Received: 15 April 2018 | Revised: 15 January 2019 | Accepted: 11 April 2019
DOI: 10.1002/for.2598

RESEARCH ARTICLE

WILEY

Modeling and forecasting the oil volatility index

João H. Gonçalves Mazzeu[1] | Helena Veiga[2,3] | Massimo B. Mariti[4]

[1]Department of Statistics, University of Campinas, Campinas-SP, Brazil
[2]Department of Statistics and Instituto Flores de Lemus, Universidad Carlos III de Madrid, Spain
[3]BRU-IUL, Instituto Universitário de Lisboa, Portugal
[4]University Milano-Bicocca, Italy

Correspondence
Helena Veiga, Department of Statistics and Instituto Flores de Lemus, Universidad Carlos III de Madrid, Spain.
Email: mhveiga@est-econ.uc3m.es

Funding information
Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Grant/Award Number: 88882.305837/2018-01; Fundação para a Ciência e a Tecnologia, Grant/Award Number: UID/GES/00315/2019; Ministerio de Economía y Competitividad, Grant/Award Number: PGC2018-096977-B-I00 and ECO2015-70331-C2-2-R

Abstract
The increase in oil price volatility in recent years has raised the importance of forecasting it accurately for valuing and hedging investments. The paper models and forecasts the crude oil exchange-traded funds (ETF) volatility index, which has been used in the last years as an important alternative measure to track and analyze the volatility of future oil prices. Analysis of the oil volatility index suggests that it presents features similar to those of the daily market volatility index, such as long memory, which is modeled using well-known heterogeneous autoregressive (HAR) specifications and new extensions that are based on net and scaled measures of oil price changes. The aim is to improve the forecasting performance of the traditional HAR models by including predictors that capture the impact of oil price changes on the economy. The performance of the new proposals and benchmarks is evaluated with the model confidence set (MCS) and the Generalized-AutoContouR (G-ACR) tests in terms of point forecasts and density forecasting, respectively. We find that including the leverage in the conditional mean or variance of the basic HAR model increases its predictive ability. Furthermore, when considering density forecasting, the best models are a conditional heteroskedastic HAR model that includes a scaled measure of oil price changes, and a HAR model with errors following an exponential generalized autoregressive conditional heteroskedasticity specification. In both cases, we consider a flexible distribution for the errors of the conditional heteroskedastic process.

KEYWORDS
forecasting oil volatility, heterogeneous autoregression, leverage, net oil price changes, scaled oil price changes

1 | INTRODUCTION

The price of oil has been fluctuating dramatically in the last decade. It reached its maximum price in July 2008, to plunge to a value of $30.28 per barrel some months later. In the last 7 years a barrel of crude oil has ranged from $125 to $30. These oil price fluctuations increase oil volatility and, consequently, the risk exposure of companies dedicated to exploring for and processing oil, and investors.

In the last decade investments in commodities have grown quickly, with oil accounting for a high percentage of these investments; see Baffes (2007). Given the importance of oil among the commodities, the Chicago Board Option Exchange (CBOE) has calculated and reported the crude oil ETF volatility index since May 2007. According to CBOE, the index “measures the market's expectation of 30-day volatility of crude oil prices and it is calculated using the market volatility index (VIX) methodology for the United States Oil Fund."[1] It is known by its ticker symbol OVX and uses real-time bid/ask quotes of nearby and second nearby options with at least 8 days to expiration, and weights these options to derive a constant, 30-day measure of expected volatility. This index is the first CBOE implied volatility index on a commodity.

CBOE has published several implied volatility indexes in the last decades that are key measures of market expectations of volatility conveyed by option prices. The best known is the volatility index of options on the S&P 500, whose ticker symbol is VIX. It is well known that VIX has long memory and reacts asymmetrically to the underlying stock market return; see Fleming, Ostdiek, and Whaley (1995), Whaley (2000, 2009), Simon (2003), Giot (2005) and Carr and Wu (2006). Although the drivers of oil markets are different from those of stock markets (Ankrim & Hensel, 1993; Belousova & Dorfleitner, 2012; Mollick & Assefa, 2013), with the process of financialization faced by commodities, it is expected that the structure of these markets has changed and caused the OVX to share some features of VIX—in particular, the long memory and the asymmetric response to positive and negative oil price changes.

Although the literature on OVX is scarce, some studies have already reported that the index has long memory; see Chen, He, and Yu (2015) and Campos, Cortazar, and Reyes (2017). Traditionally, long memory in volatility has been modeled either by autoregressive fractional integrated moving average (ARFIMA) or generalized autoregressive conditional heteroskedasticity (GARCH)-type models, such as the FIGARCH of Baillie, Bollerslev, and Mikkelsen (1996) and the FIEGARCH of Bollerslev and Mikkelsen (1996). However, models that involve fractional integrated roots have been much criticized due to the lack of economic interpretability, difficulties to extend them into the multivariate framework, and to impose a linear form of long memory that depends only on the fractional integrated parameter; see Comte and Renault (1998), Renault (1999), Abadir and Talmain (2002), and Bhardwaj and Swanson (2006) . Based on the empirical evidence that taking into account periodic intradaily dynamics is crucial for modeling volatility, Bordignon, Caporin, and Lisi (2008) propose the periodic long-memory GARCH (PLM-GARCH), which is an extension of the FIGARCH model and suffers, as does the first, from similar drawbacks that have been reported before. A different approach is to use long-memory stochastic volatility models, although these specifications are not an effective alternative due to the complexity regarding their estimation and because they are based on fractional integrated roots.

On the other hand, Campos et al. (2017) model the oil volatility with the basic heterogeneous autoregressive (HAR) model of Corsi (2009) and find that it fits the oil volatility index well. They also include in the basic HAR model financial and macroeconomic variables, as in Fernandes, Medeiros, and Scharth (2014). Contrary to Campos et al. (2017), the existence of an asymmetric response of the index is tested in this paper by fitting the asymmetric extension of the HAR model proposed by Corsi, Audrino, and Reno (2012) named HARL and new proposals based on alternative oil price asymmetric measures instead of negative oil returns. HAR models with leverage have been successfully used for forecasting the volatility of stock market returns; see, among others, Liu and Maheu (2009), Scharth and Medeiros (2009), Asai, McAleer, and Medeiros (2012), Corsi and Renò (2012), Byun and Kim (2013), Patton and Sheppard (2015) and Choi and Shin (2018, 2019). For instance, Choi and Shin (2019) propose a parametric quantile forecast strategy that focuses on forecast intervals and value-at-risk (VaR) forecasts. They show that the HARL model with errors following a skewed Student *t* exponential GARCH (EGARCH) produce out-of-sample forecast improvements of the forecast intervals and VaR forecasts of the realized volatility, and volatility indices. Furthermore, Park, Choi, and Shin (2017) forecast the VaR of volatility indices, such as the VIX, the VKOSPI, and the OVX. They show that features like long memory, conditional heteroskedasticity, asymmetry, and fat tails are crucial for obtaining accurate forecasts. In their work, the best forecasting performance is achieved by a HAR model with errors following a skewed Student *t* GARCH model. More details on the methodology and models for forecasting the realized volatility can be found in two recent surveys by Bucci (2018) and Shin (2018).

The new extensions proposed in this paper are based on the work of Hamilton (2003), the net oil price measures. Their aim is to capture how unsettling an increase/decrease in the price of oil is likely to be for the spending decisions of consumers and firms; see Ramos and Veiga (2011) for applications of this concept. According to Hamilton (2003), it is more appropriate to compare the current oil price with its value over the last year than during the previous day. Other alternative measures are those proposed by Lee, Ni, and Ratti (1995), the scaled oil price measures. According to Lee et al. (1995), what matters is how surprising an oil price decrease is for the observed changes; that is, an unexpected oil price change will have less impact when conditional variances are high because much of the change in oil prices will be regarded as transitory.

According to Chen et al. (2015), OVX allows for predicting the future oil prices and consequently hedging from potentially severe shocks; see Carr and Wu (2006), Konstantinidi, Skiadopoulos, and Tzagkaraki (2008), and

[1]For detailed information see http://www.cboe.com/publish/regcir/rg12-052.pdf.

Clements and Fuller (2012) for a similar conclusion in the context of VIX. Also, Giot (2005) and Jiang and Tian (2005) report results confirming a good forecasting performance of models based on VIX, and Corrado and Miller (2005) indicate that future volatility is better predicted with VIX rather than with historical volatility. Therefore, the second aim of this paper is to analyze the forecasting performance of the new asymmetric proposals and benchmarks in terms of both point and density forecasts. Density forecasts are very important because they provide a full description of the uncertainty of a point forecast together with risks on the upside and downside and the probability of extreme events; see Tay and Wallis (2000). They are required especially by central banks and in the field of finance, such as the area of financial risk management that is dedicated to report density forecasts of portfolio values; see Diebold, Gunther, and Tay (1998). In this paper, point forecasts are evaluated with the model confidence set (MCS) by Hansen, Lunde, and Nason (2011) and the loss functions used are the mean squared error (MSE) and the Qlike by Patton (2011). Density forecasts are evaluated with the Generalized-AutoContourR (G-ACR) tests of González-Rivera and Sun (2015).

