*Journal of Risk Model Validation* **13**(4), 1–27
DOI: 10.21314/JRMV.2019.213

Research Paper

# Value-at-risk in the European energy market: a comparison of parametric, historical simulation and quantile regression value-at-risk

**Sjur Westgaard,[1] Gisle Hoel Århus,[1] Marina Frydenberg[1] and Stein Frydenberg[2]**

[1]Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology, NTNU, NO-7491 Trondheim, Norway;
emails: sjur.westgaard@ntnu.no, gislehaa@stud.ntnu.no, marinaf@stud.ntnu.no
[2]NTNU Business School, Norwegian University of Science and Technology, NTNU, NO-7491 Trondheim, Norway; email: stein.frydenberg@ntnu.no

(Received January 30, 2018; revised February 1, 2019; accepted February 6, 2019)

## ABSTRACT

This paper examines a set of value-at-risk (VaR) models and their ability to appropriately describe and capture price-change risk in the European energy market. We make in-sample, one-day-ahead VaR forecasts using one simple parametric model, one historical simulation model and one quantile regression (QR) model. We apply our models to nine different energy futures: Brent crude oil, API2 coal, UK natural gas, and three German and Nordic power futures in the period 2007–17. The models are tested at both long and short positions. Our research suggests that the QR model is easy to implement and offers accurate VaR forecasts in the European energy market. Change OK?

**Keywords:** RiskMetrics; historical simulation (HS); quantile regression (QR); value-at-risk (VaR); European energy future markets; risk analysis.

---

Corresponding author: S. Westgaard
Print ISSN 1753-9579 | Online ISSN 1753-9587

## 1 INTRODUCTION

This paper aims to investigate the accuracy of univariate risk models in energy markets, more specifically the European energy futures markets. Companies and other actors that include energy futures in their portfolios or otherwise conduct business in energy markets rely on proper financial risk management to avoid unexpected losses. For example, an oil refinery must manage the risk arising from volatile crude oil prices as a buyer and the risk arising from volatile gasoline prices as a seller. Some energy commodities such as oil may even have an asymmetric affect on economic activity, as described by Sadorsky (1999). Other studies have also documented the impact oil has on different economies, such as China (see Zhen-xin *et al* 2011) and Canada (see Rahman and Serletis 2012). Commodity trading is also affected by global economic events, such as the financial crisis in 2007. European coal and oil prices increased sharply after the turmoil of the crisis, before dropping massively in the last part of 2008. Clearly, investors in energy markets should be wary of how energy commodities behave and should have a plan for how to handle risk.

Since the development of financial futures in the 1970s, traders have increasingly used future contracts to hedge against risk or to speculate on price fluctuations in different financial markets. The correlation between commodity futures and stocks and bonds is typically low, as one can, for example, diversify a portfolio of stocks and bonds by including commodity futures to reduce its risk, as described in Anson (2004). By including a derivative such as an energy future in a portfolio, one must be able to manage the risks that come with its use. In the 1990s, JP Morgan developed RiskMetrics (see JP Morgan 1996), a method for calculating risk. This method introduced value-at-risk (VaR) as a risk measure, which, since then, has been a popular choice for calculating and managing the risk of financial portfolios. VaR is the loss we will not exceed with a certain probability of a financial asset over a given period. Another popular risk methodology is expected shortfall (ES), which also measures the potential average losses, given that a loss occurs. However, the focus of this paper is solely on VaR models.

Energy futures have varying risk characteristics compared with other assets, such as stocks, and traditional commodities, such as gold or copper. Commodities are driven by a specific supply-and-demand relationship; they are usually more volatile than assets and show high positive price extremes. Energy commodities such as coal and gas can be stored but others such as electric power cannot. This leads to a variety of risk characteristics across energy futures. For example, coal futures does not share the same volatility or skewness as power futures. It is also important to note that commodity reserves exist in different areas of the planet and thus face different risks related to location. The differences in risk characteristics can be found in the distribution of price returns as seen in volatility, skewness, excess kurtosis and fat

Words added – OK?

Changes to sentence OK?

tails. The futures used in this study are all from Germany, the United Kingdom and the Nordic countries. This paper therefore offers valuable information for anyone interested in managing risk with regard to futures traded in these countries.

Conducting risk management using VaR models is very popular. In the European energy market, however, the literature is somewhat scarce when it comes to studies on employing VaR as a risk measure. We wish to contribute to this literature by providing a study that compares the accuracy of different types of VaR models used to manage risk in the European energy market. Our analysis consists of comparing a RiskMetrics-inspired parametric model, a historical simulation (HS) model and a quantile regression (QR) model, inspired by Steen *et al* (2015). We have adjusted our HS method using a volatility weighting technique similar to that proposed by Hull and White (1998). Previous studies have shown that electricity markets are volatile (Chan and Gray 2006; Westgaard *et al* 2014). Models that are suitable for other financial markets might not be suitable for power markets. With this in mind, we hope to find a VaR model suitable for investors trading energy futures in the European energy markets.

Changes to this sentence and the next OK?

The rest of this paper is structured as follows. Section 2 is a review of the literature regarding VaR models, and more specifically risk assessment in energy markets. Section 3 describes the VaR models used in our study as well as backtesting methodologies. Section 4 describes and presents the data we have used in our analysis, how it was gathered and any adjustments that have been made to it. Section 5 presents the empirical results found in our analysis while Section 6 summarizes and concludes our findings.

## 2 LITERATURE REVIEW

To the best of our knowledge, there are few empirical studies on modeling VaR in the European energy market. In this section, we will discuss findings that are important and relevant to our study. We have found that the energy commodity most frequently used in studies on VaR is West Texas Intermediate (WTI) crude oil. Giot and Laurent (2003), who compare the efficiency of several VaR models in different commodity markets, include Brent crude oil and WTI crude oil in their study. This finds that, over a five-year out-of-sample testing period, VaR forecasts based on a skewed Student asymmetric power generalized autoregressive conditional heteroscedasticity (APARCH) model perform best. Andriosopoulos and Nomikos (2015) discuss several suitable VaR models appropriate for capturing the dynamics in energy prices traded at the NYMEX and spot energy index, including, among others, WTI crude oil and heating oil. Here, a hybrid Monte Carlo model and a Monte Carlo simulation model are described as the more prevailing models. Hung *et al* (2008) estimate the VaRs for energy commodities, including WTI crude oil, Brent crude

Changes to sentence OK?

oil and gasoline prices. These authors use fat-tailed generalized ARCH (GARCH) models and find that one-day-ahead forecasts generated by the heavy-tailed GARCH model introduced by Politis (2004) give good estimates at both high and low confidence levels. Some studies, such as that of Marimoutou *et al* (2009), also apply certain aspects of extreme value theory (EVT) to manage risk in oil markets, namely Brent crude oil and WTI crude oil. In that paper, the authors find that a conditional generalized Pareto distribution (GPD) VaR model along with a filtered HS model (FHS) offer major improvements to VaR estimation compared with more traditional methods.

Although scarce, there are some studies that address risk management in electricity spot and futures markets. Chan and Gray (2006) examine daily aggregated electricity spot prices from five different power markets, including NordPool electricity spot prices. This study compares its own proposed EVT-based VaR model with more conventional parametric and nonparametric models. It reveals that its proposed EVT-based model performs well in forecasting out-of-sample VaR compared with conventional models. Westgaard *et al* (2014) present a quite extensive piece of research regarding risk characteristics in the European energy markets. There, it is found that risk measured in standard deviation is much higher for Nordic electricity and natural gas markets than it is for more traditional assets such as stocks and bonds. Another result of importance for our own study is that tail behavior varies for different energy futures, which can affect which risk model one should choose for each energy future contract. Byström (2005) propose an EVT approach to investigate the tails of the return distribution of electricity prices traded at the NordPool exchange. Accurate estimates and forecasts of extreme quantiles of the price change distribution are made using a GPD. That study emphasizes the benefits of EVT for risk managers and portfolio managers in electricity markets. Nowotarski and Weron (2018) give a thorough tutorial on probabilistic electricity price forecasting (EPF). The authors present guidelines for the use of methods, measures and tests in probabilistic EPF, but they argue that their study is also general enough for wind and solar power forecasting. Veka *et al* (2012) provide one of very few studies on the correlation between Nordic electricity derivatives and electricity contracts traded at the European Energy Exchange (EEX) and the Intercontinental Exchange (ICE). By using multivariate GARCH models, they find a significant time-varying correlation between the energy commodities from the three different exchanges used in the analysis, except with oil. Therefore, they argue that pricing models based on constant correlation may be misleading. Another important finding in this study for investors in the European energy market is that the strongest correlation to Nordic electricity futures was with the energy futures traded at the EEX, and this relationship increased with longer maturity contracts in all the markets.

Changes to sentence OK?

