What should be taken into consideration when forecasting oil implied volatility index?

Delis, Panagiotis and Degiannakis, Stavros and Giannopoulos, Kostantinos

26 November 2021

What should be taken into consideration when forecasting oil implied volatility index?

Panagiotis Delis[a,*], Stavros Degiannakis[a], Konstantinos Giannopoulos[a]

[a] *Department of Economics and Regional Development, Panteion University of Social and Political Sciences, 136 Syggrou Avenue, 17671, Greece*

Abstract

Crude oil is considered a key commodity in all the economies around the world. This study forecasts the oil volatility index (OVX), which is the market's expectation of future oil volatility, by incorporating information from other asset classes. The literature does not extensively test the long memory of the targeted volatility. Thus, we estimate the Hurst exponent implementing a rolling window rescaled analysis. We provide evidence for a strong long memory in the implied volatility (IV) indices which justifies the use of the HAR model in obtaining multiple days ahead OVX forecasts. We also define a dynamic model averaging (DMA) structure in the HAR model in order to allow for IV indices from other asset classes to be applicable at different time periods. The implementation of the DMA-HAR models informs forecasters to focus on the major stock market IV indices, and more specifically on the DJIA Volatility Index. Our results lead us to the conclusion that accurate OVX forecasts are obtained for short- and mid-run forecasting horizons. The evaluation framework is not limited to statistical loss functions but also embodies an options straddle trading strategy.

Keywords: crude oil, implied volatility, HAR modelling, trading strategies, dynamic model averaging, long memory

*Corresponding author

Email addresses: p.delis@panteion.gr (Panagiotis Delis), s.degiannakis@panteion.gr (Stavros Degiannakis), kwstasgiannopoulos@yahoo.gr (Konstantinos Giannopoulos)

1. Introduction

In recent years, many studies have focused on oil uncertainty due to the fact that oil price fluctuations can have large impact, not only on the global economy, but also on the stock market because of its financialization. Moreover, they focus on crude oil since it is considered one of the major inputs in the economy and more specifically because oil uncertainty could reflect sudden changes in global economic growth. Ferderer (1996) concludes that oil price volatility helps to forecast aggregate output movements in the U.S. Moreover, he provides empirical evidence that there is strong relationship between oil price changes and output growth, which can be explained by the economy's response to oil price volatility. This strong relationship is also investigated by Elder and Serletis (2010), who find a significant effect of oil price volatility on aggregate output.

Volatility, in general, has triggered the attention of investors, portfolio managers and policy makers. As far as we are concerned, the most important reasons are the following:

- Future volatility is considered an important component for pricing derivative products, such as option contracts.
- Investors are interested in future volatility predictions, since they want to estimate the limit they are willing to accept in order to make the most optimal portfolio decisions.
- The implied volatility (IV) indices allow for measuring the expectations of the financial and energy markets, which provide useful information about the outcome of the monetary policy decisions that have been made.

Recent studies that focus on volatility forecasting use realized volatility as a volatility measure. According to Andersen and Bollerslev (1998), the realized volatility is defined as the sum of the squared intraday returns. It has been widely used in many studies that investigate how uncertainty can be modelled accurately (Buncic and Gisler, 2016; Degiannakis and Filis, 2017; Lyócsa and Molnár, 2018; Delis et al., 2020). Nevertheless, IV indices are considered in many studies[1] as accurate predictors of future volatility and have been known to provide information about investor sentiment. In this paper, we investigate the dynamics of the oil volatility index (OVX), which is constructed in order to provide information on the crude oil market during volatile periods. More specifically, OVX measures the market's expectation of the 30-day volatility of crude oil

[1] For further details see Fleming et al. (1995), Busch et al. (2011), Gong and Lin (2018) and Lv (2018).

prices extracting information from the options on the United States Oil Fund (USO) for a wide range of strike prices.

With regard to the information which the different volatility measures provide, there are papers that have compared the forecasting ability of different models including these measures as explanatory variables. According to Blair et al. (2001), the IV index helps to obtain more accurate conditional volatility forecasts compared to the models including either the interday volatility or the intraday realized volatility. Koopman et al. (2005) confirmed that models using interday volatility as an explanatory variable have been outperformed by those using IV index. Furthermore, there are studies[2], which conclude that IV provides higher predictive information when we generate conditional volatility forecasts.

Apart from the aforementioned studies which investigate the predictive ability of the different volatility measures, including IV index on conditional volatility; there are studies that implement various modelling frameworks in order to directly forecast IV indices. For example, Degiannakis (2008) uses intraday data and conditional volatility estimates to forecast the VIX index[3]. He draws the conclusion that the entire predictive information is provided by VIX itself and that neither interday nor intraday volatility estimates offer incremental forecasting ability on forecasts of VIX. This leads to the conclusion that IV indices are not highly connected to volatility of the underlying asset. Furthermore, Dunis et al. (2013) investigate the predictability of intraday EUR-USD IV by exploiting intraday seasonalities such as overnight effects and they found that IV can be useful in predictions for shorter horizons, within a given day. In this paper, the main objective is to investigate whether the various modelling frameworks applied to realized volatility can offer forecasting gains in OVX, since our attention is focused on crude oil, and on which characteristics should be considered.

Methodologically, in recent years, various models have been implemented in order to generate volatility forecasts. A well-known model that has been used for this purpose is the fractionally integrated autoregressive moving average (ARFIMA) model, which captures the property of long memory in the volatility series and is considered suitable for estimating and forecasting the logarithmic transformation of volatility. In addition to the ARFIMA model, the HAR model, proposed by Corsi (2009), and its extensions are parsimonious model frameworks and capture long memory, as well. Sévi (2014) uses various HAR model specifications in order to come to a conclusion con-

[2]For example see Fleming et al. (1995), Christensen and Prabhala (1998), Giot (2003) and Frijns et al. (2010)

[3]VIX measures the market's expectation of the 30-day volatility deriving from the prices of various S&P500 options.

cerning which components provide additional predictive information in crude oil realized volatility forecasts.

The limited literature on forecasting IV indices consists of studies that have implemented not only univariate but also multivariate models. For instance, Konstantinidi et al. (2008) study the impact of certain economic variables on forecasting European and U.S. IV indices under a univariate modelling framework. In the latter study, VAR models are also implemented, which outperform the competing models. This result confirms the IV connectedness among various markets. In this regard, Awartani et al. (2016) provide evidence that volatility transmission from oil to equities, in case of IV indices, is significant and, in general, the transmission from oil to other markets has increased, since the oil price collapse in July 2014. However, the volatility transmission from all the other markets to oil is not so strong. According to Chatziantoniou et al. (2020), there is a transmission of shocks from OVX to VIX but the spillover effects between OVX and VIX do not contain significant predictive information when generating out-of-sample forecasts of OVX. In contrast, Liu et al. (2020) find that there is significant bidirectional IV spillover between the crude oil and stock markets and they draw the conclusion that there is significant positive time-varying correlation between oil and stock IV indices.

Moreover, Liu et al. (2013) investigate the short- and long-run cross-market volatility transmission implied by OVX and other important volatility indices in an in-sample analysis. The results indicate that the uncertainty transmission between the oil market and other markets is short-lived. Another conclusion of the study is the significant uncertainty in transmission from the stock market to the crude oil market, which is in line with Bašta and Molnár (2018), who conclude that the stock market IV leads the oil market IV. This could be considered of major importance for those who are interested in OVX and suggests that they should take note changes in VIX or other volatility indices of the stock market. Our paper takes into consideration the above-mentioned results and fills the gap in the literature by investigating the impact of other market IV indices on OVX in an out-of-sample forecasting framework. Moreover, the outcome of this study may provide greater evidence on the consideration of the financialization of the crude oil market. Moreover, there will be an answer on which IV indices enhance the forecasting accuracy of OVX.

Moreover, there are studies which incorporate exogenous variables as predictors in the models implemented in order to find potential incremental additional information on IV indices. For example, Fernandes et al. (2014) investigate the impact of a list of predictors including some macro-finance variables, such as the S&P500 return, the first difference of the logarithmic volume of the S&P500 index, the oil return and other variables related to the US macro-finance environment. They provide evidence that some of the exogenous variables containing mainly information from the S&P500 index seem to have large impact on VIX. However, as per the out-of-sample forecasting results, they conclude that the simple HAR model performs really well and it is difficult even for complicated model specifications to surpass it. This is justified by the fact that the persistent nature of IV indices is very strong. Moreover, they conclude that persistence is almost the only feature that matters when generating short-run forecasts. Our study focuses on the impact of IV indices of other markets on the OVX under an out-of-sample forecasting investigation. In our study, we enhance the forecasting performance of the HAR models by incorporating information coming from IV indices of other asset classes, which is considered an extension of the simple HAR model. So far, there is only one study that investigates the impact of realized volatility from several asset classes for providing forecasts of oil price realized volatility, which is the paper of Degiannakis and Filis (2017). According to the study, HAR models which include multiple asset class volatility measures outperform all the remaining models on all forecasting horizons. This result holds true not only under a statistical evaluation framework but also under trading strategies that have been implemented for comparison reasons. Moreover, it is important to note here that the aforementioned study extends the previous studies[4], which provide evidence that the HAR-RV model outperforms the competing forecasting models in the case of crude oil. Therefore, according to Degiannakis and Filis (2017) there is evidence that the impact of other markets' realized volatility on oil price realized volatility is significant. We take this into account in this study by using the HAR model specifications and include the IV indices from other asset classes.

