BANK OF GREECE
EUROSYSTEM

Working Paper

Forecasting VIX: The illusion of forecast evaluation criteria

Stavros Degiannakis
Eleftheria Kafousaki

322

JUNE 2023

WORKINGPAPERWORKINGPAPERWORKINGPAPERWORK

BANK OF GREECE
Economic Analysis and Research Department – Special Studies Division
21, E. Venizelos Avenue
GR-102 50 Athens
Tel: +30210-320 3610
Fax: +30210-320 2432

www.bankofgreece.gr

Published by the Bank of Greece, Athens, Greece

ISSN: 2654-1912 (online)
DOI: https://doi.org/10.52903/wp2023322

FORECASTING VIX: THE ILLUSION OF FORECAST EVALUATION CRITERIA

Stavros Degiannakis
Bank of Greece and Panteion University

Eleftheria Kafousaki
Panteion University

Abstract

The paper uses daily realized volatility measures in order to gain forecast accuracy over stocks' market implied volatility, as proxied by VIX Index, for forecast horizon of 1, 5, 10 and 22 days ahead. We evaluate forecast accuracy by incorporating a traditional statistical loss function, along with an objective-based evaluation criterion, that is the cumulative returns earned from the different HAR-type volatility models, through a simple yet effective trading exercise on VIX futures. Findings, illustrate how illusive the choice between the two metrics may be, as it ends in two contradicting results.

Keywords: Implied volatility forecasting, realized volatility measures, objective-based evaluation criteria

JEL classification: *C32, C53, G15*

Acknowledgements: The views expressed in this paper are those of the authors and not necessarily those of either the Bank of Greece or the Eurosystem.

Correspondence:
Stavros Degiannakis
Bank of Greece,
21 E. Venizelos Avenue, 10250
Athens, Greece
email: sdegiannakis@bankofgreece.gr

1.Introduction

And all goes around volatility, a substantial metric of risk and uncertainty, that is capable to govern the powerful transmission mechanisms of the financial system. Cult, by now, measures of dispersion, implied volatility and realized volatility, have long been consolidated, in academic research, financial policies, economic policy and trading practices (Andersen and Bollerslev, 1998; Degiannakis and Filis, 2022; Barndorff-Nielsen *et al.* 2011). Following the dramatic change in the global financial landscape, once more, we comprehend how valuable volatility forecasts, are for policy makers, governmental agencies, risk managers, traders and investors.

Over the last years, markets' dynamics have altered. Markets have been enriched with new, even more complex derivatives products, the moment that new hybrid trading platforms are set in action. New indices have been introduced to track the multiple trading strategies that are available to investors, resulting in increased trading volumes and complexity. All of the above, have substantially increased volatility. Traders and investors take a closer look at volatility, especially implied that is forward looking, in order to gain portfolio diversification, alpha generation and protection against capital loss. But they also pay special attention to another measure, the spread between implied and realized volatility, the volatility risk premium. Volatility risk premium gives a signal of the appropriate trading strategy, for long or short positions, depending on the risk exposure they are willing to dive into. Bollerslev *et al.* (2011), state that volatility risk premium is a measure of investors' implied risk aversion, interpreted as the possibility of negative events coming forward, capable of governing market decisions. After all, volatility risk premium is a priced fact, and systematically, implied volatility exceeds realized in many markets (Bakshi and Madan, 2006; Todorov, 2010), especially options markets (He *et al.*, 2015; Black and Szado, 2016). That is the reason why there are many studies that enhance their forecasting models with the addition of risk premium in order to forecast, mainly, the realized volatility of an underlying asset, or the price of that underlying asset, that can be stock, commodity as oil, exchange rate etc. (Bollerslev *et al.,* 2009; Degiannakis and Filis, 2020; Carr and Wu, 2008; Bollerslev and Todorov, 2011; Haugom *et al*., 2014). Also, there are noteworthy studies that highlight the linkage of risk premium to the macroeconomic environment, providing evidence of risk premium's multiple impact (Andersen *et al.,* 2015; Barndorff-Nielsen *et al.*, 2011; Bevilacqua *et al.,* 2018).

However, the bulk of volatility forecasting literature focuses on forecasting realized volatility, the recorded past volatility of an asset, as the availability of high frequency and ultra-high frequency data allowed the construction of model-free realized volatility measures that could replicate and approximate the concept of integrated volatility. Quadratic variation, bi-power variation, quantile-based realized variance, Min Realized variance, Med realized variance or positive and negative realized semi-variance, were some of the measures introduced able to deal efficiently with systemic bias, market frictions and microstructure noise and provide the best sampling frequency that could guarantee forecast accuracy (Andersen and Bollerslev, 1998; Andersen *et al*., 2003; Barndorff-Nielsen and Shephard, 2002;2004; Hansen and Lunde, 2006; Barndorff-Nielsen *et al.* 2008; Patton and Sheppard, 2015).