Our main empirical results are as follows: The inclusion of asymmetry in the HAR model, either in the conditional mean or in the conditional variance, is crucial for accurately forecasting the logarithm of the OVX. The asymmetric HAR models presented in this paper are quite complete since they capture the long memory, the asymmetric response of the volatility to oil price changes, and the conditional heteroskedasticity. Regarding the out-of-sample performance, the HAR-EGARCH and the HAR-EGARCH with the scaled measure report the lowest loss function values in terms of point forecasts, whereas the models that perform better in terms of density forecasting are the HAR-EGARCH and the S-HAR-GARCH.

All in all, the contributions of this paper are several: First, we propose new asymmetric specifications for the logarithm of the OVX. Second, we test their in-sample and out-of-sample performances, and third, we compare them with those obtained with benchmark models when forecasting the conditional mean and the density of the logarithm OVX.

The rest of the paper is organized as follows: Section 2 focuses on the descriptive and exploratory analysis of the OVX. Section 3 presents the models that exist in the literature and the new proposals. Section 4 reports the empirical results—that is, the in and out-of-sample analyses. Section 5 concludes the paper.

2 | EMPIRICAL FEATURES OF OVX

The sample ranges from May 5, 2007, to April 17, 2017, with a total of 2,502 daily observations. Figure 1 plots the series of OVX and its logarithm (log-OVX). Regarding the OVX, we observe that it fluctuates from 20 points to almost 100. The low values correspond to periods of low volatility and the large values of the index to periods of high volatility. The high volatility can be seen in 2009, around 2012, and in 2015. The index is more stable between 2010 and 2011 and at the end of 2012 and 2014. OVX can be interpretable as a measure of investors' fear and it seems to peak in periods of turmoil, such as the bankruptcy of Lehmann Brothers, the successive credit crunch, and the global financial crisis; see Whaley (2000) for a similar interpretation of the VIX. Taking logarithms has the advantage of decreasing the variation of OVX. Therefore, hereinafter, we analyze the series in logs.

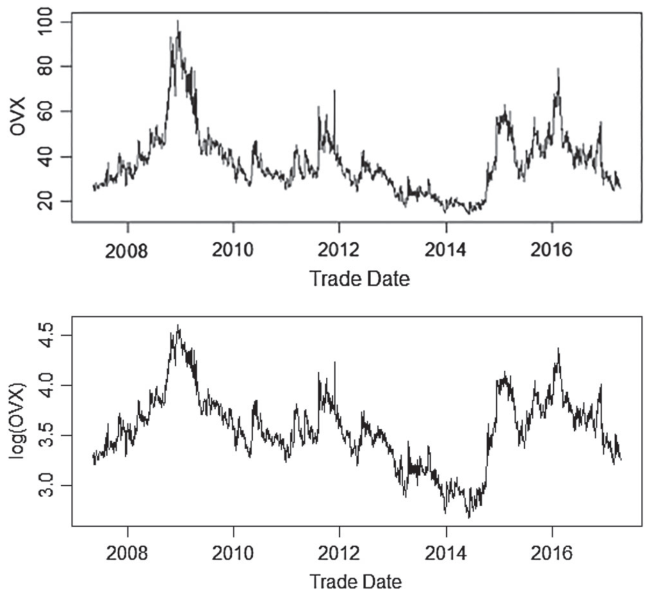

FIGURE 1 Daily OVX (top) and log-OVX (bottom) from May 5, 2007 till April 17, 2017

2.1 | Descriptive analysis

Table 1 summarizes the main sample statistical characteristics of the log-OVX. We observe that it has positive skewness and kurtosis close to that of a normal variable. The results of the tests of normality (Jarque–Bera and GSK by Lobato and Velasco (2004), for serially correlated data) confirm that log-OVX is normal.

Figure 2 plots the autocorrelation function (ACF) of the log-OVX until order 275. It suggests that it has a typical behavior of a series with long memory since the ACF decays slowly toward zero. Furthermore, the cross-correlations are positive, which means that an increase in the log-OVX today leads to an increase in its volatility tomorrow (see Figure 3). This phenomenon is known as volatility feedback. It is often presented in equity assets and can be part of the reason for the existence of an asymmetric response of oil volatility to oil price changes; see Wu (2001) for a helpful explanation of this phenomenon.

TABLE 1 Descriptive statistics

Sample statistics	
Mean	3.558
Standard deviation	0.357
Skewness	0.091
Kurtosis	3.100
Jarque & Bera	4.559 (0.992)
GSK	0.306 (0.858)

Note. Sample statistics (mean, standard deviation, skewness, kurtosis) and tests of normality for log-OVX. *p*-values (in parentheses).

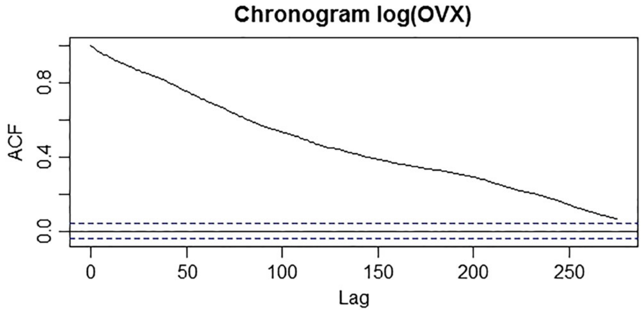

FIGURE 2 Sample autocorrelation function of the log-OVX [Colour figure can be viewed at wileyonlinelibrary.com]

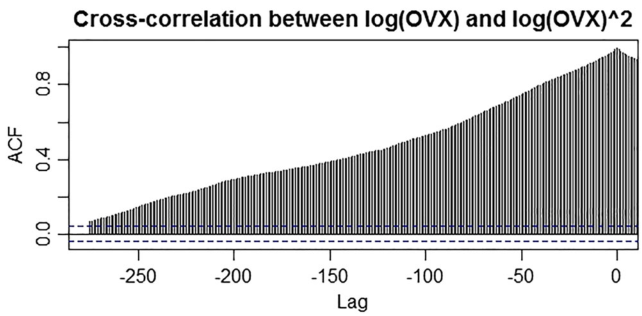

FIGURE 3 Sample cross-correlation function between log-OVX (lagged values) and log-OVX^2 [Colour figure can be viewed at wileyonlinelibrary.com]

Table 2 reports the results of the unit root and long-memory tests. Different tests have been used. The first is the traditional augmented Dickey–Fuller (ADF), whose null hypothesis is the existence of a unit root. The second is that proposed by Phillips and Perron (1988) (hereinafter PP), whose null hypothesis coincides with that of the ADF test, but considers the possibility of existence of serial correlation and heteroskedasticity in the log-OVX. In both tests, the alternative hypothesis is a stationary process with drift.[2] Regarding the long-memory tests, we have used a modified version of the R/S test proposed by Lo (1991) and a modified version of the V/S test proposed by Giraitis, Kokoszka, Leipus, and Teyssière (2000). In both tests the null hypothesis is that the process has short memory against the alternative that it has long memory. The difference is that the V/S test takes into consideration different lags, which are chosen by the BIC. The results of the unit root tests suggest that log-OVX is stationary (see Table 2). Looking at the long-memory test results, we see that both reject the null of short memory, which implies that log-OVX has long memory; see Fernandes et al. (2014) for similar results regarding the log-VIX.

TABLE 2 Unit root and long memory tests

Test	Test statistic	5% critical value
ADF	−3.070	−2.860
PP	−2.870	−2.860
R/S	2.190	1.860
V/S	7.570 (2)	1.360
	4.570 (5)	
	2.11 (10)	

Note. The table reports the statistics of different unit root and long-memory tests: ADF, PP, modified R/S and modified V/S. The values (in parentheses) for the V/S are the number of lags considered. The third column provides the tests' critical values at 5% significance level.

3 | HAR MODELS

This section presents the HAR models that coexist in the literature and proposes new extensions of the basic model that include measures of oil asymmetry that take advantage of more refined ways of measuring the impact of oil price changes on the economy. The majority of the studies use the traditional dummy variable approach (e.g., Basher & Sadorsky, 2006; Nandha & Faff, 2008; Sadorsky, 2008). In this paper, we follow Hamilton (1996), who proposes a measure that compares the current price of oil with the previous reference value; see Ramos and Veiga (2011) for the implementation of these measures. This measure is named net oil price decrease (NOPD) and seeks to capture the exogenous component of price variation. Also, Lee et al. (1995) observe that in periods of turbulence the effects of oil price changes are smaller than in periods of stability of these prices. In order to measure this properly they propose an asymmetric measure named scaled oil price decrease (SOPD).