All these studies lead us to the conclusion that forecasting IV indices remain a topic that still needs further investigation. One of the contributions of this paper is the extensive investigation of the features that are important for generating out-of-sample forecasts of OVX. In this regard, we suggest that academics and forecasters take into account the strong existence of long memory in the time series of IV indices, and more particularly of OVX, which justifies the use of the HAR model. The fact that the DMA approach can efficiently extract information from other asset classes and that it allows for parameters to change over time, which is crucial during the last decades that structural breaks occur more frequently, allow it to provide forecasters with more accurate out-of-sample forecasts of OVX.

The second contribution of this paper is to provide a detailed answer concerning which of the IV indices from other asset classes enhance the accuracy of the OVX forecasts. Moreover, the evaluation of the OVX forecasts consists of statistical loss func-

[4] For example, see Haugom et al. (2014) and Prokopczuk et al. (2016).

tions and an options straddle trading strategy. The results show that the inclusion of DJIA Volatility Index in the modelling framework is considered of major significance and enhances the predictive ability of the models implemented on short- and mid-run forecasting horizons. For longer horizons, the Energy Sector ETF Volatility Index appears to have predictive information on OVX, which is explained by the fact that both focus on energy related uncertainty. Moreover, we have concluded that the predictive ability of the HAR models is statistically significant in short- and mid-run forecasting horizons.

The remainder of the paper consists of the following sections. Section 2 gives a detailed description of the IV indices and provides further information about the dataset used. Moreover, it covers the section in which the long memory of the IV indices has been tested. In Section 3 the modelling framework has been presented maintaining both the estimation and forecasting frameworks. In Section 4 the evaluation of the generated forecasts has been analytically described, and in Section 5 we provide the results of the statistical and economic evaluation frameworks. Finally, Section 6 concludes the study.

2. Implied volatility indices

2.1. Data description

It is vital to start with a description of how IV indices are calculated. This will be useful for differentiating the different volatility measures, namely realized volatility and IV. As we have already mentioned, realized volatility reflects the actual volatility of an underlying asset in contrast with IV, which is considered by many studies a better prediction of future volatility.

In this study, we focus on the estimate of the expected 30-day volatility of crude oil as priced by USO, which is called OVX. The Chicago board of options exchange (CBOE) Volatility Index methodology, which has been applied to the OVX, uses options on USO, an ETF that is designed to track the price of West Texas Intermediate (WTI) crude oil, with a wide range of strike prices. In more detail and according to the CBOE, OVX is calculated by interpolating between two weighted sums of option midquote values [5]. OVX is therefore obtained by annualizing the interpolated value, taking its square root and expressing the result in percentage points. All the other IV indices used in this study are calculated in a similar fashion.

Our dataset consists of 14 IV indices including OVX. Starting from the major stock market's IV indices, we use the most representative indices, namely the S&P 500 Volatil-

[5] See https://www.cboe.com for further details.

ity Index (VIX), the S&P100 Volatility Index (VXO), the VXD and the Nasdaq 100 Volatility Index (VXN). Apart from the major IV indices of the stock market, we investigated the impact of other IV indices, which represent different asset classes and are the following: the Euro Currency Volatility (EVZ), the Gold Volatility Index (GVZ), the Silver ETF Volatility Index (VXSLV), the Energy Sector ETF Volatility Index (VXXLE), the Gold Miners ETF Volatility Index (VXGDX), the British Pound Volatility Index (BPVIX), the Euro Volatility Index (EUVIX), the Yen Volatility Index (JYVIX) and the 10 Year U.S. Treasury Note Volatility Index (TYVIX). Regarding the sample and the frequency of the dataset, we use data from the 16^{th} of March, 2011 up to the 1^{st} of October, 2019 and the frequency of the data is daily. The source of the data for all implied volatility indices that are used in this study is CBOE.

Figure 1 shows the evolution of OVX and the remaining IV indices used in this study over the whole sample. With regard to OVX, it is obvious that high values are observed between mid-2014 and early 2016, which was a period when the global economy faced one of the largest oil price declines. Another drop in oil prices in 2018, due to the trade war between the U.S. and China, resulted in high OVX values. Moreover, from the descriptive statistics of the entire dataset, which are reported in Table 1, it can be observed that almost all the IV indices present high values of variation, which is shown by the coefficient of variation (CV).

[TABLE 1 HERE]

[FIGURE 1 HERE]

2.2. Testing for long memory

One of the main features of IV indices is strong persistence, which is not investigated in depth in recently published papers (e.g. Degiannakis et al., 2018) which use models relying on this property. For example, not only the ARFIMA model but also the HAR model need to justify their use by providing evidence of strong long memory, which is missing from the relevant studies. In general, the statistical investigations that are performed to test long memory in time series have become a source of major controversies. In this paper, we apply the so-called rescaled range analysis, which is proposed by Mandelbrot and Wallis (1969) and used in a large number of studies, such as Weron (2002) and Sánchez Granero et al. (2008). The rescaled range analysis is the most well-known method to estimate the Hurst exponent H, which measures the intensity of long-range dependence in a time series.

The procedure can be described as follows. Let us first note that the targeted IV time series of length N is divided into a number of d time series $(X_{i,z})$ of length n each and for each sub series $z = 1, \ldots, d$:

1. Calculate the average (A_z) and the standard deviation (S_z) of each sub series.
2. Calculate the mean-adjusted series $Z_{i,z} = X_{i,z} - AV_z$ for $i = 1, \ldots, n$.
3. Calculate the cumulative series $Y_{i,z} = \sum_{j=1}^{i} Z_{j,z}$ for $i = 1, \ldots, n$.
4. Compute the range $R_z = max\{Y_{1,z}, \ldots, Y_{n,z}\} - min\{Y_{1,z}, \ldots, Y_{n,z}\}$.
5. Compute the range (R_z / S_z).
6. Calculate the average of the rescaled range ($(R/S)_n$) for all sub series of length n.

The Hurst exponent H is therefore estimated by running a simple linear regression over a sample of increasing time horizons $\log(R/S)_n = \log c + H \log n$. There is a range of values that the Hurst exponent takes, which ranges from 0 to 1. For values greater than 0.5, the volatility time series is considered persistent and for values less than 0.5, the corresponding volatility time series is anti-persistent.

However, it should be mentioned that for small n the deviation from the 0.5 slope can be significant. In this line Anis and Lloyd (1976) proposed a new formulation to enhance the performance when referring to small n. Weron (2002) shows that the Hurst exponent should be calculated as 0.5 plus the slope of $R/S_n - E(R/S)_n$, where $E(R/S)_n$ values are approximated by

$$E(R/S)_n = \begin{cases} \frac{n-\frac{1}{2}}{n} \frac{\Gamma((n-1)/2)}{\Gamma(n/2)\sqrt{\pi}} \sum_{i=1}^{n-1} \sqrt{\frac{n-i}{i}}, & \text{for } n \leqslant 340, \\ \\ \frac{n-\frac{1}{2}}{n} \frac{1}{\sqrt{n\pi/2}} \sum_{i=1}^{n-1} \sqrt{\frac{n-i}{i}}, & \text{for } n > 340, \end{cases} \tag{1}$$

where Γ is the Euler gamma function.

In this study, we estimate the Hurst exponent by applying a rolling window in order to observe the range of values that the Hurst exponent takes over the sample. Moreover, the rolling window has a fixed length of 1000 observations, which means that the first Hurst exponent is estimated by using information up to the time $t = 1000$. From Figure 2, which portrays the Hurst exponent values in a histogram, we observe that the Hurst exponent takes values from 0.8 to 1 for approximately all IV indices with OVX being one of the most persistent time series, since the Hurst exponent takes values from approximately 0.9 to 1. This indicates that the used IV time series present a long-term positive autocorrelation.

[FIGURE 2 HERE]

3. Modelling framework

3.1. Näive models

As the simplest model, we use Random Walk (RW) without a drift, which is written as follows:

$$log(OVX_t) = log(OVX_{t-1}) + \varepsilon_t, \tag{2}$$

where OVX_t is the oil volatility index at day t and ε_t is a white noise.

Apart from the RW, we will also estimate the AR(1), which represents the most näive specification of the ARIMA modelling category and it is written as:

$$log(OVX_t) = \hat{a}_0^{(t)} + \hat{a}_1^{(t)} log(OVX_{t-1}) + \varepsilon_t. \tag{3}$$

It should be noted that we use these two simple models in order to compare them with the most complicated models, which take into account properties and characteristics of the targeted time series, check whether OVX can be accurately predicted, and if so, for which forecasting horizons the näive are significantly outperformed.

3.2. HAR model specifications

As far as we are concerned, most of the papers that forecast volatility time series use a HAR model specification, which is proposed by Corsi (2009) and is capable of capturing features of financial market volatility such as long range dependence which is also known as long memory. Therefore, we can justifiably use the HAR model for predicting the time series of OVX, which is indicated to be a persistent time series with evidence of strong long-term positive autocorrelation.

