Forecasting Oil Price over 150 Years: The Role of Tail Risks
Afees A. Salisu
University of Ibadan
Rangan Gupta
University of Pretoria
Qiang Ji
University of Chinese Academy of Sciences
Working Paper: 2021-20
March 2021
Forecasting Oil Prices over 150 Years: The Role of Tail Risks

Afees A. Salisu*, Rangan Gupta** and Qiang Ji***

Abstract: In this study, we examine the predictive value of tail risks for oil returns using the longest possible data available for the modern oil industry, i.e., 1859-2020. The Conditional Autoregressive Value at Risk (CAViaR) of Engle and Manganelli (2004) is employed to generate the tail risks for both 1% and 5% VaRs across four variants (adaptive, symmetric absolute value, asymmetric slope and indirect GARCH) of the CAViaR with the best variant obtained using the Dynamic Quantile test (DQ) test and %Hits. Overall, our proposed predictive model for oil returns that jointly accommodates tail risks associated with the oil market and US financial market improves the out-of-sample forecast accuracy of oil returns in contrast with a benchmark (random walk) model as well as a one-predictor model with only its own tail risk. Our results have important implications for academicians, investors and policymakers.

Keywords: Oil returns, tail risks, forecasting, advanced equity markets

JEL Codes: C22, C53, G15, Q02

1. Introduction

The role of oil price in predicting economic activity and inflation for the United States (US) is historically well established (see, for example, Plakandaras et al., 2017; Tiwari et al., 2019; Pierdzioch and Gupta, 2020). In this regard, Hamilton (2008) indicates that nine out of ten recessions in the US since World War II have been preceded by an increase in oil price, with

* Corresponding author. Centre for Econometric and Allied Research, University of Ibadan, Ibadan, Nigeria. Email address: adebare1@yahoo.com.
** Department of Economics, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa. Email address: 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 address: jqwxnjq@163.com or jqwxnjq@casipm.ac.cn.
Hamilton (2009) going as far as arguing that a large proportion of the downturn in US output during the “Great Recession” can be attributed to the oil price shock of 2007–2008 (over and above a decline in house price). Naturally, the role of oil price movement in monetary policy decision-making cannot be denied (Bodenstein et al., 2012) In the wake of the Global Financial Crisis, which highlighted the risks associated with portfolios containing only conventional financial market assets, and due to the financialization of commodities in general, the oil market has witnessed increased participation of hedge funds, pension funds, and insurance companies, with investment in oil now being considered a profitable alternative instrument in the portfolio decisions of financial institutions (Bampinas and Panagiotidis, 2015, 2017; Bonato, 2019). Considering the importance of oil in affecting both economic and financial decisions, accurate forecasting of oil prices is of paramount importance to both policymakers and investors. Unsurprisingly, the literature on prediction of oil price, involving linear and nonlinear models and economic, financial and behavioural predictors, is massive (see, for example, Alquist et al., (2013), Bashiri Behmiri and Pires Manso (2013), Baumeister (2014), Gupta and Wohar (2017) for detailed reviews of this literature).

We aim to extend this literature on the forecastability of oil returns based on information content of tail risks by undertaking a historical perspective for the West Texas Intermediate (WTI) oil price covering the monthly period of 1859:M10 to 2020:M10, i.e., the entire modern era of the petroleum industry, when the drilling of the first oil well was done in the US at Titusville, Pennsylvania in 1859. Note that tail risk is the additional risk which, commonly observed, fat-tailed asset return distributions have relatively normal distributions (Li and Rose, 2009). At this stage, it is important to emphasize that the motivation behind our decision to look at the role of tail risks in forecasting oil returns emanates from the growing literature that relates predictability of
stock returns with their tail risks, following multiple periods of financial distress like the burst of the dot-com bubble, the Lehman default, the “Great Recession” followed by the European debt crisis and the Chinese stock market crash, and more recently the outbreak of the COVID-19 pandemic (see for example, Chevapatrakul et al., 2019; Hollstein et al., 2019; Andersen et al., 2020). Given the financialization of the oil market, it is thus reasonable to hypothesize that such extreme risks can also help in forecasting returns in the oil market, which also witnessed extreme variability during the above-mentioned episodes. Moreover, tail risks can be considered an empirical proxy for the theoretical concept of rare disaster risks, which have also been shown to affect the oil market (Demirer et al., 2018), even though the theory was developed to explain the equity premium (see Gupta et al., 2019a, b) for detailed reviews of this literature).

