



University  
of Glasgow

Psaradellis, I. and Sermpinis, G. (2016) Modelling and trading the U.S. implied volatility indices: evidence from the VIX, VXN and VXD indices. *International Journal of Forecasting*, 32(4), pp. 1268-1283.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

<http://eprints.gla.ac.uk/120240/>

Deposited on: 21 June 2016

Enlighten – Research publications by members of the University of Glasgow  
<http://eprints.gla.ac.uk>

## **Modelling and trading the U.S. implied volatility indices. Evidence from the VIX, VXN and VXD indices.**

### **Abstract**

This paper concentrates on modelling and trading three daily market implied volatility indices issued on the Chicago Board Options Exchange (CBOE) with evolving combinations of prominent autoregressive and emerging heuristics models. The motivation is to introduce an algorithm that provides a better approximation of the most popular U.S. volatility indices than the algorithms already presented in the literature and determine whether there is a capability of producing profitable trading strategies. A heterogeneous autoregressive process (HAR) is combined with a genetic algorithm-support vector regression (GASVR) model in two hybrid algorithms. The algorithms' statistical performance is benchmarked against the best forecasters on the VIX, VXN and VXD volatility indices respectively. The trading performance of the forecasts is evaluated through a trading simulation based on VIX and VXN futures contracts as well as on the VXZ exchange traded note based on the S&P 500 VIX mid-term futures index. Our findings indicate that strong nonlinearities exist in all indices examined, while the GASVR algorithm improves the statistical significance of HAR processes. The trading performance of the hybrid models reveals the possibility of economically significant profits.

JEL classification: C22; C45; C53; G12; G17;

Keywords: Implied volatility indices; Heterogeneous autoregression; Heuristics; Volatility Derivatives; Exchange Traded Notes;

## 1. Introduction-Literature

The Chicago Board Options Exchange (CBOE) implied volatility index (VIX), the so-called "investor fear gauge" (Whaley, 2000), has been widely used by academics and practitioners as a key measure of risk because it relies on the market expectations of volatility implied by the supply and demand of the S&P 500 index options. Its popularity as a hedging instrument for investors encouraged the CBOE to calculate several volatility indices measuring those expectations conveyed by option prices traded in other markets as well. Some of them are the Nasdaq-100 volatility index (VXN) and the Dow Jones Industrial Average volatility index (VXD). In particular, the VIX, VXN and VXD are forward-looking indicators that represent expected future market volatility over the next 30 calendar days. All of them are characterized by sharp increases during periods of uncertainty and turmoil in the options market (Whaley, 2009). This specific feature of the volatility indices makes them very popular tools for decision makers and financial analysts because they reveal whether some of the most-liquid markets have reached an extreme level of sentiment. Thus, the evolution of accurately predicting these specific volatility indices is of great importance not only for derivative markets but for the hedge fund industry in general. This paper concentrates on modelling the VIX, VXN and VXD by evolving combinations of prominent autoregressive and emerging heuristic techniques, which are distinguished for their forecasting potential.

Examining the empirical evidence on modelling the term structure of implied volatility, we find a considerably varied literature. Malliaris and Salchenberger (1996) and Gonzales et al. (1997) successfully apply non-parametric techniques to model the Black-Scholes implied volatility of the S&P 100 ATM call options and Ibex35 index options, respectively. They find that Neural Networks (NN's) are able to better express some characteristics of the data than traditional models. Dumas et al. (1998) use a deterministic function (DVF) for capturing the dynamic S&P 500 options' implied volatility, employing as inputs the asset prices,

moneyness ratio and expiration date of options. Their examined model does not show a significant stability across the implied volatility surface compared with a fully stochastic one. Following a similar methodology with Malliaris and Salchenberger (1996), Refenes and Holt (2001) move one step forward by not only applying a Multi-Layer Perceptron Network (MLP) in forecasting the implied volatility of Ibex35 options but also using the Durbin-Watson test on (NN) residuals for misspecification analysis purposes. Gonçalves and Guidolin (2006) express these dynamics by employing a vector autoregression (VAR) technique. They also assess the economic significance of the VAR's forecasts by constructing a variety of trading and hedging strategies. Ahn et al. (2012) follow a different and unique approach by using an artificial NN and a sliding window technique to forecast precisely the directional movements of implied volatility of KOSPI 200 options as a function of Greeks.

Other researchers such as Harvey and Whaley (1992), Fleming (1995) and Blair et al. (2001 a, b) conduct noteworthy research on the predictability of the VXO implied volatility of the S&P 100 index. The first approach demonstrates an economic variables model under the Black-Scholes assumptions. The last three methodologies demonstrate that the movements of the VXO are explained by a first-order autocorrelation model incorporating mean reversion and an ARCH model consolidating leverage effects, index returns and VIX observations. Similarly, Brooks and Oozer (2002) also use a macroeconomic variables model to forecast and trade the implied volatility derived from at-the-money options on Treasury bond futures of LIFFE.

There are numerous papers in the literature investigating the dynamics of the VIX, such as for pricing implied volatility derivatives or for predicting the directional movements of the S&P 500 index (see, e.g., Dotsis et al., 2007). However, only a limited number of studies in the literature address the question of forecasting the dynamics of the implied volatility indices directly. Ahoniemi (2006) applies a hybrid ARIMA-GARCH model for point forecasts of the

VIX index, while Konstantinidi et al. (2008) examine the predictability of a mixture of methodologies, such as an economic variables model, a vector autoregression (VAR) model and an autoregressive fractionally integrated moving average (ARFIMA) model for point and interval forecasts of several U.S. and European implied volatility indices. Both studies indicate that the ARFIMA explains the U.S. volatility indices better, and they apply out-of-sample forecasts for trading purposes. Clements and Fuller (2012) focus their study on the implementation of a long volatility hedge for an equity index, based on semi-parametric forecasts capturing increases on the VIX. Fernandes et al. (2014) apply a heterogeneous autoregressive (HAR) process (Corsi, 2009) for modelling the VIX, while considering numerous macro-finance variables from the U.S. economy. The rationale behind the use of the HAR is the long memory which characterizes the implied and realized volatility of options (Koopman et al., 2005; Bandi and Perron, 2006; Corsi, 2009). They also develop a semi-parametric model of HAR that includes a Neural Network (NN) term to capture any nonlinearities of unknown form that define the index. Their stimulus lies in the fact that some macro-finance variables (e.g., USD index) do not seem statistically significant in affecting the VIX if one controls for nonlinear dependence; conversely, they have a significant effect on the index in a linear structure.

This study employs a heterogeneous autoregressive process (Muller et al., 1997; Corsi, 2009) to predict the VIX, VXN and VXD and combines it with one of the most promising heuristic techniques, a hybrid genetic algorithm-support vector regression (GASVR) model. GASVR is a promising, fully adaptive heuristic algorithm, free from the data snooping effect and parameterization bias, and with only a small number of applications in the field of forecasting (Pai et al., 2006; Yuang, 2012; Dunis et al., 2013, Sermpinis et al. 2014). This is the first application of the GASVR in modelling option volatilities.

Financial series (particularly tradable series such as the ones under study) are vulnerable to behavioural (Froot *et. al.* (1992)) and exogenous factors such as political decisions (Frisman (2001)). These factors are impossible to capture with mathematical models and include noise to time-series estimations. Linear models (similar to the ones that dominate the relevant literature) will only partially capture the relevant underlying trend. They seem unable to help traders to generate profitable series. They have low forecasting power and volatile behaviour through time (LeBaron (2000) and Qi and Wu (2006)). For instance, the HAR process is one of the most dominant approaches to modelling and forecasting the implied volatility in a linear form, based on three past volatility components (daily, weekly and monthly). However, we find that, when considering our propose semiparametric approach, the daily component of the HAR specification is no longer statistically significant. This is a sign that the series under study exhibits a nonlinear character. By combining the best linear performers in forecasting the U.S. volatility indices with one of the most up-to-date and promising non-linear heuristic approaches, this research aims at creating a hybrid superior forecaster that will surpass the statistical and trading performance of the models presented in the relevant literature. More specifically, a HAR process (the most promising linear model according to Fernandes *et al.*, 2014) is developed and combined with a GASVR model in two hybrid models. The forecasting performance of the hybrid models would unveil if they are non-linear elements that HAR is unable to capture and whether the evolutionary concept of GASVR can actually mimic the market dynamics and is capable of producing profitable forecasts. Their performance is benchmarked against a non-linear heuristic hybrid model incorporating a HAR process and a Recurrent Neural Network (RNN), a simple HAR process, an ARFIMA model and one hybrid ARFIMA algorithm. In this study, we do not consider macro-finance variables models because they lack forecasting performance in predicting the VIX, VXN and VXD, compared to stochastic processes (see Konstantinidi *et al.*, 2008 and Fernandes *et al.*,

(2014)<sup>1</sup>. To verify the robustness of our proposed methodology, we examine two out-of-sample datasets. The first one covers the period starting from the Lehmann Brother collapse (mid-September 2008) until the end of 2009, during the period of financial crisis. The second one involves a more recent period, from the start of 2013 until April of 2014. The paper also performs a heuristic analysis based on the residuals obtained from the autoregressive models. The goal is to extract any other unknown form of nonlinearity that is not captured in the residuals of HAR and ARFIMA specifications. The aim of the paper is not to map the series under study; this is impossible with any mathematical model for any financial tradable series. Instead, this study aims to introduce an algorithm that better approximates the examined indices than the ones already presented in the literature.

