



Oil tail-risk forecasts: from financial crisis to COVID-19

Wei Kuang¹

Accepted: 2 August 2022 / Published online: 27 August 2022
© The Author(s), under exclusive licence to Springer Nature Limited 2022

Abstract

The coronavirus outbreak has caused unprecedented volatility in oil prices. This paper extends previous studies on oil Value-at-Risk (VaR) by providing extra insights into Expected Shortfall (ES) forecasting over the last decade, including several oil crises. We introduce a conditional volatility model combined with the Cornish–Fisher expansion for ES forecasting. In comparison to the widely used volatility models and innovation distributions, this approach is superior for predicting the ES of long positions but overestimates VaR for short positions. Overall, the volatility model addressing leverage effects with skewed t innovation produces the most accurate joint VaR and ES forecasting. Moreover, the magnitude of ES relative to VaR varies across models and time, implying that ES should be used in conjunction with VaR to inform timely risk management decisions. The results would be of interest to the regulatory authorities, energy companies, and financial institutions for oil tail-risk forecasting.

Keywords COVID-19 · Oil price volatility · Cornish–Fisher expansion · Value-at-Risk · Expected Shortfall · FRTB

Introduction

The COVID-19 pandemic has caused unprecedented volatility in oil prices, making energy risk management one of the primary concerns for regulators and market participants. Oil price fluctuations have a significant impact not only on the real economy (e.g., Hamilton 1983, 2003), but also on financial markets such as equity (e.g., Liu et al. 2020; Hashmi et al. 2021) and currency (e.g., Dai et al. 2020; Ding et al. 2021) with a strong tail contagion effect (Le et al. 2021). The growing integration of the oil and financial markets, particularly the large fluctuations in crude oil prices during a crisis, has necessitated the development of robust oil price risk forecasting.

✉ Wei Kuang
kuangwei8355@hotmail.com

¹ Lloyds Banking Group, 10 Gresham Street, London EC2V 7AE, UK



Value-at-risk (VaR) has been one of the most widely used risk measures in the financial industry since its adoption by the Basel Committee on Banking Supervision in 1996. VaR is defined as the portfolio's expected maximum loss over a given time horizon and with a given level of confidence. Despite its simplicity, VaR is a useful tool for monitoring risk exposure, adjusting positions to avoid large losses, and allocating capital to minimize income volatility while maximizing return on investment. Therefore, accurate VaR forecasting is critical for internal risk management and financial stability. However, during the early 2020 financial market turmoil, many financial institutions recorded excessive exceptions, i.e., the trading book losses exceeded the VaR forecasts (Risk.net 2020). This raises questions about the accuracy of the VaR models.

VaR was introduced for oil risk quantification in the 2000s under the volatile oil price environment (Sadeghi and Shavvalpour 2006). Oil price returns exhibit similar characteristics to other financial assets, such as skewness, leptokurtosis, and volatility clustering (Giot and Laurent 2003). This has motivated prior studies to forecast oil VaR based on a volatility filtering process that is carried out in two steps. To capture the volatility dynamics, returns are first filtered by the generalized autoregressive conditional heteroskedasticity (GARCH) model introduced by Engle (1982) and Bollerslev (1986), and then the standardized residuals, i.e., innovations are modeled by a specified distribution for quantile estimation. GARCH-type models with normal innovations can generate data with unconditionally fat tails, but they are not sufficiently fat to address all of the unconditional leptokurtosis observed in returns series (Kuang 2021). Alternative innovation distributions have been explored in oil VaR forecasts, including historical simulation (e.g., Costello et al. 2008) and parametric distributions such as generalized error (Fan et al. 2008), heavy-tailed (Hung et al. 2008), skewed Student's t (Giot and Laurent 2003) and other flexible skewed distributions (Lyu et al. 2017), generalized pareto (Marimoutou et al. 2009), and Pearson's type IV and Johnson's Su distributions (Patra 2021). According to these studies, choosing the appropriate distribution for the innovation process is critical because it affects the accuracy of the required quantile estimates. The best-performing models differ depending on the models compared and the time periods investigated. Overall, flexible distributions are preferred, with those that capture both skewed and fat-tailed properties outperforming normal and non-skew distributions.

Building on the research of crude oil volatility modeling and forecasting, another strand of literature investigates variants of GARCH-type models to address stylized facts of oil volatility for more accurate VaR forecasts. For example, volatility persistence is observed in the crude oil price (e.g., Choi and Hammoudeh 2009; Kang et al. 2009). This implies that price fluctuations have long-term effects on volatility. Moreover, crude oil futures have a leverage effect (e.g., Nomikos and Andriospoulos 2012; Kristoufek 2014). This suggests that downward movements (shocks) in the crude oil market are accompanied by greater volatility than similar upward movements. Wei et al. (2010) show that non-linear specification of GARCH models, allowing for long memory as well as asymmetric leverage effect in volatility, produce greater forecasting accuracy than the linear ones, especially for long-run volatility forecasting. Cheong (2009), on the other hand, reveals that while both the estimation and diagnostic evaluations favor an asymmetric long-memory autoregressive



conditional heteroskedasticity (ARCH) model, the simplest parsimonious GARCH can produce better out-of-sample forecasting performance. Despite this, models that account for long-range memory and asymmetry in the volatility process, as well as fat tails, have been shown to outperform others in predicting oil VaR (e.g., Aloui and Mabrouk 2010; Chkili et al. 2014; Youssef et al. 2015). GARCH-type models also suffer from some limitations. For example, Lux et al. (2016) discover that Markov-switching multifractal (MSM) models that capture the multiscaling, long memory, and structural breaks of oil price volatility outperform a battery of GARCH-type models in VaR forecasts. Moreover, the squares of past daily returns are commonly used in GARCH-type models, which are unbiased but inefficient estimators of volatility because they only use one price measure per period. Liu and Wan (2012) find that the intraday returns-based realized volatility model outperforms the GARCH-type models in forecasting oil price volatility. However, the MSM model is more complex than GARCH-type models, and managing high-frequency data requires more time and computational resources.

Prior studies have mainly focused on VaR forecasts. However, VaR provides no information about potential losses beyond the quantile, which is problematic when exceeded exceptions occur during a crisis, nor does it satisfy the “subadditivity” requirement, i.e., the total risk of a portfolio should be less than or equal to the sum of risks of the individual portfolio assets. The expected shortfall (ES) has recently received growing attention as an alternative measure due to the upcoming implementation of the “Fundamental review of the trading book,” a revised market risk framework proposed by the Basel Committee on Banking Supervision (BCBS 2016). ES is the conditional expectation of exceedances above a certain VaR threshold. Despite a coherent risk measure (Artzner et al. 1999), ES does not satisfy the elicibility, which is a desirable property for evaluating point forecasts (Gneiting 2011). This has prompted research into the development of ES backtesting (e.g., McNeil and Frey 2000; Fissler and Ziegel 2016; Bayer and Dimitriadis 2020).

The challenge of oil price risk management, combined with impending regulatory changes, prompted this study to look into alternative approaches for more accurate oil tail-risk forecasting without increasing model complexity. Specifically, this paper extends previous literature on oil VaR by including ES analysis for both long and short oil positions over the last decade, which covers the 2008 global financial crisis, the 2014 oil crash, and the recent COVID-19 pandemic. It considers three GARCH-type models that were previously used for oil volatility forecasting: the standard GARCH (Bollerslev 1986), the Glosten-Jagannathan-Runkle GARCH (GJR-GARCH) model of Glosten et al. (1993), and the fractionally integrated GARCH model (FIGARCH) of Baillie et al. (1996) with four innovation distributions: normal, skewed t , historical simulation, and the Cornish–Fisher expansion (CF). The VaR and ES forecasts are compared, and their performance is evaluated using traditional risk management criteria and the recently developed ES backtesting techniques. In particular, the paper contributes to the existing literature in three dimensions.

First, this research sheds light on oil ES forecasts other than VaR. It demonstrates the importance of using the ES forecast as a risk-control supplement to VaR. Tail-risk management is critical, as market participants are concerned not only about the loss



associated with a certain level of confidence but also about the magnitude of the loss that could exceed the threshold. Moreover, it illustrates the impact of shifting from VaR to ES as a risk measure, which provides insights into the performance of various models as part of the transition mandated by the financial regulators.

Second, this paper extends the GARCH model combined with the Cornish–Fisher expansion (GARCH-CF) for forecasting ES. This approach was proposed by Alexander et al. (2013) to forecast the VaR of equity, foreign exchange, and interest rates, and has recently been explored to forecast the VaR of oil (Kuang 2022). It is based on the modified VaR of Favre and Galeano (2002) but incorporates the GARCH model to capture volatility dynamics, in addition to using Cornish–Fisher to address the skewness and excess kurtosis of innovation. To the best of our knowledge, no research has been conducted on this approach for ES forecasting.

Third, to evaluate the oil ES forecast, this paper employs a novel regression-based ES test of Bayer and Dimitriadis (2020) with only the ES forecast as an input variable. This is in contrast to Youssef et al. (2015) which uses the exceedance residuals test of McNeil and Frey (2000) dependent on both VaR and ES forecasts. Furthermore, we use a joint loss function for VaR and ES proposed by Fissler and Ziegel (2016) to compare different models' forecast accuracy.

We find that the proposed GARCH-CF model is superior in forecasting oil ES for long positions. The GJR-GARCH model with skewed t provides the most accurate joint VaR and ES forecasts. The relative magnitude of the ES to VaR forecast varies across models and over time. Switching from VaR to ES as a risk measure increases risk prediction to varying degrees, with the GARCH-CF model being the least impacted.

The remainder of the article is organized as follows. “Methodology” section introduces the models and evaluation methods. “Empirical analysis” section describes the data and discusses the empirical results. “Conclusion” section concludes.

Methodology

VaR has been the standard method for quantifying and managing market risk since its adoption by the Basel Committee on Banking Supervision in 1996. VaR is defined as the maximum loss that may be incurred by a portfolio over a given time horizon h at a certain level of confidence p , i.e., $\Pr(r_{t+h} \leq \text{VaR}_{t+h}^\alpha) = \alpha$, where $\alpha = 1 - p$ and r_{t+h} is the asset's return over the period h . ES is defined as the expected loss when losses have exceeded VaR. Following the 2008 financial crisis, the Basel Committee on Banking Supervision issued the “Fundamental review of the trading book” (FRTB) (BCBS 2016) to revise the approach to calculate risk-based capital requirements for trading activities. One of the significant changes is the transition from the VaR with a confidence level of $p = 99\%$ to the ES with a confidence level of $p = 97.5\%$ as an alternative measure to quantify market risk.

Given the daily log returns of an asset following the below process:

$$r_t = \mu_t + \varepsilon_t = \mu_t + \sigma_t z_t, \quad \text{with } z_t \sim i.i.d(0, 1), \quad (1)$$



where $r_t = p(t) - p(t - 1)$ and $p(t)$ is the logarithmic asset price observed at day t . The day-ahead VaR and ES are defined as follows (Brio et al. 2020):

$$\text{VaR}_t^\alpha = \hat{\mu}_t + \hat{\sigma}_t \hat{F}_z^{-1}(\alpha) \quad (2)$$

$$\text{ES}_t^\alpha = \hat{\mu}_t + \hat{\sigma}_t \hat{S}_z(\alpha) \quad , \quad (3)$$

where $\hat{\mu}_t$ and $\hat{\sigma}_t$ are the one-day ahead conditional mean and conditional volatility forecasts, respectively; $\hat{F}_z^{-1}(\alpha)$ is the α quantile of innovation, i.e., $z_t = (r_t - u_t)/\sigma_t$ and $\hat{S}_z(\alpha) = \mathbb{E}[z|z < \hat{F}_z^{-1}(\alpha)]$. The ES estimation is generally computed by numerical integration, while close-form solutions may exist in some cases.

Forecasting VaR and ES requires the conditional mean, conditional volatility, and the quantile estimation of the innovation distribution. Conditional mean returns are difficult to forecast using available information and forecasting techniques, as daily returns are heavily influenced by specific news and events that a model often does not take into account (Christoffersen and Diebold 2006). The conditional volatility, on the other hand, is time-varying and highly predictable (Andersen and Bollerslev 1998). Empirical research on volatility modeling has exploded since the introduction of the ARCH model by Engle (1982) to capture the volatility clustering effect, i.e., the tendency for large changes in asset prices to follow large changes and small changes to follow small changes. The ARCH model was generalized into the GARCH model by Bollerslev (1986), and various extensions have been proposed to capture other stylized facts of financial asset returns.¹ In this study, we use the standard GARCH model and two of its extensions, GJR-GARCH and FIGARCH models to model and forecast the volatility of oil markets based on prior literature (e.g., Kang et al. 2009; Charles and Darné 2017). The details of the volatility models are discussed in “Volatility models” section.²

Although GARCH-type models with normal innovations can generate data with unconditionally fat tails, they are insufficient to account for all of the unconditional leptokurtosis and skewness observed in oil returns series (e.g., Lux et al. 2016; Patra 2021). As a result, in addition to the normal distribution, this study compares the non-parametric historical simulation (e.g., Westgaard et al. 2019), parametric skewed t distribution (e.g., Giot and Laurent 2003), and semi-parametric Cornish–Fisher expansion (Cornish and Fisher 1937) approaches to address the non-normality of return innovation for both VaR and ES forecasts. The details of innovation distributions are covered in “Innovation distributions” section.

¹ See Bollerslev et al. (1992), Bera and Higgins (1993), and Palm (1996) for an overview of (G)ARCH-type models.

² The parameters of the GARCH-type models are estimated using the “rugarch” package in R developed by Ghalanos (2020).