3.1 | The basic HAR model

Corsi (2009) proposes the HAR model, which is more interpretable than the ARFIMA model and captures well the long memory of volatility. It is described as an "additive

[2]The R package "urca" is used to run the unit root tests. The number of lags used in the ADF is selected using the Bayesian information criterion (BIC).

cascade model" in which the actions of market participants are considered through the volatility components. By substituting the partial volatilities recursively, the model achieved with this process is autoregressive and its components are the volatility in specific time moments. In this paper we consider the following specification of the HAR model in which the time horizons are daily, weekly, monthly, and quarterly; that is, $t = (1, 5, 22, 66)$, respectively:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 \bar{y}_{t-1:5} + \phi_3 \bar{y}_{t-1:22} + \phi_4 \bar{y}_{t-1:66} + \varepsilon_{0t}, \quad (1)$$

where $y_t = \ln(\text{OVX}_t)$, $\bar{y}_{t:i} = \frac{1}{i}\sum_{j=0}^{i-1} y_{t-j}$, and $\{\varepsilon_{0t}\}_{t=1}^{T}$ is a sequence of independent white noise disturbances.

3.2 | Asymmetric model variations

It is well known that volatility reacts asymmetrically to shocks. That is, negative shocks have more impact on the volatility than do positive shocks of the same magnitude. This effect is also known in the literature as leverage.

3.2.1 | The HAR leverage model

Corsi et al. (2012) propose an extension of the HAR model that includes, besides the traditional regressors of the HAR model, the negative oil returns at daily, weekly, monthly, and quarterly frequency.[3] The model is named HAR leverage (HARL) and is given by

$$\begin{aligned} y_t = {} & \beta_0 + \beta_1 y_{t-1} + \beta_2 \bar{y}_{t-1:5} + \beta_3 \bar{y}_{t-1:22} \\ & + \beta_4 \bar{y}_{t-1:66} + \gamma_1 r^-_{t-1} + \gamma_2 r^-_{t-1:5} + \gamma_3 r^-_{t-1:22} + \gamma_4 r^-_{t-1:66} + \varepsilon_{1t}, \end{aligned} \quad (2)$$

where $\bar{y}_{t:i}$ is given as above, daily oil returns are defined as $r_t = p_t - p_{t-1}$, p_t is the logarithm of the oil price at time t, past aggregated negative returns are given by $r^-_{t-1:t-h} = \frac{1}{h}(r_{t-1} + \cdots + r_{t-h}) I_{(r_{t-1}+\cdots+r_{t-h}<0)}$ with $I_{\{.\}}$ an indicator function that takes the value one if $r_{t-1} + \cdots + r_{t-h} < 0$ and zero otherwise, and ε_{1t} is defined as ε_{0t} in Equation 1. Some implementations of the model for the high-frequency data of stock markets can be found in Chang and McAleer (2009), Louzis, Xanthopoulos-Sisinis, and Refenes (2012), Wang and Huang (2012), and Wang, Wu, and Xu (2015), among others.

[3]In a first step, we have also considered positive oil returns because an increase in oil prices will negatively affect the economy and stock markets of oil-importing countries. Note that oil is a global input. Therefore, positive oil price changes are considered bad news, and consequently the asymmetry is expected to be of opposite direction to that found for equity; see Ramos and Veiga (2011, 2013) for some results on oil price shocks. However, these effects are stronger when we study the impact of oil price changes on stock markets and, consequently, in this paper we have considered that oil negative returns affect more the log-OVX.

3.2.2 | The Net-HAR and Scaled-HAR models

We propose two extensions of the HAR model. The first includes the NOPD, while the second takes into consideration the SOPD. Before defining the models, we present the two oil asymmetric measures. The NOPD is defined as

$$\text{NOPD}_t = \min\{0, \ln(\text{oil}_t) - \ln[\max(\text{oil}_{t-1}, \cdots, \text{oil}_{t-252})]\}, \quad (3)$$

where NOPD_t is the net price decrease at time t. NOPD can be interpreted as the extreme negative returns over the last year.

The SOPD is different from the NOPD because its main aim is to measure how surprising an oil price change decrease is. It is known that an unexpected oil price change will have less impact when the volatility is high because it will be regarded as transitory. The SOPD is defined as

$$\text{SOPD}_t = \min(0, \hat{\epsilon}^*_t), \quad (4)$$

where $\hat{\epsilon}^*_t$ is the oil standardized return at time t. The estimated conditional variance is obtained by fitting a Student EGARCH(1, 1) model to the daily oil returns. This model outperforms a set of benchmarks (GARCH and GJR both with normal, Student t and skewed Student t distributions and EGARCH with normal and skewed Student t distributions) according to the Akaike information criterion (AIC) and BIC.

Next, we propose extensions of the HAR model that include the NOPD and SOPD instead of the conventional measures of asymmetry. The first model is named Net-HAR (N-HAR) and is given by

$$\begin{aligned} y_t = {} & \beta'_0 + \beta'_1 y_{t-1} + \beta'_2 \bar{y}_{t-1:5} + \beta'_3 \bar{y}_{t-1:22} + \beta'_4 \bar{y}_{t-1:66} \\ & + \gamma'_1 \overline{\text{NOPD}}_{t-1} + \gamma'_2 \overline{\text{NOPD}}_{t-1:5} \\ & + \gamma'_3 \overline{\text{NOPD}}_{t-1:22} + \gamma'_4 \overline{\text{NOPD}}_{t-1:66} + \varepsilon_{2t}, \end{aligned} \quad (5)$$

where $\overline{\text{NOPD}}_{t:i} = \frac{1}{i}\sum_{j=0}^{i-1} \text{NOPD}_{t-j}$, while the second extension is named scaled-HAR (S-HAR) and is given by

$$\begin{aligned} y_t = {} & \beta''_0 + \beta''_1 y_{t-1} + \beta''_2 \bar{y}_{t-1:5} + \beta''_3 \bar{y}_{t-1:22} + \beta''_4 \bar{y}_{t-1:66} \\ & + \gamma''_1 \overline{\text{SOPD}}_{t-1} + \gamma''_2 \overline{\text{SOPD}}_{t-1:5} + \gamma''_3 \overline{\text{SOPD}}_{t-1:22} \\ & + \gamma''_4 \overline{\text{SOPD}}_{t-1:66} + \varepsilon_{3t}, \end{aligned} \quad (6)$$

where $\overline{\text{SOPD}}_{t:i} = \frac{1}{i}\sum_{j=0}^{i-1} \text{SOPD}_{t-j}$ and, as before, $\{\varepsilon_{2t}\}_{t=1}^{T}$ and $\{\varepsilon_{3t}\}_{t=1}^{T}$ are sequences of independent white noise errors.

4 | EMPIRICAL RESULTS

We analyze the in-sample and out-of-sample performances of the four models presented in Section 3. The in-sample analysis consists of fitting the models to the log-OVX and studying their goodness-of-fit, whereas the out-of-sample exercise consists of forecasting the log-OVX one-day-ahead and analyzing the results with predictive ability and G-ACR tests.

4.1 | In-sample results

Table 3 reports the estimates of the HAR model and its extensions. All models are estimated by ordinary least squares (OLS) and robust standard errors are reported in parentheses.[4] Regarding the estimation of the basic HAR model, we observe that the frequencies of the aggregate log-OVX are statistically significant only to the daily and monthly frequencies. For the HARL, we see that monthly and quarterly oil-negative returns are statistically significant at 5% and 10% significance levels, respectively. The estimated signs of the coefficients are negative, which means that negative returns affect the oil volatility positively, as is expected. Focusing on the new leverage HAR models, we observe that the variables related to NOPD do not show statistically significant coefficients, while SOPD is statistically significant only at the 10% level for the 1-day and quarterly frequencies. The adjusted R^2s are very similar among the HAR-type models, while the log-likelihoods reveal that the HARL fits the log-OVX the best. Finally, we compute the autocorrelations of squared residuals of order 1, 10, and 100 and we see that they are statistically significant, which suggests the presence of heteroskedasticity in the residuals.

Given the above results, we extend the HAR models given by Equations (1), (2), (5), and (6) by considering conditional heteroskedastic errors, ε_{it}, and $i = 0, 1, 2, 3$:

$$\varepsilon_{it} = \sigma_t z_t, \tag{7}$$

$$\sigma_t^2 = \omega_0 + \omega_1 \varepsilon_{it-1}^2 + \omega_2 \sigma_{t-1}^2, \tag{8}$$

or

$$\ln \sigma_t^2 = \omega_0' + \omega_1' [|z_{t-1}| - E(|z_{t-1}|)] + \omega_2' \ln \sigma_{t-1}^2 + \gamma z_{t-1}, \tag{9}$$

where z_t follows the Hansen (1994) skewed Student t $(0, 1, \nu, \lambda)$ distribution with zero mean and unit variance, and ν corresponds to the degrees of freedom and λ to the skewness.[5] Consequently, the conditional density of y_t is

$$g(y_t; \mu_t, \sigma_t, \nu, \lambda) = \begin{cases} \frac{1}{\sigma_t} bc \left[1 + \frac{1}{\nu-2}\left(\frac{bz_t + a}{1-\lambda}\right)^2\right]^{-\frac{\nu+1}{2}} & \text{if } z_t < -a/b, \\ \frac{1}{\sigma_t} bc \left[1 + \frac{1}{\nu-2}\left(\frac{bz_t + a}{1+\lambda}\right)^2\right]^{-\frac{\nu+1}{2}} & \text{if } z_t \geq -a/b, \end{cases} \tag{10}$$

[4]We use the covariance estimator robust to heteroskedasticity of White (1980).

