Forecasting the Conditional Distribution of Realized Volatility of Oil Price Returns: The Role of Skewness over 1859 to 2023[#]

Rangan Gupta[*], Qiang Ji[**], Christian Pierdzioch[***] and Vasilios Plakandaras[***]

Highlights

- The predictive value of expected skewness for oil realized volatility is tested
- A quantile predictive regression model is used considering nonlinearity and structural breaks.
- The predictive impact of expected skewness on realized volatility can be both positive and negative.
- Significant forecasting gains arise at the different horizons, particularly in the long-run.

Abstract

We examine the predictive value of expected skewness of oil returns for the corresponding realized volatility using monthly data for the entire modern history of the oil industry, covering 1859:11 to 2023:04. We utilize a quantile predictive regression model, which is able to accommodate nonlinearity and structural breaks. In-sample results show that the predictive impact of expected skewness on realized volatility can be both positive and negative, with these signs contingent on the quantiles of realized volatility. Moreover, we detect statistically significant forecasting gains that arise at the extreme ends and around the median of the conditional distribution of realized volatility, at 1-, 3-, 6- and, particularly, 12-month-ahead horizons. Our results have important implications for academics, investors and policymakers.

Keywords: Oil Returns; Expected Skewness; Realized Volatility; Quantile Regression; Forecasting
JEL Codes: C22, C53, Q02

[#] We would like to thank two anonymous referees for many helpful comments. However, any remaining errors are surely ours.
[*] Department of Economics, University of Pretoria, Private Bag X20, Hatfield, 0028, South Africa. Email: rangan.gupta@up.ac.za.
[**] Corresponding author. Institutes of Science and Development, Chinese Academy of Sciences, Beijing, China; School of Public Policy and Management, University of Chinese Academy of Sciences, Beijing, China. Email: jqwxnjq@163.com or jqwxnjq@casipm.ac.cn.
[***] Department of Economics, Helmut Schmidt University, Holstenhofweg 85, P.O.B. 700822, 22008 Hamburg, Germany. Email: macroeconomics@hsu-hh.de.
[***1]Department of Economics, Democritus University of Thrace, Komotini, 69100, Greece. Email: vplakand@econ.duth.gr.

1. Introduction

We analyze the predictability of realized volatility of oil price returns based on the information content of the expected skewness constructed from the West Texas Intermediate (WTI) oil price returns over the monthly historical period of 1859:11 to 2023:04. Our data set basically, thereby, covers the entire modern era of the petroleum industry, commencing with the drilling of the first oil well in the United States (US) in 1859 at Titusville, Pennsylvania. Expected skewness is a measure of risk of the oil market as it captures the expected asymmetry of the distribution of future realizations of returns.

Intuitively, variation in expected skewness is likely to originate from extreme directed changes in aggregate demand and supply (Yu et al., 2021; Salisu et al., 2022a), geopolitical acts and threats (Salisu et al., 2021), rare disaster events and pandemics (Demirer et al., 2018; Bouri et al., 2020; Qin et al., 2020, Salisu et al., 2021), and even financial market spillovers (Salisu et al., 2022b). Such changes are likely to impact the volatility of oil price returns via a "leverage effect" in the oil market, as initially studied by Geman and Shih (2009) and, in more recent research, by Asai et al. (2019, 2020). The original idea of the leverage effect, as proposed for the analysis of stock market fluctuations by Black (1976), implies that negative (positive) returns are generally accompanied by upward (downward) revisions of volatility.

Aboura and Chevallier (2013), however, argue that volatility may also increase following a hike in the oil price as such a hike may signal that oil consumers fear a rising oil price. In this regard, Demirer et al. (2020) argue that the volatility effect of a price change, in fact, depends on the nature of the underlying shocks. While oil supply shocks largely reflect geopolitical developments and country/region-specific surprises, demand shocks reflect unexpected changes in the aggregate, precautionary, or speculative demand for oil, which, in turn, are driven by market participants' expectations of future economic conditions and concerns regarding future supply shortfalls. For instance, positive oil price returns that are brought about by a supply shock would be associated with a subsequent slowdown of economic activity, and the resulting recessionary momentum would likely reduce demand and trading activity in the oil market, thereby leading to a leverage-effect-consistent lowering of volatility. In the case of a positive demand shock, in contrast, positive oil price returns are associated with an economic expansion, and this can either increase trading volume and volatility, or reduce volatility by dampening macroeconomic uncertainty. If, however, a demand shock is driven by an unexpected rise in precautionary demand as market participants become nervous about future supply shortfalls, or due to speculation, it could have the opposite effect, i.e., raising volatility.

In sum, positive (negative) expected skewness in oil returns, on purely theoretical grounds, could be associated with decreasing or increasing (increasing or decreasing) subsequent oil price returns volatility. Hence, the sign of the link between expected skewness and subsequent volatility can only be recovered empirically. In this regard, by covering the longest possible data available for constructing a measure of expected skewness over the historical period from 1859 to 2023, we capture various positive and negative oil shocks associated with, for example, the U.S. Civil War, the two World Wars, the West coast gas famine, the Great Depression, the Korean conflict, the Suez Crisis, the OPEC oil embargo, the Iranian revolution, the Iran-Iraq War, the Gulf War, the global financial crisis, the outbreak of the Coronavirus pandemic in 2020, and, of course, more recently the ongoing Russia-Ukraine War.[1] Moreover, upon studying the predictability of oil-price volatility due to oil risks associated with expected skewness, we avoid the issue of sample selection bias, and we provide a comprehensive answer to the question regarding the sign of the relationship between expected skewness and oil market volatility.