Only few studies forecast implied volatility through realized volatility measures, such as Christensen and Prabhala (1998), Degiannakis and Filis (2022), and Birkelund *et al.* (2015). Hence, in this study, we focus in a relation that should have been more extensively investigated, the link between future and past volatility, and add in the growing strand of volatility literature. So, we forecast stock markets' implied volatility, using Cboe Volatility Index, the VIX Index as proxy for implied volatility and for horizons spanning from 1 day to 22 days ahead. We apply a simple yet powerful model in effortlessly capturing some of volatility's stylized facts (long-memory, fat-tails, etc.), and in the same time aligning markets' participants diverse time horizons, namely the HAR framework[1] of Corsi (2009). We also apply its extensions as proposed in the works of Corsi and Renò (2012) and Degiannakis and Filis (2017; 2022), due to the addition of the extra predictors, one each time and combined, that of realized volatility of S&P 500 futures index, positive realized semi-variance[2], negative realized semi-variance of S&P 500 futures index and their risk premiums.

However, this is not the sole focus, as our work is not about just proving whether some of our HAR-type models can produce more accurate forecasts, nor is the fact

---

[1] We apply HAR framework as is the model traditionally used when high frequency data for the construction of the realized measures are available. When high frequency data are not available, realized volatility can be modeled through the classic GARCH-type models that serve as a valuable tool for modeling volatility in many distinguished studies (see Panagiotidis *et al*., 2022).

[2] We use positive and negative realized semi-variance along with the classic realized volatility, as extra predictors instead of other realized volatility estimators as Degiannakis and Filis (2022) in their work provide evidence that HAR models with positive and negative realized semi variances added as extra predictors of OVX (the 30-day volatility of the United States Oil Fund (USO) were among the best performing models according to their objective-based evaluation criteria.

whether realized volatility measures are meaningful or not for forecasting implied, compared to a benchmark as the simple autoregressive model, by simply testing forecast validity through classic loss functions, as mean square error. Our interest lies in following a distinctive path in volatility forecasting literature that lately flourishes more aggressively compared to past and aims in employing objective-based forecast evaluation criteria. Degiannakis and Filis (2022) by incorporating economic criteria, that are the cumulative returns earned from options trading strategies, in order to validate their models, along with the traditional loss functions, end in rather contradicting outcomes, stressing the point, that all depends on what is more desirable, profits or forecast accuracy?

Without indulging in the complexity, the trading of options may bare as in Degiannakis and Filis (2022) and Delis *et al.* (2023), we go through an extremely simple trading practice. We allow investors to go short or long in VIX futures based on the forecasted value of VIX we inquired for that specific day and for the forecast horizons we choose, compared to the actual price recorded that day. VIX futures are not incidentally chosen. They are among the most tradable products, since VIX index is not, and are said to have unique return drivers (Moran and Dash, 2007) and unique properties (Szado, 2018), as their returns are negatively correlated with equities but highly correlated to VIX. Most of the times their value is above spot VIX levels (in contango), but also, there are moments that is below (in backwardation), but they are not that sensitive to market movements compared to spot values. Overall, they constitute a powerful tool in traders and investors' arsenal when used and included properly in investing portfolios.

The results accomplished through this practice, are rather surprising, as according to our objective-based evaluation criteria, most of our HAR-type models, outperformed benchmark and produced superior cumulative returns for at least 1 day ahead horizon, while at the same time, the simple HAR for implied volatility outperformed benchmark for all forecast horizons.

The remainder of the paper is structured in the following manner. Sections 2 and 3 give a comprehensive description of the data used, their properties and the modelling and forecasting methodology followed. Section 4 wanders around the outcome of our research, while section 5 sets the final remarks by concluding the paper.

2. Data and descriptive statistics

2.1. Data description

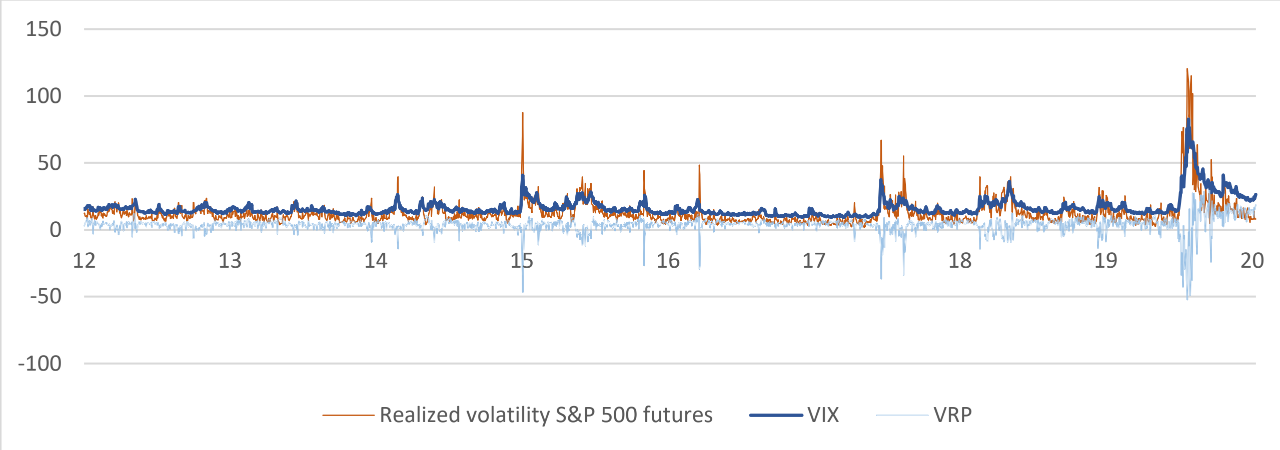
For our study we use high frequency, 5-min[3] returns of S&P 500 futures index, for constructing daily realized volatility[4], RV, as proposed by Andersen and Bollerslev (1998), daily positive realized semi variance, RSV (+) and daily negative realized semi variance, RSV (-) as proposed by Barndorff-Nielsen *et al.* (2011), which are able to capture the variation coming from positive or negative returns. As a proxy for stock's market implied volatility, we use the daily closing prices of the VIX Cboe volatility index. In our work we also include, after construction, the spreads[5] of the above realized metrics from VIX, volatility risk premium, VRP, volatility risk premium from the RSV (+), and VRP (+), volatility risk premium from RSV (-), VRP (-). High frequency data were obtained through TickData, whereas the implied volatility data were retrieved from CBOE. The sampling period, based on data availability of realized measures, spans from 22nd of August 2012 up to 31nth of August 2020.