The forecasting performance of the models under study is examined through three different predictive ability tests: the superior predictive ability (SPA) and model confidence set (MCS) tests of Hansen (2005) and Hansen et al., (2011) respectively and the Giacomini-White test (2006). Finally, we perform an out-of-sample realistic trading simulation by employing VIX and VXN futures contracts acquired from the Chicago Board Options Exchange (CBOE) volatility futures market in order to check for possible abnormal profits. For the VIX index, the trading performance is also examined through exchange traded notes (ETN), the iPath S&P 500 VIX mid-term futures index (VXZ). ETNs linked with volatility indices are highly preferred from investors as a good diversification hedge and they are available with tiny investor fee rates. The results indicate that a HAR and GASVR residual hybrid model is the only algorithm that produces statistically significant trading performance, when taking into account futures contracts. According to the trading performance of the VXZ ETN, where transaction costs are substantially lower, all HAR specifications are capable of producing

---

<sup>1</sup> We indeed find after experiments that using explanatory variables, such as the continuously compounded return on S&P 500 index, the S&P 500 volume change and the continuously compounded return on the one-month crude oil futures contract as inputs, do not improve the performance of our hybrid algorithm.

statistically significant profits. To our knowledge, this is the first time that a HAR process is employed for trading purposes except from modelling.

The remainder of the paper is structured as follows. In Section 2, a detail description of the implied volatility indices, the VIX and VXN futures contracts and the VXZ ETN is provided. Section 3 presents a synopsis of the benchmark models, the semiparametric architectures applied and the implemented combination methods. The statistical forecasting and trading performances are discussed in Sections 4 and 5, respectively. The last section presents the conclusions.

## **2. Implied Volatility Indices and Related Financial Data**

The VIX was introduced on the Chicago Board Options Exchange (CBOE) in 1993, while VXN and VXD were introduced a few years later. All three indices are settled on daily basis. VIX, VXN and VXD represent weighted indices that mixed together different types of stock index options from S&P 500, Nasdaq-100 and DJIA respectively. As already mentioned, the indices portray the expected future market volatility over the next 30 calendar days. Hence, they are forward-looking illustrations of the level of volatility expected by the market in the short term. All indices apply the VIX algorithm (a model-free implied volatility estimator (see, Jiang and Tian, 2005)) to calculate index values (see, CBOE white paper), a. Thus, it does not depend on any particular option pricing structure as Black-Scholes model (Britten-Jones and Neuberger, 2000).

In this paper, we examine two periods covered by the daily closing prices of the VIX, VXN and VXD, from August 2002 to November 2009 and January 2007 to April 2014, for robustness purposes. The datasets were separated into in-sample and out-of-sample subsets (see table 1 below). The out-of-sample subset consists of approximately the last 14 months of our dataset (292 trading days). The dataset was obtained from the CBOE website.

**Table 1.** The VIX, VXN and VXD dataset-Neural Network's and GASVR algorithm training datasets

	Name of period	Trading days	Start date	End date
<b>Dataset 1</b>	Total dataset	1830	5 August 2002	6 November 2009
	Training set	1538	5 August 2002	11 September 2008
	Out-of-sample dataset	292	12 September 2008	6 November 2009
<b>Dataset 2</b>	Total dataset	1830	3 January 2007	9 April 2014
	Training set	1538	3 January 2007	11 February 2013
	Out-of-sample dataset	292	12 February 2013	9 April 2014

The descriptive statistics of the three series are presented in Table 2.

**Table 2.** Descriptive statistics for the levels and logarithm of implied volatility indices.

	VIX	VXN	VXD
<i>Summary statistics in levels</i>			
Mean	20.6278	24.2843	19.0826
Standard deviation	9.64717	10.1325	8.88704
Skewness	2.13756	1.81777	2.08468
Kurtosis	9.26212	6.90292	8.78787
Jarque-Bera (in levels)	0.00000	0.00000	0.00000
<i>Summary statistics in logs</i>			
Mean	2.94426	3.12122	2.86711
Standard deviation	0.38630	0.35324	0.38383
Skewness	0.82264	0.84035	0.87153
Kurtosis	3.43440	3.28008	3.40348
Jarque-Bera (in logs)	0.00000	0.00000	0.00000

The examined period runs from August 5, 2002 to April 4, 2014. We report the sample mean, standard deviation, skewness and kurtosis, as well as the p-values of the Jarque-Bera test for normality.

The three series under study are non-normal (see the Jarque-Bera p-values in levels at the 99% confidence interval) and exhibit high skewness and positive kurtosis. To overcome the issues, the series were transformed into logarithms. The summary statistics of the series (in logs) are presented in Table 2 below.

The time series (in logs) were also tested for the stationarity property, a unit root and long memory through a variety of testing techniques (Table 3). The Augmented Dickey-Fuller

(ADF) and Phillips-Perron (PP) unit root tests were applied. Additionally, the KPSS test statistic for the null hypothesis of stationarity and the long memory rescaled variance test statistic (V/S) (see Giraitis et al., 2003) were employed to verify that long memory models such as ARFIMA and HAR are appropriate for modelling our data. The number of lags for the KPSS test was selected by using the quadratic spectral kernel with bandwidth choice (Andrews, 1991).

**Table 3.** Unit root, stationarity and long memory test for the logarithm of VIX, VXN and VXD.

<i>Tests</i>	VIX	VXN	VXD
ADF	0.000	0.000	0.000
PP	0.000	0.000	0.000
KPSS	0.069	0.054	0.065
V/S	5.157	5.257	5.382

The  $p$ -values of the ADF and PP tests are reported. The table also shows the values of KPSS test statistic for the stationarity property, whose critical values are 0.119, 0.146 and 0.216 at the 10%, 5% and 1% significance, respectively. Finally, the values of V/S test for long memory are reported. The critical values for the V/S test are 1.36 and 1.63 at the 5% and 1% levels, respectively.

Table 3 reports that the null hypothesis of a unit root is rejected at the 99% statistical level for the full sample according to  $p$ -values of ADF and PP tests. Likewise the KPSS test cannot reject the null of stationarity at the 1% significance level for the full sample<sup>2</sup>. Thus, the stationarity property is confirmed. The V/S test null hypothesis for short memory is rejected for both levels of significance. So our sample is characterized by long memory.

In our trading simulation, we use VIX and VXN futures contracts<sup>3</sup> from the Chicago Futures Exchange (CFE), as well as the iPath S&P 500 VIX mid-term futures ETN (VXZ)<sup>4</sup>.

The VIX and the VXN future contracts may trade up to nine near-term serial months and five months on the February quarterly cycle. The final settlement date is the Wednesday that is thirty days prior to the third Friday of the next month, when the standard S&P 500 and

<sup>2</sup> The small sample size of our sample does not allow us to distinguish reliably between long and short memory processes (Lee and Schmidt, 1996).

<sup>3</sup> We don't take into account VXD futures contracts because they have been delisted from CBOE Futures Exchange since 2009.

<sup>4</sup> The future contract and ETN specifications and settlement processes were retrieved from the CBOE and Barclays websites respectively.

Nasdaq-100 Index options expire. The contract multiplier for each future is \$ 1000. In our application, we examine seven different futures contracts traded in the second out-of-sample data set, which expire in 2013 and 2014<sup>5</sup>, respectively. We trade the contracts much closer to their expiration date, when the futures price is almost equal to the spot price, to minimize the basis risk. Finally, we roll from every future contract series to the following one, five days before their maturity, to minimize the effect of noisy data (see Dotsis et al., 2007). Table 4 presents the characteristics of VIX and VXN futures contracts considered.

**Table 4.** Volatility Indices (VIX and VXN) futures contracts

<b>Delivery month of the contract</b>	<b>Available trading days</b>
April 2013	190
June 2013	188
August 2013	186
October 2013	167
December 2013	188
February 2014	185
April 2014	187

VXZ offer to investors<sup>7</sup> a cheap alternative compared to the more expensive in terms of transaction costs, futures. Additionally, trading with ETN does not require a margin account. The VXZ ETN is designed to offer exposure to the S&P 500 VIX mid-term futures index total return. This index provides access to a daily rolling long position in the fourth, fifth, sixth and seventh month VIX futures contracts. The investor fee rates for the VXZ ETN are 0.89% per annum. The VXZ ETN is the second biggest CBOE volatility index ETF in terms of total assets. The first is the iPath S&P 500 short-term VIX futures ETN (VXX), seeking to replicate the daily rolling long position in the immediately first and second month VIX futures contracts. We choose the VXZ ETN as it is subject to less contango effect arising from the volatility forward curve<sup>6</sup> and has lower basis risk compared to the VXX<sup>7</sup>.

<sup>5</sup> VXN futures and the VXZ ETN were incepted later years than first data set period compared to VIX futures. Contracts whose trading volume of settlement prices is less than five are excluded.

<sup>6</sup> Volatility ETFs are subject to contango effect arising from the volatility forward curve, which is upward sloping, because they track VIX futures and not VIX index itself.

### 3. Forecasting Models

#### 3.1. ARFIMA model

An ARFIMA  $(1, d, 1)$  model is employed as a benchmark to capture short and long memory properties of the implied volatility index. ARFIMA  $(1, d, 1)$  performs better in forecasting the U.S. implied volatility indices compared with VAR models and other simple linear models based on economic variables (Konstantinidi et al., (2008)). A hybrid model based on the residuals of ARFIMA  $(1, d, 1)$  regressions and the GASVR algorithm is also explored. The intuition of the hybrid model is that VIX, VNX and VXD most likely follow a nonlinear pattern. The GASVR algorithm attempts to extract these non-linear elements from the residuals and combine them with the ARFIMA forecasts to present a superior forecasting model.