In order to achieve our objective, we predict both in- and out-of-sample the realized volatility (RV) of oil price returns based on expected skewness by utilizing a quantiles-based predictive regression model. In this regard, it is worth emphasizing that using RV, which in our case is captured by the square root of the sum of daily squared price returns over a month (following Andersen and Bollerslev, 1998), provides us with an observable and unconditional measure of volatility (unlike, as traditionally (see, Chan and Grant (2016)) derived from generalized autoregressive conditional heteroscedasticity (GARCH) and stochastic volatility (SV) models), which is otherwise a latent process. At the same time, a quantiles-based approach is more informative relative to a linear model (also considered), as it investigates the ability of expected skewness to predict the entire conditional distribution of RV, rather than being restricted just to the conditional-mean. This is important because looking at the conditional mean only of RV may "hide" interesting characteristics (Meligkotsidou et al., 2014) as it can lead to poor predictive performance, while it is actually valuable for predicting certain parts of the conditional distribution of RV. With the quantile-regression model, unlike in the case of the popular Markov-switching and smooth transition threshold models, we do not need to specify *ad hoc* the number of regimes of RV. Furthermore, the quantile-regression model retains the simple structure of a linear predictive regression model for any given quantile of RV but, simultaneously, renders possible to add an element of non-linearity to our empirical research

[1] The reader is referred to Hamilton (2013) for a detailed discussion of historical oil shocks from 1859 to 2009.

strategy in that the coefficients of the predictive model are allowed to vary across the different quantiles of the conditional distribution of RV. In light of the skewed distribution of RV, and statistical evidence of nonlinearity and structural breaks that we detect in its relationship with expected skewness, the derivation of our empirical results from a quantile-regression model, thus, is highly warranted.

While in recent literature realized skewness, as derived from intraday data, has been utilized to predict daily RV for over the last couple of decades only (see for example, Gkillas et al. (2020), Demirer et al. (2022), Luo et al. (2022), and references cited therein), we are the first to shed light on the importance of expected skewness in driving RV over the entire history of the WTI oil market spanning nearly 165 years. In this process, we add to the already voluminous literature on the predictability of oil market volatility based on a wide array of models, and macroeconomic, financial, behavioural, and climate patterns-related predictors (see, Bouri et al., (2021), and Salisu et al., (2022c) for detailed reviews), by considering the role of expected skewness. The only related paper to ours is one by Salisu et al. (2022a), who report that negative tail risks, derived from a Conditional Autoregressive Value at Risk (CAViaR) framework, can be utilized to the predict realized variance of oil returns over the monthly period of 1859:10 to 2020:10. Allowing skewness to capture both extreme negative and positive shocks, and utilizing a quantile-regression model to predict the entire conditional distribution of oil market volatility, is a technical improvement over the conditional mean-based negative shocks-only framework studied by Salisu et al. (2022a). At the same time, our findings should be of value to researchers, investors and policymakers, because our analysis nests Salisu et al. (2022a) in terms of a broader treatment of the shocks investigated, as well as the information on predictability involving the entire conditional distribution of RV.

Besides being an important academic question, our empirical research is important for investors and policymakers, because our findings are likely to assist in achieving "optimal" decisions. This is because, the financialization of the oil market over the last two decades or so in particular (Bampinas and Panagiotidis, 2017) has led to a substantial increase in the participation of key investment players such as hedge funds, pension funds, and insurance companies in the market, thus making oil a potentially profitable alternative investment in the asset-allocation decisions of financial institutions (Degiannakis and Filis, 2017). In addition, given that volatility, when interpreted as uncertainty, is a core input to investment decisions and portfolio choices (Poon and Granger 2003), precise predictions of oil price returns volatility are of paramount importance to participants of the oil market. Moreover, given that the second-moment movements in crude oil prices have been historically associated with a negative impact on economic activity (van Eyden et al., 2019), predicting the future path of RV is of obvious value to policymakers.

The remainder of the paper is organized as follows: Section 2 describes the data and methodology, Section 3 discusses the forecasting results, and Section 4 concludes.

2. Data and Methodology

We use daily and monthly WTI crude oil prices obtained from Global Financial Data[2], over the period of1859:11 to 2023:04. Both frequencies of the oil price data are converted into log returns in percentage, i.e., the first-difference of the natural logarithm of the price multiplied by 100. Then RV is computed as the square root of the sum of daily squared returns over a month, as per availability of daily data. Note that oil price data is available at monthly frequency only over the period 1920-1976 and, hence, we measure RV as the square root of monthly squared log-returns over this part of the overall sample.

Koenker and Bassett (1978) show that quantile regression estimators are more efficient and robust than mean regression estimators based on ordinary least squares (OLS), in cases where nonlinearities and deviations from normality exist, with both these features existing in our data (as discussed below). Hence, we consider the specific quantile regression model of the following form:

$$RV_{t+1} = {\beta_0}^{(\tau)} + {\beta_1}^{(\tau)} RV_t + {\beta_2}^{(\tau)} RV_{t-1} + {\beta_3}^{(\tau)} RV_{t-2} + {\beta_4}^{(\tau)} SKEWNESS_t + \varepsilon_{t+1} \quad (1)$$

where RV_t is the realized volatility at time t; $SKEWNESS_t$ is the expected skewness at time t (which we discuss below in detail); $\tau \in (0,1)$ and ε_{t+h} are assumed independent derived from an error distribution $g_\tau(\varepsilon)$ with the τ-th quantile equal to 0, i.e., $\int_{-\infty}^{0} g_\tau(\varepsilon) d\varepsilon = \tau$. The three lags of RV are selected based on the Schwarz Information Criterion (SIC), and capture the persistence of the process, with the intercept and the regression coefficients depending upon τ. The estimators of the parameters of the quantile regression model in Equation (1) minimize $\sum_{t=0}^{T-h} \rho_\tau (RV_{t+1} - {\beta_0}^{(\tau)} + {\beta_1}^{(\tau)} RV_t + {\beta_2}^{(\tau)} RV_{t-1} + {\beta_3}^{(\tau)} RV_{t-2} + {\beta_4}^{(\tau)} SKEWNESS_t)$, where $\rho_\tau(u) = u\big(\tau - I(u < 0)\big) = \frac{1}{2}[|u| + (2\tau - 1)u]$. The forecast of the τ-th quantile of the distribution RV at time $t+1$ is obtained as: $\widehat{RV}_{t+1}(\tau) = \widehat{{\beta_0}^{(\tau)}} + \widehat{{\beta_1}^{(\tau)}} RV_t + \widehat{{\beta_2}^{(\tau)}} RV_{t-1} + \widehat{{\beta_3}^{(\tau)}} RV_{t-2} + \widehat{{\beta_4}^{(\tau)}} SKEWNESS_t$. Our benchmark model is nested in equation (1) and obtained by setting $\widehat{{\beta_4}^{(\tau)}} = 0$.