Volatility models

Standard GARCH

Building on the ARCH model introduced by Engle (1982), Bollerslev (1986) proposed the GARCH model to address volatility clustering. The conditional volatility is expressed as a function of past returns and past conditional volatility forecasts. The GARCH(1,1) specification is defined as follows:

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \quad (4)$$

where $\omega, \alpha, \beta > 0$ and $\alpha + \beta < 1$ to ensure conditional variance σ_t^2 positive and stationary. The decay rate of the positive autocorrelation in the volatility process is governed by $\alpha + \beta$: the closer $\alpha + \beta$ is to 1, the slower the decay of the autocorrelation of σ_t .

GJR GARCH

Glosten et al. (1993) developed a GJR GARCH model to capture positive and negative shocks on the conditional variance asymmetrically using the indicator function I . The GJR GARCH has been used to address the asymmetric leverage effect in oil volatility (e.g., Charles and Darné 2017; Rizvi and Itani 2021).³ The model is defined by Wei et al. (2010):

$$\sigma_t^2 = \omega + [\alpha + \gamma I(\varepsilon_{t-1} < 0)] \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \quad (5)$$

where $I(\cdot)$ takes the value of one if the condition in the parenthesis is satisfied; γ represents the “leverage” term, i.e., for $\gamma > 0$ the impact of past negative returns on conditional variance is greater than the impact of past positive returns.

FIGARCH

Baillie et al. (1996) introduced a FIGARCH model that allows for the long-memory feature of volatility. In the FIGARCH model, shocks decay at a slower hyperbolic rate, as opposed to the GARCH model, where shocks decay at an exponential rate, or the integrated GARCH model, where shocks persist forever. Kang et al. (2009) demonstrate that the FIGARCH model is useful for modeling persistence in crude oil price volatility and outperforms the GARCH and the integrated GARCH models in volatility forecasting. The model is defined as follows (Kang et al. 2009):

$$\phi(L)(1-L)^d \varepsilon_t^2 = \omega + [1 - \beta(L)](\varepsilon_t^2 - \sigma_t^2) \quad (6)$$

³ There are other non-linear GARCH models that are used to capture the asymmetric leverage effect in oil volatility forecasts, such as the exponential GARCH of Nelson (1991) and the asymmetric power ARCH model of Ding et al. (1993). Wei et al. (2010), however, find that none of them can outperform the others for either the Brent or the WTI market across different loss functions.



or

$$\sigma_t^2 = \omega[1 - \beta(L)]^{-1} + \{1 - [1 - \beta(L)]^{-1}\phi(L)(1 - L)^d\}\varepsilon_t^2, \quad (7)$$

where L denotes the lag operator; $\phi(L) = [1 - \alpha(L) - \beta(L)](1 - L)^{-1}$; $0 \leq d \leq 1$ is the fractional difference parameter that controls the degree of long memory. The FIGARCH model is flexible to allow for a wide range of persistence via the parameter d , and accommodates the standard GARCH model with $d = 0$ and the integrated GARCH model with $d = 1$ as special cases (Kang et al. 2009). The estimation of the FIGARCH model requires a minimum number of observations. According to the standard procedure in the literature (e.g., Beine et al. 2002; Beine and Laurent 2003), the truncation order of the infinite $(1 - L)^d$ is set to 1000 lags:

$$(1 - L)^d = \sum_{k=0}^{1000} \frac{\Gamma(k - d)}{\Gamma(k + 1)\Gamma(-d)} L^k \quad (8)$$

Innovation distributions

Normal distribution

The naive assumption is that the standardized residual follows a standard normal distribution, i.e., $z_t \sim iidN(0, 1)$ with the cdf defined as

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt \quad (9)$$

The day-ahead VaR and ES forecasts are given by (Brio et al. 2020):

$$\text{VaR}_{t,N}^\alpha = \hat{\mu}_t + \hat{\sigma}_t \Phi^{-1}(\alpha) \quad (10)$$

$$\text{ES}_{t,N}^\alpha = \hat{\mu}_t + \hat{\sigma}_t \frac{1}{\alpha} \phi[\Phi^{-1}(\alpha)] \quad (11)$$

where ϕ is the pdf of standard normal distribution and $\Phi^{-1}(\alpha)$ is the corresponding α quantile.

Skewed student distribution

Fernández and Steel (1998) proposed a four-parameter skewed student distribution to account for the asymmetry and kurtosis of the returns process, and Lambert and Laurent (2001) derives its quantile function as follows.



$$c_{\alpha, \nu, \xi}^{\text{skst}} = \begin{cases} \left\{ \frac{1}{\xi} c_{\alpha, \nu}^{\text{st}} \left[\frac{\alpha}{2} (1 + \xi^2) \right] - m \right\} / s, & \text{if } \alpha < \frac{1}{1 + \xi^2} \\ \left\{ -\xi c_{\alpha, \nu}^{\text{st}} \left[\frac{1 - \alpha}{2} (1 + \xi^{-2}) \right] - m \right\} / s, & \text{if } \alpha \geq \frac{1}{1 + \xi^2} \end{cases} \quad (12)$$

where $c_{\alpha, \nu, \xi}^{\text{skst}}$ is the α th quantile of the unit variance skewed student distribution with $\nu > 2$ degrees of freedom and asymmetric parameter $\xi > 0$; $c_{\alpha, \nu}^{\text{st}}$ denotes the quantile function of the standardized Student- t density function; $m = \frac{\Gamma(\frac{\nu+1}{2})\sqrt{\nu-2}}{\sqrt{\pi}\Gamma(\frac{\nu}{2})} \left(\xi - \frac{1}{\xi} \right)$ and $s = \sqrt{\left(\xi^2 + \frac{1}{\xi^2} - 1 \right) - m^2}$ are the mean and standard deviation of the non-standardized skewed student distribution, respectively. The day-ahead VaR and ES forecasts are then given by

$$\text{VaR}_{t, \text{skst}}^\alpha = \hat{\mu}_t + \hat{\sigma}_t c_{\alpha, \nu, \xi}^{\text{skst}} \quad (13)$$

$$\text{ES}_{t, \text{skst}}^\alpha = \hat{\mu}_t + \hat{\sigma}_t \text{E} \left[z | z < c_{\alpha, \nu, \xi}^{\text{skst}} \right] \quad (14)$$

Filtered historical simulation

The filtered historical simulation (FHS) introduced by Barone-Adesi et al. (1999) is a semi-parametric approach that employs parametric models for the mean and volatility dynamics and a non-parametric estimator for the residual distribution. By assuming that the standardized residual is normally distributed, the parameters of the GARCH model defined in Eq.(4) are estimated using Quasi Maximum Likelihood and the day-ahead VaR and ES forecasts are then given by

$$\text{VaR}_{t, \text{HS}}^\alpha = \hat{\mu}_t + \hat{\sigma}_t \hat{G}^{-1}(\alpha) \quad (15)$$

$$\text{ES}_{t, \text{HS}}^\alpha = \hat{\mu}_t + \hat{\sigma}_t \text{E} \left[z | z < \hat{G}^{-1}(\alpha) \right] \quad , \quad (16)$$

where $\hat{G}^{-1}(\alpha)$ is the α quantile of the empirical distribution of the standardized residual z_t .

Cornish–Fisher expansion

Favre and Galeano (2002) proposed a four-moment “modified VaR” based on the Cornish–Fisher expansion (Cornish and Fisher 1937) for estimating the quantiles of non-normal distributions as a function of standard normal quantiles and the sample skewness and excess kurtosis. The α quantile of a distribution is defined as follows.

$$\delta_{\text{CF}}(\alpha) = Z(\alpha) + \frac{1}{6} (Z(\alpha)^2 - 1) S + \frac{1}{24} (Z(\alpha)^3 - 3Z(\alpha)) K - \frac{1}{36} (2Z(\alpha)^3 - 5Z(\alpha)) S^2 \quad , \quad (17)$$



where $Z(\alpha)$ is the α quantile of the standard normal distribution, and S and K are the skewness and excess kurtosis of standardized residual series. Alexander et al. (2013) suggest using GARCH-type model to capture volatility clustering and then the standardized residual is modeled by the Cornish–Fisher expansion for VaR forecasting. In this paper, we extend this approach to ES forecasting. The one-day ahead VaR forecast and ES are given by

$$\text{VaR}_{t,CF}^{\alpha} = \mu_t + \hat{\sigma}_t \delta_{CF}(\alpha) \quad (18)$$

$$\text{ES}_{t,CF}^{\alpha} = \hat{\mu}_t + \hat{\sigma}_t \text{E}[z|z < \delta_{CF}(\alpha)] \quad (19)$$

Evaluation criteria

Value-at-risk tests

To assess the accuracy of a VaR model, a simple approach is to use the Actual over Expected (AE) exceedance ratio, which tracks the actual number of times that the returns have exceeded the forecast VaR over the expected VaR exceptions (Laporta et al. 2018). The closer the AE ratio approaches one, the more accurately the model predicts VaR. More rigorous approaches for backtesting include unconditional coverage (Kupiec 1995), conditional coverage (Christoffersen 1998), and the dynamic quantile test (Engle and Manganelli 2004), which have been used to backtest oil VaR forecast (e.g., Hung et al. 2008; Marimoutou et al. 2009; Aloui and Mabrouk 2010; Youssef et al. 2015; Lyu et al. 2017; Westgaard et al. 2019).⁴

The unconditional coverage test LR_{uc} of Kupiec (1995) tests the null hypothesis that the exception rate is statistically equal to the expected value given a confidence level. The LR_{uc} test, however, is unable to determine whether the exceptions are distributed randomly. Christoffersen (1998) proposes a more comprehensive conditional coverage test LR_{cc} which jointly examines whether total exceptions are equal to the expected and whether exceptions are distributed independently. Similarly, Engle and Manganelli (2004) introduce a regression-based dynamic quantile (DQ) test to examine whether exceptions can be predicted by a set of explanatory variables such as the first four lags of the Hit series.⁵

Expected shortfall tests

The exceedance residuals (ER) test proposed by McNeil and Frey (2000) was one of the first and widely used ES tests. Under the null hypothesis of a correctly specified

⁴ Given that these tests are standard and well documented in previous oil VaR literature, we do not include detailed specifications in this paper.

⁵ Hit is a de-mean process of an indicator that equals one if violations of VaR occur and zero otherwise at a given significance level. We also consider including predicted VaR as an additional explanatory variable, but the results are qualitatively similar and are available from the authors on request.



risk model, the ER, defined as the difference between the realized return and the ES forecast in the event of a VaR exception, i.e., $er_t = (r_t - ES_t)I_{r_t < VaR_t}$, should have zero mean. The bootstrap method of Efron and Tibshirani (1993) is used to determine whether the expected value of the ER is zero. A one-sided test is performed against the alternative hypothesis that the mean is negative, i.e., the expected short-fall forecast is systematically underestimated. The ER test essentially compares the empirical average of returns truncated at VaR forecast to the average ES forecast whenever there is a VaR violation, so is a joint backtest for the pair VaR and ES forecasts (Bayer and Dimitriadis 2020).⁶

Bayer and Dimitriadis (2020) introduce an ES regression (ESR) test that models the conditional ES as a linear function, with returns as the response variable and ES forecasts as the explanatory variable including an intercept term. The intercept and slope parameters should be equal to zero and one, respectively, for correctly specified ES forecasts. To determine whether the ES forecast is systematically underestimated, Bayer and Dimitriadis (2020) further suggest an Intercept ESR test that limits the slope parameter to one and only estimates and tests the intercept term. The Intercept ESR test is defined as

$$r_t - \hat{e}_t = \gamma_1 + u_t^e, \quad (20)$$

where r_t denotes the daily returns; \hat{e}_t is the ES forecast; the conditional ES at α for u_t^e given the past information \mathcal{F}_{t-1} is zero, i.e., $ES^\alpha(u_t^e | \mathcal{F}_{t-1}) = 0$. The one-sided hypothesis test is given by

$$\mathbb{H}_0^{1s} : \gamma_1 \geq 0 \quad \text{against} \quad \mathbb{H}_A^{1s} : \gamma_1 < 0 \quad (21)$$

which is tested based on a one-sided t test using estimated asymptotic covariance.

Loss functions

In most cases, using the above statistical tests to evaluate the performance of individual models may not be sufficient to determine the best model. In this case, the model confidence set (MCS) procedure of Hansen et al. (2011) can be used to determine a set of “superior” models based on loss functions at a given confidence level. If a model is not in the “superior” set, it is considered to be less likely to be the best model than those are. We follow Hansen et al. (2011) by implementing MCS testing at 90% confidence on various loss functions for evaluating volatility, VaR, and the joint VaR and ES forecasts, respectively. The models with lower average loss values over the forecasting period perform better.

Volatility forecast is evaluated using the Quasi-Likelihood (QLIKE) loss function where daily squared returns are used as a proxy for realized volatility (RV). The QLIKE is an asymmetric loss function with a heavier penalty for under-forecast.

⁶ McNeil and Frey (2000) propose another version of the ER test which uses exceedance residuals standardized by a given volatility forecast, i.e., $er_t/\hat{\sigma}_t$. Bayer and Dimitriadis (2020) find that the raw and standardized ER tests produce similar results.