At this juncture, it must be pointed out that there are primarily two approaches for computing tail risks: one is associated with option implied measures, while the other is based on the underlying returns’ data (Kelly and Jiang, 2014). Understandably, due to the unavailability of such a long span of historical data on options, we take the second option, whereby we estimate tail risk with the use of Value at Risk (VaR) by employing conditional autoregressive VaR specifications, as in Engle and Manganelli (2004). In this regard, the models considered are: (i) the adaptive model, (ii) the symmetric slope model, (iii) the asymmetric slope model, and (iv) the indirect generalized autoregressive conditional heteroscedasticity (GARCH) model with an autoregressive mean. Then, the specific tail risks model that best fits the oil returns data statistically is used to forecast oil returns based on an out-of-sample forecasting exercise, given that the ultimate test of any predictive model (in terms of the econometric methodologies and the predictors used) is in its forecasting performance (Campbell, 2008). Also, given the historical evidence of the oil-stock nexus for the US (Balcilar et al., 2015; 2017), besides some recent evidence of tail risks
interconnectedness between these two markets (Mensi et al., 2017), we also analyse the forecasting ability of US tail risks for oil returns, over and above its own tail risks.

To the best of our knowledge, this is the first paper to obtain estimates of tail risks of the oil market using over 150 years of monthly data, and then incorporates its role in forecasting oil returns by controlling for tail risks of the US. The only related (working) paper that we could find is the work of Ellwanger (2017), who is using an options-based estimate of tail risks to forecast oil returns over the period of 1986–2013. In this regard, by covering the longest possible data available on the evolution of the tail risks over 1859 to 2020 (associated with events such as the US Civil War, two world wars, West coast gas famine, Great Depression, oil price shocks of 1973 and 1979, the Gulf War, Asian financial crisis, Arab Spring, besides the more recent ones mentioned above), and the corresponding forecastability of the oil returns, we avoid the issue of sample selection bias in our analysis. In terms of the forecasting analysis, we rely on a returns predictability framework following Westerlund and Narayan (2012, 2015), which allows us to account for persistence, endogeneity and conditional heteroscedasticity effects.

The remainder of the paper is organized as follows. Section 2 describes the methodology and data. Section 3 discusses the econometric results associated with the estimation of the tail risks and also the various forecasting exercises. Finally, Section 4 concludes.

2. Methodology and Data

2.1 Methodology

We construct a predictive model for oil return series that hinges on the risk-return hypothesis, such as the Capital Asset Pricing Model (CAPM) and the Arbitrage Pricing Theory, where returns respond to market (systematic) risk rather than unsystematic risk (see Fama & French, 2004, Kumar, 2016; among others). Consequently, we follow the approach of Engle and Manganelli
(2004) (technically described as conditional autoregressive value at risk [CAViaR]) to estimate
the market risk as it assumes the asymptotic form of the tail, rather than modelling the whole
distribution.\(^1\) This approach is designed to overcome the statistical problem inherent in the
standard VaR method. Since VaR is simply a particular quantile of future portfolio values,
conditional on current information, and because the distribution of portfolio returns typically
changes over time, the challenge is to find a suitable model for time-varying conditional quantiles,
an issue that is ignored in the standard VaR but incorporated in the CAViaR.\(^2\) A generic CAViaR
specification is given as:

\[
f_i(\beta) = \beta_0 + \sum_{i=1}^{q} \beta_i f_{i-1}(\beta) + \sum_{j=1}^{r} \beta_j l(x_{t-j})
\]