[5]Several distributions have been assumed for the errors of the GARCH or EGARCH models, including the normal-inverse Gaussian distribution (NIG) as suggested by Corsi, Mittnik, Pigorsch, and Pigorsch (2008). However, models' goodness-of-fit criteria (AIC and BIC) suggest the use of the skewed Student t distribution of Hansen instead. Results are available from the authors upon request.

TABLE 3 HAR-type models estimation

	HAR	HARL	N-HAR	S-HAR
Intercept	0.026**	0.045***	0.026*	0.025**
	(0.012)	(0.014)	(0.014)	(0.012)
log-OVX 1 day	0.911***	0.893***	0.889***	0.893***
	(0.036)	(0.040)	(0.041)	(0.042)
log-OVX 5 days	0.050	0.051	0.057	0.058
	(0.041)	(0.043)	(0.046)	(0.044)
log-OVX 22 days	0.041**	0.031	0.041*	0.030
	(0.020)	(0.021)	(0.024)	(0.022)
log-OVX 66 days	−0.009	0.011	0.006	0.006
	(0.010)	(0.011)	(0.013)	(0.011)
1-day (−) return		−0.083		
		(0.087)		
5-day (−) return		0.108		
		(0.200)		
22-day (−) return		−0.968**		
		(0.485)		
66-day (−) return		−1.125*		
		(0.671)		
NOPD 1 day			−0.048	
			(0.045)	
NOPD 5 days			−0.017	
			(0.057)	
NOPD 22 days			0.054	
			(0.038)	
NOPD 66 days			0.008	
			(0.019)	
SOPD 1 day				−0.003*
				(0.002)
SOPD 5 days				0.003
				(0.004)
SOPD 22 days				−0.011
				(0.010)
SOPD 66 days				−0.039*
				(0.021)
R^2	0.982	0.982	0.983	0.982
Adjusted R^2	0.982	0.982	0.983	0.982
LL	3,904.9	3,914.9	3,641.0	3,910.5
	Residuals			
ρ_1^2	0.305	0.299	0.302	0.301
	(0.000)	(0.000)	(0.000)	(0.000)
ρ_{10}^2	0.040	0.038	0.039	0.037
	(0.000)	(0.000)	(0.000)	(0.000)
ρ_{100}^2	−0.013	−0.017	−0.015	−0.015
	(0.000)	(0.000)	(0.000)	(0.000)

Note. The table reports the estimates of parameters together with robust standard errors (in parentheses), R^2, log-likelihood (LL) at the optimum, correlations of squared residuals of order k (ρ_k^2) and p-values of the Ljung–Box type test statistic (in parentheses). Asterisks indicate significance at ***1%, **5%, and *10%.

where $z_t = \frac{y_t - \mu_t}{\sigma_t}$, μ_t is the conditional mean given by the HAR equations in Equations (1), (2), (5), and (6), and σ_t is the conditional variance given by the GARCH or EGARCH model, Equations (7)–(8) or (7) and (9), respectively. The constants a, b, and c are given by

$$a = 4\lambda c\left(\frac{\nu-2}{\nu-1}\right), \quad b^2 = 1 + 3\lambda^2 - a^2,$$
$$c = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi(\nu-2)}\Gamma(\frac{\nu}{2})}.$$

The density in Equation (10) is defined for $2 < \nu < \infty$ and $-1 < \lambda < 1$. When $\lambda > 0$ the distribution is skewed to the right. When $\lambda = 0$ and $\nu = \infty$ it reduces to a normal density. Otherwise it is skewed to the left.

To guarantee the positiveness and the stationarity of the conditional variance of the GARCH model in Equations (7)–(8), we impose that $\omega_0 > 0$, $\omega_1 \geq 0$ and $\omega_2 \geq 0$, and $\omega_1 + \omega_2 < 1$, respectively. Regarding the EGARCH model in Equations (7) and (9), we have only to impose that $|\omega_2'| < 1$ to guarantee that the conditional variance is stationary. The HAR-GARCH/EGARCH skewed Student-type models are estimated by maximizing the log-likelihood function of Equation (10) with respect to the parameters of μ_t and σ_t^2 and the parameters ν and λ; that is:

$$\log L = T\ln(b) + T\ln(c) - \left(\frac{\nu+1}{2}\right) \sum_{t=1}^{T} \ln\left(1 + \frac{d_t^2}{(\nu-2)}\right) - T\ln(\sigma_t), \quad (11)$$

where $d_t = \frac{bz_t + a}{1 - \lambda s}$ and s is a sign dummy taking the value of 1 if $bz_t + a < 0$ and $s = -1$ otherwise.

The full estimation results appear in Tables 4 and 5. We see in Table 4 that, regardless of the model estimated, log-OVX depends on the previous day and month, with coefficients statistically significant at the 1% or 5% level. Moreover, the parameters of the conditional variance and those of the skewed Student t distribution are statistically significant for all HAR-GARCH-skewStudent-type models. The log-likelihoods increase substantially, suggesting an important increase in the goodness-of-fit in comparison to the OLS models. In terms of log-likelihood, the HARL-GARCH-skewStudent is still the best, although the likelihoods are quite similar among the models. Finally, in Table 5 we observe that the parameter that captures the leverage effect in the conditional variance of the errors (γ) is positive and statistically significant regardless of the model, and that the log-likelihoods in Table 5 are greater than those in Table 4, suggesting that the inclusion of asymmetry in the variance equation can also lead to improvements in models' fit.

TABLE 4 HAR-GARCH-skewStudent-type model estimation

	HAR	HARL	N-HAR	S-HAR
Conditional mean				
Intercept	0.013	0.021**	0.010	0.012
	(0.008)	(0.009)	(0.011)	(0.009)
log-OVX 1 day	0.940***	0.929***	0.939***	0.932***
	(0.020)	(0.021)	(0.021)	(0.023)
log-OVX 5 days	0.014	0.010	0.004	0.019
	(0.025)	(0.024)	(0.026)	(0.027)
log-OVX 22 days	0.047***	0.046***	0.055***	0.043**
	(0.016)	(0.015)	(0.018)	(0.016)
log-OVX 66 days	−0.004	0.008	−0.001	0.002
	(0.009)	(0.008)	(0.011)	(0.010)
1-day (−) return		−0.059		
		(0.060)		
5-day (−) return		0.171		
		(0.147)		
22-day (−) return		−1.067***		
		(0.248)		
66-day (−) return		−0.241		
		(0.176)		
NOPD 1 day			0.007	
			(0.034)	
NOPD 5 days			−0.043	
			(0.042)	
NOPD 22 days			0.041	
			(0.026)	
NOPD 66 days			−0.004	
			(0.012)	
SOPD 1 day				−0.002
				(0.002)
SOPD 5 days				0.002
				(0.004)
SOPD 22 days				−0.005
				(0.009)
SOPD 66 days				−0.012
				(0.017)
Conditional variance				
ω_0	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
ω_1	0.094***	0.091***	0.091***	0.093***
	(0.022)	(0.020)	(0.021)	(0.021)
ω_2	0.805***	0.809***	0.807***	0.805***
	(0.045)	(0.036)	(0.044)	(0.043)
ν	4.431***	4.441***	4.458***	4.412***
	(0.417)	(0.388)	(0.403)	(0.376)
λ	0.193***	0.196***	0.193***	0.192***
	(0.031)	(0.030)	(0.030)	(0.028)
LL	3,976.1	3,982.5	3,978.2	3,977.8
	Residuals			
ρ_2^2	0.004	0.005	0.004	0.005
	(0.218)	(0.206)	(0.192)	(0.216)
ρ_{10}^2	0.009	0.010	0.009	0.010
	(0.857)	(0.853)	(0.845)	(0.846)
ρ_{100}^2	−0.009	−0.010	−0.010	−0.008
	(0.181)	(0.159)	(0.177)	(0.218)

Note. The table reports the estimates of parameters together with the standard errors (in parentheses), log-likelihood (LL) at the optimum, correlations of squared residuals of order k (ρ_k^2) and corresponding p-values computed by the corrected test of Li and Mak (1994) (in parentheses). Asterisks indicate significance at ***1%, **5%, and *10%.

