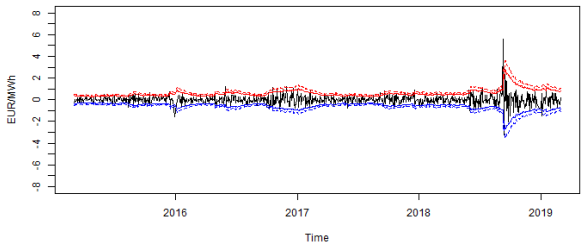
(e) DCC-n



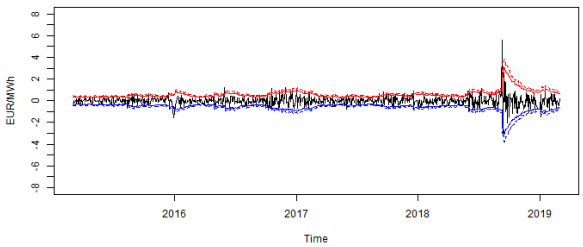
(f) DCC-t



(g) aDCC-n



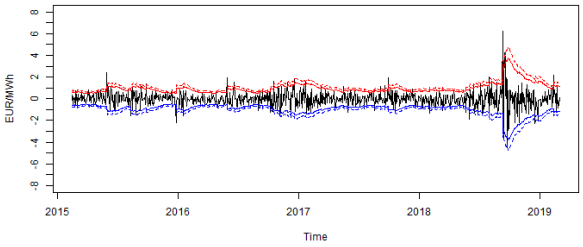
(h) aDCC-t



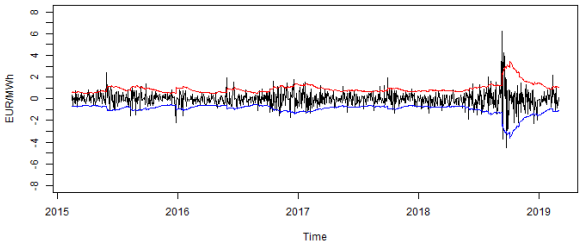
(i) DCC-EWMA

Figure 30: 95% VaR and ES forecasts with fixed estimation window for a long position (solid blue and dotted blue) and short position (solid red and dotted red), front-year

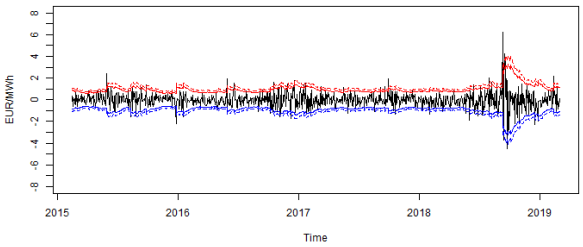
F.12 95% VaR and ES forecasts with reestimation



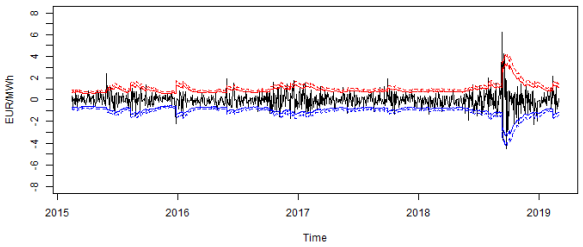
(a) RiskMetrics-n



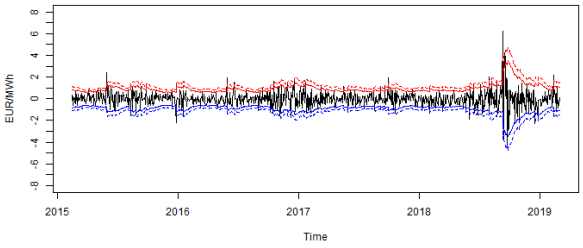
(b) Volatility-adjusted quantile regression