Previous studies on using QR in various financial markets include, among others, those of Steen *et al* (2015) and Haugom *et al* (2016). Steen *et al* (2015) compare the efficiency of VaR forecasts using twenty different commodities and three different VaR models, among them a QR approach. They find that the QR approach outperforms the HS method as well as the standard RiskMetrics approach. Haugom *et al* (2016) use a similar QR approach to that of Steen *et al* (2015). In their paper, the authors employ a parsimonious QR model to forecast one-day-ahead VaR in both commodity markets and more traditional financial assets. The novel result here is that their QR approach gives a similar performance to more complicated VaR models, such as the skewed $t$ APARCH model and the conditional autoregressive VaR (CAViaR) model subjected to coverage tests for out-of-sample VaR predictions. Another paper that investigates the usage of QR in commodity markets is by Kuralbayeva and Malone (2012). Here, the authors use a QR model to explain extreme movements in commodity prices. They investigate how various global and commodity-specific determinants explain these price movements and find that, especially since the turn of the millennium, global demand factors play a major role in explaining commodity price changes compared with commodity specific factors such as open interest.

Changes to sentence OK?

There are several studies that compare the efficiency of VaR models based on an HS numerical approach with VaR models based on a parametric approach. Füss *et al* (2010) compare the in- and out-of-sample performance of conventional VaR, Cornish–Fisher VaR, GARCH-type VaR and semi-parametric CAViaR models for commodity futures investments. They find that the choice of VaR model depends strongly on the underlying return series, but in general they find that the CAViaR and GARCH models outperform the others because of their ability to capture volatility clustering. Cabedo and Moya (2003) did a study on estimating oil price VaR in 2003. They analyzed three different VaR approaches, namely a standard HS approach, an HS approach with autoregressive moving average (HSAF) forecasts and a variance–covariance method based on ARCH model volatility forecasts. They found that the HSAF methodology provided flexible VaR estimation, fitting to the continuous oil price movements and offering efficient risk quantification. Further, Costello *et al* (2008) compare the ARMA HS VaR used by Cabedo and Moya (2003) with the semi-parametric GARCH model proposed by (Barone-Adesi *et al* 1999). Costello *et al* (2008) find that VaR forecasts based on semi-parametric GARCH models exceed the quality of those from ARMA forecasts, indicating the former is more effective as an HS technique. They argue that Cabedo and Moya's conclusion is driven by their normal distributional assumption being imposed on the future risk structure in the GARCH model. Lux *et al* (2016) study crude oil price volatility and VaR. They compare the forecasting performances of several GARCH-type models with a Markov-switching multifractal (MSM) model. Their study proves the superiority of

Change OK?

Changes to sentence OK?

Word added – OK?

the new MSM model regarding both volatility forecasting and VaR. In addition to these articles and studies, there are several useful textbooks that cover topics such as VaR and other relevant risk metrics: Alexander (2008), Jorion (2000) and Hull (2012).

## 3 THEORY AND METHODOLOGY

This section outlines the relevant theory and methods used in our analysis. We outline the techniques used to compute the VaR estimates before discussing the validation methods, eg, backtesting methods. Changes to this sentence and the next OK?

### 3.1 VaR estimation

VaR is defined as the maximum loss that will be incurred on a portfolio with a given level of confidence for a specified period. The literature mainly describes three ways of estimating VaR (see Alexander 2008; Hull 2012). These are parametric VaR, historical method and Monte Carlo simulation. Let $X_t^i$ denote the return of commodity $i$ at time $t$, defined by $X_t^i = \ln(P_t^i / P_{t-1}^i)$, where $P_t^i$ denotes the price of the commodity at time $t$, and $P_{t-1}^i$ denotes the price the previous day. Then, VaR is given by

$$\mathrm{VaR}_{h,\alpha} = \Phi^{-1}(1-\alpha)\sigma_h - \mu_h. \tag{3.1}$$

If zero mean is assumed:

$$\mathrm{VaR}_{h,\alpha} = \Phi^{-1}(1-\alpha)\sigma_h, \tag{3.2}$$

where $\Phi^{-1}(1-\alpha)$ is the inverse of the standard normal distribution, $\alpha$ is a quantile value describing the prescribed significance level, and $\mu_h$ and $\sigma_h$ are the estimated mean and the volatility over the time horizon $h$. The parametric method of calculating VaR is a simple and therefore popular approach to use as a benchmark estimate. Even so, it has several drawbacks due to its simplicity. The approach assumes that the returns are normal in addition to being identically and independently distributed (iid). This is often not the case. The distributions of returns of commodity futures often have leptokurtic and skewed shapes as well as fat tails. Not accounting for this will lead to an underestimation of risk. Moreover, using a constant volatility estimate in (3.1) does not factor in the effect of volatility clustering. One can adjust for this by using the exponentially weighted moving average (EWMA) or the ARCH proposed by Engle (1982), which was further developed into the GARCH by Bollerslev (1986). The ARCH model is a popular choice for modeling volatility because it is able to capture the clustering effect often seen in financial markets. Change OK?

## 3.2 Parametric approach

In our parametric model, we use the EWMA to model volatility, with the one-day-ahead forecast given by

$$\mathrm{VaR}_{t+1} = Z^{-1}(0, \sigma_t, 1-\alpha), \tag{3.3}$$

where $Z^{-1}(0, \sigma_t, 1-\alpha)$ is the inverse of a normal distribution at $\alpha$ quantile, with zero mean and standard deviation described by the EWMA:

$$\sigma_t = \sqrt{(1-\lambda)r_{t-1}^2 + \lambda\sigma_{t-1}^2}, \tag{3.4}$$

where $r_{t-1}^2$ is the square of the previous day's log return and $\lambda$ is the smoothing parameter that describes how the previous observations will affect the forecast of the next day's volatility. In this model, returns are assumed to be distributed normally and the smoothing parameter is set at 0.94, which is a typical assigned value.

Change OK?

## 3.3 FHS

HS is a nonparametric approach for calculating VaR and does not assume an analytical distribution of the returns. Instead, an empirical distribution of returns is used to calculate VaR. New returns are simulated by drawing randomly from past observations in order to predict tomorrow's return. However, deciding on how many returns to select from past returns to predict tomorrow's returns is problematic. One must select enough to make sure the empirical distribution is correct.

Changes to this sentence and the next OK?

The reason for this method's popularity is that it automatically captures the fat-tailed, skewness and leptokurtic effects often seen in commodity returns. It is relatively easy to implement but can be computationally challenging compared with parametric methods because it requires many simulations in order to generate accurate results. In order for this method to be suitable, an adequate number of observations is needed. Alexander (2008) advises acquiring at least 2000 observations. The number of simulations is also important: we have used 10 000 simulations to make sure the model is accurate.

When using HS, one can use volatility adjustments to improve the VaR forecast by capturing time-varying volatility movements. Hull and White (1998) suggest a volatility weighing approach. We have used a bootstrap method, in which we adjust the returns in the empirical distribution using a GARCH(1,1) volatility model. This enables us to capture the volatility clustering effect. Our adjusted HS one-day-ahead forecast is described by

Does the asterisk denote simple multiplication in this display? If so, we will delete it according to journal style, but if it denotes some other operation, we will of course retain it.

$$\mathrm{VaR}_{t+1} = \mathrm{Percentile}\left(S(r_\Omega) * \frac{\sigma_t}{(1/N_\Omega)\sum_{t=t_1}^{t_2}\sigma_t}, \alpha, N\right), \tag{3.5}$$

where $\mathrm{VaR}_{t+1}$ is the one-day-ahead forecast using $N$ simulations. $S(r_\Omega)$ is a sample function that draws randomly from the empirical data set $\Omega$ of the observed returns from start date $t_1$ until current date $t_2$ (today). $\alpha$ is the desired quantile or VaR level (for example, 5%) and $\sigma_t$ is the volatility at time $t$, calculated using a vanilla GARCH model. GARCH fits the sample returns as follows:

$$\sigma_t = \sqrt{\omega_{\text{garch}} + (\alpha_{\text{garch}})\varepsilon_{t-1}^2 + \beta_{\text{garch}}\sigma_{t-1}^2}. \tag{3.6}$$

Here, $\omega_{\text{garch}}$, $\alpha_{\text{garch}}$ and $\beta_{\text{garch}}$ are the positive GARCH parameters, $\alpha_{\text{garch}} + \beta_{\text{garch}} < 1$ ensures the unconditional variance is finite and positive, $\varepsilon_{t-1}^2$ is the previous period's squared return and $\sigma_{t-1}^2$ is the previous time step's volatility. Change OK?