According to Patton (2011), both the QLIKE and Mean Squared Error (MSE) functions are robust to the presence of noise in the volatility proxy for ranking rival volatility forecasts, but QLIKE is less sensitive to outliers than MSE.

$$\text{QLIKE}_t = \frac{\text{RV}_t}{\hat{\sigma}_t^2} - \ln \left(\frac{\text{RV}_t}{\hat{\sigma}_t^2} \right) - 1 \quad (22)$$

The tick loss function (TLF), which is used for quantile regression (Koenker and Bassett 1978), is one of the most commonly used methods for evaluating quantile forecasts (e.g., González-Rivera et al. 2004; Giacomini and Komunjer 2005). It is an asymmetric loss function that heavily penalizes observations with VaR exceptions.

$$\text{TLF}_t^\alpha = \left(\alpha - I_{r_t < \text{VaR}_t^\alpha} \right) (r_t - \text{VaR}_t^\alpha) \quad (23)$$

While ES itself is not elicitable (Gneiting 2011), Fissler and Ziegel (2016) show that the pair VaR and ES is jointly elicitable with an associated scoring function (FZL). We follow Patton et al. (2019) to use the FZL to assess each model's ability to predict VaR and ES.

$$\text{FZL}_t^\alpha = \frac{1}{\alpha \text{ES}_t^\alpha} I_{r_t < \text{VaR}_t^\alpha} (r_t - \text{VaR}_t^\alpha) + \frac{\text{VaR}_t^\alpha}{\text{ES}_t^\alpha} + \log(-\text{ES}_t^\alpha) - 1 \quad (24)$$

Empirical analysis

Data and estimation

In this study, we focus on two of the most closely watched oil markets: European Brent crude oil (Brent) and West Texas Intermediate crude oil (WTI). The daily spot prices are obtained from the Energy Information Administration (EIA) and cover the period from July 2, 2004 to August 31, 2021 for Brent and July 9, 2004 to August 31, 2021 for WTI, each with approximately 4300 observations. Figures 1 and 2 depict the daily prices and returns for Brent and WTI, respectively. Brent and WTI follow a similar pattern, with prices rapidly rising and reaching a record high above \$140 per barrel in July 2008 before collapsing to \$30 per barrel by the end of the year 2008. The markets started to recover and stabilize before the price slump in later 2014. The most severe crash was observed in early 2020 as a result of a supply glut and a drop in demand caused by the spread of Covid-19. It should be noted that the WTI fell from \$18 to \$−37 per barrel on 20 April 2020, before recovering to \$8 per barrel the next day. The extreme outlier could have a significant impact on the estimation of standard econometric models (Shi 2021). Therefore, we removed the negative price on April 20, 2020 from the WTI time series for a more robust parameters estimation. As can be seen, the impact of this extreme market shock quickly faded, as the market returned to and remain in the positive territory the following days. The



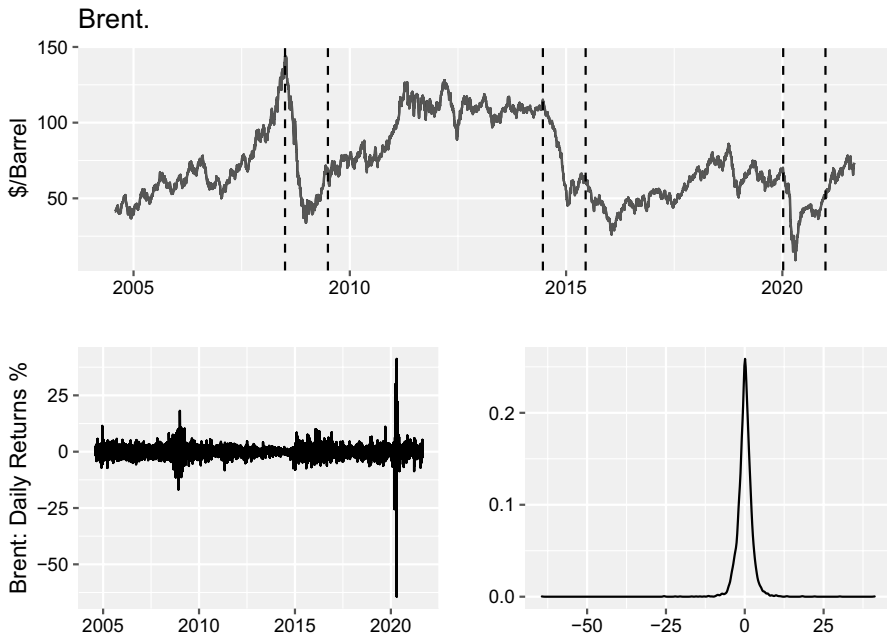


Fig. 1 Brent daily price and return. The graphs show the daily price and return series for Brent from July 9, 2004 to August 31, 2021, along with the return density graph. The three sub-sample forecast windows are denoted by dashed lines

continuously compounded daily returns are calculated as the difference in the logarithms of daily spot prices multiplied by 100, i.e., $R_t = [\ln P_t - \ln P_{t-1}] * 100$.

Each of the models is initially estimated using the first 1000 observations to produce a one-step-ahead VaR and ES forecast on both long and short positions for day 1001.⁷ After that, the estimation sample is advanced by one day, the model is re-estimated, and the forecast for day 1002 is generated, and so on until the sample is completed. As a result, the entire sample has 1000 observations for estimation and 3300 out-of-sample observations for forecast evaluation. The main advantage of using a rolling window is that it incorporates the most recent market data while discarding out-of-date observations. In order to assess whether the model performance varies across the market condition, we further investigate the VaR and ES forecasts for the long position over three sub-samples with extreme oil movement. Each sub-sample has 1250 observations, with the last 250 observations of each sample taken as the out-of-sample forecast period to evaluate the one-day ahead VaR and ES forecasting performance of each model. Table 1 contains the information about each sample period. The first two sub-samples were chosen following Lyu et al. (2017) which

⁷ The estimation window using 1000 observations enables model parameter calibration based on relevant data points, which is in line with existing oil VaR forecast studies (e.g., Marimoutou et al. 2009; Lyu et al. 2017).



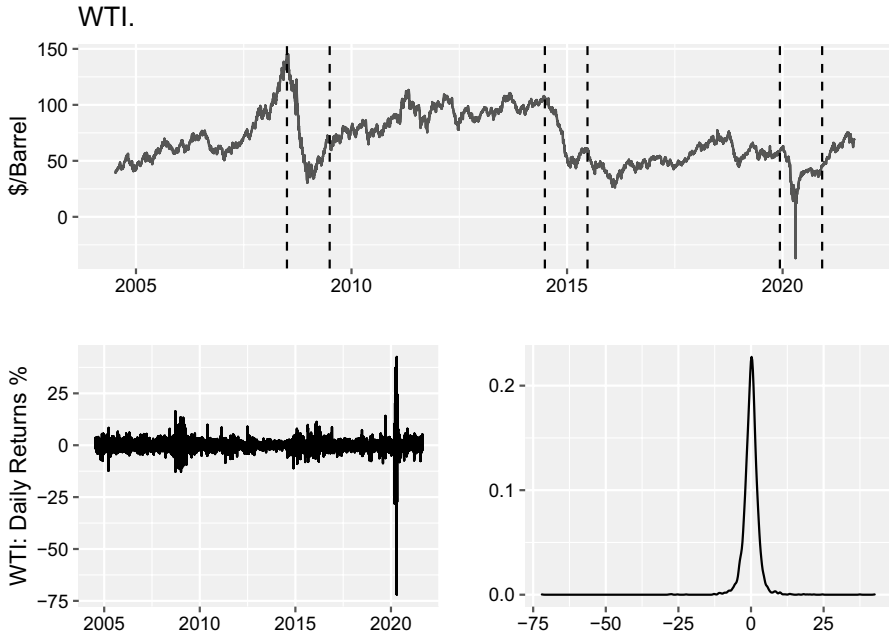


Fig. 2 WTI daily price and return. The graphs show the daily price and return series for WTI from July 9, 2004 to August 31, 2021, along with the return density graph. The three sub-sample forecast windows are denoted by dashed lines. The negative price on April 20, 2020 was removed for creating return series and the density graph

Table 1 Sample periods

	No. of obs.	Sample period	Forecast window
Panel A: Brent			
Full sample	4329	Aug 02, 2004–Aug 31, 2021	Jul 03, 2008–Aug 31, 2021
Sample I	1250	Aug 02, 2004–Jul 01, 2009	Jul 03, 2008–Jul 01, 2009
Sample II	1250	Jun 25, 2010–Jun 16, 2015	Jun 19, 2014–Jun 16, 2015
Sample III	1250	Feb 08, 2016–Dec 31, 2020	Jan 09, 2020–Dec 31, 2020
Panel B: WTI			
Full sample	4309	Jul 12, 2004–Aug 31, 2021	Jul 03, 2008–Aug 31, 2021
Sample I	1250	Jul 12, 2004–Jul 01, 2009	Jul 03, 2008–Jul 01, 2009
Sample II	1250	Jul 06, 2010–Jun 18, 2015	Jun 23, 2014–Jun 18, 2015
Sample III	1250	Jan 07, 2016–Dec 31, 2020	Jan 03, 2020–Dec 31, 2020

The table shows the number of observations, the start and end dates of each sample period, as well as the forecast window within each sample period. Full sample covers the entire data history. Sample I, II, and III are intended to assess tail-risk forecasts on long positions for the 2008 global financial crisis, the 2014 oil crash, and the COVID-19 pandemic, respectively



Table 2 Summary statistics

	Full	Sample I	Sample II	Sample III
Panel A: Brent				
Mean	0.013	0.040	- 0.017	0.037
Std	2.710	2.526	1.554	3.753
Min	- 64.370	- 16.832	- 8.245	- 64.370
Max	41.202	18.130	8.508	41.202
Skew	- 2.401	0.079	- 0.165	- 3.161
Kurt	97.563	5.495	3.156	89.466
JB	1,722,706.733 (0.000)	1581.864 (0.000)	527.831 (0.000)	420351.567 (0.000)
LM(10)	760.010 (0.000)	79.893 (0.000)	81.993 (0.000)	221.395 (0.000)
LM(20)	939.170	201.394 (0.000)	113.677 (0.000)	274.232 (0.000)
ADF	- 13.928 (0.010)	- 10.938 (0.010)	- 9.471 (0.010)	- 10.818 (0.010)
Panel B: WTI				
Mean	- 0.023	0.044	- 0.014	0.027
Std	3.264	2.803	1.854	4.225
Min	- 72.027	- 12.827	- 11.126	- 72.027
Max	42.583	16.414	9.001	42.583
Skew	- 2.423	0.020	- 0.253	- 3.000
Kurt	94.633	4.288	3.453	87.591
JB	1239137.681 (0.000)	962.837 (0.000)	638.046 (0.000)	402796.481 (0.000)
LM(10)	464.848 (0.000)	233.883 (0.000)	110.869 (0.000)	172.768 (0.000)
LM(20)	652.877 (0.000)	304.248 (0.000)	127.293 (0.000)	243.578 (0.000)
ADF	- 14.098 (0.010)	- 11.326 (0.010)	- 10.035 (0.010)	- 11.820 (0.010)

The table reports the mean, standard deviation (std), minimum (min), maximum (max), skewness (skew), kurtosis (kurt) of the daily returns in percentage terms for Brent and WTI over the full sample and three sub-sample periods. Statistics with p value in brackets are reported for the Jarque–Bera normality test, the ARCH (LM) test for autoregressive conditional heteroskedasticity of order up to 10 and 20, and the augmented Dickey–Fuller (ADF) test for stationarity. The p values for the ADF-test are less than 0.01

covers the oil crisis in later 2008 and late 2014. The third sub-sample corresponds to the recent COVID-19 period. The three sub-sample windows are denoted by dash lines in the Brent and WTI price histories shown in Figs. 1 and 2.

The returns graphs show time-varying volatility with large changes in oil price returns followed by large changes and small changes followed by small changes. During the recent Covid-19 outbreak, both Brent and WTI returns have experienced extremely high volatility, resulting in a series of large positive and negative returns in a short time. Table 2 provides the descriptive statistics of the two returns series over the full sample and sub-sample periods. The daily mean returns are close to zero. Brent and WTI have the largest standard deviation and maximum profit and loss over COVID-19 (sample III) period. The Jarque–Bera statistic rejects the normality assumption because both returns series are highly leptokurtic with negative skewness. The augmented Dickey–Fuller test indicates that both returns series are stationary. The ARCH portmanteau test for serial correlation in squared returns of



Table 3 GARCH estimation

	GARCH-ST			GJRGARCH-ST			FIGARCH-ST	
	Coef	<i>p</i> val		Coef	<i>p</i> val		Coef	<i>p</i> val
Panel A: Brent								
μ	0.036	0.247	μ	0.022	0.467	μ	0.030	0.337
ω	0.054	0.047	ω	0.058	0.092	ω	0.047	0.004
α	0.083	0.000	α	0.040	0.001	α	0.002	0.973
β	0.910	0.000	β	0.912	0.000	β	0.894	0.000
ξ	0.938	0.000	γ	0.075	0.000	δ	0.950	0.000
ν	6.269	0.000	ξ	0.934	0.000	ξ	0.936	0.000
			ν	6.844	0.000	ν	5.954	0.000
LB(10)	12.606	0.247	LB(10)	9.989	0.441	LB(10)	14.150	0.166
LM(10)	8.709	0.560	LM(10)	4.913	0.897	LM(10)	18.198	0.052
LM(20)	21.225	0.384	LM(20)	13.789	0.841	LM(20)	30.271	0.066
Panel B: WTI								
μ	0.049	0.104	μ	0.026	0.382	μ	0.047	0.124
ω	0.103	0.003	ω	0.106	0.003	ω	0.068	0.002
α	0.098	0.000	α	0.041	0.000	α	0.021	0.662
β	0.886	0.000	β	0.891	0.000	β	0.891	0.000
ξ	0.920	0.000	γ	0.098	0.000	δ	0.994	0.000
ν	6.616	0.000	ξ	0.914	0.000	ξ	0.919	0.000
			ν	7.009	0.000	ν	6.063	0.000
LB(10)	6.303	0.789	LB(10)	5.172	0.879	LB(10)	6.382	0.782
LM(10)	10.063	0.435	LM(10)	8.345	0.595	LM(10)	7.546	0.673
LM(20)	16.835	0.664	LM(20)	14.701	0.793	LM(20)	14.734	0.791

The table reports the parameter estimates and *p* values (in brackets) for the GARCH, GJRGARCH, and FIGARCH models with skewed *t* innovations for Brent and WTI returns over the full sample period. The diagnostic statistics for standardized innovations include the Ljung–Box (LB) test for autocorrelation of order up to 10, and the ARCH (LM) test for autoregressive conditional heteroskedasticity of order up to 10 and 20

order up to 20 reveals evidence of significant volatility clustering. Preliminary data analysis indicates that capturing skewness, fat-tailed, and time-varying volatility is critical for tail-risk forecasting.