The standard ARFIMA  $(p, d, q)$  process is given by

$$lrv = (1 - L)^{-d} \{\rho(L)\}^{-1} \theta(L) \varepsilon_t \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \quad (1)$$

where  $lrv$  is the logarithm of the volatility index,  $(1 - L)^d$  is the fractional difference operator with  $d$  order of fractional integration required for stationarity, which is expressed in non-integer values,  $\rho(L) = (1 - \rho_1 L - \dots - \rho_p L^p)$  and  $\theta(L) = (1 - \theta_1 L - \dots - \theta_q L^q)$  are the lagged autoregressive and moving average polynomials, respectively, and  $\varepsilon_t$  is the Gaussian error term.

#### 3.2. HAR model

---

<sup>7</sup> We have computed the basis risk for the two ETNs in the in-samples subperiods. VXZ demonstrates less basis risk than the VXX in both subperiods.

Corsi (2009) proposes the heterogeneous autoregressive model for realised volatility. He was inspired by the Heterogeneous Market Hypothesis of Müller et al. (1993), which accepts the presence of heterogeneity across traders. Specifically, he focused on the heterogeneity arising from different time horizons related to the divergent trading frequency of market agents. The notion is that there are three classes of market participants according to their trading frequency. These are classified as short-term agents (e.g., intraday traders-speculators and hedge funds) characterized by higher trading rates, usually daily, medium-term (e.g., commercial banks), which perform a weekly rebalancing of their assets and long-term (e.g., pension funds, insurance companies) defined by lower frequency transactions, usually on a monthly basis. This results in causing three different types of volatility components (daily, weekly, monthly), which create an overall pattern of volatility cascade from low to high frequencies. At each level of the cascade, the underlying volatility component consists not only of its past observation but also of the expectation of longer horizon partial volatilities. The proposed model is defined as an additive linear structure of first-order autoregressive partial volatilities able to capture long-range dependence:

$$lrv_t = \beta_0 + \beta_{(d)} lrv_{t-1}^{(d)} + \beta_{(w)} lrv_{t-1}^{(w)} + \beta_{(m)} lrv_{t-1}^{(m)} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \quad (2)$$

where  $lrv_t^{(h)} = \frac{1}{h} \sum_{j=1}^h lrv_{t-j+1}$  and  $h = (1, 5, 22)'$  is an index vector that depicts the daily, weekly and monthly components of the volatility cascade. We use the HAR specification as a second benchmark for VIX, VXN and VXD modelling because of its excellent forecastability on implied and realised volatilities (see amongst others, McAleer and Medeiros, 2008; Busch et al., 2011). In addition to this use, we employ the HAR structure to construct our semiparametric approaches involving NNs and the GASVR algorithm.

### 3.3 Neural network approach

Our third benchmark model is a semiparametric approach evolving the HAR process and a Recurrent Neural Network (RNN). As discussed above, there are several researchers in the literature who have successfully applied NNs to the task of identifying patterns in implied or realized volatilities. Usually, NN specifications have at least three layers. The first layer is called the input layer (the number of its nodes corresponds to the number of explanatory variables). The last layer is called the output layer (the number of its nodes corresponds to the number of response variables). An intermediary layer of nodes, the hidden layer, separates the input from the output layer. Its number of nodes defines the amount of complexity the model is capable of fitting. In addition, the input and hidden layer contain an extra node called the bias node. This node has a fixed value of one and has the same function as the intercept in traditional regression models. Normally, each node of one layer has connections to all the other nodes of the next layer. The training of the network (which is the adjustment of its weights such that the network maps the input value of the training data to the corresponding output value) starts with randomly chosen weights and proceeds by applying a learning algorithm called backpropagation of errors (Shapiro, 2000). The iteration length is optimized by maximizing a fitness function in the test dataset.

The RNNs have activation feedback that embodies short-term memory. In other words, the RNN architecture can provide more-accurate outputs because the inputs are (potentially) taken from all previous values. Tenti (1996) notes that RNNs need more connections and memory than do standard back-propagation networks. However, RNNs can yield better results in comparison with simple MLPs due to the additional memory inputs. For more information on RNNs, see Sermpinis et al. (2012). A similar hybrid HAR process and a simple NN model (NNHARX) performed equally in forecasting the VIX compared with

different types HAR processes (see Fernandes et al., (2014)). Straightforward modelling of the implied volatility indices with only RNNs<sup>8</sup> seems insufficient.

The hybrid HAR-RNN method is defined as follows:

$$\begin{aligned}
 lriv = & \beta_0 + \beta_{(d)} lriv_{t-1}^{(d)} + \beta_{(w)} lriv_{t-1}^{(w)} + \beta_{(m)} lriv_{t-1}^{(m)} \\
 & + \sum_{m=1}^M \frac{\lambda_m}{1 + e^{-\alpha - \beta_{(m)} lriv_{t-1}^{(d)} - \beta_{(m)} lriv_{t-1}^{(w)} - \beta_{(m)} lriv_{t-1}^{(m)}}} + \varepsilon_t
 \end{aligned} \tag{3}$$

where  $lriv_{t-1}^{(d)}$ ,  $lriv_{t-1}^{(w)}$  and  $lriv_{t-1}^{(m)}$  are the three volatility components of the HAR model,

and  $\sum_{m=1}^M \frac{\lambda_m}{1 + e^{-\alpha - \beta_{(m)} lriv_{t-1}^{(d)} - \beta_{(m)} lriv_{t-1}^{(w)} - \beta_{(m)} lriv_{t-1}^{(m)}}}$  represents the transfer sigmoid function of the

neural network. The neural network architecture is trained through the backpropagation method and the regularization parameter is optimized based on a cross-validation algorithm.

The number of  $M$  hidden units is set through a trial and error procedure in the in-sample dataset, which reveals the optimal results. In this study, the optimal number of hidden units is

3. For our NNs, we apply an objective fitness function that focuses on minimizing the Mean Squared Error (MSE) of the network's outputs. After the networks are optimized, the predictive value of each model is evaluated by applying it to the validation dataset (out-of-sample dataset).

### 3.4. HAR-GASVR framework

#### 3.4.1. The GASVR

Support vector machines (SVMs) are nonlinear algorithms used in supervised learning frameworks to solve classification problems. SVM processes belong to the general category

---

<sup>8</sup> We conduct NN experiments and a sensitivity analysis on a pool of autoregressive terms of VIX, VXN and VXD series. We find that a simple RNN approach performs poorly for both in-sample and out-of-sample datasets. The problem is most likely that a simple NN model cannot efficiently capture the long memory of implied volatilities, although it is very capable in capturing nonlinearities.

of kernel methods (Scholkopf and Smola, 2002). Their development involves sound theory first, then implementation and experiments, in contrast with the development of other heuristics that are purely atheoretic such as NNs. Their main advantage is that they can generate nonlinear decision boundaries through linear classifiers, but have a simple geometric interpretation. Additionally, the solution to an SVM is global and unique; in other words, it does not suffer from multiple local minima such as the solutions of NNs occasionally do. Another advantage is that the practitioner can apply kernel functions to data such that their vector space is not fixed in terms of dimensions. SVMs can be used in regression problems by implementing the  $\varepsilon$ -sensitive loss function by Vapnik (1995). This function established SVRs as a robust technique for constructing data-driven and nonlinear empirical regression models. Recently, SVR and its hybrid applications have become popular for time-series prediction and financial forecasting applications (see amongst others Pai et al., 2006; Yuang, 2012; Dunis et al., 2013, Sermpinis et al. 2014). Finally, they also seem able to cope well with high-dimensional, noisy and complex feature problems (Suykens et al., 2002). A theoretical framework on SVR is provided on Appendix A.

Although SVR has emerged as a highly effective technique for solving nonlinear regression problems, designing such a model can be impeded by the complexity and sensitivity of selecting its parameters. SVR performance depends on all parameters being set optimally. Numerous approaches for this optimization have been presented in the literature, such as setting  $\varepsilon$  as a non-negative constant for convenience (Trafalis and Ince, 2000), using data-driven approaches (Cherkassky and Ma, 2004), the cross-validation technique (Cao et al., 2003; Duan et al., 2003) and controlling  $\varepsilon$  with  $\nu$ -SVR (Scholkopf et al., 1999).

In this study, SVR parametrization is conducted through a Genetic algorithm (GA)<sup>9</sup>. The resultant algorithm (GASVR), genetically searches over a feature space and then provides a single optimized SVR forecast for each series under study. To perform this process, we use a simple GA in which each chromosome comprises feature genes that encode the best feature subsets, and parameter genes that encode the best choice of parameters. An RBF  $\nu$ -SVR Kernel is implemented for our hybrid approach, which in general is specified as

$$K(x_i, x) = \exp(-\gamma \|x_i - x\|^2), \gamma > 0 \quad (4)$$

where  $\gamma$  represents the variance of the kernel function. Consequently, the parameters optimized by the GA are  $C$ ,  $\nu$  and  $\gamma$ . RBF kernels are the most common in similar SVR applications (see, amongst others, Ince and Trafalis, 2006, 2008). This commonality is because they efficiently overcome overfitting and seem to excel in forecasting applications.