2.2. Descriptive statistics

Busch *et al.* (2011), in their work, have found implied volatility to systematically exceed realized in stock and bond markets and the same we observe in Figure 1. Risk premium is present, for the entire period under investigation. Furthermore, there are evident some abrupt upward movements of realized volatility, resulting in a negative risk premium, perfectly compatible with the way markets' function, during turmoil periods and volatility exposure protection strategies.

[3] Our choice to use realized volatility, positive realized semi-variance and negative realized semi-variance constructed out of 5-min returns, is governed by the strong evidence that sampling frequency of returns ranging from 5-min and 15-min to 30 min is able to effectively eliminate microstructure noise, and other systemic frictions (Andersen *et al.* 2005; Degiannakis and Filis, 2020; Liu *et al.* 2015; Bollerslev *et al.* 2011).
[4] All realized measures that appear in this paper, are constructed in annualized form, since VIX is in annualized form. Detailed description on the methodology applied for VIX calculation, is provided through https://cdn.cboe.com/resources/vix/VIX_Methodology.pdf
[5] The volatility risk premium variable for each annualized realized measure is created by simply extracting each measure from VIX.

Figure 1. The volatility risk premium existence



Note: For almost the entire period under investigation VIX index is above realized volatility of S&P 500 futures Index, resulting in a priced volatility risk premium.

Volatility measures, also come with some special properties. Andersen *et al.* (2005) have studied realized volatility at different markets and report distributions of the realized daily variances, to be skewed to the right and leptokurtic. They also state that distributions of the logarithmic transformations are approximately normal. The exact same findings we report for our metrics. Table 1 presents the descriptive statistics along with the descriptive statistics of their logarithmic transformation. Logarithmic form has indeed more desirable properties, especially for linear models, as the ones we incorporate, and is the form that will be used for our modelling framework.

Table 1. Descriptive Statistics of VIX and the realized measures of S&P500 futures

	VIX	Log VIX	RV	Log RV	RSV +	Log RSV+	RSV -	Log RSV-
Mean	16.265	2.7311	12.6678	2.3756	8.8864	2.0134	8.7681	1.9894
Median	14.3	2.667	10.1961	2.322	7.1511	1.9672	7.1534	1.9675
Max	82.69	4.415	120.48	4.79148	113.178	4.7289	97.008	4.5747
Min	9.14	2.2126	1.715	0.5394	0.9897	-0.010	0.5706	-0.561
St. Dev	7.05	0.3097	9.99	0.5254	7.5881	0.534	6.8449	0.3307
Coef of Var	43.39	11,34	78.86	22.11	85.39	26.52	78.06	28.90
Skewness	3.928	1.6328	4.8279	0.772	6.11	0.7544	4.5443	0.3195
Kurtosis	22.689	4.081	34.5917	1.8369	57.1935	2.2571	36.3448	1.3467
J-bera	48289.74**	2289.08**	108057.6**	482.15**	286018.2**	618.40**	117562.2**	185.45**
ADF	-4.8484**	-5.3386**	-6.3449**	-5.370**	-6.763**	-5.471**	-6.895**	-7.0052**
Correlation			0.8317		0.7871		0.7696	

Note: RV=realized volatility of S&P500 futures index, RSV+=positive realized semi variance, RSV- =negative realized semi variance. The columns of Log VIX, Log RV, Log RSV+ and Log RSV-, present the descriptive statistics of the logarithmic transformation of our data and Correlation at the first column, returns the correlation of each variable with VIX. All variables found to exhibit normal distribution, but all were found to be stationary. * Denotes statistical significance at 1%.

3. Methods

3.1. The autoregressive model

We start from estimating a simple benchmark model as the autoregressive with 1 lag, AR1, as it is accustomed simple models to outperform the more advanced ones. We use AR1, compared to other naive models that could be used instead, due to the close resemblance it has to the HAR framework, which incorporates extra lags in order to model volatility. The AR1 model for the logarithm of VIX, is written as follows:

$$\log(VIX_t) = w_1 + w_2 \log(VIX_{t-1}) + \varepsilon_t, \tag{1}$$

where, w_1 and w_2 are the coefficients to be estimated and ε_t the residuals that thought to be normally distributed, $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$.