The simplest and most widely used HAR model specification is the following:

$$\begin{aligned} log(OVX_t) = \hat{a}_0^{(t)} + \hat{a}_1^{(t)} log(OVX_{t-1}^{(d)}) + \hat{a}_2^{(t)} log(OVX_{t-1}^{(w)}) \\ + \hat{a}_3^{(t)} log(OVX_{t-1}^{(m)}) + \varepsilon_t, \end{aligned} \tag{4}$$

where ε_t is the error term and $\hat{a}_0^{(t)}, \hat{a}_1^{(t)}, \hat{a}_2^{(t)}, \hat{a}_3^{(t)}$ are the estimated parameters. It is also important to note that the explanatory variables used in this HAR model specification are calculated as: $log(OVX_{t-1}^{(d)}) = log(OVX_{t-1})$; $log(OVX_{t-1}^{(w)}) = \left(5^{-1} \sum_{k=1}^{5} log(OVX_{t-k})\right)$; $log(OVX_{t-1}^{(m)}) = \left(22^{-1} \sum_{k=1}^{22} log(OVX_{t-k})\right)$, which is in line with Corsi and Renò (2012).

As it is observed, the index vector of the lag components is $i = (1,5,22)'$. However, there are studies such as Fernandes et al. (2014) that also include the biweekly (10 days) and quarterly components (66 days) in the index vector of the lag components $i = (1,5,10,22,66)'$. We run the HAR models under this structure and we conclude that the biweekly and quarterly components do not provide any additional information in the models and also cannot beat the näive model's structure with the daily, weekly and monthly lag components. Therefore, the simple index vector $i = (1,5,22)'$ is used in the HAR models of this paper.

Furthermore, we focus on the impact of the different IV indices on OVX. In order to investigate this, we first implement individual HAR models by adding each IV index per model in the already defined simple HAR model specification. Let us then denote as EX the vector of the exogenous variables of HAR-EX model, which is written as follows:

$$\begin{aligned} log(OVX_t) = \hat{a}_0^{(t)} + \hat{a}_1^{(t)} log(OVX_{t-1}^{(d)}) + \hat{a}_2^{(t)} log(OVX_{t-1}^{(w)}) \\ + \hat{a}_3^{(t)} log(OVX_{t-1}^{(m)}) + \hat{\beta}_1^{(t)} log(EX_{t-1}) + \varepsilon_t, \end{aligned} \tag{5}$$

where $log(EX_{t-1})$ is the one lagged IV index used in each model in logarithmic transformation. We implement 13 individual HAR-EX models with EX being a different IV index in each one of the 13 models.

Finally, we will implement the HAR-ALL model, which includes the entire EX vector of IV indices as predictors of the OVX in order to investigate whether including all of them in one model specification enhances the predictive ability of the HAR model.

Before moving to the next part, which allows existing HAR modelling structures to apply different set of IV indices and not to be limited to one each time, we consider it more convenient to represent the models in a more general way. In this regard, we define $\mathbf{x_t}=\begin{bmatrix}1 & log(OVX_{t-1}^{(d)}) & log(OVX_{t-1}^{(w)}) & log(OVX_{t-1}^{(m)})\end{bmatrix}$ as the (1×4) vector of lag components of the simplest HAR model and $y_t = log(OVX_t)$ the dependent variable. Therefore, the simple HAR model can be written as follows:

$$y_t = \boldsymbol{x_t \alpha_t} + \varepsilon_t, \tag{6}$$

where $\boldsymbol{\alpha_t} = \begin{bmatrix}a_0^{(t)} & a_1^{(t)} & a_2^{(t)} & a_3^{(t)}\end{bmatrix}'$ is the vector of parameters. Regarding the individual HAR-EX models, we can replace Eq. (5) with the following equation:

$$y_t = \boldsymbol{x_t \alpha_t} + EX_t \beta_t + \varepsilon_t, \tag{7}$$

where β_t is the parameter reflecting the impact of each IV index included in the model specification.

After having estimated the parameters of different HAR models, we now implement the forecasting framework for generating point forecasts of OVX. These forecasts have been generated by using a direct approach, which is the most common approach when forecasting multiple periods ahead. Using the direct approach, the regression of the HAR model can be defined as:

$$y_{t+h} = \boldsymbol{x_t}\boldsymbol{\alpha}^{(\boldsymbol{h})} + \varepsilon_{t+h}. \tag{8}$$

More specifically, when referring to the benchmark HAR model, the OVX forecast is $\boldsymbol{x_t}\hat{\boldsymbol{\alpha}}_{\boldsymbol{t}}^{(\boldsymbol{h})}$, where $\hat{\boldsymbol{\alpha}}_{\boldsymbol{t}}^{(\boldsymbol{h})}$ is an estimate of $\boldsymbol{\alpha}^{(\boldsymbol{h})}$ that only relies on data up to period t. Therefore, the obtained forecast of OVX can be equal to $e^{y_{t+h}}$ [6], since the logarithmic transformation is used for the dependent variable y_{t+h}. Moreover, we use the same method for generating forecasts of OVX from the individual HAR-EX models by incorporating the corresponding IV index in Eq. (8).

3.3. Dynamic model averaging (DMA)

Apart from the use of the individual models, we also use a different modelling approach, which gives flexibility to the model used at each point in time. The first advantage of this methodology is the fact that potential structural breaks can be detected by assuming time varying parameters. In recent decades, structural breaks have become more frequent and the use of time variation in the vector of parameters provides more accurate forecasting results, which is why we implemented this modelling approach. Another advantage of the DMA approach is the fact that the set of predictors used in a HAR model does not remain constant over time. This gives us the ability to allow for K models, which use different sets of explanatory variables to be applicable at different time periods. Thus, we can say that this approach allows for time variation in both the vector of parameters and the set of predictors. The DMA approach, based on the structure of the simple HAR model[7]. It requires the following two equations in order to be implemented:

$$y_t = \boldsymbol{x}_{\boldsymbol{t}}^{(\boldsymbol{k})}\boldsymbol{\alpha}_{\boldsymbol{t}}^{(\boldsymbol{k})} + \varepsilon_t^{(k)} \quad \text{for} \quad \varepsilon_t^{(k)} \sim N(0, H_t^{(k)}), \tag{9}$$

[6] The error variance term $0.5\sigma_\varepsilon^2$ should be incorporated in the forecasting, but its impact is almost zero on the out-of-sample results.

[7] The DMA methodology works in the same way for models that include different set of predictors such as IV indices, in our case.

$$\boldsymbol{\alpha}_t^{(k)} = \boldsymbol{\alpha}_{t-1}^{(k)} + \boldsymbol{u}_t^{(k)} \quad \text{for} \quad \boldsymbol{u}_t^{(k)} \sim N(\mathbf{0}_{4\times 1}, \boldsymbol{\Sigma}_{\boldsymbol{u}_t}^{(k)}). \tag{10}$$

where $k = 1, \ldots, K$, $\boldsymbol{\alpha}_t^{(k)}$ are the parameters of the HAR model specification and the errors $\varepsilon_t^{(k)}$ and $\boldsymbol{u}_t^{(k)}$ are assumed to be mutually independent. Moreover, if there are m predictors in $\boldsymbol{x}_t^{(k)}$, the number of possible combinations of these predictors is $K = 2^m$. In the case of using the HAR-ALL model specification, we have $K = 2^{17}$ combinations. This model is estimated by using the self-perturbed Kalman filter[8] proposed by Grassi et al. (2017) and also applied by Delis et al. (2020) in the case of crude oil volatility forecasting.

In this section, our purpose is to investigate the impact of different IV indices among asset classes on forecasts of OVX. Furthermore, it is important to mention that we implement a DMA model including all the IV indices in the set of predictors. In addition, we also implement a separate DMA-STO model, which includes, apart from the constant term and the 3 lag components of OVX, only the 4 major IV indices of stock market, namely VIX, VXO, VXD, and VXN. The purpose of this section is the investigation of the predictive information that each IV index provides when forecasting OVX multiple horizons.

Figure 3 shows the probabilities of the IV indices derived by the DMA methodology through the HAR-DMA-ALL model. First of all, we can observe that the relative importance of each of the main IV indices is time varying. More specifically, the probability of VXD spans between 0% and 100% with the lower values occurring from mid- 2017 to mid- 2018. However, we can easily observe that the probability to be included in the best model is high enough for almost the entire out-of-sample period, which also holds true for the case of VIX. This can be explained due to the fact that the commodity market financialization seems to be really strong in recent years. Moreover, a plausible explanation is that a large number of investors take a speculating position with regard to crude oil market and therefore the stock market's IV indices could affect their decisions. The probabilities of the rest of the IV indices remain relatively constant with corresponding values of approximately 50% during the entire out-of-sample period.

3.4. Forecasting settings

The settings regarding the modelling framework are defined as follows. First of all, the initial sample period is T_0=1043 days, since we need 22 days from the time series of

[8] More details for each step of the estimation of the implemented DMA model can be found in the Appendix.