#### 3.4.2. The HAR-GASVR

Following the approach described above, this study combines the HAR model with a genetically optimized  $\nu$ -SVR. In this hybrid model, the  $\nu$ -SVR parameters ( $C$ ,  $\nu$  and  $\gamma$ ) are optimized through a genetic algorithm. This HAR-type Genetic Support Vector (HAR-GASVR) model is specified as follows:

$$\begin{aligned} lriv = & \beta_0 + \beta_{(d)} lriv_{t-1}^{(d)} + \beta_{(w)} lriv_{t-1}^{(w)} + \beta_{(m)} lriv_{t-1}^{(m)} \\ & + \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(lriv_i^{(h)}, lriv) + \varepsilon_t \end{aligned} \quad (5)$$

$$\text{where } K(lriv_i^{(h)}, lriv) = \exp\left(-\gamma \left\| lriv_i^{(h)} - lriv \right\|^2\right), \gamma \quad (6)$$

is an RBF Kernel function that uses as inputs the index vectors of the three volatility components.

---

<sup>9</sup> For a description of the GA algorithm see Appendix B.

For the GA optimization, we set the crossover probability to 0.9. This setting enables our model to keep some population for the next generation, hoping to create better new chromosomes from the good parts of the old ones. The mutation probability is set to 0.1 to prevent our algorithm from performing a random search, whereas the roulette wheel selection technique is applied to the selection step of the GA. Similar to NNs, our HAR-GASVR model requires training and test subsets to validate the goodness of fit of each chromosome. The population of chromosomes is initialized in the training sub-period. The optimal selection of chromosomes is achieved when their forecasts minimize the MSE in the test-sub period. Then, the optimized parameters and selected predictors of the best solution are used to train the SVR and produce the final optimized forecast, which is evaluated over the out-of-sample period.

We adjust the GA initial population to 100 chromosomes, and the maximum number of generations is regulated to 200. However, the algorithm may terminate the evolution earlier if the population is deemed converged. The population is deemed converged when the average fitness across the current population is less than 5% away from the best fitness of the current population. More specifically, when the average fitness is less than 5% away, the diversity of the population is very low, and evolving it for more generations is unlikely to produce different and better individuals than the existing ones or those already examined by the algorithm in previous generations.

### *3.5. Modelling the residuals*

Adding to the previous models, we proceed to a residual analysis over the implied volatility indices estimation approaches to express potential asymmetric effects that are unveiled among the residuals. The GASVR regression method is applied to the residuals generated from our two linear benchmarks (ARFIMA and HAR). The notion behind this HAR and

ARFIMA-type genetic support vector regression residual model (ARFIMA-GASVR(res) and HAR-GASVR (res), respectively) is to perform a heuristic analysis in the ARFIMA and HAR residuals and capture the nonlinear elements hidden in their noise. This specification should be able to forecast more accurately VIX, VXN and VXD compared with its linear counterparts.

A two-step approach is followed. The first step is to feed and train the GASVR algorithm with the series of residuals derived from the ARFIMA and HAR estimations, respectively. In the second step, the GASVR forecasted values are added to the ARFIMA and HAR forecasts. Again following the GASVR methodology described above, the main goal is to genetically optimize the  $\nu$ -SVR parameters and to minimize the mean squared error (MSE) between the residuals and residuals emerging from the SVR regression by employing the same fitness function. For this purpose, the optimization problem describing the  $\nu$ -SVR is transformed into

$$f(\varepsilon) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \exp(-\gamma \|\varepsilon_i - \varepsilon\|^2) + b, \quad 0 \leq \alpha_i, \alpha_i^* \leq \frac{c}{n}, \gamma > 0 \quad (7)$$

where  $\varepsilon_i$  are the lagged values of residuals for each benchmark model and  $\varepsilon$  is the actual ones. In the absence of any formal theory behind the selection of the inputs of a GASVR and based on the experiments during the in-sample period, we choose to feed our networks with the first five autoregressive lags of VIX, VXN and VXD estimation residuals, representing a weekly time interval<sup>10</sup>. Additionally, we keep the population and generation levels and the crossover and mutation probabilities the same as in the previous approach.

---

<sup>10</sup> We experimented with different number of lags in the in-sample (orders three to fifteen). In all cases, we obtained the best forecasted performance in the in-sample with the first five autoregressive lags. GA-SVR performance is highly sensitive to the inputs' selection (see amongst others Pai et. al. 2006, Dunis et. al. 2013 and Sermpinis et. al. 2014).

#### 4. Statistical Performance

Tables 5 and 6 present the out-of-sample statistical performance<sup>11</sup> for all the inspected models, respectively, for both periods considered. We report the root mean squared error (RMSE) and the mean absolute error (MAE) criteria to evaluate statistically our one-day-ahead forecasts. For each of those error statistics, a lower output value indicates better forecasting accuracy for each model. Apart from the examined models, we also consider the predictive ability of a random walk without drift<sup>12</sup> and an AR(1) model<sup>13</sup>.

**Table 5.** Out-of-sample performance of model specifications for each one of the implied volatility indices from September 12, 2008 to November 6, 2009.

<i>12/09/2008-06/ 11/2009</i>		<b>VIX</b>	<b>VXN</b>	<b>VXD</b>
RW	<i>MAE</i>	0.1791	0.1707	0.1874
	<i>RMSE</i>	0.2103	0.1974	0.2166
AR(1)	<i>MAE</i>	0.0517	0.0450	0.0524
	<i>RMSE</i>	0.0732	0.0620	0.0733
ARFIMA	<i>MAE</i>	0.0519	0.0456	0.0520
	<i>RMSE</i>	0.0730	0.0622	0.0725
ARFIMA-GASVR (res)	<i>MAE</i>	0.0524	0.0457	0.0518
	<i>RMSE</i>	0.0730	0.0633	0.0731
HAR	<i>MAE</i>	0.0470	0.0419	0.0472
	<i>RMSE</i>	0.0646	0.0557	0.0636
HAR-RNN	<i>MAE</i>	0.0471	0.0418	0.0473
	<i>RMSE</i>	0.0650	0.0565	0.0651
HAR-GASVR	<i>MAE</i>	0.0330	0.0418	0.0421
	<i>RMSE</i>	0.0452	0.0568	0.0579
HAR-GASVR (res)	<i>MAE</i>	0.0300	0.0392	0.0383
	<i>RMSE</i>	0.0430	0.0542	0.0521

Concerning the statistical performance of the global financial crisis out-of-sample period, we observe that the HAR-GASVR(res) displays the best statistical results according to the two

<sup>11</sup> The in-sample statistical performance for both periods considered can be provided upon request.

<sup>12</sup> Note that a random walk model with a drift has also been computed. However, incorporating the drift had a negative impact on the forecasting performance.

<sup>13</sup> In addition to the proposed models, we explored forecast combinations of the best three and the six models under study (ARFIMA, HAR, HAR-RNN, HAR-GASVR, ARFIMA-GASVR(res), HAR-GASVR(res)) with three different approaches. A simple average of the underlying forecasts, a Bayesian averaging method (Buckland et al., 1997) and a weighted average technique (Aiolfi and Timmermann, 2006). In all cases, the forecast combinations performance was inferior to the one obtained by our best model (HAR-GASVR(res)).

measures computed for all the implied volatility indices. In case of VIX index especially, HAR-GASVR(res) and HAR-GASVR models outperform by far their alternatives. For instance, the former clearly outperforms the ARFIMA model in forecasting accuracy and accomplishes even better results than the HAR and HAR-RNN methods, which have very recently been established as the most accurate techniques for forecasting the VIX (Fernandes et al., 2014). The second-best predicting ability is achieved by the HAR-GASVR method, which presents a considerably better performance than the HAR and HAR-RNN. Only in the case of VXN index the HAR-GASVR method performs equally with the above two processes. The results undoubtedly reveal the existence of nonlinearities and asymmetric effects on the implied volatility index, except from long memory and persistence. Strong evidence of this property is the recognition of HAR-GASVR(res) approach as the best forecasting model for the in-sample period. This shows that our proposed specifications have the ability to perform equally well, even in periods of turmoil.

**Table 6.** Out-of-sample performance of model specifications for each one of the implied volatility indices from February 2, 2013 to April 9, 2014.

<i>12/02/2013-09/04/2014</i>		<b>VIX</b>	<b>VXN</b>	<b>VXD</b>
RW	<i>MAE</i>	0.0809	0.0724	0.0732
	<i>RMSE</i>	0.1006	0.0850	0.0916
AR(1)	<i>MAE</i>	0.0490	0.0423	0.0451
	<i>RMSE</i>	0.0720	0.0580	0.0634
ARFIMA	<i>MAE</i>	0.0488	0.0420	0.0442
	<i>RMSE</i>	0.0716	0.0573	0.0623
ARFIMA-GASVR (res)	<i>MAE</i>	0.0480	0.0424	0.0459
	<i>RMSE</i>	0.0681	0.0580	0.0657
HAR	<i>MAE</i>	0.0489	0.0411	0.0425
	<i>RMSE</i>	0.0683	0.0545	0.0575
HAR-RNN	<i>MAE</i>	0.0490	0.0388	0.0395
	<i>RMSE</i>	0.0685	0.0543	0.0532
HAR-GASVR	<i>MAE</i>	0.0470	0.0358	0.0405
	<i>RMSE</i>	0.0610	0.0475	0.0548
HAR-GASVR (res)	<i>MAE</i>	0.0388	0.0317	0.0354
	<i>RMSE</i>	0.0522	0.0435	0.0489

Concerning the second out-of-sample statistical performance our results display almost the same picture as in the first out-of-sample period, with our proposed forecast combinations being more accurate than the ARFIMA, HAR and HAR-RNN approaches. Specifically, the HAR-type approach comprehending the GASVR error term again seems superior for the statistical measures employed in modelling the implied volatility indices. The hybrid HAR-GASVR model follows. HAR and HAR-RNN are next by presenting, almost equally, less precise out-of-sample results. However, HAR-GASVR approach shows equal performance with its previous HAR approaches in the case of VXD index.