Table 3 presents the estimation results of conditional volatility models based on standard GARCH, GJRGARCH, or FIGARCH models combined with skewed student innovations over the full sample period. It is clear that the mean returns μ are not statistically significantly different from zero in all the models. The sum of the ARCH and GARCH parameters (i.e., $\alpha + \beta$) is close to 1 which indicates the persistence of volatility. In particular, the leverage parameter γ of the GJRGARCH model is positive and significant, confirming the leverage effect, i.e., negative shocks have a larger impact on the conditional variance than the positive shocks. The long-memory parameter *d* of the FIGARCH model close to 1 suggests shocks decay at a slower hyperbolic rate rather than the exponential rate as in the standard GARCH



Table 4 Volatility forecasts accuracy (QLIKE)

	Full	Sample I	Sample II	Sample III
Panel A: Brent				
GARCH-N	2.400	3.739	3.739	3.827
GARCH-ST	2.404	3.762	3.762	3.860
GJRGARCH-N	2.377	3.696	3.696	3.726
GJRGARCH-ST	2.376	3.706	3.706	3.732
FIGARCH-N	2.409	3.712	3.712	3.934
FIGARCH-ST	2.414	3.759	3.759	3.933
Panel B: WTI				
GARCH-N	2.587	4.023	2.831	4.036
GARCH-ST	2.579	3.987	2.820	3.985
GJRGARCH-N	2.564	4.048	2.814	3.918
GJRGARCH-ST	2.558	4.054	2.804	3.882
FIGARCH-N	2.587	3.978	2.861	4.075
FIGARCH-ST	2.585	3.952	2.849	4.059

The table shows the average values of the QLIKE loss function over various sample forecast periods. Each model's volatility forecast is compared to the daily squared return which serves as a proxy for realized volatility. The MCS test is used to compare forecast accuracy at 90% confidence. The numbers in bold indicate the models that are part of the superior set of models with statistically comparable volatility forecasts

model. The skewed student distribution parameter estimates (ν and ξ) indicate fat tails and negative skewness. Finally, the standardized residuals display neither significant autocorrelation nor any ARCH effect left in all the models. Overall, the estimation results support the use of constant mean and GARCH-type volatility models, as well as skewed and fat tail innovations, for oil VaR and ES modeling.

Results

We now examine out-of-sample forecast performance. To begin, we assess the accuracy of volatility forecasting using three GARCH-type models (GARCH, GJR-GARCH, and FIGARCH) with normal and skewed t innovations.⁸ The daily squared return serves as a benchmark for comparison. The average values of the QLIKE loss function are reported in Table 4, with lower values indicating more precise volatility forecasts. The GJRGARCH model produces the most accurate volatility forecasts for both Brent and WTI, and it appears most frequently in the superior set of models based on the MCS test at 90% confidence. The differences between normal and skewed t innovations under each type of GARCH model are less evident.

⁸ The parameters of GARCH models are estimated using quasi-maximum likelihood estimation (QMLE) for innovations based on historical simulation or Cornish Fisher expansion.



The VaR and ES forecast performance is then evaluated at 97.5% and 99% confidence levels using a set of criteria that includes the AE ratios and the p values of LR_{uc} , LR_{cc} , DQ, ER, and the Intercept ESR tests. A p value of 0.05 or less is interpreted as evidence for rejecting the null hypothesis. Furthermore, the maximum absolute deviation (ADMax), as well as the tick and FZL functions are employed to compare forecast accuracy across different models.⁹ The MCS tests are run at a 90% confidence on the tick and FZL mean values, respectively. Tables 5 and 6 report the means and standard deviations of VaR and ES forecasts, as well as their backtesting performance over the full out-of-sample forecast periods for long positions in Brent and WTI, respectively. Tables 7, 8, 9, and 10 show the breakdown of performance over three sub-sample periods. Given the limited number of observations for forecast evaluation in sub-samples, which may impact test significance, we do not rely on p values tests but instead use the AE, ADMax, and the values of loss functions as indicators. The main findings for the long positions are summarized below.

First, the standard deviations of VaR and ES forecasts increase with the confidence level, and the standard deviations of ES are greater than those of VaR at the same level of confidence. This implies that as confidence levels rise, forecasts for oil tail-risk become more variable. Moreover, the COVID-19 (sample III) period has the highest mean and standard deviations of VaR and ES forecasts, followed by the 2008 GFC (sample II) period. When comparing the COVID-19 results to the GFC period, the means of VaR and ES forecasts are roughly 50% higher, with standard deviations nearly five times higher. This implies that COVID-19 has caused more uncertainty and has had a greater impact on the oil market than the GFC. VaR and ES forecasts based on the GARCH-type models are effective to track market movements via the volatility filtering process, but their relative performance requires careful examination. Furthermore, the maximum absolute deviation is significantly higher during the COVID-19 than during the GFC periods, with the increase being greater for the WTI market than the Brent market. Therefore, the model performance comparison should consider not only the frequency but also the magnitude of VaR exceptions. This is especially important during the crisis, given the AE ratio is significantly greater than one.

Second, when the performance of innovation distributions within each type of GARCH model is compared over the entire out-of-sample period, normal innovations perform the worst, with the AE ratio significantly greater than the others and the p value tests of VaR and ES being rejected in the majority of cases, indicating that the models are misspecified. The historical simulation approach is less accurate than the skewed t and the Cornish–Fisher expansion in predicting VaR at a 99% confidence level. Based on the tick and FZL functions evaluation, the skewed t produces the most accurate joint VaR and ES forecasts. Moreover, it is worth noting that the Cornish–Fisher expansion consistently produces the most conservative tail-risk forecasts even during the sub-sample stressed periods. It has the highest mean and standard deviation of VaR and ES forecasts, the lowest

⁹ The MaxAD measures the maximum absolute deviation of the violations from the VaR forecasts (McAleer and Medeiros 2008).



Table 5 Brent (long) VaR and ES results: full sample

	GARCH				GJR-GARCH				FIGARCH			
	N	ST	HS	CF	N	ST	HS	CF	N	ST	HS	CF
Panel A: Confidence level 97.5%												
mVaR	-4.350	-4.539	-4.546	-4.799	-4.328	-4.537	-4.446	-4.671	-4.410	-4.667	-4.590	-4.850
stdVaR	3.461	3.396	3.867	4.074	3.448	3.474	3.554	3.928	3.513	3.629	3.831	4.077
mES	-5.199	-6.022	-5.966	-6.367	-5.163	-5.914	-5.731	-6.035	-5.269	-6.268	-6.028	-6.410
stdES	4.128	4.695	5.124	5.699	4.112	4.725	4.793	5.332	4.191	5.087	5.187	5.631
AE	1.346	1.081	1.202	0.949	1.238	1.033	1.129	0.949	1.406	1.069	1.214	0.997
LR _{hc}	0.002	0.458	0.071	0.636	0.034	0.759	0.241	0.636	0.000	0.526	0.056	0.980
LR _{cc}	0.008	0.711	0.196	0.661	0.096	0.943	0.367	0.890	0.002	0.756	0.161	0.823
DQ	0.034	0.434	0.337	0.651	0.109	0.569	0.188	0.710	0.004	0.636	0.397	0.668
ER	0.000	0.006	0.001	0.012	0.001	0.008	0.008	0.001	0.001	0.004	0.000	0.004
ESR	0.000	0.028	0.024	0.205	0.000	0.025	0.008	0.060	0.000	0.062	0.051	0.180
ADMMax	41.615	40.529	39.095	37.985	41.556	39.707	40.570	38.991	39.578	37.956	37.623	36.017
Tick	0.171	0.171	0.170	0.171	0.166	0.165	0.165	0.165	0.172	0.172	0.171	0.172
FZL	1.786	1.751	1.754	1.754	1.736	1.708	1.720	1.714	1.797	1.761	1.765	1.762
Panel B: Confidence level 99%												
mVaR	-5.173	-5.825	-5.753	-6.281	-5.138	-5.753	-5.574	-5.979	-5.243	-6.036	-5.815	-6.328
stdVaR	4.108	4.480	4.976	5.576	4.092	4.537	4.787	5.240	4.170	4.829	5.051	5.520
mES	-5.934	-7.434	-7.282	-7.904	-5.888	-7.210	-6.937	-7.360	-6.014	-7.809	-7.321	-7.937
stdES	4.706	5.976	6.376	7.321	4.687	5.942	6.047	6.723	4.778	6.543	6.404	7.179
AE	1.892	1.232	1.262	0.961	1.532	1.021	1.202	1.081	1.892	1.232	1.292	0.961
LR _{hc}	0.000	0.195	0.145	0.821	0.004	0.902	0.257	0.641	0.000	0.195	0.106	0.821
LR _{cc}	0.000	0.356	0.291	0.590	0.008	0.654	0.422	0.637	0.000	0.356	0.233	0.590
DQ	0.000	0.508	0.463	0.568	0.009	0.632	0.543	0.648	0.000	0.491	0.411	0.568
ER	0.003	0.014	0.002	0.000	0.006	0.001	0.001	0.000	0.002	0.070	0.000	0.000



Table 5 (continued)

	GARCH			GJR-GARCH			FIGARCH					
	N	ST	HS	CF	N	ST	HS	CF	N	ST	HS	CF
ESR	0.000	0.050	0.006	0.052	0.032	0.166	0.093	0.159	0.000	0.188	0.013	0.083
ADMax	37.346	33.032	31.178	28.699	37.294	32.344	34.626	31.165	34.930	29.366	28.886	26.390
Tick	0.095	0.092	0.092	0.092	0.090	0.088	0.089	0.088	0.095	0.092	0.092	0.092
FZL	2.124	2.007	2.027	2.015	2.038	1.948	1.981	1.960	2.141	2.010	2.043	2.026

The table shows the means and standard deviations of VaR and ES forecasts at 97.5% (Panel A) and 99% (Panel B) confidence levels for long positions in Brent, as well as the backtest results for the three GARCH-type models combined with normal(N), skewed t (ST), historical simulation (HS) or Cornish-Fisher expansion (CF) over the full forecast period. The AE ratio, the p values of the LR_{opt}, LR_{CC}, DQ, one-sided ER, one-sided Intercept ESR tests, as well as the absolute maximum deviation (ADMax), and the mean values of the tick and FZL functions are all reported. The numbers in bold indicate the models that are part of the superior set of models with statistically comparable VaR forecasts based on tick loss function or joint VaR and ES forecasts based on the FZL function at 90% confidence.



Table 6 WTI (long) VaR and ES results: full sample

	GARCH				GJR-GARCH				FIGARCH			
	N	ST	HS	CF	N	ST	HS	CF	N	ST	HS	CF
Panel A: Confidence level 97.5%												
mVaR	-4.808	-4.964	-5.007	-5.558	-4.826	-5.017	-5.067	-5.525	-4.840	-5.102	-4.991	-5.563
stdVaR	4.086	3.966	4.328	5.272	4.251	4.162	4.594	5.332	4.151	4.199	4.344	5.214
mES	-5.747	-6.583	-6.740	-7.894	-5.759	-6.617	-6.665	-7.738	-5.786	-6.854	-6.764	-7.841
stdES	4.874	5.593	6.374	8.308	5.070	5.870	6.438	8.207	4.951	6.032	6.393	8.067
AE	1.209	1.149	1.161	0.931	1.233	1.137	1.112	0.967	1.306	1.100	1.161	0.907
LR _{hc}	0.062	0.181	0.148	0.521	0.038	0.218	0.309	0.762	0.007	0.363	0.148	0.384
LR _{cc}	0.098	0.403	0.275	0.804	0.116	0.228	0.574	0.717	0.020	0.629	0.275	0.574
DQ	0.186	0.514	0.419	0.650	0.095	0.272	0.524	0.696	0.005	0.665	0.238	0.575
ER	0.008	0.024	0.002	0.301	0.009	0.064	0.002	0.442	0.000	0.023	0.005	0.059
ESR	0.002	0.028	0.164	0.639	0.006	0.086	0.169	0.571	0.001	0.115	0.111	0.620
ADM _{max}	53.791	51.023	52.579	49.488	52.228	49.797	50.293	48.351	51.085	50.592	50.323	46.438
Tick	0.187	0.186	0.189	0.192	0.187	0.185	0.188	0.189	0.188	0.187	0.189	0.191
FZL	1.900	1.854	1.872	1.872	1.876	1.836	1.852	1.849	1.910	1.859	1.880	1.877
Panel B: Confidence level 99%												
mVaR	-5.718	-6.367	-6.520	-7.700	-5.731	-6.411	-6.427	-7.563	-5.757	-6.594	-6.430	-7.658
stdVaR	4.850	5.300	6.039	7.976	5.045	5.567	5.929	7.905	4.927	5.673	6.089	7.763
mES	-6.560	-8.128	-8.345	-10.225	-6.568	-8.137	-8.165	-9.942	-6.605	-8.545	-8.343	-10.111
stdES	5.557	7.223	8.412	11.410	5.780	7.573	8.287	11.139	5.645	7.902	8.359	10.985
AE	1.844	1.330	1.391	0.756	1.935	1.270	1.451	0.816	1.935	1.179	1.421	0.907
LR _{hc}	0.000	0.069	0.033	0.140	0.000	0.135	0.015	0.272	0.000	0.314	0.022	0.585
LR _{cc}	0.000	0.170	0.094	0.139	0.000	0.276	0.048	0.257	0.000	0.472	0.068	0.477
DQ	0.000	0.300	0.184	0.182	0.000	0.411	0.098	0.555	0.000	0.589	0.137	0.497
ER	0.008	0.008	0.016	0.004	0.009	0.025	0.120	0.008	0.008	0.002	0.017	0.007