[2] https://globalfinancialdata.com/.

In order to get a time-varying measure of expected skewness of oil returns (*r*), we estimate the following autoregressive quantile regression of order 1:

$$Q^{\tau}(r_t) = \gamma_0^{\tau} + \gamma_1^{\tau}(r_{t-1}) \quad (2)$$

Using the estimated model parameters from this quantile regression, we compute the one-step-ahead expected, or predicted, Kelley skewness (Kelley, 1947) as follows:

$$\mathbb{E}_{t-1}[SKEWNESS(r_t)] = \frac{\mathbb{E}_{t-1}[Q_t^{0.9}] + \mathbb{E}_{t-1}[Q_t^{0.1}] - 2\mathbb{E}_{t-1}[Q_t^{0.5}]}{\mathbb{E}_{t-1}[Q_t^{0.9}] - \mathbb{E}_{t-1}[Q_t^{0.1}]} \quad (3)$$

In equation (1), we simplify the notation by setting $\mathbb{E}_{t-1}[SKEWNESS(r_t)]$ to $SKEWNESS_t$ henceforth.

Figure A1 (Appendix) plots RV_t and $SKEWNESS_t$, with Table A1 presenting the corresponding summary statistics. The variables are non-normal distributed based on the rejection of the null hypothesis of normality under the Jarque-Bera test at the highest level of significance, with the heavy right tail of RV providing a preliminary motivation to look at a quantile predictive regression model.

3. Empirical Results

For the sake of completeness and comparability, we also estimate the conditional mean version of equation (1) using ordinary least squares (OLS) with heteroscedasticity and autocorrelation adjusted (HAC) standard errors (Newey and West, 1987). Findings (not reported in detail) show that $\widehat{\beta_4}$ is equal to -12.2154, with a *p*-value of 0.4095, i.e., the predictive effect of *SKEWNESS* is found to be negative but insignificant.[3]

In order to motivate the quantile regression approach, we now turn to statistical evidence of possible misspecification in the classical mean regression model, which may account for the insignificant predictive impact of expected skewness. In this regard, we first apply the Brock et al. (1996, BDS) test of nonlinearity on the residuals of the conditional mean model outlined above. As can be observed from Table 1 the *i.i.d.* null hypothesis is rejected for all dimensions (*m*) at the highest level of significance, thus providing strong evidence of nonlinearity in the relationship between $SKEWNESS_t$ and RV_{t+1}. In addition, we also conducted the Bai and Perron (2003) test of multiple structural breaks on the conditional-mean estimated regression. The

[3] Following Ghysels et al. (2018), we also obtained the measure of *SKEWNESS* based on a quantile-Mixed Data Sampling (MIDAS) model. However, the estimate of β_4 (=-2.6273), continued to be insignificant (with a *p*-value of 0.3086)_, with further details of this exercise are available upon request from the authors. This observation tends to suggest that the underlying way of computing the expected skewness is not necessarily the source of insignificant predictability, but, as we show below with the quantile predictive model, it is more due to misspecifications with the conditional mean framework.

UDmax and *WDmax* tests under the specification of 1 to *M* globally determined breaks revealed five breaks at: 1884:11, 1909:04, 1933:10, 1966:11, and 1991:04. Existence of nonlinearity and regime-changes confirm that the linear conditional mean-based predictive regression is misspecified, and the (insignificant) predictive impact of expected skewness on the RV cannot be considered reliable. This, in turn, leads us to consider the quantile regression model for further analyses.

Table 1. BDS Test of Nonlinearity

m	*h = 1*	*h = 3*	*h = 6*	*h = 12*
2	23.5949***	23.7299***	33.3961***	45.7055***
3	26.6924***	26.8075***	36.5895***	50.0014***
4	28.4299***	28.6670***	38.9465***	53.7848***
5	30.8409***	31.1933***	43.2056***	59.1051***
6	33.9820***	34.6113***	48.4391***	65.8200***

Note: *m* stands for the number of (embedded) dimension which embed the time series into *m*-dimensional vectors, by taking each m successive points in the series; Entries correspond to the *z*-statistic associated with the BDS test applied to the residuals recovered from the conditional mean model: $RV_{t+h} = \beta_0 + \beta_1 RV_t + \beta_2 RV_{t-1} + \beta_3 RV_{t-2} + \beta_4 SKEWNESS_t + \varepsilon_{t+h}$; *h* is the forecast horizon; *** indicates rejection of the null hypothesis of *i.i.d.* residuals involving the BDS test at 1% level of significance.

In Figure 1, we present $\widehat{\beta_4^{(\tau)}}$, at τ= 0.05, 0.10, …,0.90, 0.95, i.e., the estimated quantile-specific response of $SKEWNESS_t$ and RV_{t+1}. As it is obvious, the effect is positive until quantile 0.40, and then turns negative. We observe a significant (at 1% level) positive effect over the quantile range of 0.35-0.40, while the predictive impact is significantly negative at the highest level over τ= 0.50 to 0.95.[4] In other words, the sign and statistical significance of predictability of RV due to *SKEWNESS* is quantile-specific.

Economically speaking, in times of low risk in the oil market, as captured by the lower quantiles, a positive expected skewness causing higher returns seems to be driving more trading activity and, hence, volatility, as discussed by Fan et al. (2008), Zhang et al. (2008), Aboura and Chevallier (2013), and Demirer et al. (2020). However, beyond the median, higher returns due to a larger skewness, tend be associated with lower uncertainty, and reduce risk in the oil market via a well-established leverage effect in WTI oil prices, reported, among others, by Agnolucci (2009), Geman and Shih (2009), Larsson and Nossman (2011), Chang (2012). Put

[4] Figure A2 (Appendix) plots the predictive impact of expected kurtosis on RV. Note that, after estimating Equation (2), $\mathbb{E}_{t-1}[KURTOSIS(r_t)]$ is given by: $\frac{\mathbb{E}_{t-1}[Q_t^{0.99}]+\mathbb{E}_{t-1}[Q_t^{0.01}]}{\mathbb{E}_{t-1}[Q_t^{0.75}]-\mathbb{E}_{t-1}[Q_t^{0.25}]}$. The nature of predictability for RV_t due to $KURTOSIS_{t-1}$ is qualitatively similar to that of $SKEWNESS_{t-1}$, though in this case only statistically significant negative effect at the 1% level is detected over the quantile range of 0.45 to 0.95.