The above findings verify the relative success of the HAR method, confirming the findings of Fernandes et al. (2014). Indeed, it is obvious that every HAR type specification outperforms the ARFIMA ones. This advantage might be attributed to the special ability of the HAR method to capture strong persistence in our dataset. A persistent nature really exists in VIX, VXN and VXD, which quantify the market expectations concerning the 22-trading-days ahead risk-neutral volatility. Furthermore we find that strong nonlinearities also exist in the above indices, which makes our hybrid models perform better.

We authenticate the above results by computing the unconditional Giacomini-White (2006) test for out-of-sample predictive ability testing and forecast selection, when the model can be misspecified. The null hypothesis of the test is the equivalence in forecasting accuracy between two forecasting models. The sign of the test statistic specifies the superior model according to its forecasting performance. A positive resolution of the GW test statistic indicates that the second model is more accurate than the first one, which produces larger losses, whereas a negative resolution specifies the opposite. We calculate the test in terms of the mean squared error loss function (MSE) for each forecast for both out-of-sample periods. Tables 7-9 display the  $p$ -values of the statistic under the null hypothesis that the column

model shows equivalent performance compared with each row model, for every index separately.

**Table 7.** Giacomini-White test for the mean squared error for VIX index.

VIX	ARFIMA	ARFIMA-GASVR	HAR	HAR-RNN	HAR-GASVR
<i>12/09/2008-06/ 11/2009</i>					
ARFIMA-GASVR	0.205				
HAR	0.038**	0.048**			
HAR-RNN	0.033**	0.043**	0.225		
HAR-GASVR	0.001***	0.002***	0.000***	0.000***	
HAR-GASVR (res)	0.001***	0.000***	0.000***	0.000***	0.000***
<i>12/02/2013-09/04/2014</i>					
ARFIMA-GASVR	0.151				
HAR	0.186	0.133			
HAR-RNN	0.227	0.141	0.253		
HAR-GASVR	0.026**	0.004***	0.000***	0.001***	
HAR-GASVR (res)	0.000***	0.000***	0.000***	0.000***	0.000***

The out-of-sample periods covered run from September 12, 2008, to November 06, 2009 and from February 12, 2013, to April 9, 2014. The p-values of the GW statistic presented agree with the null hypothesis that the column model shows equivalent performance compared with each row model in terms of mean squared error. One asterisk denotes a rejection of the null hypothesis at the 10% level of significance. Two and three asterisks denote 5% and 1% levels, respectively.

From the Table 7 above, it is obvious that all HAR processes outperform the ARFIMA models, when forecasting the VIX index according to MSE loss function at the 5% and 1% significance levels. The HAR-GASVR (res) approach is superior from all the HAR processes. Similarly, only the HAR-GASVR and the HAR-GASVR(res) specifications produce significantly better forecasts compared with every competing model.

**Table 8.** Giacomini-White test for the mean squared error for VXN index.

VXN	ARFIMA	ARFIMA-GASVR	HAR	HAR-RNN	HAR-GASVR
<i>12/09/2008-06/ 11/2009</i>					
ARFIMA-GASVR	0.151				
HAR	0.030**	0.043**			
HAR-RNN	0.018**	0.031**	0.275		
HAR-GASVR	0.015**	0.028**	0.240	0.504	
HAR-GASVR (res)	0.001***	0.006***	0.024**	0.000***	0.000***
<i>12/02/2013-09/04/2014</i>					
ARFIMA-GASVR	0.137				
HAR	0.000***	0.001***			
HAR-RNN	0.040**	0.048**	0.627		
HAR-GASVR	0.000***	0.000***	0.000***	0.154	
HAR-GASVR (res)	0.000***	0.000***	0.000***	0.066*	0.000***

The out-of-sample periods covered run from September 12, 2008, to November 06, 2009 and from February 12, 2013, to April 9, 2014. The p-values of the GW statistic presented agree with the null hypothesis that the

column model shows equivalent performance compared with each row model in terms of mean squared error. One asterisk denotes a rejection of the null hypothesis at the 10% level of significance. Two and three asterisks denote 5% and 1% levels, respectively.

The picture seems the same when applying the Giacomini-White test to the predictability of VXN index. The results clearly show that the HAR-GASVR(res) model is again the best forecaster. However, our second proposed methodology, the HAR-GASVR model reveals an almost equal performance to HAR-RNN approach for both periods. The rest of the specifications are inferior to the above ones, with HAR methodology being superior to ARFIMA models.

**Table 9.** Giacomini-White test for the mean squared error for VXD index.

VXD	ARFIMA	ARFIMA-GASVR	HAR	HAR-RNN	HAR-GASVR
<b>12/09/2008-06/ 11/2009</b>					
ARFIMA-GASVR	0.255				
HAR	0.042**	0.052*			
HAR-RNN	0.011**	0.020**	0.128		
HAR-GASVR	0.002***	0.013**	0.000***	0.009***	
HAR-GASVR (res)	0.002***	0.003***	0.000***	0.000***	0.000***
<b>12/02/2013-09/04/2014</b>					
ARFIMA-GASVR	0.087*				
HAR	0.000***	0.014**			
HAR-RNN	0.000***	0.000***	0.000***		
HAR-GASVR	0.000***	0.000***	0.000***	0.139	
HAR-GASVR (res)	0.000***	0.000***	0.000***	0.000***	0.000***

The out-of-sample periods covered run from September 12, 2008, to November 06, 2009 and from February 12, 2013, to April 9, 2014. The p-values of the GW statistic presented agree with the null hypothesis that the column model shows equivalent performance compared with each row model in terms of mean squared error. One asterisk denotes a rejection of the null hypothesis at the 10% level of significance. Two and three asterisks denote 5% and 1% levels, respectively.

Taking into consideration the VXD index, Giacomini-White test provides nearly the same information with the VIX and VXN indices, with HAR-GASVR(res) method being the most accurate approach for modelling VXD.

Tables 10 and 11 exhibit some descriptive results according to Hansen's (2005) SPA test and Hansen's et al., (2011) MCS procedure to allow an equal comparison of various methodologies considered under the mean squared error (MSE) and (MAE) criteria. The SPA test focuses on a comparison of the relative forecasting performance between multiple

methodologies in a full set of models. The null hypothesis is that the benchmark forecast is not inferior to the best alternative one. Each model is used as the benchmark each time we apply the SPA test, starting with the random walk. Low p-values indicate that the respective benchmark model is inferior to at least one alternative (reject the null), whereas high p-values specify the opposite.

The MCS procedure deduces the ‘best’ models from a full set of models under specified criteria and at a given level of confidence. Actually, it is a random data-dependent set of best forecasting models because a standard confidence interval covers the population parameter although acknowledging the limitations of the data (Hansen et al. 2011). Hence, more-informative data can lead to only one best model, whilst less-informative data result in an MCS including several models because it is impossible to differentiate among the competing approaches. An equivalence test and an elimination rule are the key features of the MCS procedure. Low p-values indicate that it is unlikely for the model to belong to the set of the ‘best’ models. Therefore, p-values exceeding the usual levels of significance are preferable.

**Table 10.** Test for SPA and MCS for the out-of-sample periods for the VIX and VXN indices.

	VIX				VXN			
	SPA		MSC		SPA		MCS	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
<b><i>12/09/2008-06/ 11/2009</i></b>								
RW	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
AR(1)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ARFIMA	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0008	0.0000
ARFIMA-GASVR	0.0000	0.0000	0.0000	0.0000	0.0002	0.0000	0.0008	0.0000
HAR	0.0000	0.0000	0.0000	0.0000	0.2287	0.0101	0.2923*	0.0573
HAR-RNN	0.0000	0.0000	0.0000	0.0000	0.1542	0.0292	0.2923*	0.0573
HAR-GASVR	0.2036	0.1870	0.4140*	0.0943	0.0981	0.0377	0.2306*	0.0573
HAR-GASVR (res)	0.7964	0.9987	1.0000*	1.0000*	0.9669	0.6016	1.0000*	1.0000*
<b><i>12/02/2013- 09/04/2014</i></b>								
RW	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
AR(1)	0.0000	0.0000	0.0013	0.0002	0.0000	0.0000	0.0000	0.0000
ARFIMA	0.0031	0.0010	0.0016	0.0003	0.0000	0.0000	0.0000	0.0000
ARFIMA-GASVR	0.0002	0.0000	0.0013	0.0003	0.0000	0.0000	0.0000	0.0000
HAR	0.0040	0.0010	0.0013	0.0002	0.0000	0.0000	0.0000	0.0000
HAR-RNN	0.0000	0.0000	0.0013	0.0002	0.0812	0.0035	0.0793	0.0014
HAR-GASVR	0.0090	0.0010	0.0016	0.0002	0.0067	0.0001	0.0275	0.0009
HAR-GASVR (res)	0.7763	0.7670	1.0000*	1.0000*	0.9308	0.6303	1.0000*	1.0000*

The p-values of SPA (Hansen, 2005) and MCS (Hansen et al., 2011) tests in terms of MSE and MAE criteria are reported. Low p-values indicate that the respective benchmark model is inferior to at least one alternative or that it is unlikely for the model to belong to the set of the 'best' models, respectively. One asterisk denotes that the examined model belongs to the set of 'best' models at the 95% confidence level.