Table 6 (continued)

	GARCH				GJR-GARCH				FIGARCH			
	N	ST	HS	CF	N	ST	HS	CF	N	ST	HS	CF
ESR	0.053	0.014	0.026	0.473	0.000	0.011	0.052	0.450	0.052	0.173	0.037	0.451
ADMax	50.368	44,008	45,213	38,506	48,527	42,485	44,092	37,694	47,156	43,199	41,453	34,432
Tick	0.105	0.101	0.103	0.105	0.104	0.099	0.100	0.102	0.105	0.102	0.103	0.104
FZL	2.269	2.119	2.148	2.145	2.215	2.083	2.101	2.100	2.286	2.127	2.172	2.152

The table shows the means and standard deviations of VaR and ES forecasts at 97.5% (Panel A) and 99% (Panel B) confidence levels for long positions in WTI, as well as the backtest results for the three GARCH-type models combined with normal(N), skewed t (ST), historical simulation (HS) or Cornish-Fisher expansion (CF) over the full forecast period. The AE ratio, the p values of the LR_{nc}, LR_{cc}, DQ, one-sided ER, one-sided Intercept ESR tests, as well as the absolute maximum deviation (ADMMax) and the mean values of the tick and FZL functions are all reported. The numbers in bold indicate the models that are part of the superior set of models with statistically comparable VaR forecasts based on tick loss function or joint VaR and ES forecasts based on the FZL function at 90% confidence



Table 7 Brent (long) VaR and ES 97.5% confidence results: sub-sample

	GARCH				GJR-GARCH				FIGARCH			
	N	ST	HS	CF	N	ST	HS	CF	N	ST	HS	CF
Panel A: Sample I												
mVaR	-7.024	-7.120	-7.159	-7.283	-6.877	-7.048	-7.035	-7.167	-7.218	-7.350	-7.422	-7.462
stdVaR	2.325	2.415	2.331	2.440	2.221	2.210	2.214	2.339	2.298	2.447	2.320	2.399
mES	-8.397	-9.132	-8.894	-9.283	-8.216	-8.966	-8.698	-9.058	-8.628	-9.432	-9.104	-9.453
stdES	2.771	3.091	3.029	3.098	2.646	2.826	2.892	2.976	2.739	3.107	2.986	3.030
AE	1.920	1.920	1.920	1.920	1.760	1.760	1.600	1.600	1.920	1.760	1.600	1.760
ADMax	9.381	9.192	9.207	9.141	9.578	9.308	9.392	9.264	9.105	8.897	8.856	8.871
Tick	0.283	0.284	0.279	0.280	0.278	0.275	0.275	0.275	0.277	0.281	0.274	0.275
FZL	2.583	2.550	2.535	2.523	2.539	2.479	2.495	2.475	2.511	2.521	2.469	2.465
Panel B: Sample II												
mVaR	-3.548	-3.753	-3.796	-4.049	-3.599	-3.791	-3.860	-4.011	-3.576	-3.817	-3.773	-4.079
stdVaR	1.503	1.567	1.623	1.786	1.361	1.403	1.501	1.569	1.552	1.671	1.637	1.844
mES	-4.226	-4.914	-5.072	-5.369	-4.283	-4.899	-4.979	-5.210	-4.260	-5.028	-5.094	-5.402
stdES	1.787	2.084	2.265	2.428	1.618	1.822	1.965	2.073	1.846	2.231	2.335	2.506
AE	2.080	1.760	1.920	1.120	1.760	1.120	1.120	0.960	2.400	1.600	1.920	1.440
ADMax	4.567	4.387	4.398	4.233	4.316	4.129	4.067	3.985	4.519	4.319	4.390	4.183
Tick	0.144	0.141	0.142	0.143	0.135	0.134	0.136	0.136	0.144	0.141	0.141	0.143
FZL	1.928	1.808	1.822	1.792	1.780	1.711	1.737	1.712	1.935	1.804	1.826	1.793
Panel C: Sample III												
mVaR	-9.253	-9.446	-10.286	-10.940	-9.111	-9.728	-9.517	-10.410	-9.603	-10.349	-10.371	-11.205
stdVaR	10.047	9.687	11.350	11.957	10.249	10.112	10.519	11.749	9.989	10.149	11.007	11.719
mES	-11.054	-13.147	-13.774	-15.305	-10.868	-13.287	-12.767	-14.118	-11.472	-14.785	-14.202	-15.508
stdES	11.983	13.472	15.016	16.804	12.222	13.800	14.294	15.989	11.914	14.192	14.943	16.268
AE	2.080	1.920	1.600	1.440	1.920	1.600	1.760	1.440	1.920	1.600	1.440	1.440



Table 7 (continued)

	GARCH			GJRGARCH			FIGARCH					
	N	ST	HS	CF	N	ST	HS	CF	N	ST	HS	CF
ADMax	41.615	40.529	39.095	37.985	41.556	39.707	40.570	38.991	39.578	37.956	37.623	36.017
Tick	0.595	0.589	0.585	0.581	0.565	0.555	0.552	0.554	0.590	0.595	0.583	0.584
FZL	3.181	2.994	2.992	2.961	3.017	2.830	2.874	2.854	3.221	3.019	3.052	3.029

The table shows the means and standard deviations of VaR and ES forecasts at 97.5% confidence levels for long positions in Brent, as well as the backtest results for the three GARCH-type models combined with normal(N), skewed t (ST), historical simulation (HS) or Cornish–Fisher expansion (CF) over the sub-sample forecast periods. The AE ratio, the absolute maximum deviation (ADMax), and the mean values of the tick and FZL functions are reported



Table 8 Brent (long) VaR and ES 99% confidence results: sub-sample

	GARCH				GJR-GARCH				FIGARCH			
	N	ST	HS	CF	N	ST	HS	CF	N	ST	HS	CF
Panel A: Sample I												
mVaR	-8.356	-8.928	-8.600	-9.212	-8.175	-8.788	-8.336	-9.004	-8.585	-9.219	-8.900	-9.391
stdVaR	2.758	3.021	2.966	3.077	2.633	2.764	2.788	2.954	2.726	3.043	2.863	3.011
mES	-9.588	-10.999	-10.612	-11.222	-9.376	-10.732	-10.185	-10.883	-9.849	-11.365	-10.728	-11.377
stdES	3.158	3.720	3.709	3.736	3.015	3.398	3.469	3.595	3.122	3.717	3.628	3.640
AE	2.000	2.000	2.000	2.000	2.400	2.000	2.000	2.000	2.000	2.000	2.000	2.000
ADMax	7.969	7.319	7.784	7.236	8.210	7.497	8.229	7.367	7.644	6.963	7.379	6.926
Tick	0.158	0.156	0.159	0.155	0.154	0.149	0.156	0.148	0.152	0.151	0.151	0.149
FZL	3.057	2.923	2.996	2.890	2.998	2.828	2.962	2.817	2.923	2.839	2.856	2.785
Panel B: Sample II												
mVaR	-4.206	-4.776	-4.802	-5.300	-4.263	-4.779	-4.636	-5.158	-4.240	-4.877	-4.774	-5.333
stdVaR	1.779	2.016	2.018	2.388	1.610	1.775	1.729	2.047	1.837	2.155	2.050	2.464
mES	-4.815	-6.010	-6.148	-6.659	-4.876	-5.932	-5.939	-6.374	-4.853	-6.174	-6.172	-6.695
stdES	2.034	2.577	2.808	3.059	1.840	2.216	2.421	2.566	2.101	2.765	2.903	3.158
AE	2.800	1.600	1.600	1.200	2.000	1.600	2.000	1.600	2.800	1.600	1.600	1.200
ADMax	4.048	3.590	3.542	3.330	3.754	3.270	3.393	3.023	3.991	3.486	3.506	3.273
Tick	0.082	0.078	0.077	0.078	0.076	0.072	0.073	0.072	0.081	0.077	0.077	0.078
FZL	2.483	2.176	2.189	2.137	2.252	2.025	2.077	2.006	2.452	2.147	2.198	2.133
Panel C: Sample III												
mVaR	-10.999	-12.529	-13.483	-14.978	-10.815	-12.747	-12.833	-13.880	-11.415	-13.950	-13.920	-15.201
stdVaR	11.925	12.836	14.591	16.430	12.163	13.241	14.297	15.710	11.855	13.495	14.548	15.935
mES	-12.615	-16.777	-17.059	-19.633	-12.391	-16.733	-15.992	-17.772	-13.091	-19.209	-17.517	-19.767
stdES	13.661	17.190	18.762	21.623	13.933	17.374	18.156	20.177	13.581	18.180	18.495	20.779
AE	3.600	2.800	2.400	2.000	3.200	2.000	2.400	2.400	3.200	2.800	2.400	2.000



Table 8 (continued)

	GARCH				GJRGARCH				FIGARCH			
	N	ST	HS	CF	N	ST	HS	CF	N	ST	HS	CF
ADMax	37.346	33.032	31.178	28.699	37.294	32.344	34.626	31.165	34.930	29.366	28.886	26.390
Tick	0.404	0.376	0.365	0.360	0.379	0.347	0.363	0.354	0.402	0.372	0.362	0.358
FZL	4.298	3.668	3.724	3.579	4.035	3.390	3.571	3.431	4.426	3.652	3.858	3.695

The table shows the means and standard deviations of VaR and ES forecasts at 99% confidence levels for long positions in Brent, as well as the backtest results for the three GARCH-type models combined with normal(N), skewed t (ST), historical simulation (HS) or Cornish–Fisher expansion (CF) over the sub-sample forecast periods. The AE ratio, the absolute maximum deviation (ADMax), and the mean values of the tick and FZL functions are reported



Table 9 WTI (long) VaR and ES 97.5% confidence results: sub-sample

	GARCH				GJR-GARCH				FIGARCH			
	N	ST	HS	CF	N	ST	HS	CF	N	ST	HS	CF
Panel A: Sample I												
mVaR	-7.948	-8.121	-7.753	-8.647	-7.881	-8.121	-7.710	-8.683	-8.382	-8.849	-8.313	-8.856
stdVaR	2.976	2.966	2.911	3.409	2.881	2.847	2.798	3.363	2.936	3.051	2.960	3.204
mES	-9.501	-10.201	-9.866	-11.702	-9.417	-10.217	-9.854	-11.863	-10.020	-11.114	-10.303	-11.515
stdES	3.548	3.771	3.730	5.042	3.435	3.644	3.641	5.066	3.500	3.879	3.630	4.445
AE	1.760	1.440	2.080	1.280	1.920	2.240	2.080	1.600	1.760	1.120	1.760	1.120
ADMax	3.333	2.755	3.732	2.928	4.719	7.050	5.236	3.345	2.706	2.372	2.827	1.958
Tick	0.259	0.253	0.266	0.246	0.276	0.276	0.286	0.259	0.256	0.245	0.260	0.244
FZL	2.365	2.316	2.383	2.280	2.453	2.456	2.479	2.331	2.326	2.268	2.333	2.267
Panel B: Sample II												
mVaR	-4.541	-4.911	-5.126	-5.540	-4.622	-5.030	-4.984	-5.568	-4.636	-5.052	-5.080	-5.692
stdVaR	1.751	1.824	2.005	2.290	1.600	1.703	1.711	2.051	1.817	1.990	2.040	2.390
mES	-5.417	-6.636	-6.877	-8.082	-5.508	-6.748	-6.815	-8.001	-5.531	-6.893	-6.976	-8.339
stdES	2.085	2.469	2.795	3.400	1.904	2.303	2.488	3.019	2.163	2.708	2.900	3.562
AE	1.440	1.280	1.280	1.280	1.440	1.440	1.440	1.280	1.440	1.280	1.280	1.120
ADMax	7.841	7.578	7.488	7.281	7.530	7.231	7.237	6.899	7.863	7.574	7.658	7.280
Tick	0.186	0.185	0.188	0.191	0.193	0.190	0.192	0.193	0.188	0.186	0.190	0.196
FZL	2.342	2.172	2.216	2.193	2.333	2.181	2.220	2.165	2.397	2.218	2.278	2.249
Panel C: Sample III												
mVaR	-10.702	-10.467	-11.436	-13.889	-11.163	-11.027	-12.163	-14.093	-10.933	-11.447	-11.523	-13.923
stdVaR	11.787	11.376	12.563	15.238	12.446	12.132	13.608	15.541	11.708	11.578	12.308	14.782
mES	-12.780	-14.957	-16.867	-21.930	-13.315	-15.653	-17.037	-21.804	-13.056	-16.923	-17.116	-21.698
stdES	14.059	16.174	18.616	23.892	14.844	17.213	18.996	23.641	13.965	16.654	18.335	22.824
AE	1.760	2.080	1.600	1.280	1.760	1.760	1.600	1.280	1.920	1.600	1.600	1.280



Table 9 (continued)

	GARCH			GJRGARCH			FIGARCH					
	N	ST	HS	CF	N	ST	HS	CF	N	ST	HS	CF
ADMax	53.791	51.023	52.579	49.488	52.228	49.797	50.293	48.351	51.085	50.592	50.323	46.438
Tick	0.671	0.650	0.669	0.691	0.653	0.635	0.650	0.671	0.668	0.669	0.669	0.682
FZL	3.375	3.050	3.134	3.059	3.181	2.926	2.983	2.983	3.367	3.101	3.164	3.060