## 3.4 QR

QR, introduced by Koenker and Bassett (1978), has proven useful in forecasting electricity prices (Bunn *et al* 2016) and modeling risk in commodity markets (Steen *et al* 2015); see also Alexander (2008) on how to apply QR for financial risk management and Koenker (2005) for a full introduction to QR. The following is a simple linear QR model:

$$r_t = \alpha^q + \beta^q \sigma_t + \varepsilon_t^q, \tag{3.7}$$

where $r_t$ is the dependent variable, $\sigma_t$ is the independent variable, $\varepsilon_t^q$ is the error assumed to be iid and with a given distribution, and $q \in (0, 1)$ is the quantile of the error $\varepsilon_t^q$. The QR methodology we have used to forecast a one-day-ahead VaR is inspired by that introduced by Steen *et al* (2015):

$$\mathrm{VaR}_{t+1} = \alpha^q + \beta^q \sigma_t + \varepsilon_t^q. \tag{3.8}$$

The parameters $\alpha^q$ and $\beta^q$ need to be estimated by the optimization problem

$$\min_{\alpha,\beta} \sum_{t=1}^{T} (q - \mathbf{1}_{r \leqslant \alpha + \beta\sigma_t})(r_t - (\alpha + \beta\sigma_t)), \tag{3.9}$$

where

$$\mathbf{1}_{r \leqslant \alpha + \beta\sigma_t} = \begin{cases} 1 & \text{if } r_t \leqslant \alpha + \beta\sigma_t, \\ 0 & \text{otherwise.} \end{cases} \tag{3.10}$$

In the optimization problem, the $X_t$ is replaced by $\sigma$, which is calculated using the EWMA, as in our parametric approach (see (3.4)).

## 3.5 Model validation

A backtest looks at how the actual VaR level observed compares with the VaR predicted in the forecasted model. When comparing predicted VaR with actual VaR, the first indicator of whether your model is correct is the frequency of violations. By frequency of violations, we mean how often the observed VaR exceeds the VaR forecasts made by the model over a time horizon. The number of violations over a time horizon can be summed up or counted, and the frequency of occurrence, then, is the number of observed violations divided by how many days there are in the sample period. For example, if we look at a time horizon of 2000 days and our specified VaR level is 5%, then the number of observed violations should be approximately 100. Mathematically, one can describe this accuracy with a hit function:

Changes to sentence OK?

'$\boldsymbol{F}$' and '$\boldsymbol{t}$' are bold here, as in your submission, and '$\alpha$' is not – please confirm that's what you wanted.

$$\boldsymbol{F}(\boldsymbol{t}, \alpha) = \begin{cases} 1 & \text{if } r_t \leqslant \text{VaR}(\alpha), \\ 0 & \text{if } r_t > \text{VaR}(\alpha), \end{cases} \tag{3.11}$$

where $r_t$ is the observed log return and $\text{VaR}(\alpha)$ is the forecasted VaR for the $\alpha$ quantile, with the accuracy of the model now described by the sum of violations divided by the number of days in the sample. Kupiec (1995) introduced an unconditional coverage test to check whether the probability of observing a hit is consistent with the probability imposed by the VaR quantile. This test uses the likelihood ratio statistic

Change OK?

$$\text{LR}_{\text{UC}} = -2\log\left(\frac{p^f(1-p)^{N-f}}{(f/N)^f(1-(f/N))^{N-f}}\right), \tag{3.12}$$

where $\text{LR}_{\text{UC}}$ is the unconditional likelihood ratio, $f$ is the number of hits over the time horizon, $N$ is the total number of observations in the time horizon and $p$ is the failure rate given by the VaR level (eg, 5%). The test statistic is asymptotically distributed as a chi-square variable with one degree of freedom; see Tables 3–5 for critical values. The null hypothesis of an accurate model is rejected if the $\text{LR}_{\text{UC}}$ statistic exceeds the chi-square value at a given confidence level.

Christoffersen (1998) introduced a conditional coverage test to see if the occurrence of a hit was conditional on the previous day having a hit or not. In the null hypothesis of the Christoffersen test, it is assumed that the occurrences are independent, meaning that the probability of a hit today is independent of whether yesterday had a hit or not. The likelihood ratio is given by

$$\text{LR}_{\text{CC}} = -2\log\left(\frac{(1-p)^{f_{00}+f_{10}}(p)^{f_{01}+f_{11}}}{(1-p_0)^{f_{00}}p_0^{f_{01}}(1-p_1)^{f_{10}}p_1^{f_{11}}}\right), \tag{3.13}$$

where $\text{LR}_{\text{CC}}$ is the conditional coverage likelihood ratio, $p$ is the expected number of hits during a time horizon and $f_{ij} \in [0, 1]$ is the number of observations where an event $i$ at time $t - 1$ is followed by an event $j$ at time $t$. $p_0$ is described by $f_{01}/(f_{01} + f_{00})$ and $p_1$ is described by $f_{11}/(f_{11} + f_{10})$. These ratios show how often a nonfollowing hit occurs and how often a following hit occurs. The conditional coverage test statistic is asymptotically distributed as a chi-square variable with two degrees of freedom; see Tables 3–5 for critical values. As for the Kupiec test, the null hypothesis of the Christoffersen test (that the model is accurate) is rejected if the $\mathrm{LR}_{\mathrm{CC}}$ statistic exceeds the chi-square value at a given level of confidence.

## 4 DATA DESCRIPTION AND PRELIMINARY ANALYSIS

The data set used in our analysis consists of nine time series of daily prices for these energy futures: Brent crude oil, API2 coal, UK natural gas, German power futures and Nordic power futures. We have gathered first-month position, second-month position and third-month position contracts for both the German and Nordic power contracts as well as first-month position contracts for Brent crude oil, API2 coal and UK natural gas. The German and Nordic power futures are traded at the EEX and Nasdaq NordPool, respectively, while the other three are traded at the ICE. The series are downloaded using Montel's databases (Montel.no). The data is gathered from the period September 2007–September 2017, spanning approximately ten years and a total of 2600 trading days. Figure 1 displays the development of the prices of the futures.

We've attempted to standardize your use of front/first/1/M1, etc, to describe these position contracts. Please check this has been done correctly and mark the PDF for amendment if not.

### 4.1 Data cleansing

Roll-over returns are those that arise from holding a futures contract over the closing date. When a futures contract expires, a new one is created that replaces the old one in the set of available futures contracts. Sometimes, one can experience a price shock when a new futures position replaces the expired one. We have to shift from an expired first-month contract to a new first-month contract. This can lead to jumps in the returns series, and these jumps will induce a false volatility. This shift from a one-month position to a two-month position contract is not matched by cash changes in the margin account by the future brokerage. The holder of this futures contract does not receive the return between expired contracts and new ones; therefore, the roll-over returns must be deleted from the series. In addition, observations that are not shared are deleted. On December 24, for example, some futures – such as Brent crude oil – do trade, whereas German power futures do not. Thus, our data set is reduced from 2600 observations to 2399 observations. The roll-over returns can be found on the homepages of the exchanges where these futures are traded (www.theice.org, www.EEX.com and www.nasdaqomx.com). The roll-over returns for the futures in our data set are between the last and first day of the month. The cleansing of the data

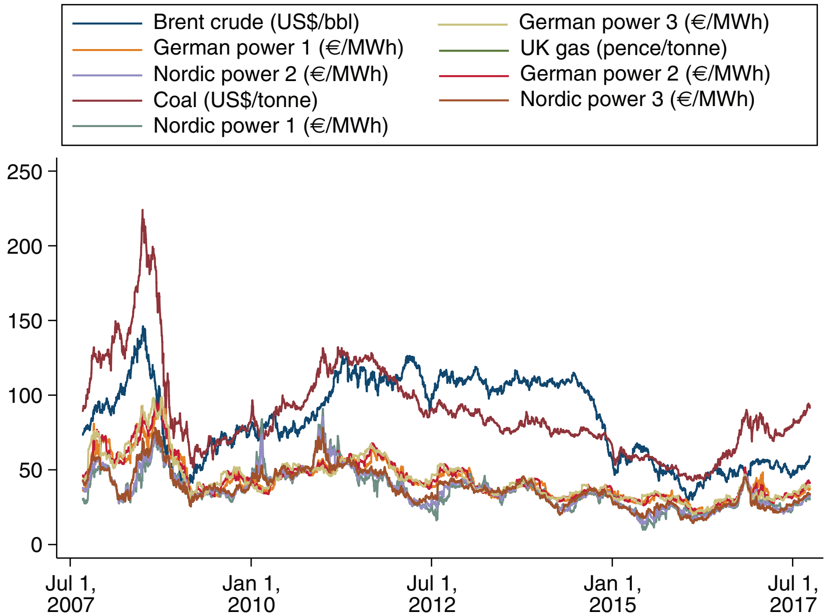
Changes to sentence OK?

Change OK?

Changes to sentence OK?

Words added – OK?

**FIGURE 1** Price of energy futures (price unit in key) for Brent crude oil; API2 coal; UK natural gas; German power positions 1, 2 and 3; and Nordic power positions 1, 2 and 3.