**Table 11.** Test for SPA and MCS for the out-of-sample periods for the VXD index.

	VXD			
	SPA		MSC	
	MSE	MAE	MSE	MAE
<b><i>12/09/2008-06/ 11/2009</i></b>				
RW	0.0000	0.0000	0.0000	0.0000
AR(1)	0.0000	0.0000	0.0000	0.0000
ARFIMA	0.0000	0.0000	0.0000	0.0000
ARFIMA-GASVR	0.0000	0.0000	0.0000	0.0000
HAR	0.0000	0.0000	0.0000	0.0000
HAR-RNN	0.0000	0.0000	0.0002	0.0000
HAR-GASVR	0.0069	0.0065	0.0129	0.0107
HAR-GASVR (res)	0.5129	0.5092	1.0000*	1.0000*
<b><i>12/02/2013- 09/04/2014</i></b>				
RW	0.0000	0.0000	0.0000	0.0000
AR(1)	0.0000	0.0000	0.0000	0.0000
ARFIMA	0.0000	0.0000	0.0000	0.0000
ARFIMA-GASVR	0.0005	0.0000	0.0001	0.0000

HAR	0.0000	0.0000	0.0001	0.0000
HAR-RNN	0.0123	0.0068	0.0237	0.0050
HAR-GASVR	0.0020	0.0002	0.0033	0.0008
HAR-GASVR (res)	0.5225	0.5130	1.0000*	1.0000*

The p-values of SPA (Hansen, 2005) and MCS (Hansen, 2011) tests in terms of MSE and MAE criteria are reported. Low p-values indicate that the respective benchmark model is inferior to at least one alternative or that it is unlikely for the model to belong to the set of the ‘best’ models, respectively. One asterisk denotes that the examined model belongs to the set of ‘best’ models at the 95% confidence level.

The results of the SPA test indicate that most of examined models are inferior at least to one of the alternatives in almost all cases. This most likely happens because the HAR-GASVR(res) model achieves the highest forecasting performance<sup>14</sup>. Only in the case of VXN index HAR processes seem to achieve the same performance according to MSE, during the global financial crisis period. In addition to that both HAR-GASVR and HAR-GASVR(res) yield the highest p-values during the same period for VIX and VXN indices, which does not make them inferior to alternatives.

The MCS findings reveal the same picture. HAR-GASVR and the HAR-GASVR(res) are the only specifications belonging to the ‘best’ set for VIX and VXN indices according to the first out-of-sample period, whereas HAR-GASVR(res) is the only superior model for the rest of the cases considered<sup>15</sup>. This allows us to conclude that the examined data are indeed informative.

## 5. Economic Significance (Out-of-Sample Trading Simulation)

In this part, a trading strategy is applied to assess the economic significance of our models by employing the time series of VIX<sup>16</sup> and VXN futures as well as the VXZ ETN<sup>17</sup> for the

<sup>14</sup> Applying the SPA test without considering the HAR-GASVR(res) approach, we find that all models are beaten by the second-best algorithm, the HAR-GASVR approach.

<sup>15</sup> The results remain the same although we set the confidence level in our application to 10%, 5% and 1%, respectively, and the number of replications to 10,000. Only when we exclude the HAR-GASVR(res) model and apply the procedure do we obtain a larger ‘best’ set.

<sup>16</sup> VIX regular calculation uses the mid-point between bid-ask of out-of-the money SPX options. VIX futures settlement price is based on actually traded prices of SPX options. This difference can lead VIX futures settlement price to diverge from the spot VIX especially in some cases where the bid ask /spread in the SPX is very wide. On the other hand, Shu and Zhang (2011) have found that in general spot VIX and VIX futures react to information synchronously.

second out-of-sample period. This is of great importance because statistical accuracy is not always synonymous with trading profitability. The trading strategy is executed separately for each of our forecasting models and involves seven different futures contracts (see Table 4) and one ETN. Transaction costs are estimated at \$ 0.5 per transaction (see CBOE specifications) for future contracts and 0.89% per annum for the VXZ ETN (see Barclay's specifications).

To evaluate the trading efficiency of our forecasts and to compare our results with previous studies, we follow a simple trading rule. The investor goes long (short) in the volatility futures and the ETN in the case in which the forecasted value of the implied volatility index is greater (smaller) than its current value.

The annualized Sharpe ratio ( $SR$ ) and the annual Leland's (1999) alpha ( $A_p$ ) are considered as measures of performance. To calculate the Sharpe ratio and Leland's alpha, the continuously compounded annual US Libor rate is used as a risk-free rate. Moreover, the  $SR$ 's and  $A_p$ 's 95% confidence intervals have been bootstrapped for each forecasting model to assess the statistical significance of the returns. Leland's (1999) alpha is applied to tackle the existence of non-normality in the distribution of the returns found at the end of the trading strategies for each model<sup>18</sup>. It is specified as

$$A_p = E(r_p) - B_p [E(r_{mkt}) - r_f] - r_f, \quad (8)$$

where,  $r_p$  is the return on trading strategy,  $r_f$  is the risk free rate,  $r_{mkt}$  is the return on market portfolio,  $B_p = \frac{cov(r_p, -(1+r_{mkt})^{-\gamma})}{cov(r_{mkt}, -(1+r_{mkt})^{-\gamma})}$  is a measure of risk similar to the CAPM's beta and

---

<sup>17</sup> It is also worth noting that the VIX and VXN futures and the VXZ ETN can be also applied as hedging tools on their respective indices. However, their efficiency is questionable (Psychoyios and Skiadopoulos, 2006, Alexander and Korovilas, 2013, Engle and Figlewski, 2015).

<sup>18</sup> The distributions of the returns of each model are found to be non-normal and far from Gaussian after performing a statistical analysis.

$\gamma = \frac{\ln[E(1+r_{mkt})]-\ln(1+r_f)}{\text{var}[\ln(1+r_{mkt})]}$  is a criterion of risk aversion (see Konstantinidi and Skiadopoulos,

2011). Additionally, the continuously compounded annual return of the S&P 500 and Nasdaq-100 index are used as a proxies for the benchmark market portfolio. The trading strategy presents an expected return over the risk adjusted degree, when  $A_p > 0$ . The trading performance of our models is presented in table 12 while in Appendix C there are the cumulative returns of the best two models over time.

**Table 12.** Trading performance of the VIX, VXN futures and the iPath S&P 500 VIX mid-term futures index ETN from February 12, 2013, to April 9, 2014.

	VIX		VXN
	Futures	ETN (VXZ)	Futures
<i>ARFIMA</i>			
Sharpe ratio	-0.046	-0.069	-0.037
95% CI	(-0.1)-0.01	(-0.12)-0.00	(-0.09)-0.02
Leland's $A_p$	-0.039	-0.016	-0.023
95% CI	(-0.09)-0.01	(-0.02)-0.00	(-0.06)-0.01
<i>ARFIMA-GASVR (res)</i>			
Sharpe ratio	-0.053	-0.017	0.006
95% CI	(-0.11)-0.0	(-0.07)-0.04	(-0.05)-0.12
$A_p$	-0.044	-0.004	0.004
95% CI	(-0.09)-0.00	(-0.01)-0.01	(-0.03)-0.04
<i>HAR</i>			
Sharpe ratio	0.088*	0.478*	0.084*
95% CI	0.02-0.14	0.41-0.54	0.02-0.014
$A_p$	0.088*	0.386*	0.060*
95% CI	0.03-0.14	0.37-0.40	0.02-0.10
<i>HAR-RNN</i>			
Sharpe ratio	-0.027	0.519*	0.087*
95% CI	(-0.08)-0.03	0.45-0.58	0.02-0.14
$A_p$	-0.024	0.398*	0.061*
95% CI	(-0.08)-0.03	0.38-0.41	0.04-0.10
<i>HAR-GASVR</i>			
Sharpe ratio	0.081*	0.096*	0.033
95% CI	0.02-0.14	0.03-0.15	(-0.02)-0.09
$A_p$	0.088*	0.294*	0.026
95% CI	0.03-0.14	0.27-0.30	(-0.02)-0.07
<i>HAR-GASVR (res)</i>			
Sharpe ratio	0.184*	0.721*	0.127*
95% CI	0.10-0.26	0.63-0.80	0.07-0.18
$A_p$	0.168*	0.451*	0.098*
95% CI	0.09-0.24	0.43-0.47	0.03-0.16

One asterisk denotes the rejection of the null hypothesis of a zero return at the 5% level of significance.

It is obvious from Table 12, that the  $SR$  and the  $A_P$  measures of VIX and VXN futures' trading performance are statistically significant for the half of the cases examined, during the most recent out-of-sample period. The rejection of the null hypothesis of a zero value at the 5% significance level is indicated by one asterisk. On the other hand, all HAR specifications are capable of producing significant profits, when taking into account the performance of the VXZ ETN.