where \(f_i(\beta) \equiv f_i(x_{t-1}, \beta_\theta)\) denotes the time \(t\) \(\theta\)-quantile of the distribution of portfolio returns
formed at \(t-1\). Note that \(\theta\) subscript is suppressed from \(\beta_\theta\) as in equation (1) for notational
convenience. Also, \(p = q + r + 1\) is the dimension of \(\beta\) and \(l\) is a function of a finite number of
lagged values of observables. The autoregressive terms \(\beta_i f_{i-1}(\beta), i = 1, \ldots, q\), ensure that the
quantile changes “smoothly” over time. The role of \(l(x_{t-j})\) is to link \(f_i(\beta)\) to observable
variables that belong to the information set. We estimate four variants of the tail risks, namely
adaptive, symmetric absolute value, asymmetric slope and indirect GARCH, and are respectively
specified as follows:

\(^1\) Several attractions to the use of Value at risk (VaR) as a standard measure of market risk are well documented in
Engle & Manganelli (2004). Chief among these attractions is its conceptual simplicity as it reduces the market risk
associated with any portfolio to a single (monetary) amount.

\(^2\) There are other approaches of modelling tail risks (see Boudoukh, Richardson & Whitelaw, 1998; Danielsson & de
Vries, 2000), however we favour the one proposed by Engle & Manganelli (2004) given the inherent shortcomings in
the previous approaches and the ability of the latter to overcome them. For instance, the approach proposed by
Danielsson and de Vries (2000) is not "extreme enough" to capture the tail of the distribution and more importantly,
the quantile models are nested in a framework of iid variables, which is not consistent with the characteristics of most
financial series, and, consequently, the risk of a portfolio may not vary with the conditioning information set (Engle
& Manganelli, 2004).
Adaptive:

\[
f_t(\beta_i) = f_{t-1}(\beta_i) + \beta_i \left[ 1 + \exp\left( G[y_{t-1} - f_{t-1}(\beta_i)] \right) \right]^{-1} - \theta \]  

(2)

Symmetric absolute value:

\[
f_t(\beta) = \beta_1 + \beta_2 f_{t-1}(\beta) + \beta_3 |y_{t-1}| \]  

(3)

Asymmetric slope:

\[
f_t(\beta) = \beta_1 + \beta_2 f_{t-1}(\beta) + \beta_3 (y_{t-1})^+ + \beta_4 (y_{t-1})^- \]  

(4)

Indirect GARCH (1,1):

\[
f_t(\beta) = \left( \beta_1 + \beta_2 f_{t-1}^2(\beta) + \beta_3 y_{t-1}^2 \right)^{1/2} \]  

(5)

where \( G \) is some positive finite number, which makes the model a smoothed version of a step function, and the last term in equation (2) converges almost surely to \( \beta_i \left[ I(y_{t-1} \leq f_{t-1}(\beta_i) \right) - \theta \] 

if \( G \to \infty \) with \( I(\cdot) \) representing the indicator function. Note that equations (3) and (5) are symmetric in nature, while equation (4) is asymmetric as the response to positive and negative returns is identical for the former category but differs for the latter. While the adaptive model has a unit coefficient on the lagged VaR, the other three are mean reverting, implying that the coefficient on the lagged VaR is not constrained to be 1.

We estimate all four variants of the CAViaR and produce results for both 1% and 5% VaRs. Thereafter, we use relevant model diagnostics, such as the Dynamic Quantile test (DQ) test and %Hits\(^3\) to determine the model that best fits the data. The results obtained are then used in the return predictability following the Westerlund & Narayan (2012, 2015) method, which allows us to account for some salient features such as persistence, endogeneity and conditional

\(^3\) These are standard test statistics for evaluating the relative performance of the alternative specifications of CAViaR test.
heteroscedasticity effects typical of most financial series and relies on the following predictive model partitioned into two variants as follows:  

**Case I:** Here, we examine the relative forecast performance of own tail risk in contrast with a driftless random walk. In other words, this is a single predictor model, as it only accommodates its own tail risk in the return predictability.

\[
oil_t = \omega + \phi tr_{t-1}^{oil} + \alpha (tr_{t-1}^{oil} - \rho tr_{t-1}^{oil}) + \varepsilon,
\]