The table shows the means and standard deviations of VaR and ES forecasts at 97.5% confidence levels for long positions in WTI, as well as the backtest results for the three GARCH-type models combined with normal(N), skewed t (ST), historical simulation (HS) or Cornish–Fisher expansion (CF) over the sub-sample forecast periods. The AE ratio, the absolute maximum deviation (ADMax), and the mean values of the tick and FZL functions are reported



Table 10 WTI (long) VaR and ES 99% confidence results: sub-sample

	GARCH				GJR-GARCH				FIGARCH			
	N	ST	HS	CF	N	ST	HS	CF	N	ST	HS	CF
Panel A: Sample I												
mVaR	-9.453	-10.028	-9.570	-11.497	-9.370	-10.040	-9.483	-11.638	-9.970	-10.923	-9.762	-11.390
stdVaR	3.530	3.691	3.545	4.877	3.418	3.561	3.502	4.887	3.483	3.796	3.441	4.344
mES	-10.846	-12.100	-11.841	-14.724	-10.747	-12.133	-11.898	-15.014	-11.439	-13.182	-12.112	-14.111
stdES	4.043	4.520	4.566	6.740	3.914	4.390	4.471	6.837	3.990	4.654	4.287	5.719
AE	2.800	2.000	2.400	0.400	3.200	1.200	3.200	0.400	2.000	0.400	2.400	0.400
ADMax	1.871	1.161	1.820	1.578	3.188	5.560	2.946	1.718	1.124	0.210	1.141	0.466
Tick	0.116	0.106	0.114	0.118	0.128	0.128	0.124	0.120	0.106	0.107	0.108	0.113
FZL	2.474	2.351	2.451	2.442	2.605	2.613	2.543	2.464	2.341	2.348	2.372	2.393
Panel B: Sample II												
mVaR	-5.390	-6.389	-6.759	-7.854	-5.482	-6.512	-6.726	-7.793	-5.504	-6.615	-6.609	-8.099
stdVaR	2.075	2.376	2.686	3.299	1.895	2.216	2.352	2.931	2.152	2.600	2.737	3.453
mES	-6.176	-8.295	-8.515	-10.629	-6.276	-8.392	-8.470	-10.432	-6.307	-8.675	-8.670	-10.991
stdES	2.375	3.093	3.571	4.518	2.168	2.883	3.199	3.997	2.463	3.405	3.739	4.744
AE	3.200	1.600	1.600	1.200	3.200	2.000	2.000	1.200	2.800	1.600	1.600	1.200
ADMax	7.229	6.478	6.325	5.701	6.862	6.044	5.949	5.165	7.253	6.429	6.540	5.675
Tick	0.110	0.105	0.108	0.112	0.112	0.104	0.104	0.108	0.111	0.108	0.112	0.115
FZL	3.123	2.626	2.728	2.601	3.038	2.585	2.644	2.539	3.226	2.705	2.886	2.684
Panel C: Sample III												
mVaR	-12.717	-14.127	-16.059	-21.055	-13.250	-14.824	-15.718	-20.990	-12.992	-15.769	-16.397	-20.869
stdVaR	13.991	15.300	17.666	22.960	14.771	16.303	17.540	22.818	13.896	15.691	17.537	21.979
mES	-14.580	-19.422	-22.142	-30.085	-15.181	-20.235	-21.820	-29.611	-14.895	-22.477	-22.307	-29.574
stdES	16.028	20.941	24.575	32.667	16.921	22.248	24.393	31.830	15.920	21.736	24.029	30.969
AE	3.600	2.400	2.400	1.600	2.800	2.400	2.400	1.600	3.200	2.000	2.400	1.200



Table 10 (continued)

	GARCH				GJRGARCH				FIGARCH			
	N	ST	HS	CF	N	ST	HS	CF	N	ST	HS	CF
ADMax	50.368	44.008	45.213	38.506	48.527	42.485	44.092	37.694	47.156	43.199	41.453	34.432
Tick	0.481	0.443	0.452	0.446	0.455	0.421	0.424	0.439	0.471	0.448	0.440	0.431
FZL	4.651	3.761	3.846	3.606	4.252	3.559	3.654	3.567	4.607	3.802	3.844	3.627

The table shows the means and standard deviations of VaR and ES forecasts at 99% confidence levels for long positions in WTI, as well as the backtest results for the three GARCH-type models combined with normal(N), skewed t (ST), historical simulation (HS) or Cornish-Fisher expansion (CF) over the sub-sample forecast periods. The AE ratio, the absolute maximum deviation (ADMax), and the mean values of the tick and FZL functions are reported



AE ratio, and the smallest maximum absolute deviations. It passes the one-sided Incept ESR tests at the 5% significance level in all cases.

Third, the differences between various GARCH-type models are investigated. Under the same type of innovation distribution, the GJRGARCH outperforms the standard GARCH and FIGARCH models in VaR and ES forecasts based on the tick and FZL functions evaluation. This is supported by the findings from the three sub-sample periods. The results are consistent with the QLIKE test reported in Table 4, which shows that the GJRGARCH model produces the most accurate volatility forecasts. Indeed, GJRGARCH combined with skewed t innovation outperforms all others in joint VaR and ES forecasts for both Brent and WTI. It is interesting to note that the FIGARCH model has a lower maximum absolute deviation than the GARCH and GJRGARCH models, which can be seen during both the GFC and COVID-19 stressed periods.

Next, we evaluate the VaR and ES forecasts for short positions. The results for Brent and WTI over the full out-of-sample forecast periods are presented in Tables 11 and 12, respectively. The results are generally in line with the long positions. In particular, the GJRGARCH model combined with skewed t innovations outperforms the others in joint VaR and ES forecasts based on the tick and FZL functions evaluation. The Cornish–Fisher expansion consistently produces the highest VaR and ES forecasts with the most variability when compared to other innovation distributions. In contrast to the findings for the long positions, the Cornish–Fisher expansion overestimates VaR forecasts for the short positions in WTI at the 99% confidence level, with a significantly low AE ratio and high tick and FZL function values. This highlights the limitation of the Cornish–Fisher expansion in considering both long and short trading positions for VaR and ES forecasts.

Finally, the relative magnitudes of VaR and ES forecasts are compared. We first follow Patra (2021) to examine ES to VaR ratio at the 99% confidence level. Table 13 reports the average of the ratio over the entire out-of-sample period and the maximum values of the ratio during each sub-sample evaluation period for Brent and WTI, respectively. The ratio is constant at 1.15 for the normal distribution of innovations, but it is significantly higher for the other non-normal distributions and varies across models and time. The highest values are observed during the COVID-19 period. For example, the ES is 30 to 40% higher than the VaR predicted by the GJRGARCH model with skewed t innovation. Furthermore, we investigate the impact of shifting risk prediction from VaR (99% confidence level) to ES (97.5% confidence level). The average of the ES to VaR ratio over the entire out-of-sample period and the maximum values of the ratio during each sub-sample evaluation period for Brent and WTI are shown in Table 14. The ratio is always greater than one, indicating that switching from VaR to ES will increase risk prediction. The increase under normal innovation is negligible, whereas the increase under historical simulation could be more than 20% for short positions in WTI during COVID-19 period. It is worth noting that the Cornish–Fisher expansion produces a more stable increase when moving from VaR to ES compared to the skewed t and historical simulation.



Table 11 Brent (short) VaR and ES results: full sample

	GARCH				GJR-GARCH				FIGARCH			
	N	ST	HS	CF	N	ST	HS	CF	N	ST	HS	CF
Panel A: Confidence level 97.5%												
mVaR	4.452	4.323	4.321	4.421	4.345	4.252	4.227	4.316	4.504	4.426	4.299	4.473
stdVaR	3.465	3.074	3.302	3.358	3.440	3.079	3.245	3.364	3.517	3.282	3.320	3.424
mES	5.300	5.621	5.494	5.706	5.181	5.388	5.355	5.460	5.363	5.818	5.457	5.755
stdES	4.133	4.130	4.207	4.530	4.104	4.032	4.200	4.419	4.194	4.459	4.259	4.567
AE	0.877	0.937	0.961	0.889	0.937	0.997	1.009	0.949	0.829	0.889	0.997	0.901
LR _{uc}	0.246	0.558	0.719	0.297	0.558	0.980	0.932	0.636	0.104	0.297	0.980	0.353
LR _{cc}	0.306	0.604	0.714	0.362	0.035	0.020	0.023	0.133	0.052	0.162	0.471	0.194
DQ	0.389	0.821	0.588	0.427	0.027	0.013	0.016	0.141	0.097	0.361	0.411	0.251
ER	0.000	0.626	0.133	0.393	0.002	0.197	0.177	0.189	0.000	0.591	0.039	0.259
ESR	0.095	0.474	0.271	0.562	0.103	0.190	0.199	0.331	0.110	0.572	0.169	0.544
ADM _{max}	8.878	11.634	9.223	8.905	8.421	8.881	8.653	8.392	11.279	13.442	12.489	11.871
Tick	0.142	0.142	0.142	0.142	0.139	0.140	0.139	0.140	0.143	0.145	0.143	0.143
FZL	1.592	1.594	1.593	1.595	1.582	1.584	1.580	1.587	1.595	1.604	1.593	1.599
Panel B: Confidence level 99%												
mVaR	5.275	5.454	5.377	5.638	5.156	5.306	5.245	5.417	5.337	5.623	5.429	5.690
stdVaR	4.113	3.961	4.104	4.437	4.084	3.915	3.975	4.351	4.174	4.259	4.213	4.483
mES	6.035	6.855	6.511	6.972	5.906	6.488	6.368	6.575	6.108	7.153	6.485	7.016
stdES	4.711	5.164	5.023	5.710	4.679	4.970	5.089	5.471	4.781	5.626	5.136	5.714
AE	1.262	1.021	1.081	0.901	1.172	1.081	1.232	1.021	1.232	0.931	1.111	0.961
LR _{uc}	0.145	0.902	0.641	0.560	0.333	0.641	0.195	0.902	0.195	0.687	0.526	0.821
LR _{cc}	0.100	0.144	0.158	0.078	0.488	0.158	0.116	0.144	0.116	0.096	0.158	0.114
DQ	0.087	0.096	0.113	0.045	0.564	0.097	0.098	0.075	0.098	0.058	0.116	0.071
ER	0.004	0.731	0.269	0.727	0.001	0.512	0.498	0.320	0.000	0.423	0.068	0.512



Table 11 (continued)

	GARCH				GJR-GARCH				FIGARCH			
	N	ST	HS	CF	N	ST	HS	CF	N	ST	HS	CF
ESR	0.074	0.504	0.358	0.607	0.023	0.218	0.187	0.321	0.065	0.499	0.276	0.490
ADMax	7.165	6.735	6.924	6.447	6.616	6.723	6.784	6.012	7.764	8.665	7.411	6,466
Tick	0.068	0.069	0.069	0.068	0.067	0.067	0.067	0.067	0.070	0.071	0.070	0.070
FZL	1.781	1.783	1.784	1.784	1.770	1.771	1.774	1.773	1.790	1.802	1.792	1.793

The table shows the means and standard deviations of VaR and ES forecasts at 97.5% (Panel A) and 99% (Panel B) confidence levels for short positions in Brent, as well as the backtest results for the three GARCH-type models combined with normal(N), skewed t (ST), historical simulation (HS) or Cornish-Fisher expansion (CF) over the full forecast period. The AE ratio, the p values of the LR_{nc}, LR_{cc}, DQ, one-sided ER, one-sided Intercept ESR tests, as well as the absolute maximum deviation (ADMax) and the mean values of the tick and FZL functions are all reported. The numbers in bold indicate the models that are part of the superior set of models with statistically comparable VaR forecasts based on tick loss function or joint VaR and ES forecasts based on the FZL function at 90% confidence



Table 12 WTI (short) VaR and ES results: full sample

	GARCH				GJR-GARCH				FIGARCH			
	N	ST	HS	CF	N	ST	HS	CF	N	ST	HS	CF
	Panel A: Confidence level 97.5%											
mVaR	4.926	4.676	4.653	5.009	4.855	4.621	4.631	4.922	4.970	4.796	4.658	5.043
stdVaR	4.090	3.496	3.630	4.168	4.248	3.575	3.788	4.345	4.153	3.677	3.695	4.262
mES	5.865	6.042	5.889	6.951	5.788	5.942	5.832	6.763	5.916	6.255	5.817	6.938
stdES	4.878	4.716	4.827	6.480	5.067	4.799	5.099	6.586	4.953	5.017	4.939	6.481
AE	0.689	0.871	0.919	0.593	0.738	0.834	0.907	0.726	0.713	0.871	0.979	0.653
LR_uc	0.002	0.223	0.449	0.000	0.011	0.116	0.384	0.008	0.005	0.223	0.849	0.001
LR_cc	0.010	0.117	0.617	0.000	0.004	0.057	0.207	0.003	0.021	0.421	0.719	0.003
DQ	0.002	0.043	0.275	0.000	0.001	0.004	0.090	0.001	0.017	0.260	0.588	0.002
ER	0.001	0.521	0.096	0.719	0.002	0.280	0.114	0.838	0.001	0.496	0.047	0.779
ESR	0.301	0.537	0.315	0.946	0.305	0.499	0.325	0.910	0.340	0.715	0.299	0.937
ADMax	11.358	14.591	15.024	10.956	11.206	12.123	11.171	11.300	10.948	14.541	14.453	10.374
Tick	0.153	0.151	0.151	0.154	0.151	0.149	0.149	0.153	0.153	0.153	0.151	0.155
FZL	1.667	1.653	1.657	1.692	1.648	1.638	1.640	1.673	1.656	1.660	1.649	1.681
Panel B: Confidence level 99%												
mVaR	5.836	5.868	5.384	6.782	5.760	5.781	5.460	6.611	5.887	6.052	5.414	6.778
stdVaR	4.854	4.510	4.224	6.198	5.043	4.595	4.487	6.327	4.929	4.774	4.327	6.222
mES	6.678	7.336	7.161	8.907	6.597	7.189	7.036	8.611	6.735	7.652	7.059	8.841
stdES	5.561	5.925	6.104	8.889	5.777	6.008	6.450	8.908	5.647	6.364	6.298	8.791
AE	0.726	0.695	0.967	0.393	0.756	0.635	0.967	0.393	0.786	0.665	1.088	0.363
LR_uc	0.095	0.062	0.850	0.000	0.140	0.024	0.850	0.000	0.199	0.039	0.615	0.000
LR_cc	0.096	0.150	0.597	0.000	0.139	0.068	0.013	0.000	0.193	0.103	0.628	0.000
DQ	0.001	0.086	0.000	0.000	0.000	0.000	0.000	0.000	0.240	0.191	0.084	0.001
ER	0.000	0.064	0.185	0.034	0.010	0.145	0.322	0.031	0.002	0.091	0.180	0.011