In particular, the findings show that the HAR and HAR-GASVR(res) methods can produce significant profits to some limited extent for VIX and VXN future contracts. However, when it comes to the trading simulation of the VXZ ETN, HAR specifications exhibit substantially larger gains. This outcome can be explained from the very small investor fee rates offered from the volatility ETNs, compared to the larger fees and margin requirements of futures contracts as described earlier. ARFIMA and ARFIMA-GASVR (res) models seem to produce losses for all products examined. The ARFIMA trading performance seems to validate the conclusion of Konstantinidi et al. (2008) and Konstantinidi and Skiadopoulos (2011), who trade VIX volatility futures with the same model.

Summarizing, the HAR-GASVR(res) approach is found superior in terms of trading performance. It produces the largest gains for futures contracts and the ETN employed. In other words, there is a noteworthy prospect of bearing economically significant profits in VIX and VXN volatility futures markets, which suggests a promise for the application of nonlinear methods and specifically of the GASVR algorithm even in trading strategies involving future contracts and ETNs.

## **6. Conclusions**

This paper examines the presence of nonlinearities in the evolution of implied volatility. In particular, it provides evidence concerning the daily settlement of three market volatility

indices, the VIX, VXN and VXD. It has been recently shown (Fernandes et. al, 2014), that a HAR process seems very prominent in forecasting the VIX due to its long-range dependence and persistent nature. Two semiparametric methodologies are introduced as a combination of HAR specification and one of the most promising heuristic techniques, a hybrid genetic algorithm-support vector regression (GASVR) model. The first semiparametric approach introduces an extra optimization term in the HAR model. Specifically, the GASVR algorithm is fed with the three volatility components (daily, weekly and monthly) of the HAR specification as inputs. The second specification performs a residual analysis to express potential asymmetric effects that are prevalent among the residuals. A heuristic regression between the residuals of HAR and its lagged values is applied to test for further persistence. The GASVR forecasted residuals are employed to develop the existing model. The performance of the proposed techniques is benchmarked with an ARFIMA model, which predicts well the U.S. implied volatility indices according to the literature (see Konstantinidi et. al, 2008), a semiparametric approach similar to our first, which uses a Recurrent Neural Network (RNN) instead of the GASVR algorithm, and a semiparametric technique focused on the residual analysis of the ARFIMA model.

The HAR-GASVR(res) approach produces undoubtedly the most accurate predictions by a significant margin compared with the other models. The second-best performance is achieved by the HAR-GASVR model. We authenticate the above results by applying the SPA test (Hansen, 2005), MCS procedure (Hansen et. al, 2011) and the Giacomini-White test (2006). However, all of the HAR processes present better predictive ability compared with the benchmark model. This justifies the findings of Fernandes et al. (2014) that this process cannot be beaten in forecasting the VIX because of its persistent feature. The forecasting superiority of hybrid models verifies the notion that the VIX, VXN and VXD indices exhibit a nonlinear nature.

Finally, the economic significance of the forecasts is assessed by implementing trading strategies with VIX and VXN futures contracts, as well as an S&P 500 VIX mid-term futures index ETN. A HAR process has been economically evaluated by using futures and ETNs for the first time. The results indicate that the HAR specifications, particularly the ones optimized with the GASVR algorithm, are to some extent capable of producing statistically significant profits in normal conditions, when trading futures contracts. On the other hand, the ETN trading performance reports that HAR specifications can achieve much higher gains because of their lower investor fee rates.

### **Acknowledgments**

The authors would like to acknowledge the gracious support of this work through the EPSRC and ESRC Centre for Doctoral Training on Quantification and Management of Risk and Uncertainty in Complex Systems and Environment. We would like also to thank Fundboiler for their support.

### **Appendix A. SVR theoretical framework**

Considering the training data  $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ , where  $x_i \in X \subseteq \mathbb{R}$ ,  $y_i \in Y \subseteq \mathbb{R}$ ,  $i=1 \dots n$  and  $n$  is the total number of training samples, then the SVR function can be specified as

$$f(x) = w^T \varphi(x) + b \quad (\text{A.1})$$

where  $w$  and  $b$  are the regression parameter vectors of the function and  $\varphi(x)$  is the nonlinear function that maps the input data vector  $x$  into a feature space in which the training data exhibit linearity.

The  $\varepsilon$ -sensitive loss  $L_\varepsilon$  function finds the predicted points that lie within the tube created by two slack variables,  $\xi_i, \xi_i^*$ :

$$L_\varepsilon(x_i) = \begin{cases} 0 & \text{if } |y_i - f(x_i)| \leq \varepsilon \\ |y_i - f(x_i)| - \varepsilon & \text{if other} \end{cases}, \varepsilon \geq 0 \quad (\text{A.2})$$

However, the lack of information on the noise of the training datasets makes the *a priori*  $\varepsilon$ -margin setting off  $\varepsilon$ -SVR a difficult task. In addition, the parameter  $\varepsilon$  takes non-negative unconstrained values, which makes the optimal setting very challenging; see Sermpinis et al. (2014). An alternative approach, the  $\nu$ -SVR, can decrease the computational task and simplify the parametrization.

The  $\nu$ -SVR approach encompasses the  $\varepsilon$  parameter in the optimization process and controls it with a new parameter  $\nu \in (0, 1)$ . The optimization problem transforms to

$$\text{Minimize } C[\nu\varepsilon + \frac{1}{n} \sum_{i=1}^n (\xi_i + \xi_i^*)] + \frac{1}{2} \|w\|^2 \quad (\text{A.3})$$

$$\text{subject to } \begin{cases} \xi_i \geq 0 \\ \xi_i^* \geq 0 \\ C \geq 0 \end{cases} \text{ and } \begin{cases} y_i - w^T \varphi(x) - b \leq +\varepsilon + \xi_i \\ w^T \varphi(x) + b - y_i \leq +\varepsilon + \xi_i^* \end{cases}$$

The above quadratic optimization problem is transformed into a dual problem, and its solution is based on the introduction of two Lagrange multipliers  $\alpha_i, \alpha_i^*$  and mapping with kernel function  $K(x_i, x)$ :

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x) + b \text{ where } 0 \leq \alpha_i, \alpha_i^* \leq \frac{C}{n} \quad (\text{A.4})$$

The application of the kernel function transforms the original input space into one with more dimensions, in which a linear decision border can be identified. Factor  $b$  is computed following Karush–Kuhn–Tucker conditions. A detailed mathematical explanation of the above solution can be found in Vapnik (1995). Support Vectors (SVs) ( $x_i$  in equation (7))

lie outside the  $\varepsilon$ -tube<sup>19</sup>, whereas non-SVs lie within the  $\varepsilon$ -tube. Increasing  $\varepsilon$  leads to less SV selection, whereas decreasing it results in more ‘flat’ estimates. The term  $[v\varepsilon + \frac{1}{n}\sum_{i=1}^n(\xi_i + \xi_i^*)]$  is the training error, as specified by slack variables. In particular, in the ‘ $\nu$ -trick’, as presented by Scholkopf et al. (1999), increasing  $\varepsilon$  leads to a proportional increase of the first term (training error) in equation (7), whereas its second term decreases proportionally to the fraction of side the  $\varepsilon$ -tube. Hence,  $\nu$  can be considered the upper bound on the fraction of errors. Conversely, decreasing  $\varepsilon$  leads again to a proportional change of the first term, but the change in the second term is also proportional to the fraction of SVs. In other words,  $\varepsilon$  will shrink as long as the fraction of SVs is smaller than  $\nu$ ; therefore,  $\nu$  is also the lower band in the fraction of SVs. For a more detailed mathematical analysis of the above solutions, see Vapnik (1995). The norm term  $\|w\|^2$  characterizes the complexity (flatness) of the model and the term. Consequently, the introduction of parameter  $C$  satisfies the need to trade model complexity for training error and vice versa (Cherkassky and Ma, 2004). In general, both terms cannot be minimal or close to zero at the same time. The SVR algorithm estimates the  $w$  and  $b$  of the linear function of equation 4 with the predefined  $\varepsilon$  and  $C$  for the resulting regression function to achieve good generalization ability. This result should not be too complex and at the same time avoids many training errors. If this balance is achieved, then the SVR offers a solution to the overfitting problem.

## Appendix B. GA theoretical framework

GAs, introduced by Holland (1995), are search algorithms inspired by the principle of natural selection. They are useful and efficient if the search space is large and complicated or there is not any available mathematical analysis of the problem. A population of candidate solutions,

---

<sup>19</sup> An SV is either a boundary vector ( $(\alpha_i - \alpha_i^*) \in [-C/n, C/n]$ ,  $\xi_i = \xi_i^* = 0$ ) or an error vector ( $\alpha_i = \alpha_i^* = \frac{C}{n}$  and  $\xi_i, \xi_i^* > 0$ ).

called chromosomes, is optimized via a number of evolutionary cycles and genetic operations, such as crossovers or mutations<sup>20</sup>. Chromosomes consist of genes, which are the optimizing parameters. At each iteration (generation), a fitness function is used to evaluate each chromosome, measuring the quality of the corresponding solution, and the fittest chromosomes are selected to survive. This evolutionary process is continued until some termination criteria are met. In general, GAs can address large search spaces and do not become trapped in local optimal solutions as do other search algorithms.