(6)

**Case II:** We extend equation (6) to include US stock tail risk. This is motivated by the strong connection between oil and US stock (see Salisu and Oloko, 2015; Salisu, Raheem and Ndako, 2019; Salisu, Swaray and Oloko, 2019).

\[
oil_t = \omega + \phi tr_{t-1}^{oil} + \alpha_t (tr_{t-1}^{oil} - \rho_{oil} tr_{t-1}^{oil}) + \phi_{oil} tr_{t-1}^{as} + \alpha_t (tr_{t-1}^{as} - \rho_{oil} tr_{t-1}^{as}) + \varepsilon,
\]

(7)

where \(\oil_t\) is the log return of stock price indexes at period \(t\), computed as \(100 \times \Delta \log(p_t)\), \(p_t\) being the price data; \(\omega\) is the intercept; \(tr\) is the tail risk obtained as the one that best fits the data while \(\varepsilon\) is the zero mean idiosyncratic error term. Note that the superscript on the tail risk defines the return series used in calculating it, thus, superscripts “\(oil\)” and “\(us\)” denote the tail risks for the oil return series (i.e. the dependent variable) and US stock returns, respectively. With reference to equation (6), we include an additional term - \(\alpha (tr_{t-1}^{oil} - \rho tr_{t-1}^{oil})\) in the predictive model in addition to the lagged predictor series - \(\phi tr_{t-1}^{oil}\) in order to resolve any inherent endogeneity bias resulting from the correlation between the predictor series and the error term as well as any potential

---

4 See Westerlund and Narayan (2015) for computational details while several applications are evident in the literature as regards the use of this methodology for stock return predictability (see for example, Bannigidadmath and Narayan, 2015; Narayan and Bannigidadmath, 2015; Narayan and Gupta, 2015; Phan, Sharma, and Narayan, 2015; Devpura, Narayan, and Sharma, 2018; Salisu, Raheem and Ndako, 2019; Salisu, Swaray and Oloko, 2019, among others).
persistence effect.\textsuperscript{5} This same procedure is followed for equation (7), while the technical details justifying this inclusion are well documented in Westerlund and Narayan (2012, 2015). Finally, given the use of monthly frequency over centuries of data, accounting for conditional heteroscedasticity effect becomes necessary, and this is implemented by pre-estimating the predictive model with the conventional GARCH-type model and pre-weighting all the data with the inverse of standard deviation obtained from the latter. The resulting equation is estimated with the OLS to obtain the Feasible Quasi Generalized Least Squares estimates.\textsuperscript{6}

For the forecast evaluation, we essentially focus on the out-of-sample forecast performance since the literature is replete with studies on in-sample predictability whose outcome cannot be used to generalize for the out-of-sample predictability and, more importantly, forecast accuracy of return series is better determined with out-of-sample forecasts (see Narayan and Gupta, 2015; Salisu, Swaray and Oloko, 2019). As it is conventional for time series forecasting of financial series, we use the driftless random walk as the benchmark, and its forecast performance is compared with the tail risk-based predictive models. We employ both single (Root Mean Square Forecast Error) and pairwise forecast measures using the Clark & West (2007) while the 75:25 data split is respectively used to split the data into 75\% of the full sample for the in-sample estimation and the remaining 25\% for the out-of-sample forecast.\textsuperscript{7}

\textsuperscript{5} Some preliminary tests are rendered in this regard to establish the presence of these effects and the results can be provided by the authors upon request.

\textsuperscript{6} A similar approach is followed in a related study by Salisu, Rangan and Ogbonna (2021) albeit with a focus on stock return predictability of advanced economies.

\textsuperscript{7} Note that there is no theoretical guidance in the literature for data splitting in forecast analysis, however, studies have adopted 25:75, 50:50 and 75:25 respectively between the in-sample and out-of-sample forecasts (see Narayan and Gupta, 2015) and the outcome is observed to be insensitive to the choice of data split (see Narayan and Gupta, 2015; Salisu, Swaray and Oloko, 2019).
2.2 Data sources

The data used in this paper are monthly WTI crude oil prices and the S&P 500 stock price index covering 1859:M10 to 2020:M10 and are obtained from Global Financial Data. Both series are converted into log returns in percentage, i.e., the first difference of the natural logarithm of the indices multiplied by 100.