TABLE 5 HAR-EGARCH-skewStudent-type model estimation

	HAR	HARL	N-HAR	S-HAR
Conditional mean				
Intercept	0.017**	0.023***	0.010	0.016**
	(0.007)	(0.009)	(0.009)	(0.008)
log-OVX 1 day	0.959***	0.948***	0.959***	0.951***
	(0.020)	(0.022)	(0.019)	(0.021)
log-OVX 5 days	−0.003	−0.006	−0.013	0.003
	(0.024)	(0.026)	(0.019)	(0.025)
log-OVX 22 days	0.044***	0.047***	0.052***	0.041**
	(0.016)	(0.018)	(0.015)	(0.016)
log-OVX 66 days	−0.005	0.004	−0.001	0.000
	(0.008)	(0.009)	(0.009)	(0.009)
1-day (−) return		−0.059		
		(0.060)		
5-day (−) return		0.140		
		(0.209)		
22-day (−) return		−0.941**		
		(0.386)		
66-day (−) return		−0.095		
		(0.334)		
NOPD 1 day			0.009	
			(0.030)	
NOPD 5 days			−0.041	
			(0.036)	
NOPD 22 days			0.036	
			(0.024)	
NOPD 66 days			−0.001	
			(0.011)	
SOPD 1 day				−0.002
				(0.002)
SOPD 5 days				0.001
				(0.004)
SOPD 22 days				−0.003
				(0.009)
SOPD 66 days				−0.010
				(0.016)
Conditional variance				
ω'_0	−0.851***	−0.893***	−0.869***	−0.855***
	(0.201)	(0.268)	(0.160)	(0.161)
ω'_1	0.179***	0.189***	0.172***	0.180***
	(0.033)	(0.036)	(0.032)	(0.032)
ω'_2	0.861***	0.854***	0.858***	0.860***
	(0.033)	(0.043)	(0.026)	(0.026)
γ	0.116***	0.108***	0.121***	0.115***
	(0.026)	(0.026)	(0.024)	(0.024)
ν	4.458***	4.454***	4.492***	4.439***
	(0.406)	(0.424)	(0.380)	(0.403)
λ	0.208***	0.208***	0.209***	0.206***
	(0.030)	(0.029)	(0.029)	(0.030)
LL	3,985.1	3,990.0	3,987.5	3,986.3
Residuals				
ρ_1	0.001	0.001	0.001	0.001
	(0.396)	(0.400)	(0.367)	(0.398)
ρ_{10}	0.027	0.028	0.026	0.027
	(0.825)	(0.795)	(0.822)	(0.807)
ρ_{100}	−0.011	−0.011	−0.011	−0.010
	(0.401)	(0.279)	(0.379)	(0.413)

(Continous)

TABLE 5 *(Continued)*

Note. The table reports the estimates of parameters together with the standard errors (in parentheses), log-likelihood (LL) at the optimum, correlations of squared residuals of order k (ρ_k^2) and corresponding p-values computed by the corrected test of Li and Mak (1994) (in parentheses). Asterisks indicate significance at ***1%, **5% and *10%.

In the next subsection we analyze the forecasting performance of each HAR-GARCH/EGARCH-skewStudent model.

4.2 | Forecasting results

We use a fixed rolling window scheme with 1,000 log-OVX observations. In total, for all models, we obtain 1,339 out-of-sample one-step-ahead forecasts.

The models' forecasting performance is evaluated using either the MCS procedure by Hansen et al. (2011) and programmed in R by Catania and Bernardi (2015) and the G-ACR test proposed by González-Rivera and Sun (2015). Hansen et al.'s procedure consists of a sequence of statistical tests to construct a set of models called "superior set model" (SSM). Models in the SSM have statistically the same predictive ability and are ranked by the value of the loss function. The test statistic is calculated for an arbitrary loss function and evaluates point forecasts. On the other hand, the G-ACR test evaluates the adequacy of the conditional forecast density model based on the probability integral transforms (PITs). Its advantage is the possibility of obtaining a graph named "autocontour" based on the PITs that suggests the reasons of the failure of the model.

4.3 | Point forecasts

Table 6 displays the SSM of the log-OVX forecasting models under study. The loss functions are the squared error (SE) that is defined as $\left(\ln(\text{OVX}_{t+1}) - \ln(\widehat{\text{OVX}}_{t+1})\right)^2$ and the Qlike of Patton (2011), which is a robust loss function to the presence of noise in the volatility proxy and is defined as Qlike $\equiv \frac{\ln(\text{OVX}_{t+1})}{\ln(\widehat{\text{OVX}}_{t+1})} - \ln\left(\frac{\ln(\text{OVX}_{t+1})}{\ln(\widehat{\text{OVX}}_{t+1})}\right) - 1$. According to both loss functions and confidence levels, the HAR-EGARCH always ranks first, while the S-HAR-EGARCH ranks second. On the other hand, the HAR-GARCH always ranks last. However, all the models have a similar predictive ability since all belong to the SSM.

TABLE 6 Model confidence set results

	Loss functions			
	95%		80%	
Model	**SE**	**QLIKE**	**SE**	**QLIKE**
HAR-GARCH	8	8	8	8
HARL-GARCH	3	4	3	4
N-HAR-GARCH	5	6	5	6
S-HAR-GARCH	7	7	7	7
HAR-EGARCH	1	1	1	1
HARL-EGARCH	4	3	4	3
N-HAR-EGARCH	6	5	6	5
S-HAR-EGARCH	2	2	2	2
p-value	0.788	0.815	0.782	0.804

Note. The table reports the rankings of the log-OVX forecasters with different loss functions.The statistical tests are done at 95% and 80% confidence levels. 5,000 bootstrap samples are used.

4.4 | Density forecasts

For evaluating the density forecasts we use the G-ACR test of González-Rivera and Sun (2015), which is based on PITs. The test is based on the idea that if the proposed conditional forecast density for y_t—that is, $g(y_t|\Omega_{t-1})$—coincides with the true one $f(y_t|\Omega_{t-1})$ (H_0), then the sequence of PITs $\{u_t\}_{t=1}^{H}$, where $u_t = \int_{-\infty}^{y_t} g(v_t|\Omega_{t-1})dv_t$ and H is the total number of out-of-sample one-step-ahead forecasts, must be i.i.d. $U(0,1)$. The PITs can be plotted in the plane (u_t, u_{t-k}) such that the square with $\sqrt{\alpha_i}$-side and origin at $(0,0)$ contains $\alpha_i\%$ of observations, that is:

$$\text{G-ACR}_{k,\alpha_i} = \{B(u_t, u_{t-k}) \subset \Re^2 | 0 \leq u_t \leq \sqrt{\alpha_i} \text{ and } 0 \leq u_{t-k} \leq \sqrt{\alpha_i}, \text{s.t. : } u_t \times u_{t-k} \leq \alpha_i\}. \tag{12}$$

The proportion of PITs inside the cube defined in Equation (12) is given by

$$\hat{\alpha}_{k,i} = \frac{\sum_{t=k+1}^{H} I_t^{k,\alpha_i}}{H-k},$$

where I_t^{k,α_i} is an indicator that takes the value one if $(u_t, u_{t-k}) \in \text{G-ACR}_{k,\alpha_i}$ and zero otherwise. Therefore, the test statistic is given by

$$t_{k,\alpha_i} = \frac{\hat{\alpha}_{k,i} - \alpha_i}{\sigma_{\alpha_i}},$$

where σ_{α_i} is obtained by bootstrap. Note that this statistic is asymptotically standard normal distributed under the

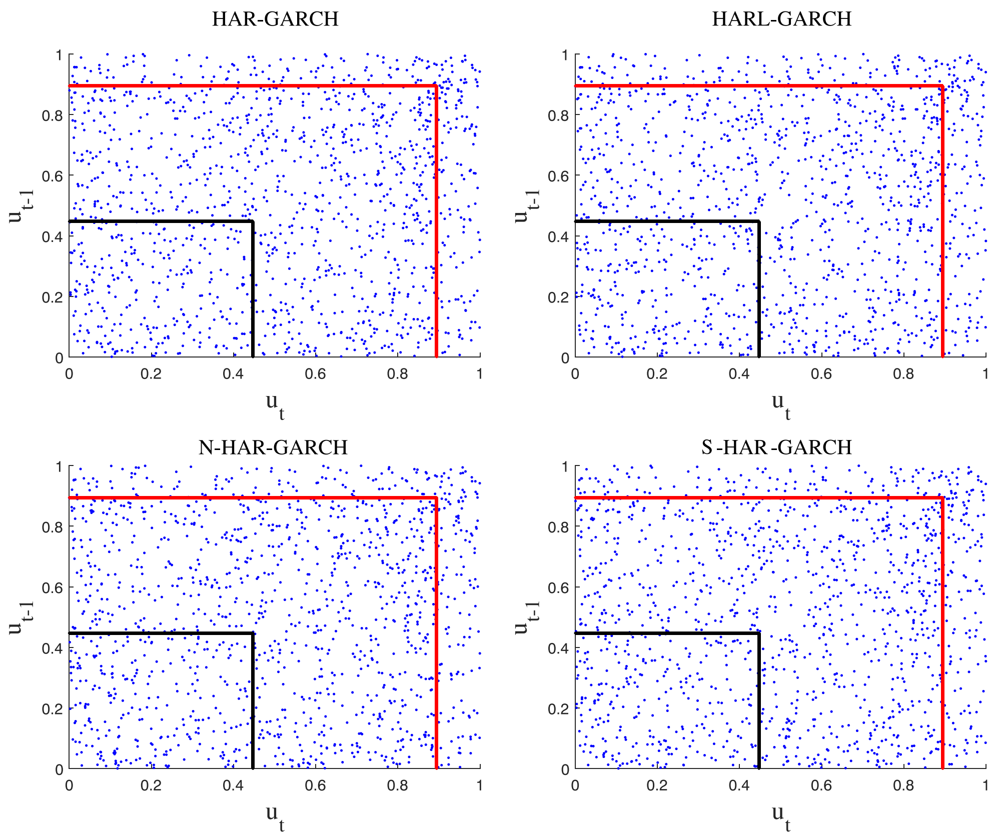

FIGURE 4 Pairs (u_t, u_{t-1}) and autocontours for the studied models. The PITS are obtained assuming that the errors follow a skewed Student t distribution; see Hansen (1994). $\text{ACR}_{20\%,1}$ corresponds to the black box and $\text{ACR}_{80\%,1}$ to the red box [Colour figure can be viewed at wileyonlinelibrary.com]