differently, we observe that the leverage effect causes a negative link between expected skewness and in the volatility of oil price returns only beyond the conditional median of the latter.[5]

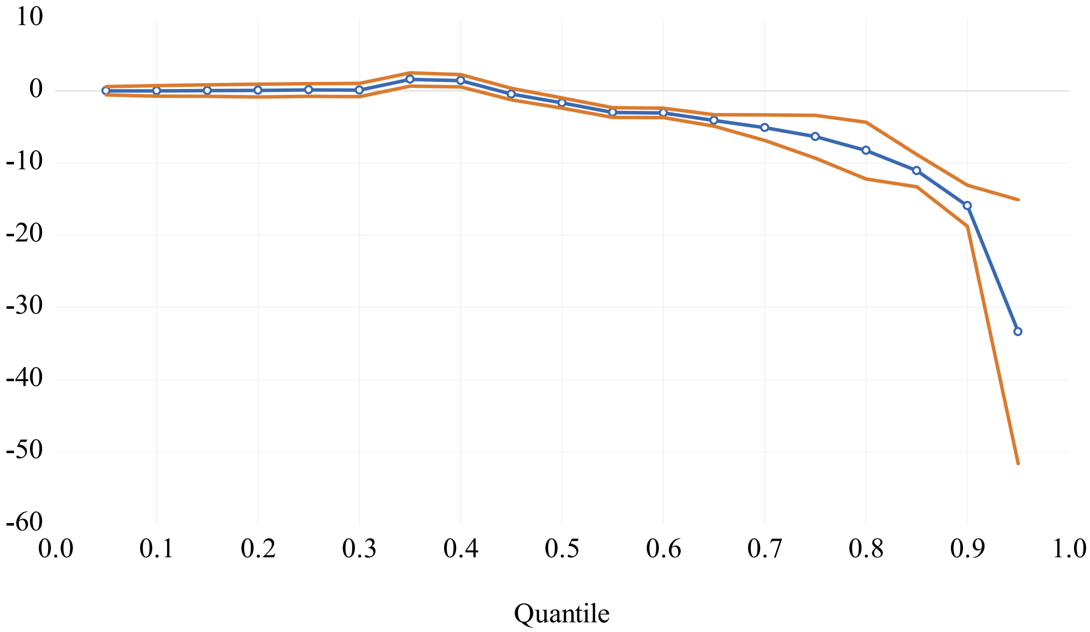
Given the well-established close linkages between the first- and second-moment of oil and stock prices (Degiannakis et al., 2018; Smyth and Narayan, 2018), we can also provide a possible explanation to the above findings in an indirect manner. As reported by Antonakakis et al. (2017), the correlation between oil and stock returns is in general negative when the former depicts low levels of volatility, whereas the sign becomes positive in periods of high oil returns volatility, a pattern that has become more pronounced following the financialization of the oil market since the turn of this century. Putting things together, when oil market uncertainty is low (high), an increase in the expected skewness of oil returns produces a higher oil price, and stock returns would decline (increase), and this decline would translate into higher (lower) realized stock market volatility via the leverage effect (Mei et al., 2017), and eventually spillover to result in increased (lower) variability of oil returns.

Using the longest available data sample, we are thus able to add another layer to the volatility-skewness nexus of oil returns in the sense that not only do we find positive and negative impacts of the latter on the former, but these signs are contingent on the states, as defined by the quantiles, of realized volatility.[6]

[5] Based on the suggestion of an anonymous referee, we constructed an alternative measure of expected skewness by estimating the asymmetric conditional autoregressive value at risk by regression (CAViaR) quantiles model of Engle and Manganelli (2004). We found the impact of this alternative metric of expected skewness on *RV* be quantitatively and qualitatively similar to that from the one *AR(1)* quantile regression (as reported in Figure 1), suggesting the robustness of our results. Complete details of this finding are available upon request from the authors.

[6] In this regard, note that, Kristoufek (2014) had indicated that the existence of the leverage effect is time-scale-dependent.

Figure 1. In-sample Predictive Impact of Expected Skewness (*SKEWNESS*) on Realized Volatility (*RV*)



Note: The figure plots the estimated parameter $\widehat{\beta_4{}^{(\tau)}}$ (in blue) for the equation: $RV_{t+1} = \beta_0{}^{(\tau)} + \beta_1{}^{(\tau)}RV_t + \beta_2{}^{(\tau)}RV_{t-1} + \beta_3{}^{(\tau)}RV_{t-2} + \beta_4{}^{(\tau)}SKEWNESS_t + \varepsilon_{t+1}$ at the quantiles τ = 0.05, 0.10,…0.90, 0.95, with lower and upper 95% confidence bands (in orange).

Since in-sample predictability does not guarantee out-of-sample forecasting gains, we conduct a one-step-ahead (h = 1) forecasting exercise that starts in 1884:11 and corresponds to the first identified structural break date. The quantile regression is recursively estimated over the out-of-sample period of 1884:11-2023:04, with an in-sample of 1859:11-1884:10, for each of the 19 considered quantiles. In line to make our results comparable with those reported in the extant literature of using (Root) Mean Square Errors ((R)MSE) as the forecast performance statistic involving quantile regressions in forecasting movements in financial and macroeconomic variables (Gupta et al., 2017), in Table 2, we present the *MSE-F* test statistic of McCracken (2007). The *MSE-F* statistic tentatively tests whether the MSEs for the model in Equation (1) with lagged *SKEWNESS* are lower than those produced by its nested benchmark equation with $\beta_4 = 0$, i.e., without $SKEWNESS_{t-1}$ (basically an autoregressive model of order 3: *AR(3)*), in a statistically significant manner.[7] As can be seen from the table, forecasting gains due to

[7] The *MSE-F* test is designed to accommodate for nestedness across the two competing models. The statistic is formally given as: $(T\text{-}R\text{-}h\text{+}1)\times(MSE_0/MSE_1\text{-}1)$, where MSE_0 (MSE_1) is the MSE from the restricted or benchmark (unrestricted or full (including *SKEWNESS*)) model, T is the total sample size, R is number of observations used for estimation of the model from which the first forecast is formed (i.e. the in-sample portion of the total number of observations), and h the forecasting horizon.