3.2. The HAR-IV and HAR-IV-X frameworks for in-sample estimation

Heterogeneous Autoregressive model, HAR, a model aligned with market's fractal structure, was originally proposed for realized volatility by Corsi (2009). Corsi and Reno (2012), Degiannakis and Filis (2017) among others, extended it to allow for additional regressors. In line with predecessors, we employ HAR, an additive linear combination of indicators of volatility components at different time horizons, 1-day, 5-days and 22-days. The simple HAR for the in-sample estimation of implied volatility, HAR-IV model[6] is given by:

$$\begin{aligned}\log(VIX_t) = w_0 + w_1 \log(VIX_{t-1}) + w_2\left(5^{-1}\sum_{k=1}^{5}\log(VIX_{t-k})\right) \\ + w_3\left(22^{-1}\sum_{k=1}^{22}\log(VIX_{t-k})\right) + \varepsilon_t,\end{aligned} \tag{2}$$

where, w_0, w_1, w_2 and w_3 are the coefficients to be estimated and ε_t is normally distributed, $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$.

[6] See also Degiannakis *et al.* (2022).

The rest HAR-type models with the extra exogenous regressors, HAR-IV-X, alone and in pairs that of realized volatility, HAR-IV-RV, positive realized semi variance, HAR-IV-RSV(+), negative realized semi variance, HAR-IV-RSV(-), volatility risk premium, HAR-IV-VRP, volatility risk premium from RSV(+), HAR-IV-VRP(+), volatility risk premium from RSV(-), HAR-IV-VRP(-) and their combinations HAR-IV-RV-VRP, HAR-IV-RSV(+)-VRP(+) and HAR-IV-RSV(-)-VRP(-) are obtained by[7]:

$$\begin{aligned}\log(VIX_t) &= w_0 + w_1\log(VIX_{t-1}) + w_2(5^{-1}\sum_{k=1}^{5}\log(VIX_{t-k})) \\ &+ w_3(22^{-1}\sum_{k=1}^{22}\log(VIX_{t-k})) + w_4\log(RV_{t-k}) \\ &+ w_5(5^{-1}\sum_{k=1}^{5}\log(RV_{t-k})) + w_6(22^{-1}\sum_{k=1}^{22}\log(RV_{t-k})) + \varepsilon_t\end{aligned} \tag{3}$$

and

$$\begin{aligned}\log(VIX_t) &= w_0 + w_1\log(VIX_{t-1}) + w_2(5^{-1}\sum_{k=1}^{5}\log(VIX_{t-k})) \\ &+ w_3(22^{-1}\sum_{k=1}^{22}\log(VIX_{t-k})) + w_4(VRP_{t-k}) \\ &+ w_5(5^{-1}\sum_{k=1}^{5}(VRP_{t-k})) + w_6(22^{-1}\sum_{k=1}^{22}(VRP_{t-k})) + \varepsilon_t.\end{aligned} \tag{4}$$

For the case of HAR-IV-X[8], pairs combination, HAR-IV-RSV (+)-VRP (+) and HAR-IV-RSV (-)-VRP (-), extra w_7, w_8 and w_9 coefficients have to be estimated. In total we construct eleven (11) forecast models, including the AR1 as well.

3.3 Generating real out-of-sample forecasts

For our forecasting exercise, we generate real out-of-sample forecasts by incorporating a rolling window approach. The initial data sample consists of 2020

[7] By RV in Eq. 3, and VRP in Eq. 4 and the equations that are about to follow, we denote all the different combinations of realized volatility measures and volatility risk premiums that we test. And the same goes for the combined pairs. Representation would be rather space consuming without adding new information.

[8] It should be mentioned that combined models were test for multicollinearity as volatility risk premium used is the outcome of the difference of VIX_t and RV_t. Fortunately, although correlated, no severe detection spotted to allow for extra methods to be applied or outcome of models to be jeopardized.

trading days. We use the first 1000 observations for the in-sample estimation, that is from 20th of August 2012 up to 12th of August 2016 and the rest 1020 for producing real out-of-sample iterated forecasts for 1, 5, 10 and 22 days ahead horizon, with a fixed rolling window of 1000 observations. Since we have used the log form of VIX for estimating our models, forecasts for the 1-day-ahead horizon are given by Eq. 5 and 6:

$$VIX_{t+1|t} = \exp\left(\hat{w}_0 + \hat{w}_1 \log(VIX_t) + \hat{w}_2\left(5^{-1}\sum_{k=1}^{5}\log(VIX_{t-k+1})\right) + \hat{w}_3\left(22^{-1}\sum_{k=1}^{22}\log(VIX_{t-k+1})\right) + {}^1\!/_2\,\hat{\sigma}_{\varepsilon}^2\right), \tag{5}$$

and

$$VIX_{t+1|t}^{\tau} = \exp\left(\hat{w}_0 + \hat{w}_1 \log(VIX_t) + \hat{w}_2\left(5^{-1}\sum_{k=1}^{5}\log(VIX_{t-k+1})\right) + \hat{w}_3\left(22^{-1}\sum_{k=1}^{22}\log(VIX_{t-k+1})\right) + \hat{w}_4 \log(RV_t) + \hat{w}_5\left(5^{-1}\sum_{k=1}^{5}\log(RV_{t-k+1})\right) + \hat{w}_6\left(22^{-1}\sum_{k=1}^{22}\log(RV_{t-k+1})\right) + {}^1\!/_2\,\hat{\sigma}_{\varepsilon}^2\right). \tag{6}$$