OVX in order to construct the maximum of the lag components, which is the monthly one and $\max(h)-1$ days for implementing direct forecasting approach. The remaining 1000 days of the initial sample period T_0 is the fixed length that we use for the rolling window estimation. A more detailed description of the sample that we use for estimating our models and obtaining OVX forecasts is that we choose data from the 1^{st} to the 1000^{th}. We estimate the parameters of the relevant model and we construct the forecasts afterwards, then we re-estimate the parameters using data from the 2^{nd} to the 1001^{st} and we implement the forecasting methodology again. The remaining out-of-sample period is used for evaluating the forecasts of OVX and is defined as T_{OOS}.

4. Evaluation framework

4.1. Loss functions and the model confidence set

This section consists of two evaluation techniques, namely the loss functions and the model confidence set (MCS). The first category utilizes two well-known loss functions, the Mean Squared Predicted Error (MSPE) and the Mean Absolute Error (MAE), which are defined as:

$$MSPE^{(h)} = \frac{1}{T_{OOS}} \sum_{t=1}^{T_{OOS}} (OVX_{t+h|t} - OVX_{t+h})^2, \tag{11}$$

and

$$MAE^{(h)} = \frac{1}{T_{OOS}} \sum_{t=1}^{T_{OOS}} \mid OVX_{t+h|t} - OVX_{t+h} \mid, \tag{12}$$

where $\text{OVX}_{t+h|t}$ is the h-days-ahead forecast of OVX, OVX_{t+h} is the OVX at time $t+h$ and T_{OOS} is the number of the out-of-sample data points.

However, the two statistical loss functions are not adequate to draw conclusions as to whether the forecasts of OVX from a model are more accurate than those of other models. Therefore, we also use the MCS test proposed by Hansen et al. (2011) in order to further evaluate our predictions. This specific procedure identifies the set of the best models, as these are defined in terms of a specific statistical loss function, which is MSPE in our study [9].

[9] We also implemented this procedure by using MAE as a loss function, which provides qualitatively similar results.

The target of the MCS test is to investigate which set of models remain until the end, under an elimination algorithm, at a predefined level of significance a. At the beginning of the process, the full set of models $M = M_0 = \{1, \ldots, m_0\}$ is used and the following null hypothesis of equal predictive ability is repeatedly tested:

$$H_{0,M} : E(d_{i,i*,t}) = 0, \;\; \forall \; i, i^* \in M, \tag{13}$$

where $d_{i,i*,t} = \Psi_{i,t} - \Psi_{i*,t}$ is defined as the evaluation differential for $i,\; i^* \in M_0$ and $\Psi_{i,t} = (OVX_{t+h|t} - OVX_{t+h})^2$, where $OVX_{t+h|t}$ is the h-days-ahead forecast of OVX obtained by the i^{th} model. This process is repeated until the null is not rejected anymore. We define the level of significance as $a = 0.1$. Another predefined setting of the MCS procedure is the block bootstrap with 10,000 bootstrap replications[10].

4.2. Option straddles trading strategy

Apart from the statistical evaluation of the obtained forecasts, which is based on the loss functions, we also investigate the forecasting performance of the applied models by using a trading strategy as an economic criterion. As far as we are concerned, this options straddle trading strategy has been used by Angelidis and Degiannakis (2008), Andrada-Félix et al. (2016) and Degiannakis and Filis (2017) and under this trading strategy we allow investors to go long (short) in a straddle when the forecast of OVX at time $t+h$ is higher than OVX at time t.

Moreover, this trading strategy, which relies on the purchase (or sale) of both a call and a put option with the same day of maturity is implemented. Therefore, the straddle holder's rate of return is affected only by the volatility changes and not by the changes in the underlying asset price.

In detail, the computation of the expected price of a straddle on a \$1 share of the USO on the next trading day with h days to maturity and \$1 exercise price comes from the following equation:

$$S_{t+1|t} = 2N\left(\frac{rf_t\sqrt{h}}{\overline{OVX}_{t+h|t}} + \frac{\overline{OVX}_{t+h|t}\sqrt{h}}{2}\right) - 2e^{-rf_t h}N\left(\frac{rf_t\sqrt{h}}{\overline{OVX}_{t+h|t}} - \frac{\overline{OVX}_{t+h|t}\sqrt{h}}{2}\right) + e^{-rf_t h} - 1, \tag{14}$$

where N(.) denotes the cumulative normal distribution function, $\overline{OVX}_{t+h|t} = \frac{1}{h-1}\sum_{i=1}^{h}\left(\frac{OVX_{t+i|t}}{\sqrt{252}}\right)$ is the average forecast of OVX during the life of the option and rf_t is the risk-free

[10] For further details see Hansen et al. (2011).

interest rate. The daily profit from holding the straddle is then calculated as $\pi_{t+1} = max(e^{y_{t+1}} - e^{rf_{t+1}}, e^{rf_{t+1}} - e^{y_{t+1}})$, for y_t denoting the USO daily returns in logarithmic transformation.

Let us say that three investors are assumed to trade according to their forecasts. Each investor i prices the straddles, $S^{(i)}_{t+1|t}$, every day of the out-of-sample period. A trade between two investors, i and j, is executed at the average of their forecasting prices, yielding to the investors i a profit of:

$$\pi_t^{(i,j)} = \begin{cases} \pi_{t+1} - (S^{(i)}_{t+1|t} + S^{(j)}_{t+1|t}), & \text{if } S^{(i)}_{t+1|t} > S^{(j)}_{t+1|t} \\ \\ (S^{(i)}_{t+1|t} + S^{(j)}_{t+1|t}) - \pi_{t+1}, & \text{if } S^{(i)}_{t+1|t} < S^{(j)}_{t+1|t} \end{cases} \tag{15}$$

We finally define the cumulative returns as $\pi = \frac{1}{2}\sum_{t=1}^{T_{OOS}}\sum_{j=1}^{2}\pi_t^{(i,j)}$ in order to evaluate the OVX forecasts.

5. Empirical findings

5.1. Loss functions results

The results of the two statistical loss functions that are used, namely the MSPE and MAE, can be found in Tables 2 and 3, respectively. First of all, regarding the MSPE results, we observe that the two näive models are outperformed by models including all IV indices, such as the HAR-ALL, HAR-DMA-STO and HAR-DMA-ALL, with HAR-DMA-STO model the one with the smallest MSPE values, for short-run horizons. It is noteworthy that the HAR-VXD model from the individual model specifications outperforms all the remaining models including only one IV index and presents results comparable to models that have the entire set of IV indices as exogenous variables. This provides us with evidence that DJIA Volatility Index significantly affects OVX, when referring to a 1-day ahead forecasting horizon. The results are also considered qualitatively similar in the case of MAE loss function.

Furthermore, the fact that there is short-term impact of the used IV indices on OVX can be justified due to the strong financialization of crude oil. The largest short-term impact comes from the major IV indices of the U.S. stock market, namely the VIX, VXO, VXD and VXN. We can also conclude, under the statistical loss functions of the evaluation framework, that the DMA methodology is of major importance for retrieving predictive information from a high number of IV indices in forecasting OVX with its contribution limited to short-run horizons.

With regard to mid-run horizons, we observe that the difference of loss functions among competing models is not so high, which means that the impact of the exogenous variables, namely IV indices, has been reduced. However, the HAR-VXD and

HAR-DMA-STO seem to perform better when compared to the rest of the models but with a lower impact for both 5 and 10 days ahead. It is important to mention that the simple HAR model, which is widely used not only for obtaining realized volatility forecasts but also IV forecasts, does not perform well and it is even outperformed by an AR(1) model specification. As has been mentioned in a previous section of the paper, we also implement a HAR model with a different structure by incorporating the biweekly and the quarterly components but the results are qualitatively similar [11].

Finally, regarding the longest forecasting horizon (22-days ahead), it is shown that for both the MSPE and the MAE loss functions there is no impact of IV indices on OVX and the simple AR(1) outperforms all the competing models, which can be considered evidence for the efficient market hypothesis. However, the impact of HAR-VXXLE model on 5-, 10- and 22-days ahead forecasts is noteworthy. From MSPE and MAE results, we observe that the corresponding values of HAR-VXXLE model are slightly lower than those of the AR(1) and the HAR models. This provides evidence that VXXLE, which estimates the expected 30-day volatility of the price of the Energy Sector ETF, has a small effect on OVX for mid- and long-run horizons.

5.2. Model Confidence Set procedure results

Apart from the loss functions, another test has been applied in order to evaluate the OVX forecasts. This is the so called MCS test and the results for the entire set of models that have been implemented can be found in Table 4. From this table, first of all, we observe that from the individual models, only HAR-VXD belongs to the set of the best models for all forecasting horizons, which is in line with results of the two loss functions presented in the previous section. Regarding models that include all IV indices as exogenous variables, only the HAR-DMA-STO model, which includes only the major IV indices capturing the main movements of the U.S. stock market, belongs to the set of the best models for all forecasting horizons.

The DMA models belong to the set of the best models, especially for short- and mid-run forecasting horizons, contrary to the simple AR(1) and HAR models that belong to the set of the best models at mid- and long-run forecasting horizons. This provides evidence that more complicated models can be used for generating short-term forecasts and that it is difficult to surpass näive model specification for longer-run forecasting horizons. The latter could be considered extra evidence for efficient market hypothesis. Finally, the results of the MCS test are in line with the results of the two loss functions and we draw the conclusion that the impact of the most representative IV indices, and more particularly the predictive information of VXD on OVX, is limited

[11] These results can be shared upon request.

to short-run forecasting horizons. In contrast to the short-run impact, VXXLE seems to have predictive information for mid- and long-run forecasting horizons, which is further investigated in the trading strategy that is implemented in order to evaluate the forecasts of OVX from an economic point of view.