expected skewness for *RV* are primarily statistically significant, consistently at the 1% level, over the range τ= 0.60 to 0.95 (with the exceptions of the quantiles 0.70 and 0.85), which is in line with the strongly significant leverage effect observed in Figure 1, associated with in-sample predictability. Having said this, significant forecasting gains, at the 5% level, are also observed over the range τ = 0.05 to 0.15, i.e., at the lower quantiles, suggesting that there is not necessarily a one-to-one correspondence between in- and out-of-sample predictability results. Given that, the ultimate test of any predictive model (in terms of econometric methodologies and predictors used) is in its forecasting performance (Campbell, 2008), in Table 2, we present the *MSE-F* statistics for the longer forecasting horizons of 3-, 6- and 12-month-ahead over the out-of-sample period of 1884:11 to 2023:04. The quantile-regression model continues to be the preferred model in light of strong evidence of nonlinearity at the longer forecast horizons, as reported in Table 1. Furthermore, for h = 3-, 6-, and 12-month-ahead, five break dates are detected for each case,[8] and, hence, warrant the usage of the quantiles-based approach. The results are in line with the one-month-ahead i.e., h = 1 findings, especially for h = 3, with the effect being slightly weaker under h = 6, but quite stronger under h = 12, in terms of the number of quantiles for which statistically significant forecasting gains are observed over the benchmark *AR(3)* model (primarily at the 1% level), due to *SKEWNESS* for *RV*.[9] From these results, we can, in general, conclude that expected skewness forecasts the extreme ends and around the median of the conditional distribution of oil market volatility, especially in the longer-run.[10] In the process, just like intraday data-based results spanning the last two decades or so, skewness is shown to contain information for producing accurate forecasts of *RV* on oil price returns, but only in a quantile-specific manner.[11]

[8] The breaks were at: 1884:11, 1909:04, 1933:10, 1966:11, 1991:04; 1885:04, 1909:09, 1934:03, 1974:05, 1998:10, and; 1885:08, 1909:12, 193408, 1961:12, 1986:04 for h = 3-, 6-, and 12-month-ahead, respectively.

[9] Note that, for $h > 1$, RV_{t+h} is the average value of *RV* over t to $t+h$, and then the same static-forecasting approach over the out-of-sample period.

[10] Interestingly, the *MSE-F* statistics for the conditional mean-based model was also found to be statistically significant at the 1% level for h =1, 3, and 6, but not 12, with the stats being respectively, 20.6915, 20.6667, 5.9406, and -4.7934. The finding that the misspecified model produced forecasting gains for 3 of the 4 horizons, could be due to the fact that the recursive estimation of the conditional mean model over the out-of-sample period starts on the first break date, and hence, is able to accommodate, at least partially, for nonlinearity and regime changes, unlike in the case of the full-sample predictability test.

[11] We must point out that, the quantile-MIDAS-based estimate of *SKEWNESS* as outlined in Ghysels et al. (2018) performed relatively weakly in the sense that it produced significant forecasting gains for *RV* in 23 cases compared to the 33 derived under the standard *AR(1)* quantile regression approach in Equation (2), for the four forecasting horizons considered. As far as the results of the CAViaR-based expected skewness was concerned, the performance was virtually the same as obtained from the *AR(1)* model, with weaker performance at the longest forecasting horizon considered. Complete details of these results are available upon request from the authors.

Table 2. Out-of-sample Forecasting Results of Realized Volatility (*RV*) from Expected Skewness (*SKEWNESS*)

Quantile Regression (τ)	h = 1	h = 6	h = 9	h = 12
0.05	5.9417***	5.9345***	4.0521***	0.3690
0.10	4.0595***	4.0546***	33.4010***	1.7292**
0.15	1.9969**	1.9945**	-22.9396	6.0962***
0.20	-25.7436	-25.7127	-7.9035	-18.5431
0.25	0.3743	0.3738	-42.9027	-52.7130
0.30	-4.4436	-4.4383	-34.5828	-14.2317
0.35	-23.1330	-23.1052	-53.6389	-20.3499
0.40	-5.8396	-5.8326	-25.9456	7.5521***
0.45	-1.8121	-1.8099	-18.4494	109.4094***
0.50	-0.6746	-0.6738	-10.1691	25.0393***
0.55	-0.5909	-0.5902	16.6940***	20.2168***
0.60	15.2664***	15.2480***	-9.6048	-73.3776
0.65	89.4419***	89.3345***	-77.1870	112.2568***
0.70	-3.1969	-3.1931	-9.9420	97.4393***
0.75	120.6791***	120.5343***	-276.7012	-7.7630
0.80	120.6431***	120.4983***	-10.3778	16.5276***
0.85	-9.1227	-9.1117	-67.0148	7.8648***
0.90	14.8585***	14.8406***	13.9713***	-101.2330
0.95	47.4214***	47.3644***	-113.5328	12.9182***

Note: The entries correspond to the *MSE-F* test statistic of McCracken (2007), which tests whether the MSE of the unrestricted model ($RV_{t+h} = \beta_0{}^{(\tau)} + \beta_1{}^{(\tau)}RV_t + \beta_2{}^{(\tau)}RV_{t-1} + \beta_3{}^{(\tau)}RV_{t-2} + \beta_4{}^{(\tau)}SKEWNESS_t + \varepsilon_{t+h}$) is statistically lower than that of the restricted model ($RV_{t+h} = \beta_0{}^{(\tau)} + \beta_1{}^{(\tau)}RV_t + \beta_2{}^{(\tau)}RV_{t-1} + \beta_3{}^{(\tau)}RV_{t-2} + \varepsilon_{t+h}$) for a specific quantile τ and forecast horizon h; *** and ** indicates 1% and 5% levels of significance respectively of the *MSE-F* statistic with critical values of 3.9510 and 1.5180.