For generating forecasts for horizons of t+2 days-ahead and up to 22 days ahead, we proceed in line with Degiannakis and Filis (2017;2022) and Degiannakis *et al.* (2022) and produce real out-of-sample forecasts that enhance a rather mis-specified element in forecasting toolbox, dealing with the information set of exogenous variables that is not available to forecasters at t+2 days ahead and on. We produce forecasts of the t+s days ahead values of our exogenous variables, using satellite HAR models and insert them back to our forecasting framework. Eq. 7 and 8 give an indication of how forecasted and constructed exogenous variables' information enters forecasting equation for horizon t+2 and on:

$$VIX_{t+s|t} = \exp\Big(\widehat{w}_0 + \widehat{w}_1 \log\left(VIX_{t+s-1|t}\right) + \widehat{w}_2\left(s^{-1}\sum_{k=1}^{s-1} log(VIX_{t-k+s|t}) + (5-s)^{-1}\sum_{k=s}^{5}\log\left(VIX_{t-k+s}\right)\right) + \widehat{w}_3\left(s^{-1}\sum_{k=1}^{s-1} log(VIX_{t-k+s|t}) + (22-s)^{-1}\sum_{k=s}^{22}\log\left(VIX_{t-k+s}\right)\right) + {}^{1}\!/_{2}\,\hat{\sigma}_{\varepsilon}^{2}\Big). \tag{7}$$

And for the multivariate HAR-IV-X models, the appropriate form is:

$$\begin{aligned} VIX_{t+s|t} = \exp\Big(&\widehat{w}_0 + \widehat{w}_1 \log\left(VIX_{t+s-1|t}\right) \\ &+ \widehat{w}_2\left(s^{-1}\sum_{k=1}^{s-1} log(VIX_{t-k+s|t}) + \left(5-s)^{-1}\sum_{k=s}^{5}\log\left(VIX_{t-k+s}\right)\right)\right) \\ &+ \widehat{w}_3\left(s^{-1}\sum_{k=1}^{s-1} log(VIX_{t-k+s|t}) + (22-s)^{-1}\sum_{k=s}^{22}\log\left(VIX_{t-k+s}\right)\right) \\ &+ \widehat{w}_4 \log\left(RV_{X,t-k+s|t}\right) \\ &+ \widehat{w}_5\left(s^{-1}\sum_{k=1}^{s-1} log(RV_{t-k+s|t}) + \left(5-s)^{-1}\sum_{k=s}^{5}\log\left(RV_{t-k+s}\right)\right)\right) \\ &+ \widehat{w}_6\left(s^{-1}\sum_{k=1}^{s-1} log(RV_{t-k+s|t}) + (22-s)^{-1}\sum_{k=s}^{22}\log\left(RV_{t-k+s}\right)\right) \\ &+ {}^{1}\!/_{2}\,\hat{\sigma}_{\varepsilon}^{2}\Big). \end{aligned} \tag{8}$$

When the investigated models are the combinations of the realized measures, then in Eq. 8, we add the extra fitted components.

3.4 The forecast evaluation criteria

We incorporate two different evaluation criteria in order to access the forecast accuracy of the generated values out of the eleven different forecast models. We use the classical loss function of mean squared error (MSE), in place of the statistical criterion and the cumulative returns accomplished from the proposed models, by following a trading practice, in place of the economic criterion. Eq. 9, returns MSE:

$$MSE = T^{-1} \sum (VIX_{i,t+s|t} - VIX_{t+s})^2, \tag{9}$$

where, $\mathrm{VIX}_{t+s|t}$ and VIX_{t+s} are the forecasted and the actual values of VIX volatility index, respectively. T is the out-of-sample forecast period and i denotes the different forecast models, where i=1, 2, …, 11.

The trading practice we incorporate, follows a simple yet powerful trading rule:

1. If $\mathrm{VIX}_{i,\,t+s|t} > \mathrm{VIX}_t$ then go short on VIX futures.
2. If $\mathrm{VIX}_{i,\,t+s|t} < \mathrm{VIX}_t$ then go long on VIX futures.

Short or long positions are translated into selling or purchasing VIX futures respectively. So, depending on the outcome generated out by the trading rule, we calculate the cumulative returns for each model.

4. Findings

In Table 2 and 3 we present the values for the statistical and the economic criteria we applied in our study in order to evaluate the forecasting performance of the proposed models. Additionally, since we have a multiple comparisons problem, we also incorporate a "benchmark-free" process, with several advances over other identical tests, the model confidence set, MCS, of Hansen *et al.* (2011)[9]. MCS will help us decide over the final outcome and whether HAR-type models proposed outperformed or not AR1. The p-values of MCS test are reported to the columns next to the ones of the MSE and of cumulative returns for the 1, 5, 10 and 22 days ahead forecast horizon.