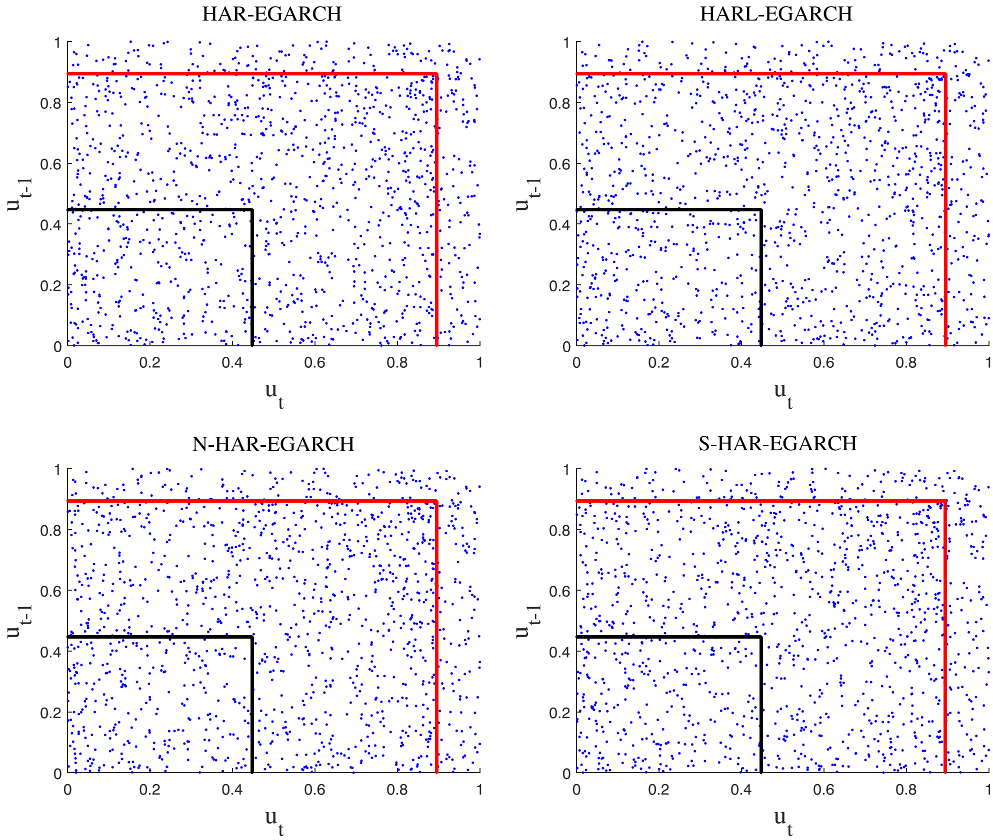

FIGURE 5 Pairs (u_t, u_{t-1}) and autocontours for the studied models. The PITS are obtained assuming that the errors follow a skewed Student *t* distribution; see Hansen (1994). $ACR_{20\%,1}$ corresponds to the black box and $ACR_{80\%,1}$ to the red box [Colour figure can be viewed at wileyonlinelibrary.com]

null hypothesis. Moreover, it is constructed for a single fixed autocontour, α_i, and a single fixed lag, k. One can go further and use a test statistic that considers a set of lags with a fixed autocontour. Let $L_{\alpha_i} = (\ell_{1,\alpha_i}, \cdots, \ell_{K,\alpha_i})'$ be a vector with element $\ell_{k,\alpha_i} = \widehat{\alpha}_{k,i} - \alpha_i$. Under the null hypothesis $L'_{\alpha_i}\Lambda^{-1}_{\alpha_i}L_{\alpha_i}$ is asymptotically χ^2_K distributed, where Λ_{α_i} is the asymptotic variance–covariance matrix. Finally, we can also consider a test statistic for several autocontours with a fixed lag. Let C_k be a vector such that $C_k = (c_{k,1}, \cdots, c_{k,C})'$ with $c_{k,i} = \widehat{\alpha}_{k,i} - \alpha_i$. The test statistic is $C'_k\Omega^{-1}_kC_k$, where Ω_k is the asymptotic variance–covariance matrix. The test statistic follows asymptotically a χ^2_C distribution (see González-Rivera & Sun, 2015, for details on these test statistics).

Figures 4 and 5 plot the PITs calculated for the selected models, which are computed assuming the density in Equation (10). The PITs of the HAR-GARCH-type models are more or less uniformly distributed except for a slight concentration at the area of the upper right corner. This may suggest that the proposed models are not fitting perfectly the true one-step-ahead forecast densities; see Mazzeu, González-Rivera, Ruiz, and Veiga (2017) for the interpretability of this graphical tool. Nevertheless, the concentration of points at the upper right corner seems to be lower for the HAR-EGARCH-type models.

Tables 7 and 8 report the results of the test statistics presented above. We have implemented a parametric bootstrap procedure for approximating the asymptotic variance and covariance matrix of the tests in order to incorporate the parameter uncertainty as suggested by González-Rivera and Sun (2015). The benefit of the tests is to distinguish among the HAR-GARCH/EGARCH-type models. Therefore, we compute the proportions $\widehat{\alpha}_{k,i}$, for $k = 1, \cdots, 5$ and 13 autocontours. Looking at both tables we observe that the coverages $(\hat{\alpha}_{1,i})$ are below the nominal levels for the autocontours 0.4, 0.5, 0.6, and 0.7. Regarding the individual t_{1,α_i} statistics, we observe that the null hypothesis is rejected for some of those autocontours, but the S-HAR-GARCH and HAR-EGARCH models have seldom rejected autocontours. The number of rejections increases when we consider the portmanteau statistics $L^5_{\alpha_i}$. The HAR-GARCH-type models show more rejections than their HAR-EGARCH counterparts. Note that the HAR-GARCH-type models report 7 or 8 rejected autocontours, while the

TABLE 7 G-ACR test results

α_i	0.01	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
HAR-GARCH													
$\widehat{\alpha}_{1,i}$	0.011	0.058	0.104	0.203	0.301	0.383	0.464	0.568	0.664	0.803	0.903	0.948	0.993
$\lvert t_{1,\alpha_i}\rvert$	|0.402|	|1.232|	|0.381|	|0.181|	|0.076|	|0.971|	|1.963|**	|1.804|*	|2.159|**	|0.233|	|0.247|	|0.287|	|0.625|
$L^5_{\alpha_i}$	6.034	5.883	9.116	13.561**	12.754**	13.177**	12.244**	7.763	17.294***	10.484*	10.729*	3.997	0.394
C^{13}_1	18.107												
HARL-GARCH													
$\widehat{\alpha}_{1,i}$	0.013	0.068	0.106	0.204	0.300	0.383	0.468	0.567	0.673	0.804	0.904	0.949	0.993
$\lvert t_{1,\alpha_i}\rvert$	|0.913|	|2.670|***	|0.616|	|0.286|	|0.028|	|0.972|	|1.771|*	|1.909|*	|1.597|	|0.283|	|0.372|	|0.102|	|0.616|
$L^5_{\alpha_i}$	8.092	10.943*	8.863	9.403*	16.431***	17.510***	13.516**	12.380**	15.573***	11.544**	8.722	4.094	0.383
C^{13}_1	22.223*												
N-HAR-GARCH													
$\widehat{\alpha}_{1,i}$	0.013	0.058	0.108	0.197	0.287	0.388	0.466	0.564	0.674	0.815	0.908	0.951	0.993
$\lvert t_{1,\alpha_i}\rvert$	|0.899|	|1.159|	|0.874|	|0.201|	|0.822|	|0.728|	|1.953|*	|2.075|**	|1.589|	|1.087|	|0.719|	|0.083|	|0.563|
$L^5_{\alpha_i}$	2.799	4.995	10.614*	12.963**	12.098**	20.891***	15.221***	12.109**	15.616***	12.082**	4.850	0.819	0.323
C^{13}_1	24.240**												
S-HAR-GARCH													
$\widehat{\alpha}_{1,i}$	0.013	0.058	0.110	0.199	0.306	0.387	0.466	0.570	0.673	0.805	0.903	0.951	0.993
$\lvert t_{1,\alpha_i}\rvert$	|0.839|	|1.123|	|0.957|	|0.082|	|0.351|	|0.711|	|1.858|*	|1.710|*	|1.602|	|0.330|	|0.254|	|0.083|	|0.692|
$L^5_{\alpha_i}$	5.776	5.399	11.995**	6.825	15.119***	15.140***	12.083**	8.131	12.403**	9.500*	10.848*	3.993	0.505
C^{13}_1	16.574												

Note. The table reports the results of the G-ACR tests for the HAR-GARCH-skewStudent-type models fitted to log-OVX. The tests are computed assuming that the errors follow Hansen's 1994 skewed Student *t* distribution. Asterisks indicate that H_0 is rejected at *10%, **5%, and ***1% levels of significance.