4. Conclusions

We analyse the predictive value of the expected skewness of oil returns for the realized volatility of oil price returns over the historical monthly period of 1859:11 to 2023:04. Our in-sample analysis shows that, while predictability cannot be detected by the misspecified conditional mean-based predictive model, due to nonlinearity and regime-changes, a causal impact does indeed exist in a quantile-regression model, that is able to accommodate for such misspecifications. Specifically, we observed that the predictive effect of skewness can be both positive and negative, with the sign contingent on the states (quantiles) of realized volatility. Based on an out-of-sample forecasting analysis, we also found statistically significant forecasting gains not only at the short one-step-ahead horizon, but also at medium- to long-run forecasting horizons of three-, six-, and twelve-month-ahead. Hence, expected skewness forecasts the extreme ends and around the median of the conditional distribution of oil market volatility, particularly in the long-run.

Given the importance of real-time forecasts of oil price volatility for both investors and policy authorities, our results have important implications for these two groups of economic agents. Our findings can be used by policy authorities to make better inferences regardingthe future path of the volatility of oil price returns due to associated expected skewness at various forecasting horizons. This knowledge, in turn, could be useful to predict the momentum of macroeconomic fluctuations in general and economic recessions in partiular, given that oil-price volatility is known to negatively impact real economic activity. Moreover, with volatility being a key input in portfolio decisions, the predictability of the volatility of oil price returns due to expected skewness should be of vital importance to oil investors, especially over the investment horizon of one year.

Our results add a new aspect to the debate of whether oil price returns skewness impacts the volatility of the market negatively or positively, by showing that both these effects are possible. As an extension of our study, it is interesting to investigate the role of own- and/or cross-market(s) expected skewness in predicting historical volatility of other important asset markets, contingent on the availability of data. At this stage, we must point out that when shapes of quantile curves are nonlinear, the linear quantile regression model, which we employ in our study, does not always suffice to adequately express the relationship between covariates (expected skewness) and quantile functions of the response variable, i.e., realized volatility of oil returns (Balcilar et al., 2022). Hence, to tackle this limitation of our work, future research could be devoted to the usage of a nonparametric quantile model, which is expected to provide more reliable and robust inferences, to check for the robustness of our findings.

References

Aboura, S., and Chevallier, J. (2013). Leverage vs. Feedback: Which Effect Drives the Oil Market? Finance Research Letters, 10(3), 131-141.

Agnolucci, P. (2009). Volatility in crude oil futures: A comparison of the predictive ability of GARCH and implied volatility models. Energy Economics, 31(2), 316-321.

Andersen T.G., and Bollerslev T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review, 39(4), 885-905.

Antonakakis, N., Gupta, R., Kollias, C., and Papadamou, S. (2017). Geopolitical risks and the oil-stock nexus over 1899-2016. Finance Research Letters, 23, 165-173.

Asai, M. Gupta, R., and McAleer, M. (2019). The Impact of Jumps and Leverage in Forecasting the Co-Volatility of Oil and Gold Futures. Energies, 12, 3379.

Asai, M. Gupta, R., and McAleer, M. (2020). Forecasting Volatility and co-volatility of crude oil and gold futures: Effects of leverage, jumps, spillovers, and geopolitical risks. International Journal of Forecasting, 36(3), 933-948.

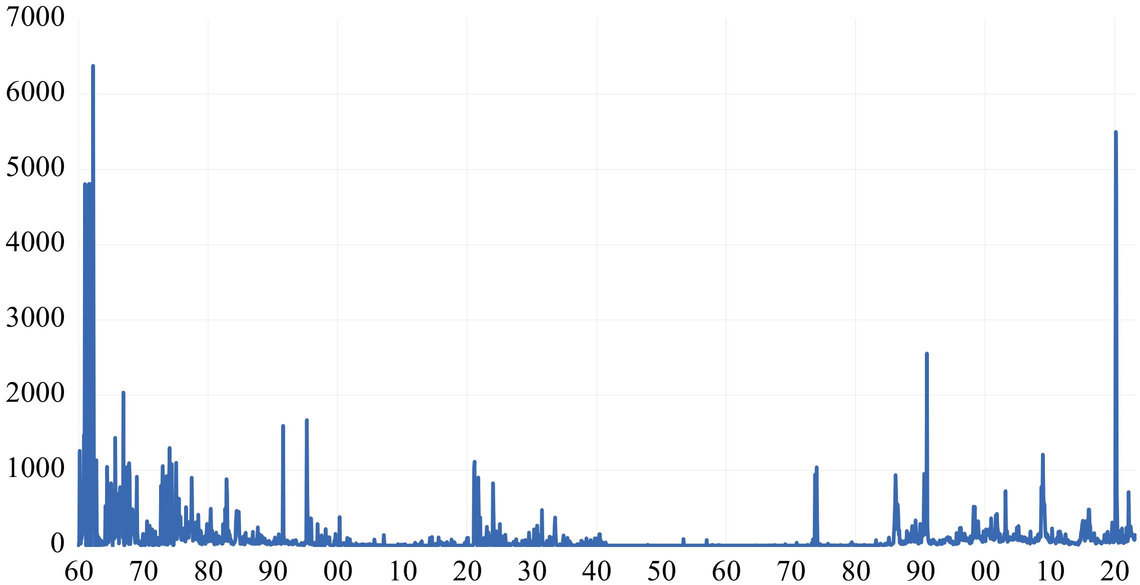
Bai, J., and Perron, P. (2003). Computation and analysis of multiple structural change models. Journal of Applied Econometrics, 18, 1-22.
Balcilar, M., Gupta, R., and Nel, J. (2022). Rare disaster risks and gold over 700 years: Evidence from nonparametric quantile regressions. Resources Policy, 79, 103053.
Bampinas, G., and Panagiotidis, T. (2017). Oil and stock markets before and after financial crises: A local Gaussian correlation approach. The Journal of Futures Markets, 37(12), 1179-1204.
Black, F. (1976). Studies of stock price volatility changes. In Proceedings of the 1976 Meeting of the Business and Economic Statistics Section; American Statistical Association: Washington, DC, USA, 177-181.
Bouri, E., Gkillas, K., Gupta, R., and Pierdzioch, C. (2020). Infectious diseases, market uncertainty and realized volatility of oil. Energies, 13(16), 4090.
Bouri, E., Gupta, R., Pierdzioch, C., and Salisu, A.A. (2021). El Niño and forecastability of oil-price realized volatility. Theoretical and Applied Climatology, 144(22), 1173-1180.
Brock, W.A., Scheinkman, J.A., Dechert, W.D., and LeBaron, B. (1996). A test for independence based on the correlation dimension. Econometric Reviews, 15(3), 197-235.
Campbell, J.Y. (2008). Viewpoint: Estimating the equity premium. Canadian Journal of Economics, 41(1), 1-21.
Chan, J.C.C., and Eisenstat, E. (2015). Marginal likelihood estimation with the cross-entropy method. Econometric Reviews, 34(3), 256-285.
Chang, K.L. (2012). Volatility regimes, asymmetric basis effects and forecasting performance: An empirical investigation of the WTI crude oil futures market. Energy Economics, 34(1), 294-306.
Degiannakis, S., and Filis, G. (2017). Forecasting oil price realized volatility using information channels from other asset classes. Journal of International Money and Finance, 76(C), 28-49.
Degiannakis, S., Filis, G., and Arora, V. (2018). Oil Prices and Stock Markets: A Review of the Theory and Empirical Evidence. The Energy Journal, 39(5), 85-130.
Demirer, R., Gkillas, K., Gupta, R., and Pierdzioch, C. (2022). Risk aversion and the predictability of crude oil market volatility: A forecasting experiment with random forests. Journal of the Operational Research Society, 73(8), 1755-1767.
Demirer, R., Gupta, R., Pierdzioch, C., and Shahzad, S.J.H. (2020). The predictive power of oil price shocks on realized volatility of oil: A note. Resources Policy, 69(C), 101856.
Demirer, R., Gupta, R., Suleman, M.T., and Wohar, M.E. (2018). Time-varying rare disaster risks, oil returns and volatility. Energy Economics, 75(C), 239-248.
Engle, R.F., and Manganelli, S. (2004). CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. Journal of Business & Economic Statistics, 22(4), 367-381.
Fan, Y., Zhang, Y.-J., Tsai, H.-T. , and Wei, Y.-M. (2008). Estimating 'Value at Risk' of crude oil price and its spillover effect using the GED–GARCH approach. Energy Economics, 30(6), 3156-3171
Geman, H., and Shih, Y.F. (2009). Modeling Commodity Prices under the CEV Model. Journal of Alternative Investments, 11(3), 65-84.
Ghysels, E., Iania, L., and Striaukas, J. (2018). Quantile-based Inflation Risk Models. Working Paper Research 349, National Bank of Belgium.
Gkillas, K., Gupta, R., and Pierdzioch, C. (2020). Forecasting realized oil-price volatility: The role of financial stress and asymmetric loss. Journal of International Money and Finance, 104(C), 102137.
Gupta, R., Majumdar, A., and Wohar, M.E. (2017). The Role of Current Account Balance in Forecasting the US Equity Premium: Evidence From a Quantile Predictive Regression Approach. Open Economies Review, 28(1), 47-59.