Table 2. The MSE and MCS p-values

Model	FORECASTING HORIZON							
	1 day		5 days		10 days		22 days	
	MSE	MCS	MSE	MCS	MSE	MCS	MSE	MCS
AR (1)	4.7999	0.935*	19.174	1.000*	37.247	0.961*	67.561	0.703
HAR-IV	4.7915	1.000*	19.353	0.992*	37.391	0.961*	65.332	1.000*
HAR-IV-RV	4.7958	0.935*	19.712	0.504	38.181	0.484	66.330	0.929*
HAR-IV-VRP	4.9883	0.648	22.297	0.633	41.169	0.606	66.624	0.929*
HAR-IV-RV-VRP	5.0995	0.752	20.541	0.519	40.831	0.530	73.133	0.394
HAR-IV-RSV (+)	4.7931	0.958*	19.263	0.992*	37.169	1.000*	66.100	0.929*
HAR-IV-VRP (+)	4.8531	0.648	20.242	0.633	38.876	0.606	66.271	0.929*
HAR-IV-RSV (+)-VRP (+)	5.0549	0.669	25.219	0.633	62.457	0.606	143.51	0.394
HAR-IV-RSV (-)	4.8130	0.818	19.849	0.633	38.410	0.606	65.722	0.929*

[9] Model Confidence Set test of Hansen *et al.* (2011) returns a set of best performing models among the ones proposed depending on the user-defined power of the test.

HAR-IV-VRP (-)	4.8546	0.686	20.504	0.633	38.926	0.606	65.931	0.929*
HAR-IV-RSV (-)-VRP (-)	4.8095	0.878	20.519	0.633	39.555	0.620	71.310	0.693

Note: results of mean square error are for the forecasted values of VIX in their original form not the logarithms. * Denotes the models that are included in the set of the best performing models based to model confidence set test. Only models over 0.9 are included in MCS.

According to Table 2, the models that are chosen from MCS based on their MSE for 1 day ahead are HAR-IV, HAR-IV-RSV+, HAR-IV-RV and AR1. For the horizon of 5 and 10 days ahead the chosen ones were HAR-IV, HAR-IV-RSV (+) and AR1. For the 22 days ahead horizon, surprisingly, come HAR-IV, HAR-IV-RV, HAR-IV-VRP, HAR-IV-RSV (+), HAR-IV-VRP (+), HAR-IV-RSV (-) AND HAR-IV-VRP (-), while AR1 is not even included. The results of Table 3, give a more profound outcome. Most of the HAR-type models overperformed benchmark by generating higher cumulative returns, especially for 1 day ahead horizon. Of course, only HAR model included in MCS test set due to the high p-value we imposed, but it certainly gives us a strong indication of HAR's superiority, as is the model included in MCS for all forecast horizons. Another model that was included for the 5 and 10 days ahead, was the HAR-IV-VRP, that also ended with higher cumulative returns. For 10 days ahead horizon, there is also another model included in the set, the HAR-IV-RSV (-)-VRP (-). From Table 3, we deduce that simple HAR for VIX index in the best performing model out of all with cumulative returns by far exceeding the ones gained by AR1. Most of our HAR-type models generated superior cumulative returns compared to benchmark, no matter if they were eventually included or not in the MCS. So, the inclusion of the realized measures of the S&P 500 futures index and their respective risk premiums as extra regressors in the incorporated HAR framework, resulted in models generating higher cumulative returns at least for the short-run horizon and the inclusion of VRP premium and VRP (+) for longer horizons. Overall, according to the statistical criterion the inclusion of realized measures and their risk premiums and even the use of HAR framework in order to forecast stock market's implied volatility is of no use. But economically, refraining from using HAR framework and including the realized measures and their risk premiums an investor would also refrain from alpha generation.

Table 3. The cumulative returns and MCS p-values

Model								
	FORECASTING HORIZON							
	1 day		5 days		10 days		22 days	
	Cum. Returns	MCS	Cum. Returns	MCS	Cum. Returns	MCS	Cum. Returns	MCS
AR(1)	303%	0.765	209%	0.886	226%	0.921*	206%	0.859
HAR-IV	487%*	1.000*	291%*	0.917*	270%*	1.000*	271%*	1.000*
HAR-IV-RV	352%*	0.812	133%	0.356	194%	0.743	102%	0.678
HAR-IV-VRP	428%*	0.812	273%*	0.917*	220%	0.921*	230%*	0.859
HAR-IV-RV-VRP	326%*	0.812	208%	0.895	32%	0.350	42%	0.547
HAR-IV-RSV+	295%	0.608	191%	0.417	153%	0.435	36%	0.547
HAR-IV-VRP+	373%*	0.812	179%	0.626	185%	0.780	246%*	0.859
HAR-IV-RSV(+)-VRP+	234%	0.277	22%	0.626	115%	0.780	23%	0.547
HAR-IV-RSV-	265%	0.277	162%	0.576	201%	0.921*	169%	0.790
HAR-IV-VRP-	355%*	0.812	187%	0.562	181%	0.810	151%	0.790
HAR-IV-RSV(-)-VRP(-)	350%*	0.812	387%*	1.000*	184%	0.831	-3%	0.547

Note: Cumulative returns are in percentage form and appear in left column of each forecast horizon. * Denotes the models that are included in the set of the best performing models according to model confidence set test.