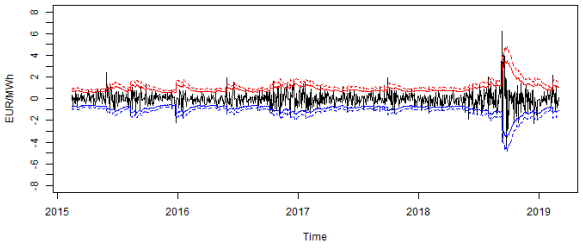
(c) GARCH(1,1)-n



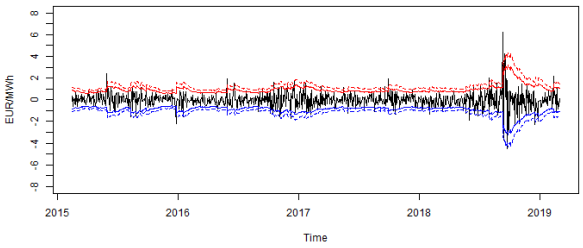
(d) GJR-GARCH(1,1)-n



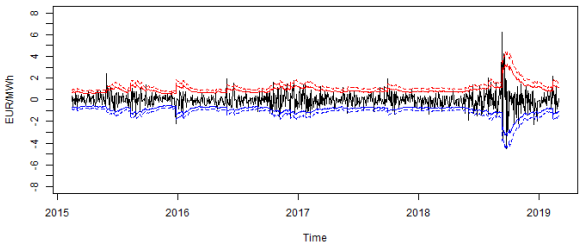
(e) GARCH(1,1)-t



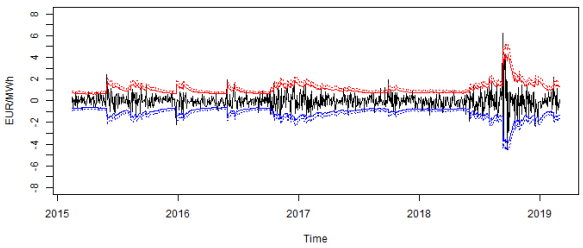
(f) GJR-GARCH(1,1)-t



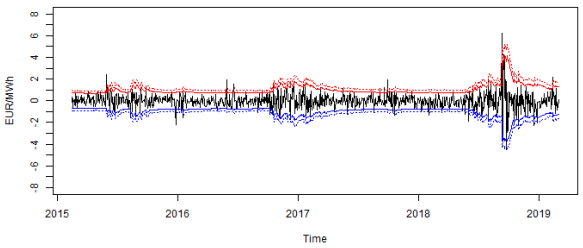
(g) GARCH(1,1)-skew t



(h) GJR-GARCH(1,1)-skew t

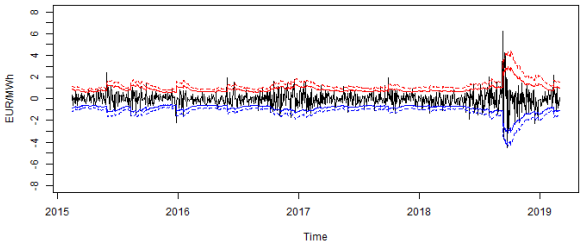


(i) MSGARCH(1,1)-n

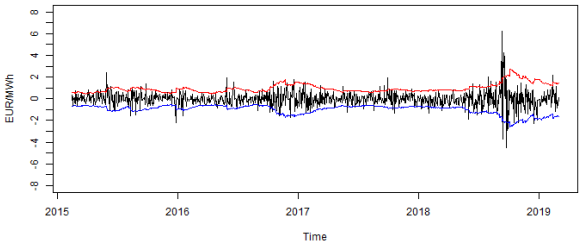


(j) MSGARCH(1,1)-t

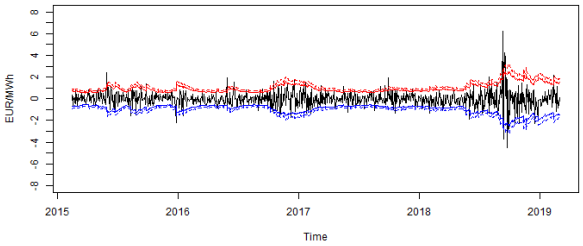
Figure 31: 95% VaR and ES forecasts with reestimation for a long position (solid blue and dotted blue), short position (solid red and dotted red), front-quarter