TABLE 8 G-ACR test results

α_i	**0.01**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**	**0.99**
HAR-EGARCH													
$\widehat{\alpha}_{1,i}$	0.016	0.061	0.108	0.202	0.295	0.386	0.471	0.566	0.677	0.813	0.912	0.960	0.993
$\lvert t_{1,\alpha_i}\rvert$	|1.887|*	|1.560|	|0.744|	|0.130|	|0.299|	|0.824|	|1.605|	|1.917|*	|1.362|	|0.867|	|0.996|	|1.153|	|0.583|
$L^5_{\alpha_i}$	9.940*	5.530	13.558**	11.753**	13.070**	7.860	9.613*	11.776**	5.655	3.154	5.186	4.671	0.344
C^{13}_1	17.486												
HARL-EGARCH													
$\widehat{\alpha}_{1,i}$	0.016	0.064	0.108	0.206	0.297	0.385	0.469	0.573	0.677	0.820	0.912	0.960	0.993
$\lvert t_{1,\alpha_i}\rvert$	|2.135|**	|1.990|**	|0.770|	|0.395|	|0.159|	|0.872|	|1.714|*	|1.501|	|1.352|	|1.339|	|0.981|	|1.143|	|0.561|
$L^5_{\alpha_i}$	12.976**	6.588	9.516*	11.285**	12.279**	11.484**	10.367*	7.534	7.960	4.035	6.394	4.619	0.318
C^{13}_1	24.140**												
N-HAR-EGARCH													
$\widehat{\alpha}_{1,i}$	0.016	0.059	0.102	0.195	0.291	0.380	0.468	0.560	0.674	0.825	0.912	0.957	0.993
$\lvert t_{1,\alpha_i}\rvert$	|1.939|*	|1.398|	|0.170|	|0.365|	|0.577|	|1.222|	|1.807|*	|2.362|**	|1.592|	|1.740|*	|1.012|	|0.775|	|0.481|
$L^5_{\alpha_i}$	6.430	5.166	10.977*	12.024**	16.307***	16.787***	11.513**	12.430**	7.595	6.270	2.805	1.937	0.234
C^{13}_1	33.049***												
S-HAR-EGARCH													
$\widehat{\alpha}_{1,i}$	0.015	0.066	0.111	0.200	0.296	0.380	0.466	0.567	0.674	0.813	0.913	0.958	0.993
$\lvert t_{1,\alpha_i}\rvert$	|1.622|	|2.330|**	|1.070|	|0.032|	|0.251|	|1.121|	|1.902|*	|1.854|*	|1.549|	|0.895|	|1.101|	|0.973|	|0.581|
$L^5_{\alpha_i}$	7.856	8.755	12.261**	10.179*	12.024**	11.935**	12.625**	11.639**	9.137	1.775	5.243	8.693	0.342
C^{13}_1	21.514*												

Note. The table reports the results of the G-ACR tests for the HAR-EGARCH-skewStudent-type models fitted to log-OVX. The tests are computed assuming that the errors follow Hansen's 1994 skewed Student *t* distribution. Asterisks indicate that H_0 is rejected at *10%, **5%, and ***1% levels of significance.

HAR-EGARCH-type models report 6 rejected autocontours in total. Considering the $L^5_{\alpha_i}$ statistics at the 5% and 1% significance levels, the N-HAR-GARCH is the most rejected model, while the HAR-EGARCH and HARL-EGARCH models are the least rejected models. Finally, regarding the portmanteau C^{13}_1 statistic, the HARL-GARCH/EGARCH, N-HAR-GARCH/EGARCH and the S-HAR-EGARCH are the only rejected models. On the other hand, the S-HAR-GARCH and HAR-EGARCH models show the lowest C^{13}_1 statistics.

All in all, the S-HAR-GARCH and HAR-EGARCH models with errors following a skewed Student *t* distribution seem to do a good job at forecasting the one-step-ahead density of the log-OVX. These results suggest that, apart from addressing asymmetry in the error distribution, it is also important to address asymmetry in the conditional mean or in the conditional variance of HAR models, if the aim is to forecast accurately the one-step-ahead density of the log-OVX.

5 | CONCLUSION

This paper examines the daily time series properties of the crude oil ETF volatility index. Our study shows that it has a dependence structure similar to the CBOE's market volatility index; that is, it is stationary but its autocorrelation function decays hyperbolically toward zero, which suggests that the index displays long memory. We model this feature of the oil volatility index using four different HAR models: a basic HAR model and its asymmetric version named HAR leverage as benchmark models and two new extensions that consider oil net and scaled measures of oil price changes instead of negative returns. The first extension seeks to measure how unsettling a decrease in the price of oil is likely to be for the spending decisions of consumers and firms, while the second measures how surprising an oil price decrease is for the observed oil price changes. Note that an unexpected oil price change will have less of an impact when conditional variances are high because much of the change in oil prices will be regarded as transitory. Given the existence of conditional heteroskedasticity in the residuals of the HAR-type models, we estimate HAR-GARCH/EGARCH-type models considering flexible distributions for the errors of the GARCH/EGARCH process.

The out-of-sample forecasting analysis shows that, in terms of point forecasts under both the squared error and Qlike loss functions, the HAR-EGARCH-skewStudent ranks first and the S-HAR-EGARCH-skewStudent ranks the second in all cases. However, none of the models is eliminated from the set of superior models, which indicates that all specifications show a similar predictive ability under the two loss functions. Regarding the density forecasting, the tests show that the models that are less rejected are the S-HAR-GARCH-skewStudent and the HAR-EGARCH-skewStudent.

All in all, the results confirm that the inclusion of asymmetry in the conditional mean or in the conditional variance of HAR models improves both the goodness-of-fit and the forecasting performance of the models, suggesting that the leverage is an important feature to be considered if the aim is an accurate modeling and forecasting of the logarithm of the crude oil ETF volatility index.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author, Helena Veiga, upon request.

ACKNOWLEDGEMENT

The first author acknowledges financial support from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) grant 88882.305837/2018-01. The second author acknowledges financial support from the Spanish Ministry of Economy and Competitiveness, research projects ECO2015-70331-C2-2-R and PGC2018-096977-B-I00, and from Fundação para a Ciência e a Tecnologia, grant UID/GES/00315/2019.

REFERENCES

Abadir, K., & Talmain, G. (2002). Aggregation, persistence and volatility in a macro model. *Review of Economic Studies*, *69*(4), 749–779.

Ankrim, E. M., & Hensel, C. R. (1993). Commodities in asset allocation: A real-ass et alternative to real estate? *Financial Analysts Journal*, *49*(3), 20–29.

Asai, M., McAleer, M., & Medeiros, M. C. (2012). Asymmetry and long memory in volatility modeling. *Journal of Financial Econometrics*, *10*(3), 495–512.

Baffes, J. (2007). Oil spills on other commodities. *Resources Policy*, *32*(3), 126–134.

Baillie, R. T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *74*, 3–30.

Basher, S. A., & Sadorsky, P. (2006). Oil price risk and emerging stock markets. *Global Finance Journal*, *17*(2), 224–251.

Belousova, J., & Dorfleitner, G. (2012). On the diversification benefits of commodities from the perspective of euro investors. *Journal of Banking and Finance*, *36*(9), 2455–2472.

Bhardwaj, G., & Swanson, N. R. (2006). An empirical investigation of the usefulness of ARFIMA models for predicting macroeconomic and financial time series. *Journal of Econometrics*, *131*(1), 539–578.

Bollerslev, T., & Mikkelsen, H. O. (1996). Modeling and pricing long memory in stock market volatility. *Journal of Econometrics*, *73*, 151–184.

Bordignon, S., Caporin, M., & Lisi, F. (2008). Periodic long-memory GARCH models. *Econometric Reviews*, *28*(1–3), 60–82.

Bucci, A. (2018). Forecasting realized volatility: A review. *Journal of Advanced Studies in Finance*, *8*(2), 94–138.

Byun, S. J., & Kim, J. S. (2013). The information content of risk-neutral skewness for volatility forecasting. *Journal of Empirical Finance*, *23*, 142–161.

Campos, I., Cortazar, G., & Reyes, T. (2017). Modeling and predicting oil VIX: Internet search volume versus traditional variables. *Energy Economics*, *66*, 194–204.

Carr, P., & Wu, L. (2006). A tale of two indices. *Journal of Derivatives*, *13*(3), 13–29.