Data collected by Montel.no through the ICE, EEX and Nasdaq OMX, following the period September 4, 2007–September 19, 2017, for a total of 2399 observations.

may change the mean of the returns, but it also reduces the standard deviation and the tail distribution or fat-tailed effect. Figure 2 displays the log returns for all nine contracts.

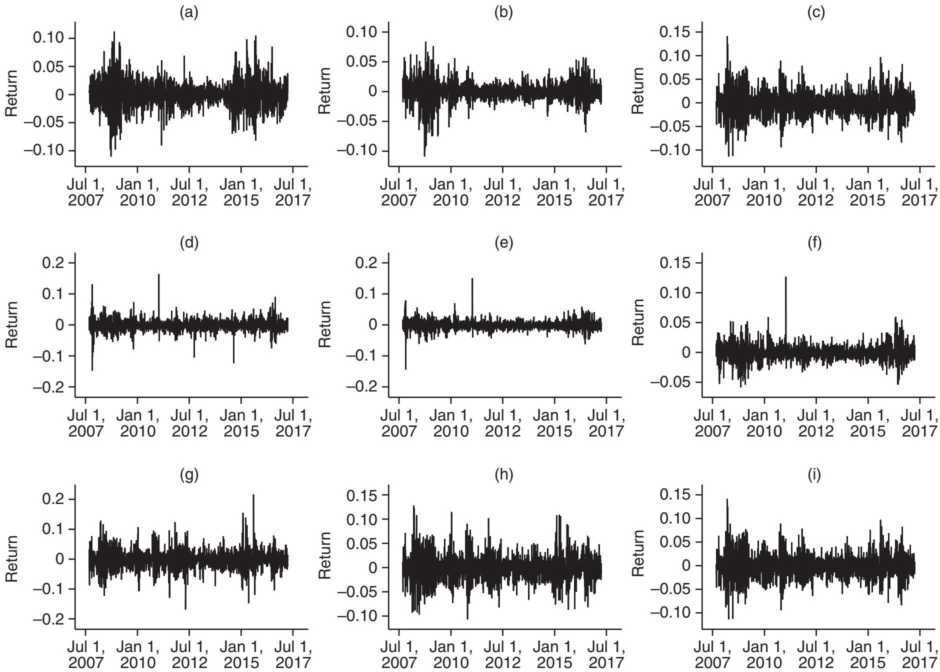
Please check changed wording of (a)–(i) part labels – all OK? I tried to ensure consistency with later use in tables.

## 4.2 Statistics

Changes to sentence OK?

Figure 1 shows that oil and coal prices rose for a time following the financial crisis of 2007–8 before experiencing a large drop and a subsequent recovery until roughly 2012, before again decreasing in 2015. Whereas API2 coal and Brent crude oil depend on global economy conditions, power futures and natural gas tend to be more spiky in their behavior and depend more on local supply-and-demand conditions. They also exhibited high volatility after the financial crisis of 2007–8. Table 1 displays the general statistics of the log returns of the data set used in this paper. Figures 2 and 3 describe the log returns over the time horizon and the distribution of the log returns, respectively. As can be seen in Figure 3, all of the energy commodities exhibit the typical leptokurtic shape of the distribution of returns as well as fat

**FIGURE 2** Returns (log of price changes) of energy futures.




(a) Brent crude first-month contract. (b) API2 coal first-month contract. (c) UK natural gas first-month contract. (d) German power first-month contract. (e) German power second-month contract. (f) German power third-month contract. (g) Nordic power first-month contract. (h) Nordic power second-month contract. (i) Nordic power third-month contract. Data collected by Montel.no through the ICE, EEX and Nasdaq OMX, following the period September 5, 2007–September 19, 2017, for a total of 2398 observations.

tails. All of the commodities reject the null hypothesis of the Jarque–Bera normality test. The mean of the distributions of all futures is centered around zero, which is expected, while the daily volatility varies across the futures. It is evident that oil and coal have a volatility clustering around 2009, when the turmoil of the financial crisis hit Europe the hardest, as well as around 2016, when the dollar rose and the Organization of the Petroleum Exporting Countries (OPEC) responded with an oversupply. The gas and power futures show further spiky trends, with some very extreme returns, eg, when the German power futures spiked on March 15, 2011. This might be because it was announced on this date that Angela Merkel would be shutting down seven of Germany's power plants (see Harding 2011). Gas and power prices tend to exhibit clustering around supply-and-demand shocks; Westgaard *et al* (2014) show similar findings.

Changes to this sentence and the next OK?

Change OK?

All of the series are more or less symmetrical when it comes to maximum/

**TABLE 1** Descriptive statistics for Brent crude oil; API2 coal; UK natural gas; German power positions 1, 2 and 3; and Nordic power positions 1, 2 and 3.

| Futures | Mean (%) | SD (%) | Median (%) | Min. (%) | Max. (%) | Skew | Kurtosis | JB test |
|---|---|---|---|---|---|---|---|---|
| Brent crude oil (ICE) | –0.01 | 2.19 | –0.01 | –10.95 | 11.13 | 0.09 | 2.96 | 0.00 |
| API2 coal (ICE) | 0.01 | 1.57 | 0.00 | –10.82 | 8.32 | –0.57 | 6.47 | 0.00 |
| UK natural gas (ICE) | –0.04 | 2.32 | 0.00 | –11.28 | 14.04 | 0.10 | 2.43 | 0.00 |
| German power 1 (EEX) | –0.07 | 1.79 | –0.09 | –14.61 | 16.27 | 0.10 | 8.86 | 0.00 |
| Nordic power 1 (Nasdaq OMX) | –0.08 | 2.96 | 0.00 | –16.71 | 21.51 | 0.19 | 3.94 | 0.00 |
| German power 2 (EEX) | –0.05 | 1.43 | –0.06 | –14.23 | 14.89 | 0.39 | 11.38 | 0.00 |
| Nordic power 2 (Nasdaq OMX) | –0.05 | 2.48 | 0.00 | –10.59 | 12.71 | 0.18 | 2.26 | 0.00 |
| German power 3 (EEX) | –0.04 | 1.26 | –0.06 | –5.78 | 12.61 | 0.50 | 6.73 | 0.00 |
| Nordic power 3 (Nasdaq OMX) | –0.04 | 2.32 | 0.00 | –11.28 | 14.04 | 0.10 | 2.43 | 0.00 |

Data collected by Montel.no through the ICE, EEX and Nasdaq OMX, following the period September 4, 2007–September 19, 2017, for a total of 2398 observations. These are adjusted for roll-over returns and nonsharing dates, such as December 24. “JB test” refers to the Jarque–Bera normality test.

minimum returns, with the exception of the Nordic power first-month position and the German power third-month position. These have absolute positive returns that are larger than their negative returns. The only future that has a higher negative tail risk is coal. All of the futures are positively skewed except for the coal futures, which have the largest skewness in absolute terms. In addition, all of the futures show leptokurtic shapes but their magnitude varies, with the German power second-month position having the largest. Table 1 shows that the daily volatility of the energy futures ranges from 1.26% to 2.96%, or approximately 19–47% on an annual basis. Natural gas and Nordic power futures are the most volatile, so care should be taken when handling these futures; adequate risk management is very important here.

Changes to sentence OK?

It is also important to take time-varying risk characteristics into account when one develops risk models. Table 2 lets us investigate how the distributions of various energy futures change over time. We have split our data set into four equally sized sample periods: September 5, 2007– March 18, 2010; March 19, 2010–September 17, 2012; September 17, 2012– March 26, 2015; and March 27, 2015– September 26, 2017. One can see from Table 2 that all of the futures have their highest standard deviations during the period including the financial crisis (2007–8). The Nordic power first-month position also has a very high daily standard deviation over all the periods studied. It is interesting to see that the kurtosis for the German power futures changes drastically in period 2. This is due to the mentioned spike in prices, which took place on March 15, 2015. This also skews the distribution positively in this period. Regarding volatility, one can see that Brent crude oil is more affected by global activities (the financial crisis in period 1 and the OPEC oil oversupply in period 4) than the German power futures, which have a more even distribution across the four periods. It is also worth mentioning that the tail behavior also changes over time. One can see that the difference between the maximum and minimum is highest in period 1 for all of the energy futures except for the Nordic power first-month future, which has the highest gap in period 4, and the German power third-month future, which has the highest gap in period 2.

Our examination of the empirical risk characteristics of European energy markets from the ICE, EEX and Nasdaq OMX can be summarized as follows. Risk measured in volatility is highest for the Nordic power market and the natural gas market; thus, our findings are similar to those of Westgaard *et al* (2014). The skewness and kurtosis vary over time, and the German power market is prone to national political decisions. The tail risk is positive for all futures with the exception of coal, which has a higher negative tail risk. All futures are more or less symmetrical, but the German and Nordic power markets tend to have greater positive tail risk. Volatility clustering is evident for all energy futures, but particularly in oil and coal, while the power and gas markets tend to be more spiky in terms of their behavior. The distributions also vary a lot over time for all of the futures.