We present some descriptive statistics for oil price returns in Table 1. On average, the crude oil market has higher volatility and lower returns than the US stock market, judging by both standard deviation (Std. Dev.) and coefficient of variation (CV) while both series are heavy tailed given the leptokurtic nature of the kurtosis statistics and are also negatively skewed. It is therefore not surprising why they are non-normal judging by the Jarque-Bera test. Hence, limiting the measurement of the market to the distribution of the tail rather than the whole distribution is justified. Nonetheless, we offer additional empirical support in the next section.

Table 1. Summary statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>CV</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>JB test</th>
<th>Nobs</th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
<td>0.2504</td>
<td>3.8523</td>
<td>15.385</td>
<td>-0.6695</td>
<td>15.0064</td>
<td>16722.92***</td>
<td>2750</td>
</tr>
<tr>
<td>Oil</td>
<td>0.0351</td>
<td>9.2902</td>
<td>264.678</td>
<td>-0.3286</td>
<td>15.9951</td>
<td>13636.05***</td>
<td>1933</td>
</tr>
</tbody>
</table>

Note: Nobs = Number of observations; *** denotes significance at 1% level; Std. Dev. = Standard Deviation; CV = Coefficient of Variation.

3. Empirical results

On the choice of best tail risk for each return series, we estimate the four CAViaR specifications described in the preceding section and thereafter use the DQ test and %Hits to select the tail risk that best fits the return series being examined. The analyses are rendered for both 1% and 5% VaRs for robustness, and the results are presented in Tables 2 and 3, respectively. We

---

8 https://globalfinancialdata.com/
9 The computational and theoretical procedures for the implementation of the four variants of the CAViaR test are well presented in the Engle & Manganelli (2004). We are also grateful to these authors for providing useful Matlab odes for CAViaR estimation.
expect the %Hits to be relatively 1% for 1% VaR and 5% for 5% VaR while DQ test statistic is not expected to be significant. In cases where more than one tail risk is statistically insignificant in terms of the DQ test, we consider the tail risk with the closest value to the expected value for the %Hits. In the same vein, where all the tail risks are statistically significant, then the %Hits becomes a major criterion except where some distinctions can still be made with the significant DQ test statistics. For easy reference, the best choice for each oil return series as well as US stock return is put in bold, and we find that the choice of VaR matters as the performance seems to differ between 1% VaR and 5% VaR. Thus, our out-of-sample forecast analysis is carried out for both VaRs to further test whether the same conclusion would be obtained for the tail risks’ predictability. Nonetheless, the return series exhibit volatility clustering as measured by the statistically significant coefficient (Beta2) on the autoregressive term. This outcome further confirms that the phenomenon of clustering of volatilities is relevant also in the tails (see also Engle & Manganelli, 2004), and thus, pre-weighting the data with the inverse of standard deviation obtained from the conventional GARCH-type model is justified. Some graphical illustrations are provided in Figures 1 and 2 for the 1% and 5% VaRs, respectively, and some relative co-movements can be teased out between the tail risks and the return series, also to the volatile nature of the series.

Table 2: Estimates and Relevant statistics for the CAViaR specification of oil returns