5.3. Trading strategy results

As an additional evaluation method of the generated out-of-sample forecasts of OVX, we use an options straddle trading strategy, which is used as an economic criterion. The results of this trading strategy are the cumulative returns coming from the profit holding the straddle and the trade between the investors.

From Figure 4 we observe that models, including all IV indices, perform really well and offer positive returns for almost the entire out-of-sample period, which is consistent with the results of the statistical evaluation framework. The median of their cumulative returns is approximately 40% with a large number of outliers[12], mainly positive in the case of HAR-DMA-ALL model. It is noteworthy the fact that HAR-VXD is one of the models providing high values of cumulative returns with the median of the cumulative returns to be more than 60%, which explains the predictive information that VXD provides to OVX in short-run forecasting horizons.

With regard to the mid-run forecasting horizons, and more specifically referring to Figures 5 and 6, the results are qualitatively similar to those extracted by the statistical loss functions. The forecasts generated by HAR-VXD offer the highest cumulative returns compared to the remaining models. We also observe that cumulative returns are always positive, the median value is approximately 50% and the highest approaches 80% of predictive gains coming from the options straddle trading strategy. The performance of the remaining models, apart from the HAR-VXXLE and HAR-DMA-STO models, is worse when compared to the cumulative returns of RW or a simple AR(1) with most of the median values being negative. This means that there is no impact of IV indices, apart from VXD and VXXLE, on OVX when referring to mid-run forecasting horizons, namely 5- and 10-days ahead.

Finally, from Figure 7, we see that only forecasts obtained by the HAR-VXXLE and HAR-DMA-STO models offer the highest predictive gains from the options straddle trading strategy. The median value of cumulative returns coming from the use of forecasts of the aforementioned models is approximately 40%. These results are consistent with those of the statistical evaluation with the only exception the performance of the simple AR(1) model, which does not outperform the rest of the models in the longest-run forecasting horizon (22-days ahead).

---

[12]An outlier is considered a data point, which is 1.5 times outside the interquartile range.

6. Conclusion

In this paper, we focused primarily on how OVX could be predicted more accurately. More specifically, we investigate the impact of IV indices that represent different asset classes on the crude oil IV index. In this regard, we fill in the gap in the existing literature by extending the modelling framework with an out-of-sample investigation. There is a number of studies that draw conclusions relying only on in sample analysis for identifying relations between oil and other markets (e.g. volatility spillovers and causality). However, in this study, we investigate the predictive information of IV indices from other asset classes to OVX when generating out-of-sample forecasts up to 22-days ahead. The evaluation of the obtained forecasts has been done by applying not only statistical loss functions but also an option straddle trading strategy.

The first contribution of this paper lies in the preliminary analysis that is implemented, namely the rolling estimation of the Hurst exponent, which justifies the use of the HAR model. The results of the rescaled range analysis show that there is a strong existence of long memory in the implied volatility time series, especially in the time series of OVX. Moreover, regarding the modelling framework, we conclude that forecasters should consider using the DMA approach which enhances the performance of HAR models by allowing not only for parameters to change over time but also for different set of predictors to be applied over time. These are the features that anyone who is interested in generating out-of-sample forecasts of OVX should take into account. However, we found that even if the DMA methodology can offer predictive gains for short- and mid-run forecasting horizons, this is not the case for a forecasting horizon 22-days ahead.

Regarding the inclusion of IV indices that represent other asset classes in the HAR model, this paper makes the following contributions. First of all, from both the statistical and the economic results, we conclude that the HAR models including VXD, outperform the remaining ones for short- and mid-run forecasting horizons. From the evaluation of the 22-days ahead forecasts of OVX, we conclude that there is no significant predictive information of IV indices. The only exception is that of VXXLE, which seems to provide higher predictive information compared to models including other IV indices. The latter statement can be justified by the high cumulative returns coming from the options straddle trading strategy but not from the statistical evaluation results. These results provide further evidence for justifying the efficient market hypothesis and more particularly the fact that it is impossible to "beat the market" consistently on a risk-adjusted basis since market prices should only react to new information.

The results of this paper might be considered inspiring for both academics and investors and open new avenues for further research. More trading strategies could be implemented in order to evaluate the corresponding volatility forecasts generated by different models. Finally, a MIDAS modelling framework could be used with its main purpose the investigation of the impact of several policy uncertainty indicators on IV indices.

Acknowledgements

This research has been co-financed by the Operational Program "Human Resources Development, Education and Lifelong Learning" and is co-financed by the European Union (European Social Fund) and Greek national funds. The authors would like to express their gratitude for the comments and suggestions of Ms. Helga Stefansson.

References

Andersen, T.G., Bollerslev, T., 1998. Answering the Skeptics : Yes , Standard Volatility Models do Provide Accurate Forecasts. Int. Econ. Rev. 39, 885–905.

Andrada-Félix, J., Fernández-Rodríguez, F., Fuertes, A.M., 2016. Combining nearest neighbor predictions and model-based predictions of realized variance: Does it pay? Int. J. Forecast. 32, 695–715.

Angelidis, T., Degiannakis, S., 2008. Volatility forecasting: Intra-day versus inter-day models. J. Int. Financ. Mark. Inst. Money. 18, 449–465.

Anis, A., Lloyd, E., 1976. The expected value of the adjusted rescaled hurst range of independent normal summands. Biometrika 63, 111–116.

Awartani, B., Aktham, M., Cherif, G., 2016. The connectedness between crude oil and financial markets: Evidence from implied volatility indices. J. Commod. Mark. 4, 56–69.

Bašta, M., Molnár, P., 2018. Oil market volatility and stock market volatility. Financ. Res. Lett. 26, 204–214.

Blair, B.J., Poon, S.H., Taylor, S.J., 2001. Forecasting S&P 100 volatility: The incremental information content of implied volatilities and high-frequency index returns. J. Econom. 105, 5–26.

Buncic, D., Gisler, K.I., 2016. Global equity market volatility spillovers: A broader role for the United States. Int. J. Forecast. 32, 1317–1339.

Busch, T., Christensen, B.J., Nielsen, M.Ø., 2011. The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. J. Econom. 160, 48–57.

Chatziantoniou, I., Degiannakis, S., Delis, P., Filis, G., 2020. Forecasting oil price volatility using spillover effects from uncertainty indices. Financ. Res. Lett. , 101885.

Christensen, B.J., Prabhala, N.R., 1998. The relation between implied and realized volatility. J. financ. econ. 50, 125–150.

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. J. Financ. Econom. 7, 174–196.

Corsi, F., Renò, R., 2012. Discrete-Time Volatility Forecasting With Persistent Leverage Effect and the Link With Continuous-Time Volatility Modeling. J. Bus. Econ. Stat. 30, 368–380.

Degiannakis, S., 2008. Forecasting VIX. J. Money, Invest. Bank. 4, 5–19.

Degiannakis, S., Filis, G., 2017. Forecasting oil price realized volatility using information channels from other asset classes. J. Int. Money Financ. 76, 28–49.

Degiannakis, S., Filis, G., Hassani, H., 2018. Forecasting global stock market implied volatility indices. J. Empir. Financ. 46, 111–129.

Delis, P., Degiannakis, S.A., Filis, G., 2020. What Matters When Developing Oil Price Volatility Forecasting Frameworks? SSRN Electron. J. .

Dunis, C., Kellard, N.M., Snaith, S., 2013. Forecasting EUR-USD implied volatility: The case of intraday data. J. Bank. Financ. 37, 4943–4957.

Elder, J., Serletis, A., 2010. Oil price uncertainty. J. Money, Credit Bank. 42, 1137–1159.

Ferderer, J.P., 1996. Oil price volatility and the macroeconomy. J. Macroecon. 18, 1–26.

Fernandes, M., Medeiros, M.C., Scharth, M., 2014. Modeling and predicting the CBOE market volatility index. J. Bank. Financ. 40, 1–10.

Fleming, J., Ostdiek, B., Whaley, R.E., 1995. Predicting stock market volatility: A new measure. J. Futur. Mark. 15, 265–302.

Frijns, B., Tallau, C., Tourani-Rad, A., 2010. The information content of implied volatility: Evidence from australia. J. Futur. Mark. 30, 134–155.

Giot, P., 2003. The information content of implied volatility in agricultural commodity markets. J. Futur. Mark. 23, 441–454.

Gong, X., Lin, B., 2018. The incremental information content of investor fear gauge for volatility forecasting in the crude oil futures market. Energy Econ. 74, 370–386.

Grassi, S., Nonejad, N., De Magistris, P.S., 2017. Forecasting With the Standardized Self-Perturbed Kalman Filter. J. Appl. Econom. 32, 318–341.

Hansen, P., Lunde, A., Nason, J.M., 2011. The Model Confidence Set. Econometrica 79, 453–497.

Haugom, E., Langeland, H., Molnár, P., Westgaard, S., 2014. Forecasting volatility of the U.S. oil market. J. Bank. Financ. 47, 1–14.