Changes to sentence OK?

**TABLE 2** Summary statistics of energy futures over time. [Table continues on next page.]

(a) Period 1

| | SD | Min. | Max. | Skewness | Kurtosis |
|---|---|---|---|---|---|
| Brent crude oil | 2.84 | −10.95 | 11.13 | −0.02 | 1.63 |
| API2 coal | 2.39 | −10.82 | 8.32 | −0.71 | 2.94 |
| UK natural gas | 2.99 | −11.28 | 14.04 | 0.12 | 1.68 |
| German power 1 | 2.26 | −14.61 | 12.98 | −0.24 | 5.51 |
| Nordic power 1 | 3.27 | −12.10 | 12.74 | 0.10 | 1.32 |
| German power 2 | 1.86 | −14.23 | 7.77 | −0.46 | 6.95 |
| Nordic power 2 | 3.07 | −9.58 | 12.71 | 0.17 | 1.24 |
| German power 3 | 1.57 | −5.78 | 5.17 | −0.21 | 1.26 |
| Nordic power 3 | 2.99 | −11.28 | 14.04 | 0.12 | 1.68 |

(b) Period 2

| | SD | Min. | Max. | Skewness | Kurtosis |
|---|---|---|---|---|---|
| Brent crude oil | 1.68 | −8.96 | 6.81 | −0.42 | 2.28 |
| API2 coal | 1.07 | −4.50 | 5.55 | 0.47 | 2.36 |
| UK natural gas | 2.20 | −9.32 | 8.83 | 0.02 | 1.38 |
| German power 1 | 1.51 | −5.07 | 16.27 | 2.31 | 23.66 |
| Nordic power 1 | 3.17 | −16.71 | 12.22 | −0.16 | 2.55 |
| German power 2 | 1.32 | −4.21 | 14.89 | 2.83 | 28.19 |
| Nordic power 2 | 2.41 | −10.59 | 10.08 | 0.09 | 1.62 |
| German power 3 | 1.16 | −3.24 | 12.61 | 2.61 | 23.95 |
| Nordic power 3 | 2.20 | −9.32 | 8.83 | 0.02 | 1.38 |

(c) Period 3

| | SD | Min. | Max. | Skewness | Kurtosis |
|---|---|---|---|---|---|
| Brent crude oil | 1.54 | −6.88 | 7.56 | 0.08 | 4.42 |
| API2 coal | 0.96 | −4.10 | 3.81 | 0.06 | 1.50 |
| UK natural gas | 1.52 | −5.87 | 6.16 | −0.07 | 0.84 |
| German power 1 | 1.48 | −12.20 | 4.83 | −1.41 | 10.69 |
| Nordic power 1 | 2.03 | −8.85 | 7.51 | −0.17 | 1.27 |
| German power 2 | 0.87 | −3.83 | 2.78 | −0.04 | 0.77 |
| Nordic power 2 | 1.63 | −6.57 | 4.94 | −0.08 | 0.67 |
| German power 3 | 0.76 | −2.62 | 3.19 | 0.16 | 0.78 |
| Nordic power 3 | 1.52 | −5.87 | 6.16 | −0.07 | 0.84 |

**TABLE 2** Continued.

(d) Period 4

| | SD | Min. | Max. | Skewness | Kurtosis |
|---|---|---|---|---|---|
| Brent crude oil | 2.42 | –8.11 | 10.42 | 0.35 | 1.59 |
| API2 coal | 1.44 | –6.74 | 5.68 | 0.21 | 2.31 |
| UK natural gas | 2.34 | –8.31 | 9.61 | 0.08 | 1.41 |
| German power 1 | 1.80 | –7.41 | 8.90 | 0.27 | 2.83 |
| Nordic power 1 | 3.18 | –14.59 | 21.51 | 0.67 | 6.59 |
| German power 2 | 1.49 | –6.07 | 5.74 | 0.28 | 1.80 |
| Nordic power 2 | 2.60 | –8.59 | 10.80 | 0.23 | 2.21 |
| German power 3 | 1.40 | –5.30 | 5.88 | 0.25 | 2.55 |
| Nordic power 3 | 2.34 | –8.31 | 9.61 | 0.08 | 1.41 |

This table lists descriptive statistics for Brent crude oil; API2 coal; UK natural gas; German power positions 1, 2 and 3; and Nordic power positions 1, 2 and 3. The table displays time-changing statistics for four periods: September 5, 2007–March 18, 2010; March 19, 2010–September 17, 2012; September 17, 2012–March 26, 2015; and March 27, 2015–September 26, 2017. The data is adjusted for roll-over returns and nonsharing dates, such as December 24.

## 5 EMPIRICAL RESULTS

Figure 1 displays the price development of the futures prices for all of the energy commodities. The figure displays the large price fluctuations of 2007–8 as well as price development until September 2017. We calculate the log returns of the prices of the energy commodities and plot them in Figure 2. The price fluctuation can also be seen in these plots.

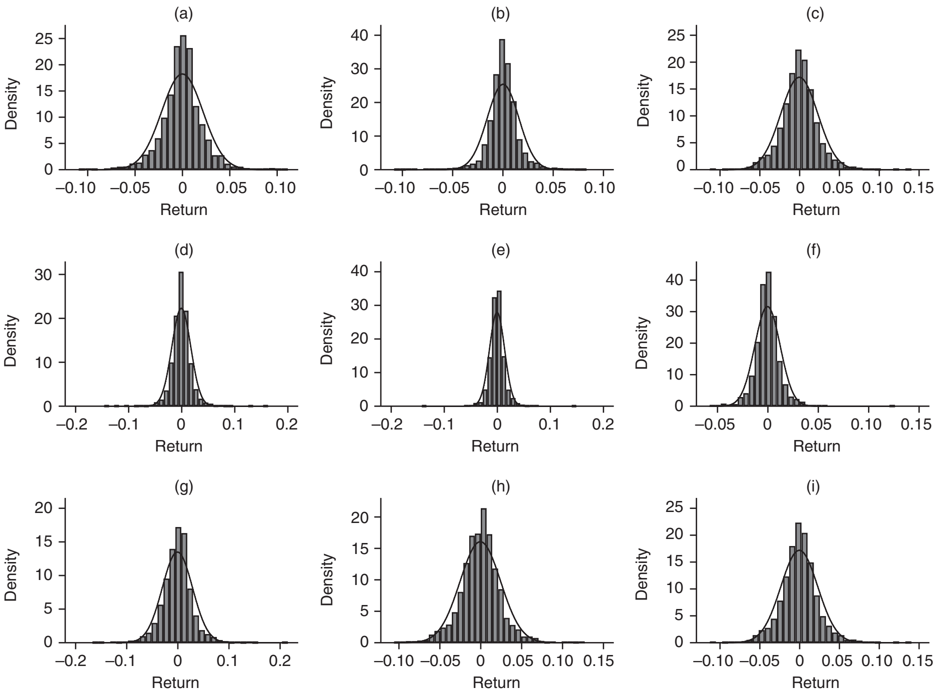
To evaluate the accuracy of our VaR models, we use the methodology described in Section 3 to make in-sample one-day-ahead VaR forecasts over the entire sample period. We evaluate our models on both long and short positions, with three different VaR levels on each side, and with a total of six confidence levels at 99%, 95%, 90%, 10%, 5% and 1%. The log returns are then compared with the VaR forecasts, and the Kupiec and Christoffersen test statistics are calculated. The null hypothesis is rejected if the test statistic is higher than a given value; this implies a wrongly specified model. The critical values are gathered in the final row of Tables 3–5. The critical values for the Kupiec and Christoffersen tests are, respectively, 6.63 and 9.21 for 99% VaR, 3.84 and 5.99 for 95% VaR, and 2.71 and 4.61 for 90% VaR. This is also the case for short positions.

Changes to sentence OK?

Table 3 displays the test statistics for the Kupiec and Christoffersen tests on the VaR forecast with the parametric model, while Table 4 and Table 5 offer those for the FHS and QR forecasts. The models that fail the tests at the given confidence level are marked with bold text, whereas those that pass are left unmarked (normal text). From

Change OK?