Hamilton, J.D. (2013). Historical Oil Shocks. In Routledge Handbook of Major Events in Economic History, edited by Randall E. Parker and Robert Whaples, 239-265, Routledge Taylor and Francis Group, New York, USA.
Kelley, T.L. (1947). Fundamentals of Statistics. Harvard University Press, Harvard, USA.
Koenker, R., and Bassett Jr, G. (1978). Regression quantiles. Econometrica, 46(1), 33-50.
Kristoufek, L. (2014). Leverage effect in energy futures. Energy Economics, 45(C), 1-9.
McCracken, M.W. (2007). Asymptotics for out of sample tests of Granger causality. Journal of Econometrics, 140, 719-752.
Larsson, K., and Nossman, M. (2011). Jumps and stochastic volatility in oil prices: Time series evidence. Energy Economics 33(3), 504-514.
Mei, D., Liu, J., Ma, F., and Chen, W. (2017). Forecasting stock market volatility: Do realized skewness and kurtosis help? Physica A: Statistical Mechanics and its Applications, 481(C), 153-159.
Meligkotsidou, L., Panopoulou, E., Vrontos, I.D., and Vrontos, S.D. (2014). A Quantile Regression Approach to Equity Premium Prediction. Journal of Forecasting, 33(7), 558-576.
Newey, W.K., and West, K.D. (1987). A simple positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix. Econometrica, 55, 703–08.
Poon, S-H, and Granger, C.W.J. (2003). Forecasting volatility in financial markets: A review. Journal of Economic Literature, 41(2), 478-539.
Qin, M., Qiu, L-H., Tao, R., Umar, M., Su, C-W., and Jiao, W. (2020). The inevitable role of El Niño: a fresh insight into the oil market. Economic Research-Ekonomska Istraživanja, 33(1), 1943-1962.
Salisu, A.A., Gupta, R., Bouri, E., and Ji, Q. (2022c). Mixed-frequency forecasting of crude oil volatility based on the information content of global economic conditions. Journal of Forecasting, 41(1), 134-157.
Salisu, A.A., Gupta, R., and Ji, Q. (2022b). Forecasting oil prices over 150 years: The role of tail risks. Resources Policy, 75(C), 102508.
Salisu, A.A., Pierdzioch, C., and Gupta, R. (2021). Geopolitical risk and forecastability of tail risk in the oil market: Evidence from over a century of monthly data. Energy, 235, 121333.
Salisu, A.A., Pierdzioch, C., and Gupta, R. (2022a). Oil Tail Risks and the Forecastability of the Realized Variance of Oil-Price: Evidence from Over 150 Years of Data. Finance Research Letters, 46(B), 102378.
Smyth, R., and Narayan, P.K. (2018). What do we know about oil prices and stock returns? International Review of Financial Analysis, 57, 148-156.
van Eyden R., Difeto, M., Gupta, R., and Wohar, M.E. (2019). Oil price volatility and economic growth: Evidence from advanced economies using more than a century of data. Applied Energy, 233, 612-621.
Yu, X., Wang, X., Wang, Y., and Li, Y. (2022). The effects of skewness on hedging decisions: an application of the skew-normal distribution in WTI and Brent futures, Economic Research-Ekonomska Istraživanja, 35(1), 3099-3118.
Zhang, Y.-J., Fan, Y., Tsai, H.-T., and Wei, Y.-M. (2008). Spillover effect of us dollar exchange rate on oil prices. Journal of Policy Modeling, 30(6), 973-991.