Konstantinidi, E., Skiadopoulos, G., Tzagkaraki, E., 2008. Can the evolution of implied volatility be forecasted? Evidence from European and US implied volatility indices. J. Bank. Financ. 32, 2401–2411.

Koop, G., Korobilis, D., 2012. Forecasting inflation using dynamic model averaging. Int. Econ. Rev. 53, 867–886.

Koopman, S.J., Jungbacker, B., Hol, E., 2005. Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. J. Empir. Financ. 12, 445–475.

Liu, M.L., Ji, Q., Fan, Y., 2013. How does oil market uncertainty interact with other markets? Anempirical analysis of implied volatility index. Energy 55, 860–868.

Liu, Z., Tseng, H.K., Wu, J.S., Ding, Z., 2020. Implied volatility relationships between crude oil and the U.S. stock markets: Dynamic correlation and spillover effects. Resour. Policy 66.

Lv, W., 2018. Does the OVX matter for volatility forecasting? Evidence from the crude oil market. Phys. A Stat. Mech. its Appl. 492, 916–922.

Lyócsa, Š., Molnár, P., 2018. Exploiting dependence: Day-ahead volatility forecasting for crude oil and natural gas exchange-traded funds. Energy 155, 462–473.

Mandelbrot, B.B., Wallis, J.R., 1969. Some long-run properties of geophysical records. Water Resour. Res. 5, 321–340.

Prokopczuk, M., Symeonidis, L., Wese Simen, C., 2016. Do Jumps Matter for Volatility Forecasting? Evidence from Energy Markets. J. Futur. Mark. 36, 758–792.

Raftery, A.E., Kárný, M., Ettler, P., 2010. Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. Technometrics , 52–66.

Sánchez Granero, M.A., Trinidad Segovia, J.E., García Pérez, J., 2008. Some comments on Hurst exponent and the long memory processes on capital markets. Phys. A Stat. Mech. its Appl. 387, 5543–5551.

Sévi, B., 2014. Forecasting the volatility of crude oil futures using intraday data. Eur. J. Oper. Res. 235, 643–659.

Weron, R., 2002. Estimating long-range dependence: Finite sample properties and confidence intervals. Phys. A Stat. Mech. its Appl. 312, 285–299, arXiv:0103510.

	OVX	VIX	VXO	VXD	VXN	EVZ	GVZ
Mean	33.19	16.25	15.75	15.60	18.45	9.38	16.67
Median	31.35	14.82	14.41	14.37	17.08	8.73	16.01
Maximum	78.97	48.00	50.13	41.45	46.61	19.87	39.95
Minimum	14.50	9.14	6.32	7.58	10.31	4.69	8.88
Std. Dev.	10.80	5.40	5.78	4.69	5.19	2.82	4.91
Skewness	0.82	2.05	1.80	1.99	1.75	0.91	1.17
Kurtosis	3.53	8.46	7.50	8.10	6.84	3.48	5.07
CV	0.33	0.33	0.37	0.30	0.28	0.30	0.29

	VXSLV	VXXLE	VXGDX	BPVIX	EUVIX	JYVIX	TYVIX
Mean	28.83	22.85	36.06	8.93	9.23	9.91	5.23
Median	27.26	20.97	34.13	8.38	8.59	9.47	5.07
Maximum	80.64	57.47	63.95	29.10	20.25	18.37	9.57
Minimum	14.89	11.71	15.40	4.33	3.99	4.29	3.16
Std. Dev.	9.83	6.95	9.38	2.64	2.79	2.51	1.16
Skewness	1.45	1.49	0.41	2.22	0.91	0.50	0.89
Kurtosis	6.14	5.54	2.24	13.70	3.58	2.78	3.57
CV	0.34	0.30	0.26	0.30	0.30	0.25	0.22

Table 1: This table presents the descriptive statistics of the IV indices including the targeted variable, namely the OVX.

MSPE

Days ahead	**1**	**5**	**10**	**22**
RW	3.907	15.707	29.950	56.433
AR(1)	3.877	15.289	28.576	52.314
HAR	3.880	15.515	29.026	53.073
HAR-VIX	3.869	15.522	29.633	56.222
HAR-VXO	3.866	15.551	29.772	56.608
HAR-VXD	3.771	15.053	28.733	55.275
HAR-VXN	3.887	15.715	30.136	56.129
HAR-EVZ	3.887	15.637	29.370	53.929
HAR-GVZ	3.891	15.686	29.353	53.950
HAR-VXSLV	3.890	15.573	29.051	53.136
HAR-VXXLE	3.862	15.399	28.456	53.036
HAR-VXGDX	3.889	15.656	29.093	52.473
HAR-BPVIX	3.886	15.565	29.055	53.713
HAR-EUVIX	3.889	15.677	29.376	53.417
HAR-JYVIX	3.890	15.623	29.517	55.072
HAR-TYVIX	3.889	15.593	29.235	54.387
HAR-ALL	3.724	16.081	32.657	59.656
HAR-DMA-STO	3.658	15.089	28.566	55.413
HAR-DMA-ALL	3.681	15.887	39.518	72.181

Table 2: The results of the MSPE loss function for different forecasting horizons.

MAE				
Days ahead	**1**	**5**	**10**	**22**
RW	1.274	2.697	3.778	5.147
AR(1)	1.271	2.670	3.743	4.950
HAR	1.268	2.709	3.777	4.962
HAR-VIX	1.271	2.726	3.826	5.197
HAR-VXO	1.271	2.734	3.841	5.236
HAR-VXD	1.269	2.712	3.778	5.098
HAR-VXN	1.276	2.739	3.864	5.164
HAR-EVZ	1.266	2.706	3.779	4.946
HAR-GVZ	1.265	2.719	3.787	5.080
HAR-VXSLV	1.260	2.690	3.719	4.937
HAR-VXXLE	1.265	2.705	3.727	5.004
HAR-VXGDX	1.261	2.688	3.701	4.900
HAR-BPVIX	1.267	2.704	3.763	5.000
HAR-EUVIX	1.267	2.710	3.774	4.940
HAR-JYVIX	1.276	2.740	3.829	5.107
HAR-TYVIX	1.269	2.734	3.855	5.159
HAR-ALL	1.241	2.761	3.952	5.279
HAR-DMA-STO	1.231	2.654	3.707	5.119
HAR-DMA-ALL	1.226	2.681	4.176	5.948

Table 3: The results of the MAE loss function for different forecasting horizons.

MCS

Days ahead	**1**	**5**	**10**	**22**
RW	0.005	**0.401**	**0.218**	0.015
AR(1)	0.006	**0.729**	**0.989**	**1.000**
HAR	0.003	**0.699**	**0.704**	**0.432**
HAR-VIX	0.005	**0.699**	**0.457**	0.015
HAR-VXO	0.006	**0.699**	**0.457**	0.015
HAR-VXD	**0.244**	**1.000**	**0.942**	**0.406**
HAR-VXN	0.003	**0.401**	**0.218**	0.015
HAR-EVZ	0.003	**0.401**	**0.457**	**0.432**
HAR-GVZ	0.005	**0.401**	**0.704**	**0.432**
HAR-VXSLV	0.006	**0.704**	**0.925**	**0.865**
HAR-VXXLE	0.014	**0.704**	**1.000**	**0.910**
HAR-VXGDX	0.006	**0.699**	**0.925**	**0.910**
HAR-BPVIX	0.005	**0.699**	**0.853**	**0.432**
HAR-EUVIX	0.003	**0.401**	**0.457**	**0.432**
HAR-JYVIX	0.003	**0.401**	**0.457**	**0.206**
HAR-TYVIX	0.003	**0.401**	**0.457**	0.015
HAR-ALL	**0.244**	**0.401**	**0.218**	0.015
HAR-DMA-STO	**1.000**	**0.916**	**0.989**	**0.432**
HAR-DMA-ALL	**0.244**	**0.401**	**0.167**	0.001

Table 4: The results of the MCS test for different forecasting horizons. Figures in bold denote the model that belongs to the confidence set of the best performing models.