<table>
<thead>
<tr>
<th>Oil returns</th>
<th>1% VaR</th>
<th>5% VaR</th>
<th>1% VaR</th>
<th>5% VaR</th>
<th>1% VaR</th>
<th>5% VaR</th>
<th>1% VaR</th>
<th>5% VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta1</td>
<td>0.6428</td>
<td>0.0894</td>
<td>1.3322</td>
<td>0.2309</td>
<td>1.0007</td>
<td>0.0000</td>
<td>20.1375</td>
<td>4.0488</td>
</tr>
<tr>
<td>Standard errors</td>
<td>0.2149</td>
<td>0.0331</td>
<td>0.3324</td>
<td>0.0944</td>
<td>0.7132</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>P values</td>
<td>0.0014</td>
<td>0.0035</td>
<td>0.0000</td>
<td>0.0072</td>
<td>0.0803</td>
<td>0.4858</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Beta2</td>
<td>0.8364</td>
<td>0.8668</td>
<td>0.7977</td>
<td>0.7416</td>
<td>0.7519</td>
<td>0.7529</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Standard errors</td>
<td>0.0258</td>
<td>0.0386</td>
<td>0.0254</td>
<td>0.0377</td>
<td>0.0102</td>
<td>0.0021</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>P values</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Beta3</td>
<td>0.5609</td>
<td>0.2934</td>
<td>0.2415</td>
<td>0.2059</td>
<td>1.7883</td>
<td>0.7438</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Standard errors</td>
<td>0.0939</td>
<td>0.0882</td>
<td>0.0675</td>
<td>0.0687</td>
<td>0.4637</td>
<td>0.0061</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>P values</td>
<td>0.0000</td>
<td>0.0004</td>
<td><strong>0.0002</strong></td>
<td>0.0014</td>
<td>0.0001</td>
<td>0.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>--------</td>
<td>--------</td>
<td>----------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta4</td>
<td>0.7180</td>
<td>0.7376</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard errors</td>
<td>0.1603</td>
<td>0.0892</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P values</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RQ</td>
<td>392.256</td>
<td>1311.754</td>
<td><strong>357.088</strong></td>
<td>1235.017</td>
<td>379.332</td>
<td><strong>1363.500</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hits in-sample (%)</td>
<td>1.0695</td>
<td>5.0505</td>
<td><strong>1.0101</strong></td>
<td>5.0505</td>
<td>1.2478</td>
<td>4.6940</td>
<td>0.6536</td>
<td><strong>3.9810</strong></td>
</tr>
<tr>
<td>Hits out-of-sample (%)</td>
<td>2.0000</td>
<td>9.2000</td>
<td><strong>1.6000</strong></td>
<td>10.4000</td>
<td>2.8000</td>
<td>8.0000</td>
<td>2.0000</td>
<td><strong>6.0000</strong></td>
</tr>
<tr>
<td>DQ in-sample (P-values)</td>
<td>0.0002</td>
<td>0.0000</td>
<td><strong>0.1135</strong></td>
<td>0.0001</td>
<td>0.2793</td>
<td>0.0000</td>
<td>0.0034</td>
<td><strong>0.0000</strong></td>
</tr>
<tr>
<td>DQ out-of-sample (P-values)</td>
<td>0.0018</td>
<td>0.0000</td>
<td><strong>0.9625</strong></td>
<td>0.0011</td>
<td>0.0475</td>
<td>0.0000</td>
<td>0.0024</td>
<td><strong>0.0000</strong></td>
</tr>
</tbody>
</table>

Note: SAV = Symmetric Absolute Value; ASY = Asymmetric slope; GARCH = Indirect GARCH; ADAPT = Adaptive. The tail risk that best "fits" the return series is put in bold. The criteria used are the DQ test and %Hits for Out-of-Sample. For the “best” tail risk variant, we expect the %Hits to be 1% for 1% VaR and 5% for 5% VaR, while the DQ test statistic is not expected to be significant. In cases where more than one tail risk is statistically insignificant in terms of the DQ test, we consider the tail risk with the closest value to the expected value for the %Hits. In the same vein, where all the tail risks are statistically significant, then the %Hits becomes a major criterion except where some distinctions can still be made with the significant DQ test statistics.