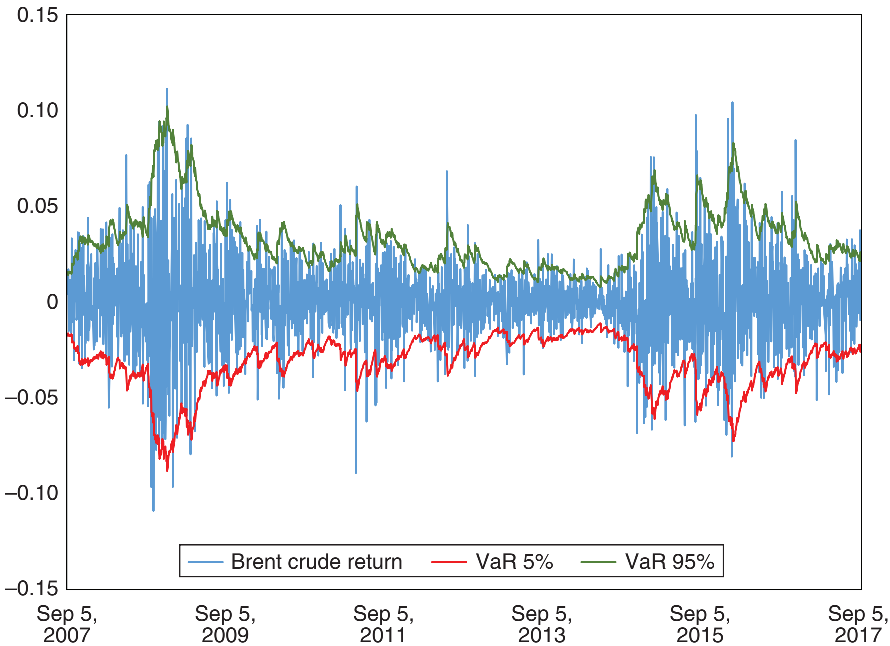
**FIGURE 3** Histograms of distribution of log returns of (a) Brent crude, (b) API2 coal, (c) UK natural gas, (d) German power 1, (e) German power 2, (f) German power 3, (g) Nordic power 1, (h) Nordic power 2 and (i) Nordic power 3.




Tables 3–5, it can be seen that at the 1% and 99% significance levels there is at least one model that is correctly specified for all energy futures concerning unconditional and conditional coverage. We can also see that the most difficult future to assign a VaR model to is the coal future; here, the QR model prevails as the most accurate in terms of both conditional and unconditional coverage. Changes to sentence OK?

It can be seen from Table 7 that the parametric VaR model fails the null hypothesis of Kupiec's test at the 1% VaR level for both long and short positions (eg, VaR 99%) quite often. This comes as no surprise due to its assumption of a normal distribution of returns. This means that the expected number of violations is not close to the actual number of violations, eg, the number of times the actual returns exceed the VaR forecast is not reflective of how many times the returns are expected to exceed the forecast. This leads to a belief that the parametric model is not very suitable to accurately describe risk at the ends of the tails of the return distribution. However, it is quite accurate at the 10% VaR level for long positions as well as at the 5% Is this correct? Table 7 is mentioned before Table 6?

**FIGURE 4** ICE Brent crude oil with 2398 observations of returns (blue line), including the financial crisis of 2007–8.




This figure also includes the 5% left-tail (red line) and 5% right-tail (green line) estimations using the QR VaR model. Violations occur when the blue observations cross the red line. The model responds well to the volatile period of the financial crisis. In this figure, the model cannot be rejected using the Kupiec test at either tail for either confidence level.

VaR level for short positions. One should be careful when using the RiskMetrics approach, because – even though it adjusts for volatility clustering – the assumption of normally distributed returns is incorrect and will lead to wrong VaR forecasts. It was shown in Table 2 and Table 1 that the returns are skewed and have excess kurtosis, and thus the assumption of a normal distribution is a simplification. The parametric VaR model can be improved if one finds a more suitable analytical distribution that better fits the actual returns. One could try a Student $t$ distribution, as seen, for example, in the work of Marimoutou *et al* (2009). One can see in Table 6 that the conditional coverage of the parametric VaR model is by no means obtained. Looking at Table 3, the model fails most often at the leftmost and rightmost VaR levels; however, a large portion of failures also occur within the first- and second-month positions of the German and Nordic power futures at the 10% and 5% VaR levels for short positions (the right-hand tail). One can see in Table 5 that the total violations of conditional coverage for the parametric VaR model are twenty-four out of fifty-four. This means that in twenty-four test statistics the observed violations tend not to be independent. The parametric VaR model is not able to obtain adequate conditional coverage and unconditional coverage. The results regarding the parametric model are consistent with the findings of Steen *et al* (2015). The unconditional coverage failures occur a bit more on the left-hand tail than on the right-hand tail, and the conditional coverage failures occur more on the right-hand tail than on the left-hand tail.

Overall, the FHS VaR model performs better than the parametric VaR model. It obtains better conditional and unconditional coverage, as can be seen in Tables 6 and 7. It can also be seen that while the parametric model performs poorly near the tails, the FHS model performs relatively well. This is because the distribution is not assumed to be normal but is drawn empirically instead, as described in Section 3. Still, the FHS model also has higher failure rates at extreme tail VaR levels (1% and 99%). It can also be seen that the futures that have these failures are coal and the German power futures. In Section 4, we described the data and noted that the spike found in the German power series drives up the kurtosis and positive skewness of the German power futures. It is possible that somehow this spike renders the forecasts made from the empirical distribution inaccurate. Another reason why the FHS VaR model is not more accurate at the tails is because we have only 2398 observations (approximately). This means we have only about twenty-four observations at the 1% extreme. Table 4 shows that the FHS model exhibits Kupiec test violations in seven out of eighteen cases at the 1% VaR level, and in two out of eighteen cases at the 5% and 10% VaR levels. Meanwhile, the Christoffersen test is violated in eight out of eighteen cases at the 1% VaR level, five out of eighteen cases at the 5% VaR level and six out of eighteen cases at the 10% VaR level. A total failure of conditional coverage Changes to this sentence and the next OK? for the FHS VaR model is seen in nineteen out of fifty-four cases, which is only five better than the parametric VaR model. The unconditional coverage measured by Kupiec's test is better, as it only fails in twelve cases. As was seen for the parametric method, the conditional coverage failures occur more on the right-hand tail than on the left-hand tail, while the unconditional coverage failures occur symmetrically on the left- and right-hand tails.

VaR forecasts made by the QR model provide better conditional and unconditional coverage than forecasts made by either the parametric model or the FHS VaR model. The unconditional coverage has zero violations, which means that the expected number of violations is correct compared with the actual number of violations. The reason the QR model is so accurate when it comes to unconditional coverage is that it is able to capture volatility clustering effects and time-varying volatility as well as the correct distribution of returns. The QR model also has the most conditional coverage. It experiences no failures at the 1% VaR level, either long or short. The right-hand tail has more conditional coverage failures than the left-hand tail, as was the case both for the parametric model and for the FHS model. The total conditional and unconditional coverage failures (see Table 6) are 13 and 0, which is the best of all the VaR models.

The empirical results and their implications can be summarized as follows.

- The forecasts made by the parametric VaR model are poor. The model fails in approximately half of the cases over all the confidence levels with regard to conditional coverage, and in roughly 32% of the cases with regard to unconditional coverage. This is to be expected due to the underlying assumption of normal distributed returns, although we do not recommend using this simple VaR model when accurate VaR measures are needed. Change OK?

- The FHS approach is better than the parametric approach. The model gives a quite accurate VaR forecast at the 10% and 5% VaR levels, but this forecast is not quite so accurate when the tail values are extreme. This is believed to be because there is not enough data to describe the extreme tail values. Changes to sentence OK?

- The QR VaR model performs superiorly to both the parametric and the FHS VaR models with regard to both conditional and unconditional coverage. The QR model is the only model that gives an accurate unconditional VaR forecast at all specified confidence levels.

- There is at least one VaR model at all the confidence levels that possesses unconditional coverage. This is not, however, the case when it comes to conditional coverage for all energy futures at all confidence levels. At the 1% VaR level for long and short positions, there is at least one model that has both conditional coverage and unconditional coverage for all energy futures. Changes to this sentence and the next OK?

- The right-hand tail is the most difficult to model. Most failures occur on the right-hand tails of the return distributions for all energy futures. This means that the risk models are more adequate for investors with portfolios including long positions in the European energy futures markets.

- Coal is the most difficult future to model. The Nordic power third-month position and oil are the easiest to model. The coal futures have the most violations of any model in terms of both conditional and unconditional coverage. Oil and the Nordic power third-month position have the fewest. Changes to sentence OK?

- The conditional coverage is affected by the clustering of occurrence. The observed occurrence of violations is not independent for many of the futures contracts. Models should be improved concerning this issue.