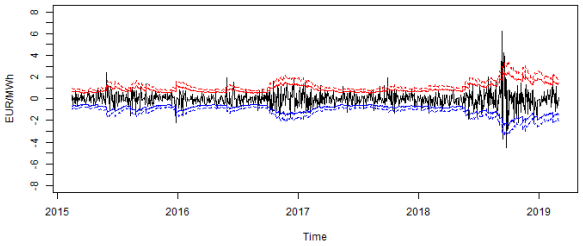
(a) GARCH(1,1)-EVT



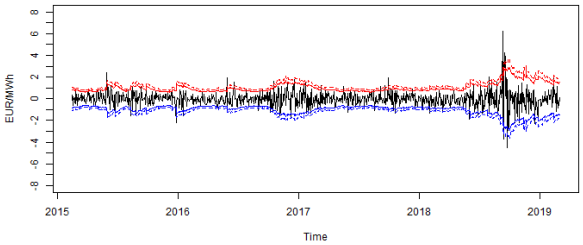
(b) Vol-corr-adj qreg



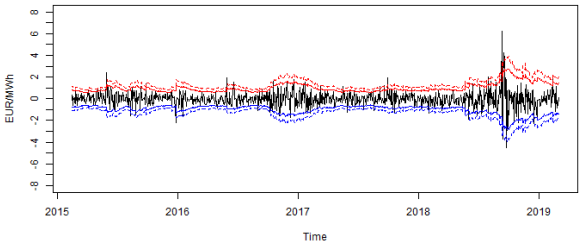
(c) CCC-n



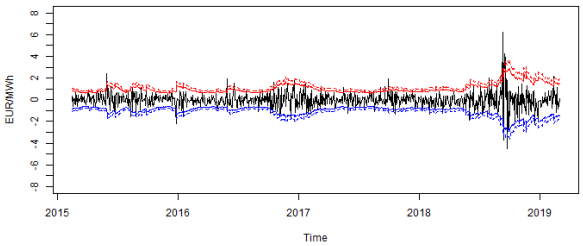
(d) CCC-t



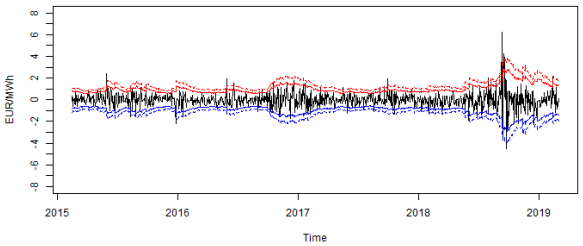
(e) DCC-n



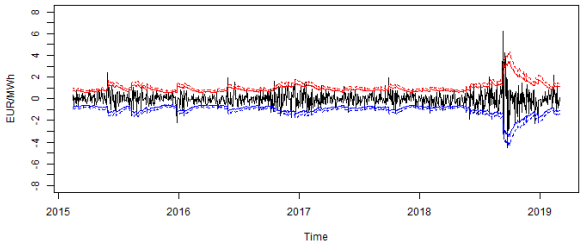
(f) DCC-t



(g) aDCC-n

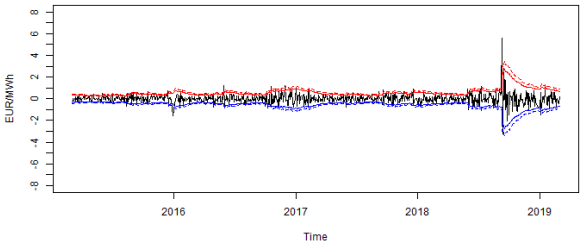


(h) aDCC-t

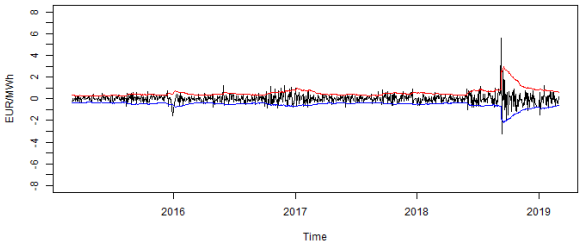


(i) DCC-EWMA

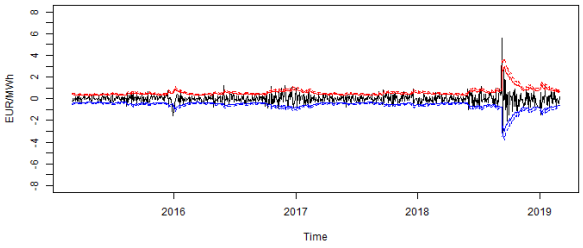
Figure 32: 95% VaR and ES forecasts with reestimation for a long position (solid blue and dotted blue), short position (solid red and dotted red), front-quarter