### Table 3: Estimates and Relevant statistics for the CAViaR specification of US stock returns

<table>
<thead>
<tr>
<th></th>
<th><strong>SAV</strong></th>
<th><strong>ASY</strong></th>
<th><strong>GARCH</strong></th>
<th><strong>ADAPTIVE</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
<td>1% VaR</td>
<td>5% VaR</td>
<td>1% VaR</td>
<td>5% VaR</td>
</tr>
<tr>
<td>Beta1</td>
<td>0.2037</td>
<td>0.3888</td>
<td>0.3725</td>
<td><strong>0.4731</strong></td>
</tr>
<tr>
<td>Standard errors</td>
<td>0.1275</td>
<td>0.0781</td>
<td>0.1765</td>
<td><strong>0.0936</strong></td>
</tr>
<tr>
<td>P values</td>
<td>0.0551</td>
<td>0</td>
<td>0.0174</td>
<td>0</td>
</tr>
<tr>
<td>Beta2</td>
<td>0.8495</td>
<td>0.8365</td>
<td>0.81</td>
<td><strong>0.7802</strong></td>
</tr>
<tr>
<td>Standard errors</td>
<td>0.0228</td>
<td>0.059</td>
<td>0.0222</td>
<td><strong>0.0344</strong></td>
</tr>
<tr>
<td>P values</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Beta3</td>
<td>0.6528</td>
<td>0.2587</td>
<td>0.3388</td>
<td><strong>0.1227</strong></td>
</tr>
<tr>
<td>Standard errors</td>
<td>0.1362</td>
<td>0.1465</td>
<td>0.1087</td>
<td><strong>0.0489</strong></td>
</tr>
<tr>
<td>P values</td>
<td>0</td>
<td>0.0387</td>
<td>0.0009</td>
<td><strong>0.0061</strong></td>
</tr>
<tr>
<td>Beta4</td>
<td>0.9072</td>
<td>0.4843</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard errors</td>
<td>0.0996</td>
<td><strong>0.1251</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P values</td>
<td>0</td>
<td>0</td>
<td>0.001</td>
<td><strong>0.0001</strong></td>
</tr>
<tr>
<td>RQ</td>
<td>317.42</td>
<td>1000.243</td>
<td>303.08</td>
<td><strong>959.557</strong></td>
</tr>
<tr>
<td>Hits in-sample (%)</td>
<td>1.0667</td>
<td>5.0667</td>
<td>1.0222</td>
<td><strong>5.0222</strong></td>
</tr>
<tr>
<td>Hits out-of-sample (%)</td>
<td>2.0000</td>
<td>9.0000</td>
<td>1.6</td>
<td><strong>5.2</strong></td>
</tr>
<tr>
<td>DQ in-sample (P-values)</td>
<td>0.0061</td>
<td>0</td>
<td>0.4206</td>
<td><strong>0.9228</strong></td>
</tr>
<tr>
<td>DQ out-of-sample (P-values)</td>
<td><strong>0.9923</strong></td>
<td>0.8609</td>
<td>0.8295</td>
<td><strong>0.7885</strong></td>
</tr>
</tbody>
</table>

Note: see table 2.
On the out-of-sample predictive value of the tail risk, we follow a three-step procedure. First, select the best tail risks for oil and stock returns, as previously mentioned. Second, we estimate two models: one-predictor (best own (oil) tail risk) and two-predictor (both best own (oil) and US tail risks) models using the approach of Westerlund and Narayan (2012, 2015). Third, we evaluate the out-of-sample performance of the two forecast models relative to the benchmark (driftless random walk) model over multiple forecast horizons of 6, 12 and 24 months using both RMSFE and the Clark and West (2007) test statistics. Our results presented in Table 4 show that the two
proposed models (one-predictor and two-predictor) at both 1% and 5% VaRs offer better forecast outcomes than the benchmark model, suggesting that the significance of market risk measured as tail risk in the predictability of oil returns. In other words, the oil market is influenced by both global (own) risk and the level of economic activity. This outcome aligns with the studies of Salisu and Oloko (2015), which found bidirectional spillover volatility transmission between oil and the US stock market, albeit with a focus on in-sample predictability. A closer comparison of the one-predictor and two-predictor tail risk-based models indicates the superior out-of-sample forecast performance of the latter over the former for 1% VaR, while the reverse is the case for 5% VaR. The forecast prowess for both 1% and 5% VaRs improves over a short period (within 12 months), while it tends to decline over a long forecast horizon (over 12 months).