Catania, L., & Bernardi, M. (2015). *MCS: Model Confidence Set Procedure*. R package version 0.1.1.

Chang, C.-L., & McAleer, M. (2009). Daily tourist arrivals, exchange rates and volatility for Korea and Taiwan. *Korean Economic Review*, *25*(2), 241–267.

Chen, Y., He, K., & Yu, L. (2015). The information content of OVX for crude oil returns analysis and risk measurement: Evidence from the Kalman filter model. *Annals of Data Science*, *2*(4), 471–487.

Choi, J.-E., & Shin, D. W. (2018). Forecasts for leverage heterogeneous autoregressive models with jumps and other covariates. *Journal of Forecasting*, *37*(6), 691–704.

Choi, J.-E., & Shin, D. W. (2019). Quantile forecasts for financial volatilities based on parametric and asymmetric models. *Journal of the Korean Statistical Society*, *48*(1), 68–83.

Clements, A. E., & Fuller, J. (2012). *Forecasting Increases in the VIX: A Time-Varying Long Volatility Hedge for Equities.* (NCER Working Paper Series 88) Brisbane. Australia: National Centre for Econometric Research.

Comte, F., & Renault, E. (1998). Long memory in continuous-time stochastic volatility models. *Mathematical Finance*, *8*(4), 291–323.

Corrado, C. J., & Miller, T. W. Jr (2005). The forecast quality of CBOE implied volatility indexes. *Journal of Futures Markets*, *25*(4), 339–373.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174–196.

Corsi, F., Audrino, F., & Reno, R. (2012). HAR modeling for realized volatility forecasting, In L. Bauwens C. Hafner, & S. Laurent (Eds.), *Handbook of volatility models and their applications*. (pp. 363–382). Hoboken, NJ: Wiley.

Corsi, F., Mittnik, S., Pigorsch, C., & Pigorsch, U. (2008). The volatility of realized volatility. *Econometric Reviews*, *27*(1–3), 46–78.

Corsi, F., & Renò, R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business and Economic Statistics*, *30*(3), 368–380.

Diebold, F. X., Gunther, T. A., & Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, *39*(4), 863–883.

Fernandes, M., Medeiros, M. C., & Scharth, M. (2014). Modeling and predicting the CBOE market volatility index. *Journal of Banking and Finance*, *40*, 1–10.

Fleming, J., Ostdiek, B., & Whaley, R. E. (1995). Predicting stock market volatility: A new measure. *Journal of Futures Markets*, *15*(3), 265–302.

Giot, P. (2005). Implied volatility indexes and daily value at risk models. *Journal of Portfolio Management*, *31*(3), 92–100.

Giraitis, L., Kokoszka, P., Leipus, R., & Teyssière, G. (2000). Semiparametric estimation of the intensity of long memory in conditional heteroskedasticity. *Statistical Inference for Stochastic Processes*, *3*(1), 113–128.

González-Rivera, G., & Sun, Y. (2015). Generalized autocontours: Evaluation of multivariate density models. *International Journal of Forecasting*, *31*(3), 799–814.

Hamilton, J. D. (1996). This is what happened to the oil price–macroeconomy relationship. *Journal of Monetary Economics*, *38*(2), 215–220.

Hamilton, J. D. (2003). What is an oil shock? *Journal of Econometrics*, *113*(2), 363–398.

Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*, *35*, 705–730.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*, 453–497.

Jiang, G. J., & Tian, Y. S. (2005). The model-free implied volatility and its information content. *Review of Financial Studies*, *18*(4), 1305–1342.

Konstantinidi, E., Skiadopoulos, G., & Tzagkaraki, E. (2008). Can the evolution of implied volatility be forecasted? Evidence from European and US implied volatility indices. *Journal of Banking and Finance*, *32*(11), 2401–2411.

Lee, K., Ni, S., & Ratti, R. A. (1995). Oil shocks and the macroeconomy: The role of price variability. *Energy Journal*, *16*(4), 39–56.

Li, W. K., & Mak, T. K. (1994). On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. *Journal of Time Series Analysis*, *15*(6), 627–636.

Liu, C., & Maheu, J. M. (2009). Forecasting realized volatility: A Bayesian model-averaging approach. *Journal of Applied Econometrics*, *24*(5), 709–733.

Lo, A. W. (1991). Long-term memory in stock market prices. *Econometrica*, *59*(5), 1279–1313.

Lobato, I. N., & Velasco, C. (2004). A simple test of normality for time series. *Econometric Theory*, *20*(4), 671–689.

Louzis, D. P., Xanthopoulos-Sisinis, S., & Refenes, A. P. (2012). Stock index realized volatility forecasting in the presence of heterogeneous leverage effects and long range dependence in the volatility of realized volatility. *Applied Economics*, *44*(27), 3533–3550.

Mazzeu, J. H. G., González-Rivera, G., Ruiz, E., & Veiga, H. (2017). *A Bootstrap Approach for Generalized Autocontour Testing: Implications for VIX Forecast Densities*. (Working Paper 201709). Spain: Universidad Carlos III de Madrid.

Mollick, A. V., & Assefa, T. A. (2013). US Stock returns and oil prices: The tale from daily data and the 2008–2009 financial crisis. *Energy Economics*, *36*, 1–18.

Nandha, M., & Faff, R. (2008). Does oil move equity prices? a global view. *Energy Economics*, *30*(3), 986–997.

Park, S.-K., Choi, J.-E., & Shin, D. W. (2017). Value at risk forecasting for volatility index. *Applied Economics Letters*, *24*(21), 1613–1620.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, *160*(1), 246–256.

Patton, A. J., & Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility full access. *Review of Economics and Statistics*, *97*(3), 683–697.

Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, *75*(2), 335–346.

Ramos, S. B., & Veiga, H. (2011). Risk factors in oil and gas industry returns: International evidence. *Energy Economics*, *33*(3), 525–542.

Ramos, S. B., & Veiga, H. (2013). Oil price asymmetric effects: Answering the puzzle in international stock markets. *Energy Economics*, *38*, 136–145.

Renault, E. (1999). La mémoire longue en économie: Discussion et commentaires. *Journal de la Société Française de Statistique*, *40*(2), 79–89.

Sadorsky, P. (2008). Assessing the impact of oil prices on firms of different sizes: It's tough being in the middle. *Energy Policy*, *36*(10), 3854–3861.

Scharth, M., & Medeiros, M. C. (2009). Asymmetric effects and long memory in the volatility of Dow Jones stocks. *International Journal of Forecasting*, *25*(2), 304–327.

Shin, D. W. (2018). Forecasting realized volatility: A review. *Journal of the Korean Statistical Society*, *47*(4), 395–404.

Simon, D. P. (2003). The Nasdaq volatility index during and after the bubble. *Journal of Derivatives*, *11*(2), 9–24.

Tay, A. S., & Wallis, K. F. (2000). Density forecasting: A survey. *Journal of Forecasting*, *19*(4), 235–254.

Wang, T., & Huang, Z. (2012). The relationship between volatility and trading volume in the Chinese stock market: A volatility decomposition perspective. *Annals of Economics and Finance*, *13*(1), 211–236.

Wang, X., Wu, C., & Xu, W. (2015). Volatility forecasting: The role of lunch-break returns, overnight returns, trading volume and leverage effects. *International Journal of Forecasting*, *31*(3), 609–619.

Whaley, R. E. (2000). The investor fear gauge. *Journal of Portfolio Management*, *26*(3), 12–17.

Whaley, R. E. (2009). Understanding the VIX. *Journal of Portfolio Management*, *35*(3), 98–105.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*(4), 817–838.

Wu, G. (2001). The determinants of asymmetric volatility. *Review of Financial Studies*, *14*(3), 837–859.

AUTHOR BIOGRAPHIES

João H. Gonçalves Mazzeu is currently a Postdoctoral Researcher at the Statistics Department of University of Campinas, Brazil. He has a doctor's degree in Business and Quantitave Methods by the Statistics Department of Universidad Carlos III de Madrid, Spain. His research interests are time series forecasting, volatility models, construction and evaluation of prediction densities under model uncertainty and dynamic factor models for high-dimensional data.

Helena Veiga is currently an Associate Professor of Econometrics at the Statistic Department of the University Carlos III of Madrid. Her research is on financial econometrics, time series, panel data, empirical finance, in particular, experimental finance, commodities markets and financial investments in the energy sector. She has published her work in several high ranked journals, such as, the Journal of Banking and Finance, Journal of Empirical Finance, European Review of Economics, Energy Economics, Experimental Economics, Computational Statistics and Data Analysis, Journal of Productivity Analysis, European Journal of Operation Research, among others. Helena Veiga holds a PhD. in Economics from Universitat Autònoma de Barcelona.

Massimo B. Mariti was a Student of the University Milano-Bicocca and this work is part of his master thesis.

How to cite this article: Mazzeu JHG, Veiga H, Mariti MB. Modeling and forecasting the oil volatility index. *Journal of Forecasting*. 2019;38:773–787. https://doi.org/10.1002/for.2598