**TABLE 3** Test statistics for both conditional and unconditional coverage of the one-day-ahead VaR forecasts made by the parametric model.

| | Parametric method: Christoffersen test | | | | | | Parametric method: Kupiec test | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **1% VaR** | **5% VaR** | **10% VaR** | **90% VaR** | **95% VaR** | **99% VaR** | **1% VaR** | **5% VaR** | **10% VaR** | **90% VaR** | **95% VaR** | **99% VaR** |
| Oil | **16.01** | 4.40 | 0.85 | 0.44 | 0.75 | 2.72 | **14.80** | 3.37 | 0.18 | 0.36 | 0.31 | 1.90 |
| Natural gas | **17.97** | **11.78** | **6.16** | **16.00** | **14.79** | **18.32** | **10.06** | 0.23 | 0.36 | 0.55 | 0.87 | **14.80** |
| Coal | **17.97** | 4.84 | 3.74 | 4.11 | 4.26 | 5.61 | **10.06** | **4.80** | 1.50 | 2.48 | 0.43 | 5.27 |
| German power 1 | 4.73 | 0.16 | 1.72 | **39.86** | **26.18** | **42.47** | 3.74 | 0.01 | 0.36 | 2.94 | 0.43 | **36.72** |
| Nordic power 1 | **19.34** | **6.55** | **7.78** | **17.12** | **7.66** | 6.41 | **17.44** | **4.06** | 0.08 | **5.53** | 0.23 | 6.13 |
| German power 2 | 4.19 | 1.35 | 2.78 | **24.83** | **11.12** | 9.15 | 3.74 | 0.71 | 0.02 | 1.34 | 0.43 | **7.03** |
| Nordic power 2 | **12.74** | 3.74 | 0.24 | **8.76** | **6.50** | **10.83** | **11.17** | 3.71 | 0.08 | **5.53** | 0.04 | **9.00** |
| German power 3 | 8.19 | 2.92 | 4.33 | 4.49 | 3.03 | **14.87** | **7.99** | 2.75 | 0.24 | 1.50 | 0.87 | **13.54** |
| Nordic power 3 | **17.97** | 4.84 | 3.74 | 4.11 | 4.26 | 5.61 | **10.06** | **4.80** | 1.50 | 2.48 | 0.43 | 5.27 |
| Critical value | 9.21 | 5.99 | 4.61 | 4.61 | 5.99 | 9.21 | 6.63 | 3.84 | 2.71 | 2.71 | 3.84 | 6.63 |

Bold numbers mean that the null hypothesis is rejected and that the model is wrongly specified. Critical values for the chi-squared statistics are presented in the final row.

**TABLE 4** Test statistics for both conditional and unconditional coverage of the one-day-ahead VaR forecasts made by the FHS model.

| | FHS method: Christoffersen test | | | | | | FHS method: Kupiec test | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **1% VaR** | **5% VaR** | **10% VaR** | **90% VaR** | **95% VaR** | **99% VaR** | **1% VaR** | **5% VaR** | **10% VaR** | **90% VaR** | **95% VaR** | **99% VaR** |
| Oil | 5.09 | 3.83 | 1.57 | 1.44 | 6.34 | **12.94** | 4.93 | 3.68 | 0.04 | 1.50 | **5.84** | **7.40** |
| Natural gas | **25.80** | **7.86** | **16.70** | **17.63** | **9.71** | **14.97** | **25.80** | 0.15 | **7.14** | 2.64 | 0.43 | **8.89** |
| Coal | 6.45 | 3.55 | 1.85 | 0.68 | 6.18 | **10.62** | 0.17 | 2.03 | 1.85 | 0.04 | 2.32 | **10.55** |
| German power 1 | **14.55** | 0.71 | 3.08 | **29.71** | **20.27** | 4.84 | **12.41** | 0.56 | 0.12 | 2.22 | 0.13 | 2.28 |
| Nordic power 1 | 5.10 | 1.51 | 3.84 | **5.82** | **7.05** | 5.09 | 3.92 | 0.32 | 0.68 | 1.85 | 0.31 | 4.93 |
| German power 2 | **16.88** | 1.42 | 2.07 | **28.33** | 2.77 | **9.66** | **14.50** | 0.87 | 1.19 | **4.91** | 0.43 | 6.09 |
| Nordic power 2 | 1.93 | 5.16 | 0.66 | 5.15 | **6.86** | 6.22 | 1.12 | **4.48** | 0.68 | 1.19 | 1.76 | 6.09 |
| German power 3 | **19.49** | 3.97 | 3.75 | **6.95** | 1.88 | 4.00 | **16.84** | 2.32 | 2.03 | 0.31 | 0.07 | 1.65 |
| Nordic power 3 | 2.61 | 3.16 | 1.85 | 0.58 | 5.20 | 10.62 | 0.39 | 1.76 | 1.85 | 0.02 | 2.63 | **10.55** |
| Critical value | 9.21 | 5.99 | 4.61 | 4.61 | 5.99 | 9.21 | 6.63 | 3.84 | 2.71 | 2.71 | 3.84 | 6.63 |

Bold numbers mean that the null hypothesis is rejected and that the model is wrongly specified. Critical values for the chi-squared statistics are presented in the final row.

**TABLE 5** Test statistics for both conditional and unconditional coverage of the one-day-ahead VaR forecasts made by the QR model.

| | QR model: Christoffersen test | | | | | | QR model: Kupiec test | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **1% VaR** | **5% VaR** | **10% VaR** | **90% VaR** | **95% VaR** | **99% VaR** | **1% VaR** | **5% VaR** | **10% VaR** | **90% VaR** | **95% VaR** | **99% VaR** |
| Oil | 0.49 | 0.22 | 2.00 | 0.49 | 0.92 | 0.57 | 0.04 | 0.01 | 0.02 | 0.00 | 0.01 | 0.04 |
| Natural gas | 5.58 | **13.81** | **8.99** | **18.96** | **10.63** | 0.57 | 0.04 | 0.01 | 0.00 | 0.00 | 0.01 | 0.04 |
| Coal | 5.58 | 0.61 | 0.86 | 1.61 | **8.59** | 0.49 | 0.04 | 0.01 | 0.00 | 0.01 | 0.01 | 0.00 |
| German power 1 | 0.49 | 0.17 | 2.64 | **38.30** | **29.34** | 1.30 | 0.04 | 0.01 | 0.00 | 0.01 | 0.01 | 0.04 |
| Nordic power 1 | 0.49 | 0.75 | 3.79 | **13.83** | 9.20 | 1.30 | 0.00 | 0.01 | 0.00 | 0.01 | 0.01 | 0.04 |
| German power 2 | 1.56 | 0.67 | 3.19 | **27.33** | **10.63** | 4.97 | 0.04 | 0.00 | 0.00 | 0.02 | 0.01 | 0.04 |
| Nordic power 2 | 1.56 | 0.01 | 0.08 | **8.84** | 5.07 | 0.57 | 0.04 | 0.01 | 0.00 | 0.00 | 0.01 | 0.04 |
| German power 3 | 0.49 | 0.67 | 1.85 | **5.58** | 2.36 | 1.30 | 0.04 | 0.00 | 0.00 | 0.00 | 0.01 | 0.04 |
| Nordic power 3 | 5.58 | 0.61 | 0.86 | 1.61 | **8.59** | 0.49 | 0.04 | 0.01 | 0.00 | 0.01 | 0.01 | 0.00 |
| Critical value | 9.21 | 5.99 | 4.61 | 4.61 | 5.99 | 9.21 | 6.63 | 3.84 | 2.71 | 2.71 | 3.84 | 6.63 |

Bold numbers mean that the null hypothesis is rejected and that the model is wrongly specified. Critical values for the chi-squared statistics are presented in the final row.

**TABLE 6** Accuracy of VaR models: Christoffersen test.

| | Parametric VaR | FHS VaR | QR VaR |
|---|---|---|---|
| Total failures (54) | 24 | 19 | 13 |
| Failures at 1% long position (9) | 6 | 4 | 0 |
| Failures at 5% long position (9) | 2 | 1 | 1 |
| Failures at 10% long position (9) | 2 | 1 | 1 |
| Failures at 10% short position (9) | 5 | 5 | 6 |
| Failures at 5% short position (9) | 5 | 4 | 5 |
| Failures at 1% short position (9) | 4 | 4 | 0 |

This table presents the total accuracy of each model with regard to conditional coverage. This table also lists the accuracy of conditional coverage for all confidence levels, long and short, for all VaR models. The values represent how many of the test variables were rejected. The values in parentheses represent the total number of test variables.

**TABLE 7** Accuracy of VaR models: Kupiec test.

| | Parametric VaR | FHS VaR | QR VaR |
|---|---|---|---|
| Total failures (54) | 17 | 12 | 0 |
| Failures at 1% long position (9) | 7 | 4 | 0 |
| Failures at 5% long position (9) | 3 | 1 | 0 |
| Failures at 10% long position (9) | 0 | 1 | 0 |
| Failures at 10% short position (9) | 2 | 1 | 0 |
| Failures at 5% short position (9) | 0 | 1 | 0 |
| Failures at 1% short position (9) | 5 | 4 | 0 |

This table presents the total accuracy of each model with regard to unconditional coverage. This table also lists the accuracy of conditional coverage for all confidence levels, long and short, for all VaR models. The values represent how many of the test variables were rejected. The values in parentheses represent the total number of test variables.