Table 4: Out-of-sample forecast evaluation for oil returns

<table>
<thead>
<tr>
<th>Oil</th>
<th>One predictor vs. Random walk 1%</th>
<th>Two predictors vs. Random walk 1%</th>
<th>One predictor vs. Two predictors 1%</th>
<th>One predictor 5%</th>
<th>Two predictors 5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>h = 6</td>
<td>12.0820 ( ^a ) 14.3581 ( ^a ) 18.5348 ( ^a ) 24.8348 ( ^a ) 10.3451 ( ^c ) 21.9663 ( ^a ) 15.9565 9.7134 13.7844 16.7957</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.0108) (4.5480) (2.3871) (3.1801) (1.4887) (2.8376)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h = 12</td>
<td>12.0246 ( ^a ) 14.3326 ( ^a ) 18.4588 ( ^a ) 24.7407 ( ^a ) 10.3025 ( ^c ) 21.8608 ( ^a ) 15.9472 9.6993 13.7721 16.7846</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.0089) (4.5578) (2.3867) (3.1804) (1.4885) (2.8351)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h = 24</td>
<td>10.6522 ( ^a ) 16.6589 ( ^a ) 17.9334 ( ^a ) 24.5222 ( ^a ) 10.1219 ( ^c ) 20.8808 ( ^a ) 16.0262 9.7505 13.8406 16.8149</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.7799) (4.7968) (2.3338) (3.1672) (1.4722) (2.7153)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: For the Clark and West test, the null hypothesis of equal forecast accuracy between the benchmark and the proposed models is rejected if the t-statistic is greater than +1.282 (for a one-sided 0.10 test), +1.645 (for a one-sided 0.05 test), and +2.00 for 0.01 test (for a one-sided 0.01 test) (see Clark and West, 2007), and are denoted by \( ^c, ^b \) and \( ^a \), respectively; and the values of the t-statistic are denoted in parentheses. CW denotes Clark and West (2007) test. For the comparison of a one-predictor model with the two-predictor model, a rejection of the null hypothesis implies the superior out-of-sample forecast performance of the latter over the former, while a non-rejection implies equal forecast accuracy between the two models. RMSFE is the Root Mean Square Forecast Error. One predictor here only accommodates own tail, while two predictors involve both own and US tail risk.

4. Conclusion

In this study, we examine the predictive value of tail risks for oil returns using centuries of data. Relying on the risk-return, we construct a predictive model for oil returns where the market risks are measured as tail risks following the CAViaR of Engle and Manganelli (2004). We
consider four variants (adaptive, symmetric absolute value, asymmetric slope and indirect GARCH) of the CAViaR framework estimated for both 1% and 5% VaRs, and the best variant is selected for under each VaR using the Dynamic Quantile test (DQ) test and %Hits. We forecast oil returns with two models: one-predictor (own (oil) tail risk) and two-predictor (both own (oil) and US tail risks) models. The estimation procedure follows the approach of the Westerlund and Narayan (2012, 2015) method, which allows us to account for some salient features such as persistence, endogeneity and conditional heteroscedasticity effects. Our findings reveal that the model that accommodates the tail risks outperforms the benchmark (random walk) model, while a two-predictor model (with own tail risk and risk associated with the US financial market) outperforms the one-predictor model with own tail risk. The forecast prowess for both 1% and 5% VaRs improves over a short period (within 12 months), while it tends to decline over a long forecast horizon (over 12 months).

Understandably, our results highlight that investors need to account for tail risks, not only oil, but also the equity market, when producing forecasts of oil returns to draw optimal portfolio decisions. Also, from the perspective of academicians, our results suggest that oil markets are at least weakly inefficient, and the role of own and cross-market tail risks must be incorporated into asset pricing models. Finally, with oil market movements being a well-known predictor of the real economy, policy authorities would need to closely monitor tail risks in the oil and equity markets to get an understanding of the future movements in output and inflation, and accordingly design monetary policy responses.

As an extension of this study, based on data availability, would be to investigate the role of historical tail risks in predicting other asset and commodity markets.
References