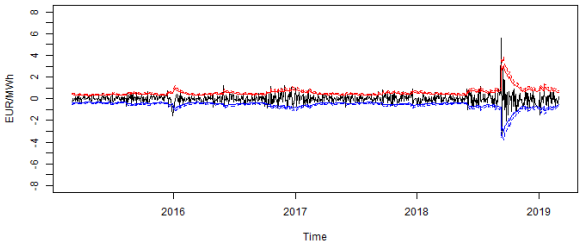
(a) RiskMetrics-n



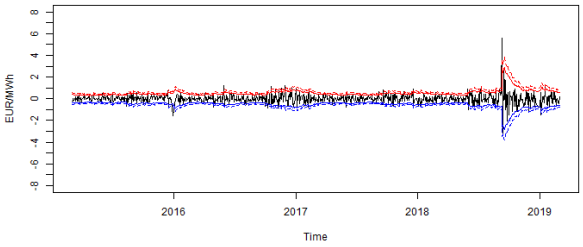
(b) Volatility-adjusted quantile regression



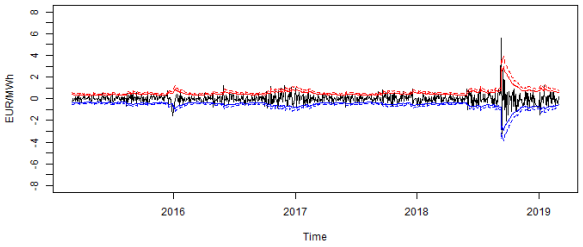
(c) GARCH(1,1)-n



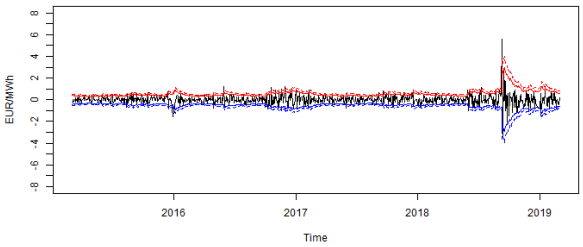
(d) GJR-GARCH(1,1)-n



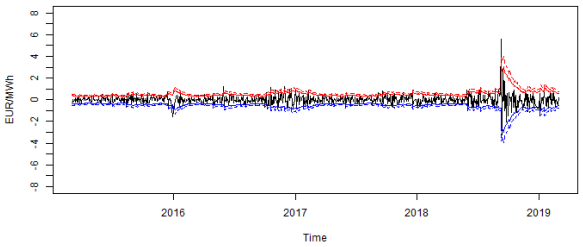
(e) GARCH(1,1)-t



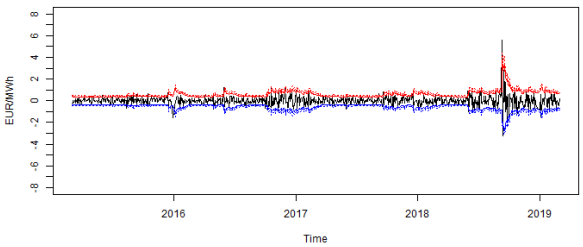
(f) GJR-GARCH(1,1)-t



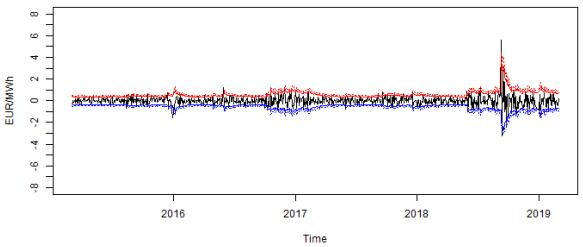
(g) GARCH(1,1)-skew t



(h) GJR-GARCH(1,1)-skew t

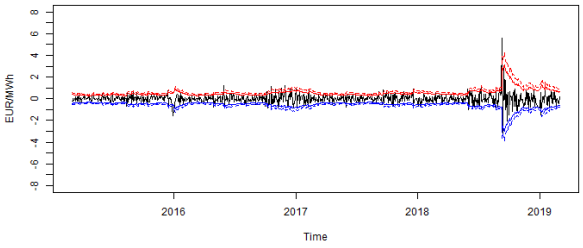


(i) MSGARCH(1,1)-n

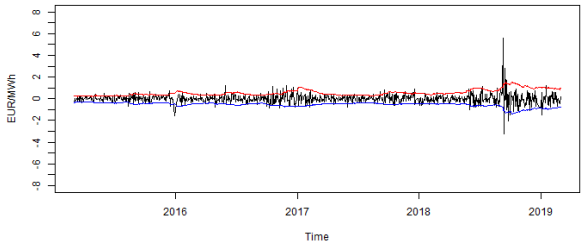


(j) MSGARCH(1,1)-t

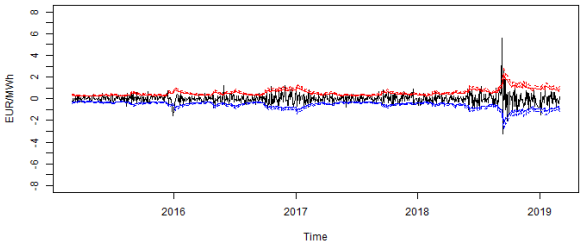