## 6 CONCLUSIONS

In this paper, we have examined risk in the European energy markets, more precisely future contracts traded at the ICE, EEX and Nasdaq OMX (NordPool). Energy markets are a special class of commodity markets. Due to the difficulties involved in storing energy, one often sees very volatile behavior in price return distributions. We investigate risk for both short and long positions at the 1%, 5% and 10% confidence levels for crude oil, coal, natural gas and the German and Nordic power commodities. Risk, measured in terms of volatility, is generally highest for the Nordic power and natural gas market, and lowest for coal. It is evident that the empirical distribution is time dependent. Oil and coal show typical volatility clustering behavior, while natural gas and the power futures exhibit behavior that is more spiky. Our findings

Changes to sentence OK?

Changes to sentence OK?

with regard to statistical risk characteristics in this paper are similar to those found by Westgaard *et al* (2014)

Energy commodities are a nonhomogeneous asset class with nonnormal price return distributions. It is therefore important to be careful when selecting which risk model one wishes to use to manage risk. To help answer this, we have made use of three different VaR models – a simple parametric model, an HS model and a QR model – in order to make one-day-ahead VaR forecasts. These forecasts are then evaluated by two backtesting methods, namely Kupiec's unconditional coverage test and Christoffersen's conditional coverage test.

We are able to accurately predict in-sample one-day-ahead VaR forecasts for each commodity at all specified confidence levels for both short and long positions in the period 2007–17. We are not able to obtain complete conditional coverage for all commodities at all confidence intervals due to clustering effects when violations occur. The parametric model shows clear weaknesses because of the assumption of normal distributed price change returns. The HS model is the better performer of the two, but because there is not enough sample data describing extreme tails, it is thought that VaR forecasts at extreme tails are somewhat inaccurate. The QR-based VaR model provides accurate VaR forecasts for all futures at all confidence intervals. This method is able to capture the empirical distribution of price change returns as well as volatility clustering through an EWMA. Even so, the model is not perfect because it does not obtain complete conditional coverage.

Changes to sentence OK?

Energy futures are a challenging asset class. One should be careful when selecting a risk model for a portfolio consisting of such assets. Investors should bear in mind that the risk characteristics of energy futures change over time and are also different for long and short positions. In future, risk models should be applied to assess out-of-sample VaR. Other risk metrics should also be applied, such as ES. Our VaR models should also be improved to lessen clustering effects from occurring.

Change OK?

## DECLARATION OF INTEREST

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

## REFERENCES

Alexander, C. (2008). *Value-at-Risk Models*. Wiley.

Andriosopoulos, K., and Nomikos, N. (2015). Risk management in the energy markets and value-at-risk modelling: a hybrid approach. *European Journal of Finance* **21**(7), 548–574 (https://doi.org/10.1080/1351847X.2013.862173).

Anson, M. (2004). Managing downside risk in return distributions using hedge funds, managed futures and commodity futures.

Details?

Barone-Adesi, G., Giannopoulos, K., and Vosper, L. (1999). VaR without correlations for portfolio of derivative securities. Technical Report, Università della Svizzera Italiana (https://doi.org/10.1002/(SICI)1096-9934(199908)19:5<583::AID-FUT5>3.0.CO;2-S).

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**(3), 307–327 (https://doi.org/10.1016/0304-4076(86)90063-1).

Bunn, D., Andresen, A., Chen, D., and Westgaard, S. (2016). Analysis and forecasting of electricity price risks with quantile factor models. *Quarterly Journal of the IAEE's Energy Economics Education Foundation* **37**(1), 101–122 (https://doi.org/10.5547/01956574.37.1.dbun).

Byström, H. N. (2005). Extreme value theory and extremely large electricity price changes. *International Review of Economics and Finance* **14**(1), 41–55 (https://doi.org/10.1016/S1059-0560(03)00032-7).

Cabedo, J. D., and Moya, I. (2003). Estimating oil price "value at risk" using the historical simulation approach. *Energy Economics* **25**(3), 239–253 (https://doi.org/10.1016/S0140-9883(02)00111-1).

Chan, K. F., and Gray, P. (2006). Using extreme value theory to measure value-at-risk for daily electricity spot prices. *International Journal of Forecasting* **22**(2), 283–300 (https://doi.org/10.1016/j.ijforecast.2005.10.002).

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review* **39**(4), 841–862 (https://doi.org/10.2307/2527341).

Costello, A., Asem, E., and Gardner, E. (2008). Comparison of historically simulated VaR: evidence from oil prices. *Energy Economics* **30**(5), 2154–2166 (https://doi.org/10.1016/j.eneco.2008.01.011).

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society* **50**(4), 987–1007 (https://doi.org/10.2307/1912773).

Füss, R., Adams, Z., and Kaiser, D. G. (2010). The predictive power of value-at-risk models in commodity futures markets. *Journal of Asset Management* **11**(4), 261–285 (https://doi.org/10.1057/jam.2009.21).

Giot, P., and Laurent, S. (2003). Market risk in commodity markets: a VaR approach. *Energy Economics* **25**(5), 435–457 (https://doi.org/10.1016/S0140-9883(03)00052-5).

Harding, L. (2011). Angela Merkel switches off seven nuclear power plants. Details?

Haugom, E., Ray, R., Ullrich, C. J., Veka, S., and Westgaard, S. (2016). A parsimonious quantile regression model to forecast day-ahead value-at-risk. *Finance Research Letters* **16**, 196–207 (https://doi.org/10.1016/j.frl.2015.12.006).

Hull, J. (2012). *Risk Management and Financial Institutions*, Volume 733. Wiley.

Hull, J., and White, A. (1998). Incorporating volatility updating into the historical simulation method for value-at-risk. *The Journal of Risk* **1**(1), 5–19 (https://doi.org/10.21314/JOR.1998.001).

Hung, J.-C., Lee, M.-C., and Liu, H.-C. (2008). Estimation of value-at-risk for energy commodities via fat-tailed GARCH models. *Energy Economics* **30**(3), 1173–1191 (https://doi.org/10.1016/j.eneco.2007.11.004).

Jorion, P. (2000). *Value at Risk*, 2nd edn. McGraw-Hill.

JP Morgan (1996). *Riskmetrics: Technical Document*. Morgan Guaranty Trust Company of New York. Changes to bibliography entry OK?

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press (https://doi.org/10.1017/CBO9780511754098).

Koenker, R., and Bassett, G., Jr. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society* **46**(1), 33–50 (https://doi.org/10.2307/1913643).

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives* **3**(2), 73–84 (https://doi.org/10.3905/jod.1995.407942).

Kuralbayeva, K., and Malone, S. W. (2012). The determinants of extreme commodity prices. 2012 Annual Meetings Paper, Midwest Finance Association (https://doi.org/10.2139/ssrn.1929043).

Lux, T., Segnon, M., and Gupta, R. (2016). Forecasting crude oil price volatility and value-at-risk: evidence from historical and recent data. *Energy Economics* **56**, 117–133 (https://doi.org/10.1016/j.eneco.2016.03.008).

Marimoutou, V., Raggad, B., and Trabelsi, A. (2009). Extreme value theory and value at risk: application to oil market. *Energy Economics* **31**(4), 519–530 (https://doi.org/10.1016/j.eneco.2009.02.005).

Nowotarski, J., and Weron, R. (2018). Recent advances in electricity price forecasting: a review of probabilistic forecasting. In *Renewable and Sustainable Energy Reviews*, Volume 81, Part 1, pp. 1548–1568 (https://doi.org/10.1016/j.rser.2017.05.234).

Politis, D. N. (2004). A heavy-tailed distribution for ARCH residuals with application to volatility prediction. Details?

Rahman, S., and Serletis, A. (2012). Oil price uncertainty and the Canadian economy: evidence from a VARMA, GARCH-in-mean, asymmetric BEKK model. *Energy Economics* **34**(2), 603–610 (https://doi.org/10.1016/j.eneco.2011.08.014).

Sadorsky, P. (1999). Oil price shocks and stock market activity. *Energy Economics* **21**(5), 449–469 (https://doi.org/10.1016/S0140-9883(99)00020-1).

Steen, M., Westgaard, S., and Gjolberg, O. (2015). Commodity value-at-risk modeling: comparing RiskMetrics, historic simulation and quantile regression. *The Journal of Risk Model Validation* **9**(2), 49–78 (https://doi.org/10.21314/JRMV.2015.146).

Veka, S., Lien, G., Westgaard, S., and Higgs, H. (2012). Time-varying dependency in European energy markets: an analysis of Nord Pool, European Energy Exchange and Intercontinental Exchange energy commodities. *The Journal of Energy Markets* **5**(2), 3–32 (https://doi.org/10.21314/JEM.2012.072).

Westgaard, S., Veka, S., Haugom, E., and Lien, G. (2014). A note on the risk characteristics of european energy futures markets. *Beta* **28**(1), 6–19.

Zhen-xin, W., Bing, X., and Shu-ping, W. (2011). The impact of oil price volatility on China's economy based on VaR model. *Chinese Journal of Management Science* **19**(1), 21–28.