Table 12 (continued)

	GARCH				GJR-GARCH				FIGARCH			
	N	ST	HS	CF	N	ST	HS	CF	N	ST	HS	CF
ESR	0.170	0.435	0.284	0.829	0.188	0.463	0.277	0.800	0.243	0.506	0.306	0.812
ADMax	9.946	9.016	10.339	9.348	10.245	9.745	10.251	9.684	9.231	8.781	10.642	8.735
Tick	0.077	0.076	0.076	0.081	0.074	0.074	0.074	0.079	0.076	0.077	0.076	0.080
FZL	1.908	1.898	1.900	1.981	1.877	1.870	1.877	1.950	1.881	1.898	1.885	1.958

The table shows the means and standard deviations of VaR and ES forecasts at 97.5% (Panel A) and 99% (Panel B) confidence levels for short positions in WTI, as well as the backtest results for the three GARCH-type models combined with normal(N), skewed t (ST), historical simulation (HS) or Cornish-Fisher expansion (CF) over the full forecast period. The AE ratio, the p values of the LR_{nc}, LR_{cc}, DQ, one-sided ER, one-sided Intercept ESR tests, as well as the absolute maximum deviation (ADMax) and the mean values of the tick and FZL functions are all reported. The numbers in bold indicate the models that are part of the superior set of models with statistically comparable VaR forecasts based on tick loss function or joint VaR and ES forecasts based on the FZL function at 90% confidence



Table 13 ES99 vs VaR99 ratio

	GARCH				GJRGARCH				FIGARCH			
	N	ST	HS	CF	N	ST	HS	CF	N	ST	HS	CF
Panel A: Brent												
Long												
Full period	1.147	1.269	1.261	1.250	1.146	1.247	1.242	1.225	1.147	1.283	1.256	1.247
Sample I	1.150	1.251	1.282	1.243	1.150	1.239	1.288	1.244	1.150	1.253	1.261	1.237
Sample II	1.149	1.273	1.342	1.281	1.148	1.263	1.366	1.254	1.149	1.279	1.364	1.282
Sample III	1.151	1.385	1.314	1.337	1.148	1.356	1.320	1.307	1.151	1.479	1.330	1.325
Short												
Full period	1.144	1.251	1.209	1.230	1.145	1.221	1.206	1.210	1.144	1.265	1.191	1.227
Sample I	1.145	1.235	1.310	1.244	1.145	1.272	1.323	1.243	1.145	1.237	1.319	1.240
Sample II	1.147	1.254	1.245	1.253	1.149	1.244	1.234	1.228	1.147	1.260	1.236	1.254
Sample III	1.146	1.340	1.299	1.316	1.146	1.314	1.391	1.282	1.146	1.428	1.325	1.301
Panel B: WTI												
Long												
Full period	1.148	1.266	1.259	1.301	1.146	1.258	1.248	1.289	1.148	1.279	1.280	1.296
Sample I	1.150	1.234	1.350	1.341	1.150	1.236	1.427	1.354	1.150	1.244	1.367	1.294
Sample II	1.149	1.333	1.299	1.398	1.147	1.324	1.318	1.381	1.150	1.351	1.367	1.404
Sample III	1.150	1.433	1.441	1.468	1.148	1.421	1.475	1.465	1.150	1.553	1.420	1.459
Short												
Full period	1.144	1.245	1.331	1.293	1.145	1.238	1.283	1.282	1.144	1.257	1.297	1.286
Sample I	1.146	1.223	1.279	1.332	1.146	1.224	1.257	1.344	1.146	1.232	1.233	1.283
Sample II	1.146	1.298	1.366	1.402	1.147	1.289	1.345	1.382	1.146	1.313	1.316	1.409
Sample III	1.145	1.374	1.651	1.488	1.146	1.363	1.592	1.490	1.146	1.484	1.584	1.471

The table shows the ratio of ES (99% confidence) to VaR (99% confidence) forecasts for both long and short positions in Brent (Panel A) and WTI (Panel B) under the three GARCH-type models combined with normal(N), skewed t (ST), historical simulation (HS) or Cornish–Fisher expansion (CF). The average of the ratios over the full forecast period, as well as the maximum ratio values over the sub-sample periods are reported

Discussion

First, this paper complements the recent study of Rizvi and Itani (2021) which compares the oil market volatility during the COVID-19 pandemic, GFC, and the SARS outbreak of 2002–2004. They confirm the presence of volatility clustering in oil prices during the crisis periods and show a more severe negative skewness, positive kurtosis, and volatility leverage effect in COVID-19 than in the other crisis. Indeed, COVID-19 has had a greater impact on the oil market than the GFC (e.g., Jebabli et al. 2021; Zhang and Hamori 2021). Our analysis reveals a significantly higher number of VaR exceptions than expected during the crisis periods, as well as a higher level and variability of oil tail-risk prediction over the COVID-19 period in comparison to the GFC and the 2014 oil crash.



Table 14 ES97.5 vs VaR99 ratio

	GARCH				GJRGARCH				FIGARCH			
	N	ST	HS	CF	N	ST	HS	CF	N	ST	HS	CF
Panel A: Brent												
Long												
Full period	1.005	1.032	1.036	1.012	1.005	1.026	1.031	1.008	1.005	1.036	1.037	1.012
Sample I	1.005	1.027	1.078	1.012	1.005	1.024	1.074	1.012	1.005	1.028	1.049	1.011
Sample II	1.005	1.033	1.094	1.017	1.005	1.030	1.116	1.013	1.005	1.034	1.107	1.017
Sample III	1.005	1.061	1.045	1.026	1.005	1.053	1.056	1.021	1.005	1.089	1.061	1.024
Short												
Full period	1.005	1.029	1.022	1.011	1.005	1.016	1.017	1.007	1.005	1.033	1.004	1.010
Sample I	1.005	1.025	1.070	1.012	1.005	1.067	1.086	1.012	1.005	1.026	1.076	1.011
Sample II	1.005	1.030	1.051	1.016	1.005	1.027	1.060	1.011	1.005	1.031	1.048	1.016
Sample III	1.005	1.053	1.069	1.027	1.005	1.057	1.127	1.021	1.005	1.078	1.077	1.024
Panel B: WTI												
Long												
Full period	1.005	1.031	1.029	1.021	1.005	1.029	1.028	1.019	1.005	1.035	1.050	1.020
Sample I	1.005	1.023	1.088	1.028	1.005	1.024	1.126	1.030	1.005	1.026	1.113	1.020
Sample II	1.005	1.048	1.035	1.037	1.005	1.045	1.047	1.034	1.005	1.052	1.096	1.038
Sample III	1.005	1.075	1.084	1.047	1.005	1.071	1.130	1.048	1.005	1.111	1.084	1.046
Short												
Full period	1.005	1.028	1.094	1.022	1.005	1.026	1.066	1.020	1.005	1.032	1.070	1.020
Sample I	1.005	1.022	1.086	1.028	1.005	1.028	1.067	1.030	1.005	1.024	1.055	1.019
Sample II	1.005	1.042	1.114	1.040	1.005	1.041	1.120	1.037	1.005	1.046	1.083	1.041
Sample III	1.005	1.063	1.238	1.055	1.005	1.060	1.211	1.055	1.005	1.096	1.204	1.052

The table shows the ratio of ES (97.5% confidence) to VaR (99% confidence) forecasts for both long and short positions in Brent (Panel A) and WTI (Panel B) under the three GARCH-type models combined with normal(N), skewed t (ST), historical simulation (HS) or Cornish–Fisher expansion (CF). The average of the ratios over the full forecast period, as well as the maximum ratio values over the sub-sample periods are reported

Second, this paper extends the oil VaR literature by incorporating ES performance for comparison under different models and across various stressed periods. We show that volatility models and innovation distributions are important not only for VaR but also for ES forecasting. In particular, the proposed GARCH-type models with the Cornish Fisher expansion are superior for ES forecast and have the lowest maximum absolute deviation for long positions while being overly conservative for VaR forecasting in short positions. The findings would be of interest to financial institutions concerned about a capital shortfall as well as regulators seeking to maintain financial system stability in the event of a crisis.

Last but not least, this paper broadens the previously limited understanding of the relationship between VaR and ES forecast. We demonstrate that the ES to VaR ratio varies across models and over time. Given the variability of this ratio, it is beneficial for institutions to monitor it closely especially during times of stress to inform timely



risk management decisions. Furthermore, we find that switching from VaR (99% confidence) to ES (97.5 % confidence) increases tail-risk prediction to varying degrees depending on the models. The Cornish–Fisher expansion leads to the most stable increase compared to the parametric and non-parametric distributions. The findings would be useful to financial institutions and regulators in preparing for the transition of risk measures mandated by the FRTB.

Conclusion

Forecasting and managing oil risk has become more important and challenging than ever before due to the growing financialization of the commodity market. Given the limitations of the VaR measure and the increasing significance of ES, this paper extends previous studies on oil VaR forecasting by shedding light on some interesting aspects of ES forecasts over the last decade. We introduce a GARCH-type model combined with the Cornish–Fisher expansion to address volatility clustering, skewness, and excess kurtosis in oil return series, and compare its VaR and ES forecast performance to some widely used GARCH-type models and innovation distributions based on recently developed VaR and ES backtesting methodology.

Four main findings are presented. First, the GJRGRACH with skewed t innovation generally outperforms the others in joint VaR and ES forecasting. Second, the Cornish–Fisher expansion produces more accurate ES forecasting than the normal, skewed t , and historical simulation, especially for long positions vulnerable to extreme oil crashes. Third, the magnitude of expected loss exceeding VaR relative to the VaR forecast varies across models and over time, with an increase observed during the COVID-19 period. Finally, switching from VaR to ES as a risk measure at comparable confidence levels increases risk prediction to varying degrees depending on the model, and the Cornish–Fisher expansion produces a more stable change than the other innovation distributions studied.

Our analysis suggests that ES, in addition to VaR, should be used to inform timely oil risk management decisions. Furthermore, we demonstrate the value of using Cornish–Fisher expansion in conjunction with GARCH volatility models to forecast ES. In view of the recent unprecedented volatility in the oil market, the findings could help regulators, energy companies, and financial institutions quantify and manage oil price risk.

Acknowledgements The author thanks the editor and an anonymous referee for their detailed review and valuable comments that greatly improved the paper.

Declarations

Conflict of interest The authors declared that they have no conflict of interest.