5. Conclusion

Through this work we examine whether the inclusion of different realized volatility measures and their risk premiums can provide more accurate forecasts for stock market's implied volatility utilizing the HAR framework compared to a benchmark model. But that is not the main and the only target. We also investigate whether the forecast performance of the competing eleven models, is accessed differently under different evaluation criteria and show how this is linked with profits or possible losses. We use VIX Cboe volatility index as a proxy for the implied volatility and 5 min returns of S&P 500 futures index in order to construct the 3 realized volatility measures and their risk premiums. We also use the classic loss function of MSE in place of the statistical evaluation criterion and the cumulative returns, gained from the different forecast models under a simple trading practice, in place of the objective-based evaluation criterion.

The findings of our study are rather interesting as they highlight how contradictive, the outcome accomplished through the two different evaluation criteria, can actually be. According to the statistical criterion none of our models were able to outperform benchmark model. Neither the simple HAR, nor the inclusion of realized measures, their risks premiums or both, were capable of generating more accurate forecasts over AR1 model. But based on the economic criterion and the outcome presented in Table 3, the simple HAR model for implied volatility was by far the best performing, for all forecast horizons, followed by the HAR with the addition one each time of VRP, VRP(+), VRP(-), RV and their combinations of RV-VRP, RSV(-)-VRP(-) for the short-run horizon, as well as VRP for longer horizon. All resulting in noticeable excess returns compared to AR1 model.

Concluding, let us stress out that by following a naïve trading practice, as the one used in our study, we replicate the way markets indeed function. Every time an index ticks a new sentiment is formed and new trading strategies instantaneously are set in action. For a market participant, whether she/he is a trader, investor, asset management fund etc., excess returns and portfolio diversification for capital loss protection are essential decision drivers. The VIX futures incorporated here, are not plain assets but powerful hedging instruments and of course markets provide a plentiful of such hedging instruments. So, what we actually state here is that for a market participant who desires to engage in trading activities using a model-based trading strategy, the choice of evaluation criteria can have a huge impact on the final outcome. Infinitesimal small differences between forecasted values of different models, that statistically may be indifferent, can actually lead to a contradicting outcome when object-based evaluation criteria are used. So, the model and the accuracy tool used must be carefully chosen, depending on what is more desireful, as they can be rather deceptive.

References

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

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica 71 (2)*, 579–625.

Andersen, T. G., Bollerslev, T. and Meddahi, N. (2005) Correcting the errors: volatility forecast evaluation using high-frequency data and realized volatilities, *Econometrica,* 73 (1), 279–296.

Andersen, T. G., Fusari, N. and Todorov, V. (2015) Parametric inference and dynamic state recovery from options panels, *Econometrica,* 83, 1081-1145

Bakshi, G., and Madan, D. (2006) A theory of volatility spreads, *Management Science,* 52(12), 1945-1956.

Barndorff-Nielsen, O. E., Hansen, P.R., Lunde, A. and Shephard, N. (2008) Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise, *Econometrica,* 76 (6), 1481–1536.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A. and Shephard, N. (2011) Subsampling realized kernels, *Journal of Econometrics,* 160 (1), 204–219.

Barndorff-Nielsen, O.E. and Shephard, N. (2002) Estimating Quadratic Variation Using Realized Variance, *Journal of Applied Econometrics*, 17, 457-477.

Barndorff-Nielsen, O. and Shephard, N., (2004) Power and bipower variation with stochastic volatility and jumps. *J. Financial Econometrics 2 (1)*, 1–37.

Bevilacqua, M., Morelli, D. and Tunaru, R. (2018) The Determinants of the Model-Free Positive and Negative Volatilities, *Journal of International Money and Finance*, 92, 1-24.

Birkelund, O., Haugom, E., Molnár, P., Opdal, M., and Westgaard, S. (2015) A comparison of implied and realized volatility in the Nordic power forward market, *Energy Economics*, 48, 288-294.

Black, K. and Szado, E. (2016) Performance Analysis of options-based equity Mutual Funds, Closed-End Funds, and Exchange-Traded Funds", *Journal of Wealth Management,* 19(1), 51-69.

Bollerslev, T., Gibson, M. and Zhou, H. (2011) *Dynamic Estimation of Volatility Risk Premia and Investor Risk Aversion from Option-Implied and Realized Volatilities*, CREATES Research Paper.

Bollerslev, T., Tauchen, G., and Zhou, H. (2009) Expected stock returns and variance risk premia, *Review of Financial Studies,* 22(11*),* 4463–4492.

Bollerslev, T. and Todorov, V. (2011) Tails, Fears and risk premia, *Journal of Finance,* 66, 2165-2211.

Busch, T., Christensen, B. J., and Nielsen, M. Ø. (2011) The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Econometrics,* 160 (1), 48–57.

Carr, P. and Wu, L. (2008) Variance risk premiums, *Review of Financial Studies,* 22(3), 1311–1341.

Christensen, B. J. and Prabhala, N. R. (1998) The relation between implied and realized volatility, *Journal of Financial Economics*, 50, 125–150.