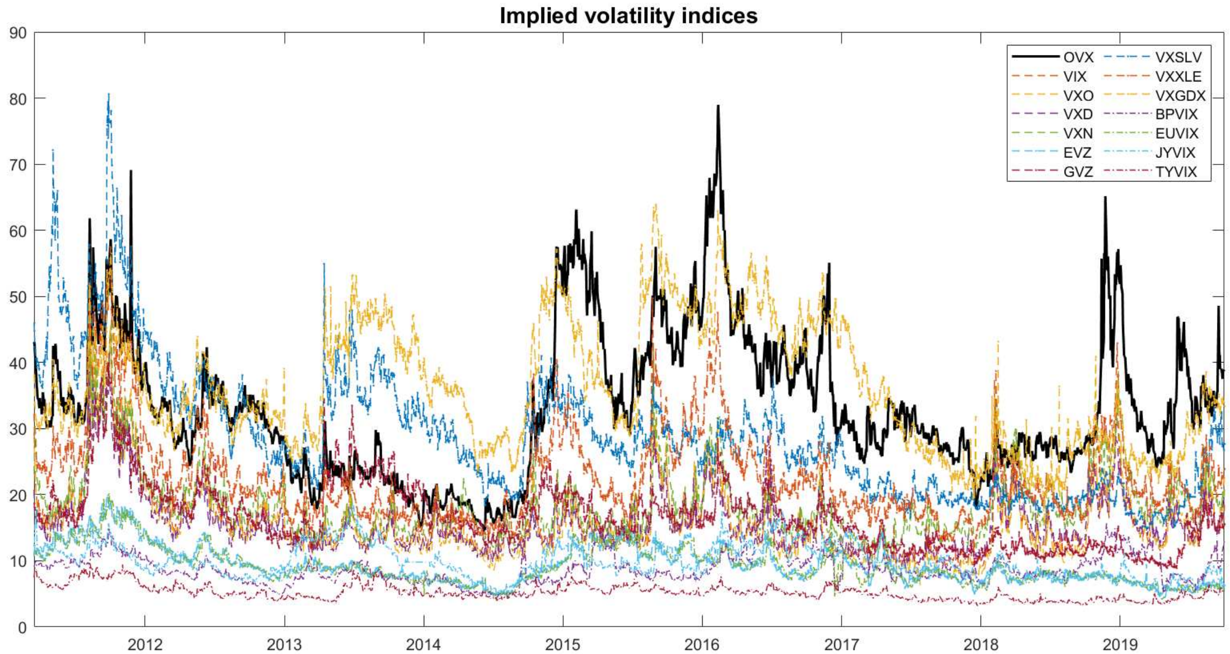

Figure 1: Evolution of the IV indices over the entire sample.

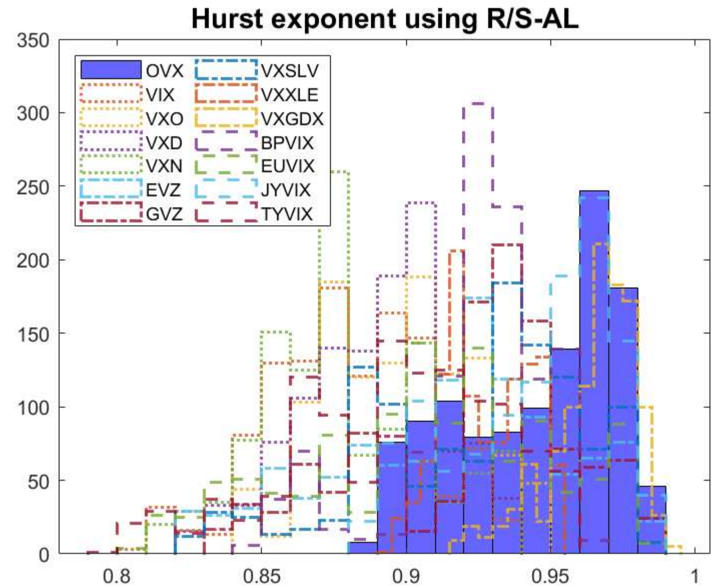

Figure 2: This histogram shows the values of the Hurst exponent under the rolling window implementation for all the IV indices focusing on the OVX time series.

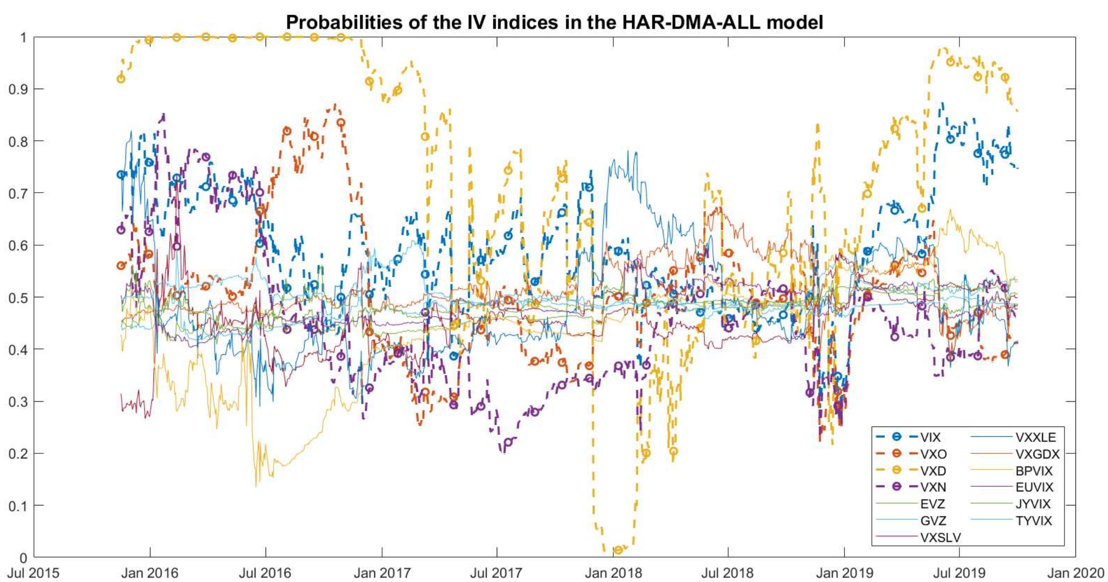

Figure 3: This figure plots the probabilities of the IV indices under the DMA methodology over the out-of-sample period. It is also important to mention that these probabilities are derived by forecasting OVX 1-day ahead.

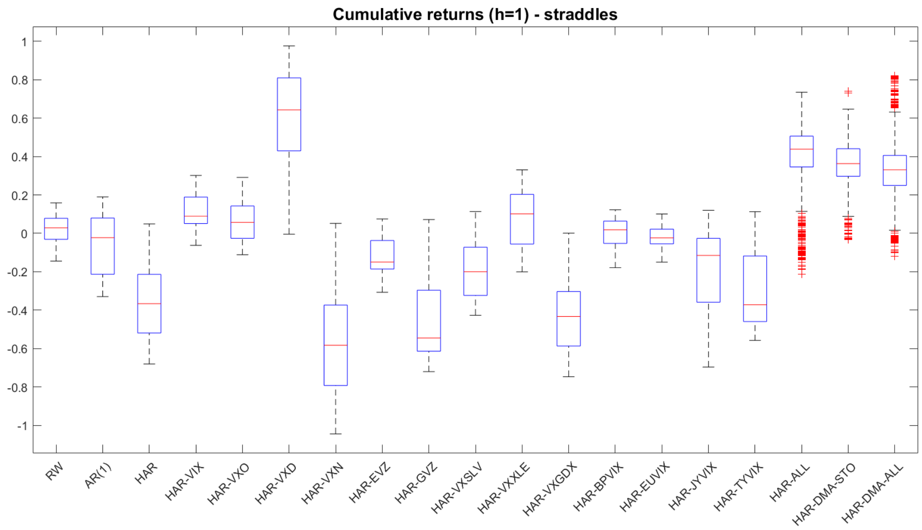

Figure 4: This figure presents the cumulative returns for the different models with the use of box plot. In this figure, the 1-day ahead forecasts of OVX have been used in the options straddle trading strategy. The outliers are considered data points, which are outside 1.5 times the interquartile range above the upper quartile and below the lower quartile.

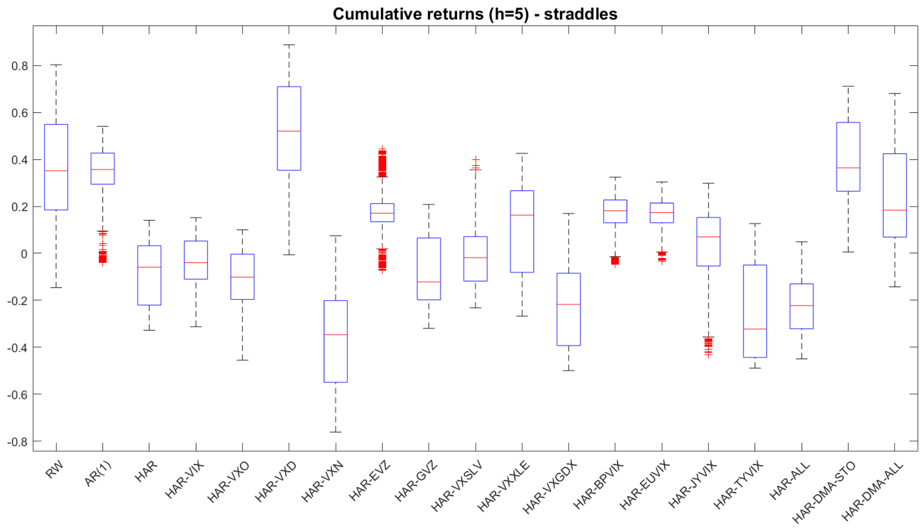

Figure 5: This figure presents the cumulative returns for the different models with the use of box plot. In this figure, the 5-days ahead forecasts of OVX have been used in the options straddle trading strategy. The outliers are considered data points, which are outside 1.5 times the interquartile range above the upper quartile and below the lower quartile.