APPENDIX:

Figure A1. Data Plot

(a). Realized Volatility (RV)



(b). Expected Skewness (SKEWNESS)

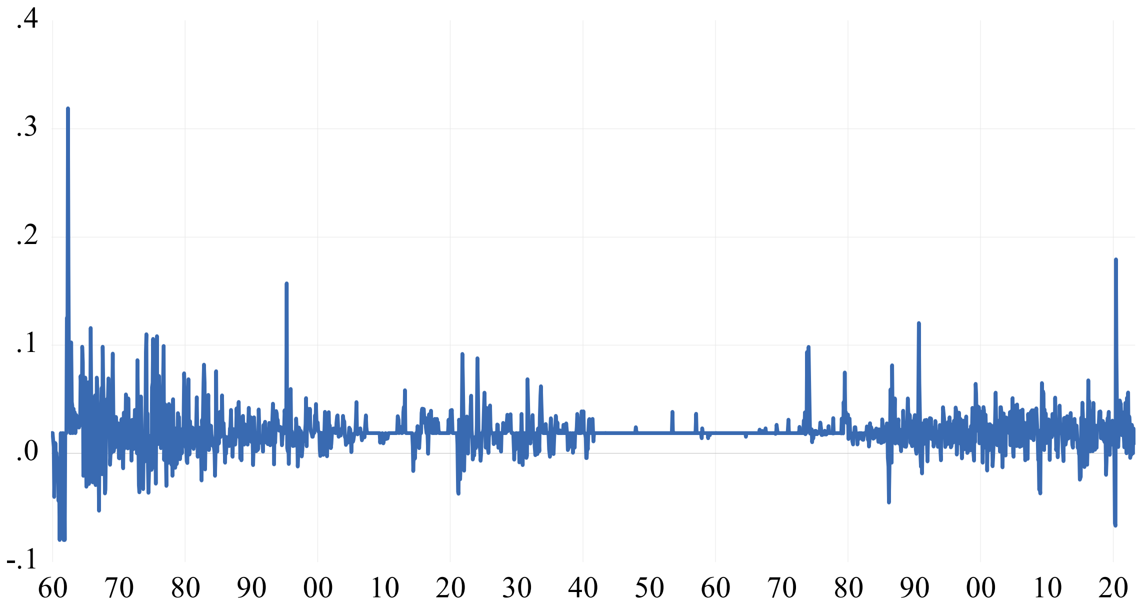
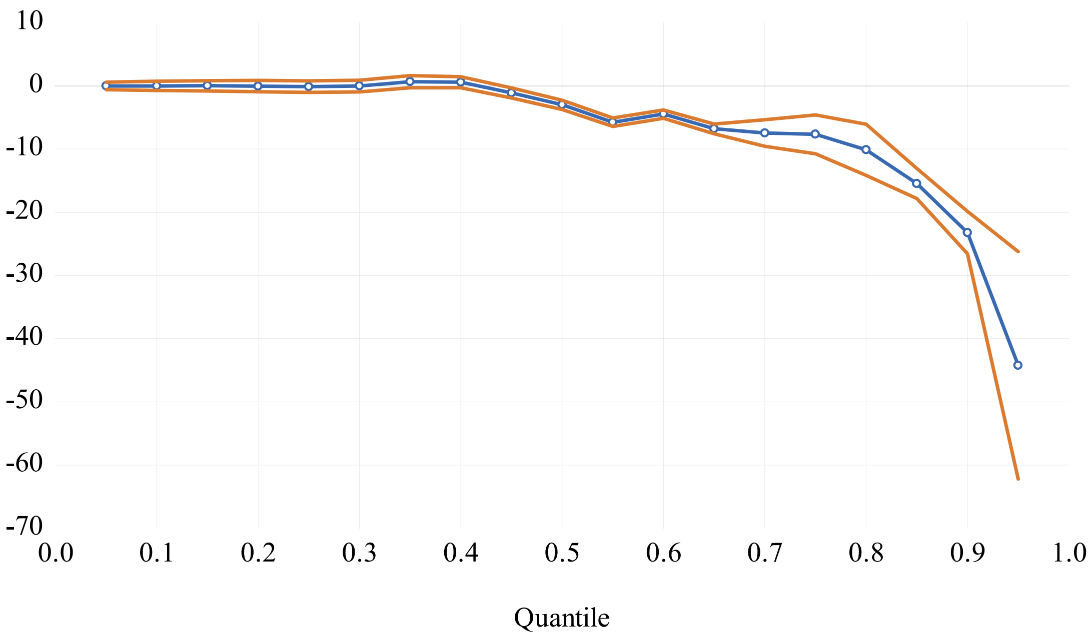


Table A1. Summary Statistics

STATISTIC	*RV*	*SKEWNESS*
Mean	101.1786	0.0197
Median	18.8682	0.0186
Maximum	6376.1450	0.3189
Minimum	0.0000	-0.0803
Std. Dev.	351.3405	0.0198
Skewness	10.7726	2.7898
Kurtosis	151.1006	38.7215
Jarque-Bera	1831036.0000	106860.1000
p-value	0.0000	0.0000
Obs.	1962	1962

Note: *RV* and *SKEWNESS* denotes realized volatility and skewness of oil returns respectively; Std. Dev. symbolizes the Standard Deviation; *p*-value corresponds to the test of normality based on the Jarque-Bera test.

Figure A2. In-sample Predictive Impact of Expected Kurtosis (*KURTOSIS*) on Realized Volatility (*RV*)



Note: The figure plots the estimated parameter $\widehat{\beta_4^{(\tau)}}$ (in blue) for the equation: $RV_{t+1} = \beta_0^{(\tau)} + \beta_1^{(\tau)} RV_t + \beta_2^{(\tau)} RV_{t-1} + \beta_3^{(\tau)} RV_{t-2} + \beta_4^{(\tau)} KURTOSIS_t + \varepsilon_{t+1}$ at the quantiles τ = 0.05, 0.10,…0.90, 0.95, with lower and upper 95% confidence bands (in orange).