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

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

Degiannakis, S. and Filis, G. (2017) Forecasting oil price realized volatility using information channels from other asset classes, *Journal of International Money and Finance,* 76, 28–49.

Degiannakis, S. and Filis, G. (2020) Oil price assumptions for macroeconomic policy, MPRA Paper 100705, University Library of Munich, Germany.

Degiannakis, S. and Filis G. (2022) Oil price volatility forecasts: What do investors need to know? *Journal of International Money and Finance,* 18(5), 449–465.

Degiannakis, S., Filis, G., Klein, T. and Walther, T. (2022) Forecasting Realized Volatility of Agricultural Commodities, *International Journal of Forecasting*, 38(1), 74-96.

Delis, P., Degiannakis, S. and Giannopoulos, C. (2023) What should be taken into consideration when Forecasting Oil Implied Volatility Index? *Energy Journal*, 44(5), forthcoming.

Hansen, P. R. and Lunde, A. (2006) Realized variance and market microstructure noise, *Journal of Business and Economic Statistics,* 24(2), 127–161.

Hansen, P. R., Lunde, A. and Nason, J. M. (2011) The model confidence Set, *Econometrica*, 79(2), 453-497.

Haugom, E., Langeland, H., Molnár, P. and Westgaard, S. (2014). Forecasting volatility of the US oil market. *J. Bank. Finance 47*, 1–14.

He, D. X., Usu, C. J. and Rue, N. (2015) Option-Writing Strategies in a Low Volatility Framework, *Journal of Investing,* 24(3), 116-28.

Liu, Y. L., Patton, J. A. and Sheppard, K. (2015) Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes, *Journal of Econometrics,* 187, 293-311.

Moran, M. T. and Dash, S. (2007) VIX Futures and Options: Pricing and Using Volatility Products to Manage Downside Risk and Improve Efficiency in Equity Portfolios, *The Journal of Trading*, 2(3), 96 - 105.

Panagiotidis, T., Papapanagiotou, G., and Stengos, T. (2022) On the volatility of cryptocurrencies. *Research in International Business and Finance, 62*, 101724.

Patton, A.J. and Sheppard, K. (2015) Good volatility, bad volatility: signed jumps and the persistence of volatility. *Rev. Econ. Statistics 97 (3)*, 683–697

Szado, E. (2018) The Distinctive Characteristics of VIX Futures and Options, Working Paper.

Todorov, V. (2010) Variance risk-premium dynamics: The role of Financial Studies, *The Review of Financial Studies*, 23, 345-383.

BANK OF GREECE WORKING PAPERS

304. Kotidis, A., D. Malliaropulos and E. Papaioannou, "Public and private liquidity during crises times: evidence from emergency liquidity assistance (ELA) to Greek banks", September 2022.

305. Chrysanthakopoulos, C. and A. Tagkalakis, "The effects of fiscal institutions on fiscal adjustments", October 2022.

306. Mavrogiannis, C. and A. Tagkalakis, "The short term effects of structural reforms and institutional improvements in OECD economies", October 2022.

307. Tavlas, S. G., "Milton Friedman and the road to monetarism: a review essay", November 2022.

308. Georgantas, G., Kasselaki, M. and Tagkalakis A., "The short-run effects of fiscal adjustment in OECD countries", November 2022

309. Hall G. S., G. S. Tavlas and Y. Wang, "Drivers and spillover effects of inflation: the United States, the Euro Area, and the United Kingdom", December 2022.

310. Kyrkopoulou, E., A. Louka and K. Fabbe, "Money under the mattress: economic crisis and crime", December 2022.

311. Kyrtsou, C., "Mapping inflation dynamics", January 2023.

312. Dixon, Huw, T. Kosma and P. Petroulas, "Endogenous frequencies and large shocks: price setting in Greece during the crisis", January 2023.

313. Andreou P.C, S. Anyfantaki and A. Atkinson, "Financial literacy for financial resilience: evidence from Cyprus during the pandemic period", February 2023.

314. Hall S. G, G.S. Tavlas and Y. Wang, "Forecasting inflation: the use of dynamic factor analysis and nonlinear combinations", February 2023.

315. Petropoulos A., E. Stavroulakis, P. Lazaris, V. Siakoulis and N. Vlachogiannakis, "Is COVID-19 reflected in AnaCredit dataset? A big data - machine learning approach for analysing behavioural patterns using loan level granular information", March 2023.

316. Kotidis, A. M. MacDonald, D. Malliaropulos, "Guaranteeing trade in a severe crisis: cash collateral over bank guarantees", March 2023.

317. Degiannakis, S. "The D-model for GDP nowcasting", April 2023.

318. Degiannakis, S., G. Filis, G. Siourounis, L. Trapani, "Superkurtosis", April 2023.

319. Dixon, H. T. Kosma, and P. Petroulas, "Explaining the endurance of price level differences in the euro area", May 2023.

320. Kollintzas, T. and V. Vassilatos, "Implications of market and political power interactions for growth and the business cycle II: politico-economic equilibrium", May 2023.

321. Bragoudakis, Z. and I. Krompas "Greek GDP forecasting using Bayesian multivariate models", June 2023.