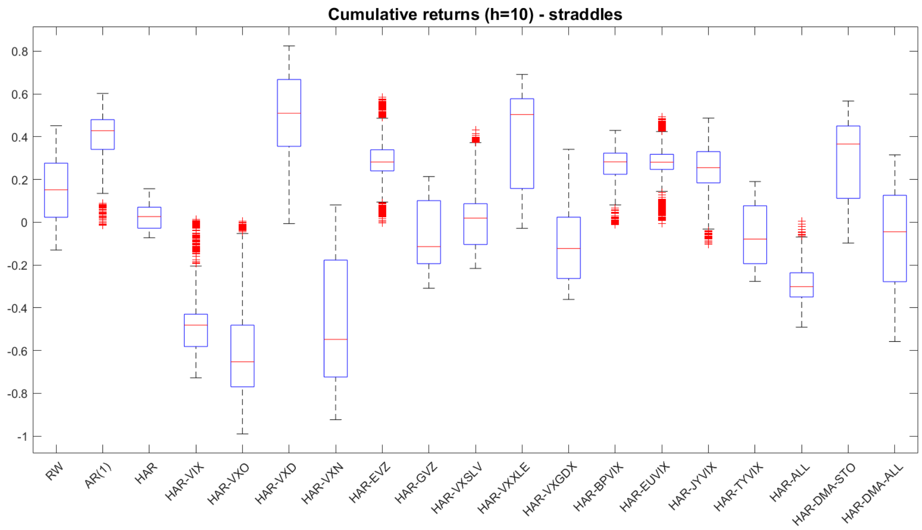

Figure 6: This figure presents the cumulative returns for the different models with the use of box plot. In this figure, the 10-days ahead forecasts of OVX have been used in the options straddle trading strategy. The outliers are considered data points, which are outside 1.5 times the interquartile range above the upper quartile and below the lower quartile.

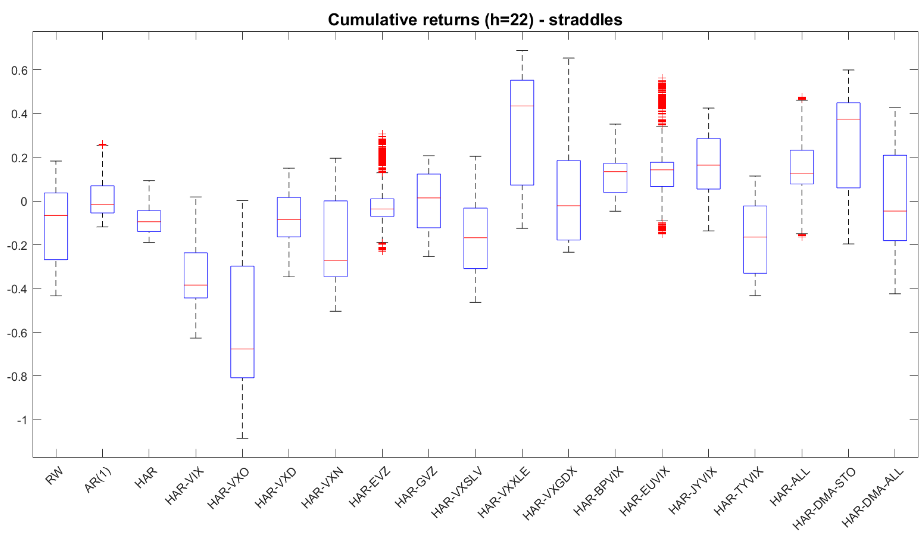

Figure 7: This figure presents the cumulative returns for the different models with the use of box plot. In this figure, the 22-days ahead forecasts of OVX have been used in the options straddle trading strategy. The outliers are considered data points, which are outside 1.5 times the interquartile range above the upper quartile and below the lower quartile.

Appendix. DMA methodology

In this part, we concentrate on one-step ahead forecasting procedure in order to show the updating steps of DMA method in detail. Regarding multi-period ahead forecasting, the idea is similar because the whole framework of DMA is based on Eq. (8), which describes how the direct forecasts are generated.

The main methodological approach of the updating equations of a time varying parameter (TVP) model is based on Kalman filter, which begins with the result:

$$(\boldsymbol{\alpha_{t-1}} \mid y^{t-1}) \sim N(\boldsymbol{\hat{\alpha}_{t-1}}, \boldsymbol{\Sigma_{t-1|t-1}}), \tag{A.1}$$

Kalman filtering process proceeds as follows:

$$(\boldsymbol{\alpha_{t}} \mid y^{t-1}) \sim N(\boldsymbol{\hat{\alpha}_{t-1}}, \boldsymbol{\Sigma_{t|t-1}}), \tag{A.2}$$

where $\boldsymbol{\Sigma_{t|t-1}} = \boldsymbol{\Sigma_{t-1|t-1}} + \boldsymbol{\Sigma_{u_t}}$.

Since we are motivated by the proposed approach of Grassi et al. (2017), the updating equation of $\boldsymbol{\Sigma_{t|t-1}}$ is perturbed by a function of the squared prediction errors, which is shown in the updating steps. At this step, we assume the following:

$$\boldsymbol{\Sigma_{t|t-1}} = \boldsymbol{\Sigma_{t-1|t-1}}. \tag{A.3}$$

At this point, we have to mention that due to the fact that we use the aforementioned approach, we no longer have to estimate $\boldsymbol{\Sigma_{u_t}}$. Kalman filter procedure is completed by the updating equation:

$$(\boldsymbol{\alpha_{t}} \mid y^{t}) \sim N(\boldsymbol{\hat{\alpha}_{t}}, \boldsymbol{\Sigma_{t|t}}), \tag{A.4}$$

where

$$\boldsymbol{\hat{\alpha}_{t|t}} = \boldsymbol{\hat{\alpha}_{t|t-1}} + \boldsymbol{\Sigma_{t|t-1}} \boldsymbol{x'_t} (\hat{H}_t + \boldsymbol{x_t} \boldsymbol{\Sigma_{t|t-1}} \boldsymbol{x'_t})^{-1} (y_t - \boldsymbol{x_t} \boldsymbol{\hat{\alpha}_{t-1}}), \tag{A.5}$$

and

$$\boldsymbol{\Sigma_{t|t}} = \boldsymbol{\Sigma_{t|t-1}} - \boldsymbol{\Sigma_{t|t-1}} \boldsymbol{x'_t} (\hat{H}_t + \boldsymbol{x_t} \boldsymbol{\Sigma_{t|t-1}} \boldsymbol{x'_t})^{-1} \boldsymbol{x_t} \boldsymbol{\Sigma_{t|t-1}} + \beta \cdot max\left[0, FL\left(\frac{\varepsilon_t^2}{\hat{H}_t} - 1\right)\right] \cdot I, \tag{A.6}$$

where $\varepsilon_t = y_t - \boldsymbol{x_t}\boldsymbol{\hat{\alpha}_{t-1}}$ and the estimated error variance is calculated by the following[13]:

$$\hat{H}_t = \kappa \hat{H}_{t-1} + (1-\kappa)\varepsilon_t^2. \tag{A.7}$$

Recursive forecasting is implemented by using the predictive distribution,

$$(y_t \mid y^{t-1}) \sim N(\boldsymbol{x_t}\boldsymbol{\hat{\alpha}_{t-1}}, \hat{H}_t + \boldsymbol{x_t}\boldsymbol{\Sigma_{t|t-1}}\boldsymbol{x'_t}). \tag{A.8}$$

After having estimated each individual model of the K combinations under the TVP modelling approach, which is explained analytically in the previous part, DMA averages the forecasts obtained by the individual models using $\pi_{t|t-1,k}$ as weights for $k = 1, \ldots, K$ over the out-of-sample period. Those DMA forecasts can be expressed as:

$$E(y_t \mid y^{t-1}) = \sum_{k=1}^{K} \pi_{t|t-1,k} \boldsymbol{x}^{\boldsymbol{(k)}}_{\boldsymbol{t-1}} \boldsymbol{\hat{\alpha}}^{\boldsymbol{(k)}}_{\boldsymbol{t-1}} \tag{A.9}$$

where $\boldsymbol{\hat{\alpha}}^{\boldsymbol{(k)}}_{\boldsymbol{t-1}}$ are Kalman filter estimates of the state-space model at time $t-1$.

At this point, probability in the forecasting model has to be determined. As proposed by Raftery et al. (2010), the relation between $\pi_{t|t-1,k}$ and $\pi_{t-1|t-1,k}$ is described as:

$$\pi_{t|t-1,k} = \frac{\pi^{\alpha}_{t-1|t-1,k}}{\sum_{l=1}^{K} \pi^{\alpha}_{t-1|t-1,l}} \tag{A.10}$$

where $0 < \alpha \leqslant 1$ is a forgetting factor[14], which is constant and smaller than 1.

The updating equation is defined as follows:

$$\pi_{t|t,k} = \frac{\pi_{t|t-1,k} f_k(y_t \mid y^{t-1})}{\sum_{l=1}^{K} \pi_{t|t-1,l} f_l(y_t \mid y^{t-1})} \tag{A.11}$$

where $f_k(y_t \mid y^{t-1})$ is the predictive density of model k. The main idea of this updating equation is that a model, which had a better forecasting performance in the past, will receive higher weight at time t.

[13]The design parameters β and κ are set as 1e-10 and 0.94, respectively.

[14]In this study, we follow Koop and Korobilis (2012) in setting $\alpha = 0.99$.