Figure 33: 95% VaR and ES forecasts with reestimation for a long position (solid blue and dotted blue), short position (solid red and dotted red), front-year



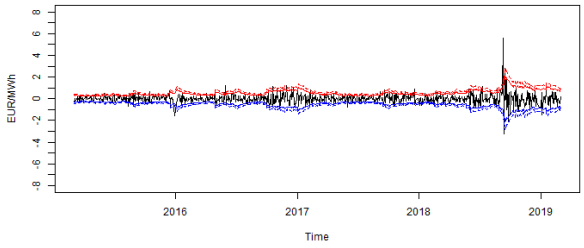
(a) GARCH(1,1)-EVT



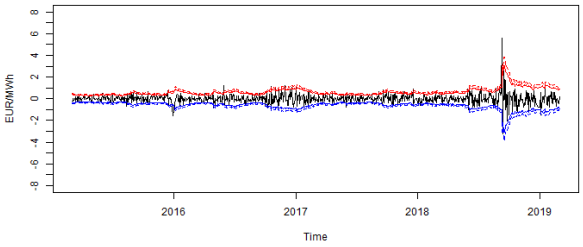
(b) Vol-corr-adj qreg



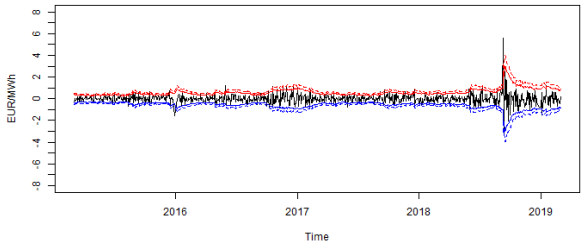
(c) CCC-n



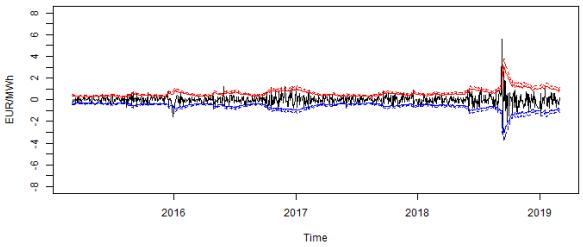
(d) CCC-t



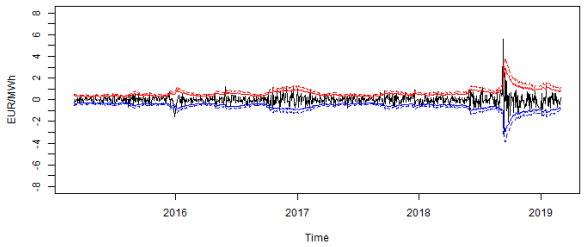
(e) DCC-n



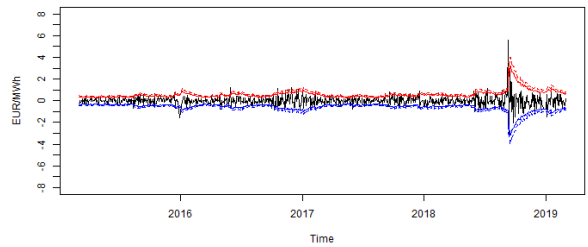
(f) DCC-t



(g) aDCC-n



(h) aDCC-t



(i) DCC-EWMA

Figure 34: 95% VaR and ES forecasts with reestimation for a long position (solid blue and dotted blue), short position (solid red and dotted red), front-year