References

- Alexander, C., E. Lazar, and S. Stanescu. 2013. Forecasting var using analytic higher moments for garch processes. *International Review of Financial Analysis* 30: 36–45. <https://doi.org/10.1016/j.irfa.2013.05.006>.
- Aloui, C., and S. Mabrouk. 2010. Value-at-risk estimations of energy commodities via long-memory, asymmetry and fat-tailed garch models. *Energy Policy* 38: 2326–2339. <https://doi.org/10.1016/j.enpol.2009.12.020>.
- Andersen, T.G., and T. Bollerslev. 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39: 885–905. <https://doi.org/10.2307/2527343>.
- Artzner, P., F. Delbaen, J.M. Eber, and D. Heath. 1999. Coherent measures of risk. *Mathematical Finance* 9: 203–228. <https://doi.org/10.1111/1467-9965.00068>.
- Baillie, R.T., T. Bollerslev, and H.O. Mikkelsen. 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 74: 3–30. [https://doi.org/10.1016/S0304-4076\(95\)01749-6](https://doi.org/10.1016/S0304-4076(95)01749-6).
- Barone-Adesi, G., K. Giannopoulos, and L. Vosper. 1999. Var without correlations for portfolios of derivative securities. *Journal of Futures Markets* 19: 583–602. [https://doi.org/10.1002/\(SICI\)1096-9934\(199908\)19:5<583::AID-FUT5>3.0.CO;2-S](https://doi.org/10.1002/(SICI)1096-9934(199908)19:5<583::AID-FUT5>3.0.CO;2-S).
- Bayer, S., and T. Dimitriadis. 2020. Regression-based expected shortfall backtesting. *Journal of Financial Econometrics*. <https://doi.org/10.1093/jffnec/nbaa013>.
- BCBS. 2016. *Minimum capital requirements for market risk*. Technical Report. Basel Committee on Banking Supervision.
- Beine, M., A. Bénassy-Quéré, and C. Lecourt. 2002. Central bank intervention and foreign exchange rates: New evidence from figarch estimations. *Journal of International Money and Finance* 21: 115–144. [https://doi.org/10.1016/S0261-5606\(01\)00040-7](https://doi.org/10.1016/S0261-5606(01)00040-7).
- Beine, M., and S. Laurent. 2003. Central bank interventions and jumps in double long memory models of daily exchange rates. *Journal of Empirical Finance* 10: 641–660. [https://doi.org/10.1016/S0927-5398\(03\)00009-4](https://doi.org/10.1016/S0927-5398(03)00009-4).
- Bera, A.K., and M.L. Higgins. 1993. Arch models: Properties, estimation and testing. *Journal of Economic Surveys* 7: 305–366. <https://doi.org/10.1111/j.1467-6419.1993.tb00170.x>.
- Bollerslev, T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31: 307–327. [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1).
- Bollerslev, T., R.Y. Chou, and K.F. Kroner. 1992. Arch modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics* 52: 5–59. [https://doi.org/10.1016/0304-4076\(92\)90064-X](https://doi.org/10.1016/0304-4076(92)90064-X).
- Brio, E.B.D., A. Mora-Valencia, and J. Perote. 2020. Risk quantification for commodity etfs: Backtesting value-at-risk and expected shortfall. *International Review of Financial Analysis* 70: 101163. <https://doi.org/10.1016/j.irfa.2017.11.007>.
- Charles, A., and O. Darné. 2017. Forecasting crude-oil market volatility: Further evidence with jumps. *Energy Economics* 67: 508–519. <https://doi.org/10.1016/j.eneco.2017.09.002>.
- Cheong, C.W. 2009. Modeling and forecasting crude oil markets using arch-type models. *Energy Policy* 37: 2346–2355. <https://doi.org/10.1016/j.enpol.2009.02.026>.
- Chkili, W., S. Hammoudeh, and D.K. Nguyen. 2014. Volatility forecasting and risk management for commodity markets in the presence of asymmetry and long memory. *Energy Economics* 41: 1–18. <https://doi.org/10.1016/j.eneco.2013.10.011>.
- Choi, K., and S. Hammoudeh. 2009. Long memory in oil and refined products markets. *The Energy Journal* 30: 97–116. <https://doi.org/10.5547/ISSN0195-6574-EJ-Vol30-No2-5>.
- Christoffersen, P.F. 1998. Evaluating interval forecasts. *International Economic Review* 39: 841–862. <https://doi.org/10.2307/2527341>.
- Christoffersen, P.F., and F.X. Diebold. 2006. Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science* 52: 1273–1287. <https://www.jstor.org/stable/20110599>.
- Cornish, E.A., and S.R.A. Fisher. 1937. Moments and cumulants in the specification of distributions. *Review of the International Statistical Institute*. <https://doi.org/10.2307/1400905>.
- Costello, A., E. Asem, and E. Gardner. 2008. Comparison of historically simulated var: Evidence from oil prices. *Energy Economics* 30: 2154–2166. <https://doi.org/10.1016/j.eneco.2008.01.011>.



- Dai, X., Q. Wang, D. Zha, and D. Zhou. 2020. Multi-scale dependence structure and risk contagion between oil, gold, and us exchange rate: A wavelet-based vine-copula approach. *Energy Economics* 88: 104774. <https://doi.org/10.1016/j.eneco.2020.104774>.
- Ding, Q., J. Huang, and J. Chen. 2021. Dynamic and frequency-domain risk spillovers among oil, gold, and foreign exchange markets: Evidence from implied volatility. *Energy Economics* 102: 105514. <https://doi.org/10.1016/j.eneco.2021.105514>.
- Ding, Z., C.W. Granger, and R.F. Engle. 1993. A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1: 83–106. [https://doi.org/10.1016/0927-5398\(93\)90006-D](https://doi.org/10.1016/0927-5398(93)90006-D).
- Efron, B., and R. Tibshirani. 1993. *An introduction to the bootstrap*. New York: Chapman and Hall. <https://doi.org/10.1201/9780429246593>.
- Engle, R.F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica* 50: 987–1007. <https://doi.org/10.2307/1912773>.
- Engle, R.F., and S. Manganelli. 2004. Caviar. *Journal of Business & Economic Statistics* 22: 367–381. <https://doi.org/10.1198/073500104000000370>.
- Fan, Y., Y.J. Zhang, H.T. Tsai, and Y.M. Wei. 2008. Estimating value at risk of crude oil price and its spillover effect using the ged-garch approach. *Energy Economics* 30: 3156–3171. <https://doi.org/10.1016/j.eneco.2008.04.002>.
- Favre, L., and J.A. Galeano. 2002. Mean-modified value-at-risk optimization with hedge funds. *The Journal of Alternative Investments* 5: 21–25. <https://doi.org/10.3905/jai.2002.319052>.
- Fernández, C., and M.F.J. Steel. 1998. On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association* 93: 359–371. <https://doi.org/10.1080/01621459.1998.10474117>.
- Fissler, T., and J.F. Ziegel. 2016. Higher order elicibility and osband's principle. *The Annals of Statistics* 44: 1680–1707.
- Ghalanos, A. 2020. Introduction to the rugarch package (Version 1.4-3). Technical Report.
- Giacomini, R., and I. Komunjer. 2005. Evaluation and combination of conditional quantile forecasts. *Journal of Business & Economic Statistics* 23: 416–431. <https://doi.org/10.1198/073500105000000108>.
- Giot, P., and S. Laurent. 2003. Market risk in commodity markets: A var approach. *Energy Economics* 25: 435–457. [https://doi.org/10.1016/S0140-9883\(03\)00052-5](https://doi.org/10.1016/S0140-9883(03)00052-5).
- Glosten, L.R., R. Jagannathan, and D.E. Runkle. 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance* 48: 1779–1801. <https://doi.org/10.1111/j.1540-6261.1993.tb05128.x>.
- Gneiting, T. 2011. Making and evaluating point forecasts. *Journal of the American Statistical Association* 106: 746–762. <https://doi.org/10.1198/jasa.2011.r10138>.
- González-Rivera, G., T.H. Lee, and S. Mishra. 2004. Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of Forecasting* 20: 629–645.
- Hamilton, J.D. 1983. Oil and the macroeconomy since world war ii. *Journal of Political Economy* 91: 228–248. <https://doi.org/10.1086/261140>.
- Hamilton, J.D. 2003. What is an oil shock? *Journal of Econometrics* 113: 363–398. [https://doi.org/10.1016/S0304-4076\(02\)00207-5](https://doi.org/10.1016/S0304-4076(02)00207-5).
- Hansen, P.R., A. Lunde, and J.M. Nason. 2011. The model confidence set. *Econometrica* 79: 453–497. <https://doi.org/10.3982/ECTA5771>.
- Hashmi, S.M., B.H. Chang, and N.A. Bhutto. 2021. Asymmetric effect of oil prices on stock market prices: New evidence from oil-exporting and oil-importing countries. *Resources Policy* 70: 101946. <https://doi.org/10.1016/j.resourpol.2020.101946>.
- Hung, J.C., M.C. Lee, and H.C. Liu. 2008. Estimation of value-at-risk for energy commodities via fat-tailed garch models. *Energy Economics* 30: 1173–1191. <https://doi.org/10.1016/j.eneco.2007.11.004>.
- Jebabli, I., N. Kouaissah, and M. Arouri. 2021. Volatility spillovers between stock and energy markets during crises: A comparative assessment between the 2008 global financial crisis and the covid-19 pandemic crisis. *Finance Research Letters*. <https://doi.org/10.1016/j.flr.2021.102363>.
- Kang, S.H., S.M. Kang, and S.M. Yoon. 2009. Forecasting volatility of crude oil markets. *Energy Economics* 31: 119–125. <https://doi.org/10.1016/j.eneco.2008.09.006>.
- Koenker, R., and G. Bassett. 1978. Regression quantiles. *Econometrica* 46: 33–50. <https://doi.org/10.2307/1913643>.



- Kristoufek, L. 2014. Leverage effect in energy futures. *Energy Economics* 45: 1–9. <https://doi.org/10.1016/j.eneco.2014.06.009>.
- Kuang, W. 2021. Dynamic var forecasts using conditional pearson type iv distribution. *Journal of Forecasting* 40: 500–511. <https://doi.org/10.1002/for.2726>.
- Kuang, W. 2022. Oil value-at-risk forecasts: A filtered semiparametric approach. *Journal of Energy Markets* 15 (1): 47–83. <https://doi.org/10.21314/JEM.2022.011>.
- Kupiec, P.H. 1995. Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives* 3: 73–84. <https://doi.org/10.3905/jod.1995.407942>.
- Lambert, P., and Laurent, S. 2001. Modelling financial time series using GARCH-type models with a skewed student distribution for the innovations. Working Paper 0125. University of Liege. Mimeo. <http://hdl.handle.net/2078.1/91014>.
- Laporta, A.G., L. Merlo, and L. Petrella. 2018. Selection of value at risk models for energy commodities. *Energy Economics* 74: 628–643. <https://doi.org/10.1016/j.eneco.2018.07.009>.
- Le, T.H., H.X. Do, D.K. Nguyen, and A. Sensoy. 2021. Covid-19 pandemic and tail-dependency networks of financial assets. *Finance Research Letters* 38: 101800. <https://doi.org/10.1016/j.frl.2020.101800>.
- Liu, L., and J. Wan. 2012. A study of shanghai fuel oil futures price volatility based on high frequency data: Long-range dependence, modeling and forecasting. *Economic Modelling* 29: 2245–2253.
- Liu, Z., H.K. Tseng, J.S. Wu, and Z. Ding. 2020. Implied volatility relationships between crude oil and the U.S. stock markets: Dynamic correlation and spillover effects. *Resources Policy* 66: 101637. <https://doi.org/10.1016/j.resourpol.2020.101637>.
- Lux, T., M. Segnon, and R. Gupta. 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>.
- Lyu, Y., P. Wang, Y. Wei, and R. Ke. 2017. Forecasting the var of crude oil market: Do alternative distributions help? *Energy Economics* 66: 523–534. <https://doi.org/10.1016/j.eneco.2017.06.015>.
- Marimoutou, V., B. Raggad, and A. Trabelsi. 2009. Extreme value theory and value at risk: Application to oil market. *Energy Economics* 31: 519–530. <https://doi.org/10.1016/j.eneco.2009.02.005>.
- McAleer, M., and M.C. Medeiros. 2008. Realized volatility: A review. *Econometric Reviews* 27: 10–45. <https://doi.org/10.1080/07474930701853509>.
- McNeil, A.J., and R. Frey. 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance* 7: 271–300. [https://doi.org/10.1016/S0927-5398\(00\)00012-8](https://doi.org/10.1016/S0927-5398(00)00012-8).
- Nelson, D.B. 1991. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59: 347–370. <https://doi.org/10.2307/2938260>.
- Nomikos, N., and K. Andriopoulos. 2012. Modelling energy spot prices: Empirical evidence from nymex. *Energy Economics* 34: 1153–1169. <https://doi.org/10.1016/j.eneco.2011.10.001>.
- Palm, F. 1996. 7 GARCH models of volatility. In *Handbook of Statistics*, ed. G.S. Maddala and C.R. Rao, vol. 14, pp. 209–240. Statistical Methods in Finance. Amsterdam: Elsevier. [https://doi.org/10.1016/S0169-7161\(96\)14009-8](https://doi.org/10.1016/S0169-7161(96)14009-8).
- Patra, S. 2021. Revisiting value-at-risk and expected shortfall in oil markets under structural breaks: The role of fat-tailed distributions. *Energy Economics* 101: 105452. <https://doi.org/10.1016/j.eneco.2021.105452>.
- Patton, A.J. 2011. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* 160: 246–256. <https://doi.org/10.1016/j.jeconom.2010.03.034>.
- Patton, A.J., J.F. Ziegel, and R. Chen. 2019. Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics* 211: 388–413. <https://doi.org/10.1016/j.jeconom.2018.10.008>.
- Risk.net. 2020. Covid-19: Pandemic risk-Special report 2020. Risk.net.
- Rizvi, S.K.A., and R. Itani. 2021. Oil market volatility: Comparison of covid-19 crisis with the sars outbreak of 2002 and the global financial crisis of 2008. *Economic Research—Ekonomika Istraživanja*. <https://doi.org/10.1080/1331677X.2021.1927788>.
- Sadeghi, M., and S. Shavvalpour. 2006. Energy risk management and value at risk modeling. *Energy Policy* 34: 3367–3373. <https://doi.org/10.1016/j.enpol.2005.07.004>.
- Shi, Y. 2021. A discussion on the robustness of conditional heteroskedasticity models: Simulation evidence and applications of the crude oil returns. *Finance Research Letters*. <https://doi.org/10.1016/j.frl.2021.102053>.



- Wei, Y., Y. Wang, and D. Huang. 2010. Forecasting crude oil market volatility: Further evidence using GARCH-class models. *Energy Economics* 32: 1477–1484. <https://doi.org/10.1016/j.eneco.2010.07.009>.
- Westgaard, S., G.H. Arhus, M. Frydenberg, and S. Frydenberg. 2019. Value-at-risk in the european energy market: A comparison of parametric, historical simulation and quantile regression value-at-risk. *Journal of Risk Model Validation* 13: 1–27. <https://doi.org/10.21314/JRMV.2019.213>.
- Youssef, M., L. Belkacem, and K. Mokni. 2015. Value-at-risk estimation of energy commodities: A long-memory garch-evt approach. *Energy Economics* 51: 99–110. <https://doi.org/10.1016/j.eneco.2015.06.010>.
- Zhang, W., and S. Hamori. 2021. Crude oil market and stock markets during the covid-19 pandemic: Evidence from the US, Japan, and Germany. *International Review of Financial Analysis* 74: 101702. <https://doi.org/10.1016/j.irfa.2021.101702>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