The GA uses the *one-point crossover* and the *mutation operators*. The one-point crossover creates two offspring from every two parents. The parents and a crossover point  $c_x$  are selected at random. The two offspring are made by concatenating the genes that precede  $c_x$  in the first parent with those that follow (and include)  $c_x$  in the second parent. The probability of selecting an individual as a parent for the crossover operator is called *crossover probability*. The offspring produced by the crossover operator replace their parents in the population. Conversely, the mutation operator places random values in randomly selected genes with a certain probability named mutation probability. This operator is very important for avoiding local optima and exploring a larger surface of the search space. For the selection step of the GA, *the roulette wheel selection process* is used (Holland, 1995). In roulette wheel selection, chromosomes are selected according to their fitness. The better the chromosomes, the more chances they have to be selected. Usually, elitism is used to raise the evolutionary pressure in better solutions and to accelerate the evolution. Thus, we ensure that the best solution is copied without changes to the new population so that the best solution found can survive at the end of every generation.

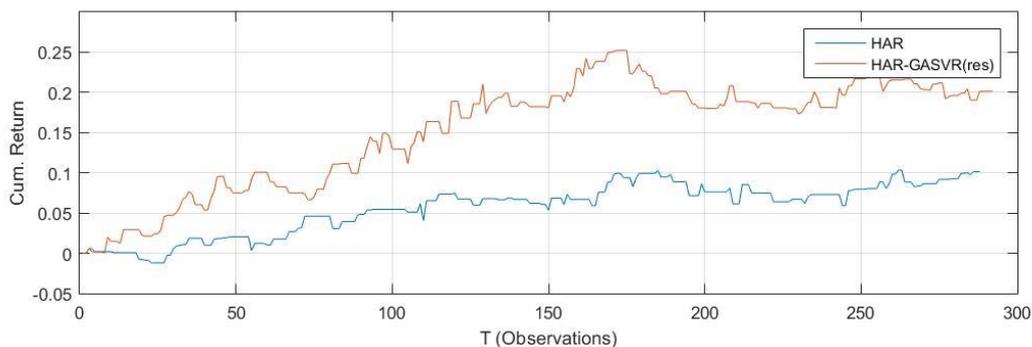
---

<sup>20</sup> The specifications of GA were based on the guidelines of Koza (1992).

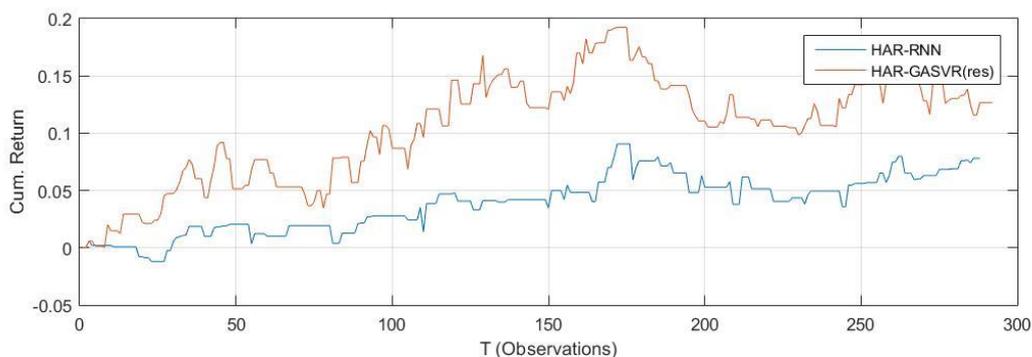
## Appendix C. Cumulative return

In figures C.1, C.2 and C.3 the cumulative returns of the best two models in terms of profitability over time for the VIX futures, VXN futures and VXZ ETN is presented.

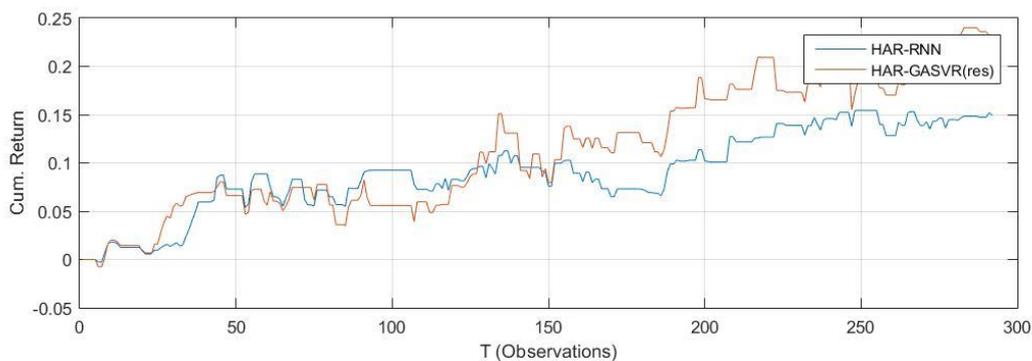
**Figure C.1.** Cumulative return of HAR and HAR-GASVR(res) in the out-of-sample for VIX futures.



**Figure C.2.** Cumulative return of HAR-RNN and HAR-GASVR(res) in the out-of-sample for VXN futures.



**Figure C.3.** Cumulative return of HAR and HAR-GASVR(res) in the out-of-sample for VXZ ETN.



From the figures above, we note that all models strategies present a relatively stable performance in terms of profitability with no large drawdowns.

### References

Ahn, J. J., Kim, D. H., Oh, K. J., & Kim, T. Y., 2012. Applying option Greeks to directional forecasting of implied volatility in the options market: An intelligent approach. *Expert Systems with Applications*, 39(10), 9315–9322.

Ahoniemi, K., 2006. Modeling and forecasting implied volatility: An econometric analysis of the VIX index. *Working paper, Helsinki School of Economics*.

Aiolfi, M., Timmermann, A., 2006. Persistence in forecasting performance and conditional combination strategies. *Journal of Econometrics*, 135 31-35.

Alexander, C., Korovilas, D., 2013. Volatility Exchange-Traded Notes: Curse or Cure? *The Journal of Alternative Investments*, vol. 16, no.2, 52-70.

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Bandi, F.M., Perron, B., 2006. Long memory and the relation between implied and realized volatility. *Journal of Financial Econometrics* 4, 636–670.

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. *Journal of Econometrics*, 105 (1), 5–26.

Blair, B. J., Poon, S.H., & Taylor, S. J., 2001. Modelling S&P 100 volatility: The information content of stock returns. *Journal of Banking & Finance*, 25 (9), 1665–1679.

Britten-Jones, M., Neuberger, A., 2000. Option prices, implied price processes, and stochastic volatility. *Journal of Finance* 55, 839-866.

Brooks, C., Oozeer, M.C., 2002. Modeling the implied volatility of options on long gilt futures. *Journal of Business Finance and Accounting* 29, 111–137.

Buckland, S.T., Burnham, K.P. & Augustin, N.H., 1997. Model selection: an integral part of inference, *Biometrics* 53 (2), 603–618.

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. *Journal of Econometrics*, 160(1), 48–57.

Cao LJ, Chua KS, Guan LK., 2003. C-ascending support vector machines for financial time series forecasting. In 2003 IEEE International Conference on Computational Intelligence for Financial Engineering. IEEE: New York; 317–323.

- Cherkassky V, Ma Y., 2004. Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks* 17:113–126.
- Clements, A.C., Fuller, J., 2012. Forecasting Increases in the VIX: A Time-Varying Long Volatility Hedge for Equities. *NCER Working Paper Series*.
- Corsi, F., 2009. A simple approximate long memory model of realized volatility. *Journal of Financial Econometrics* 7, 174–196.
- Dotsis, G., Psychoyios, D., & Skiadopoulos, G., 2007. An empirical comparison of continuous-time models of implied volatility indices. *Journal of Banking and Finance*, 31,3584–3603.
- Duan K, Keerthi SS, Poo AN., 2003. Evaluation of simple performance measures for tuning SVM hyperparameters. *Neurocomputing* 51:41–59.
- Dumas, B., Fleming, J., Whaley, R.E., 1998. Implied volatility functions: Empirical tests. *Journal of Finance* 53, 2016–2059.
- Dunis, C., Likothanassis, S., Karathanasopoulos, A., Sermpnis, G. & Theofilatos, K., 2013. A hybrid genetic algorithm–support vector machine approach in the task of forecasting and trading. *Journal of Asset Management*.
- Engle, R., Figlewski S., 2015. Modelling the Dynamics of Cprrelations among Implied Volatilities. *Review of Finance*, 19: pp. 991-1018
- Fernandes, M., Medeiros, M. C., & Scharth, M., 2014. Modeling and predicting the CBOE market volatility index. *Journal of Banking & Finance*, 40, 1–10.
- Fleming, J., Ostdiek, B., Whaley, R.E., 1995. Predicting stock market volatility: a new measure. *Journal of Futures Markets* 15, 265–302.
- Frisman, R., 2001. Estimating the Value of Political Connections. *The American Economic Review* 91, 4, 1095-1102.
- Froot, K., Schafferstein, D., Stein, J., 1992 Herd on the street: Informational inefficiencies in a market with short-term speculation, *Journal of Finance*, 47, 4, 1461–1484.
- Giacomini, R., White, H., 2006. Tests of conditional predictive ability. *Econometrica* 74, 1545–1578.
- Giraitis, L., Kokoszka, P., Leipus, R., Teyssière, G., 2003. Rescaled variance and related tests for long memory in volatility and levels. *Journal of Econometrics* 112, 265–294.
- Gonçalves, S., Guidolin, M., 2006. Predictable dynamics in the S&P 500 index options implied volatility surface. *Journal of Business* 79, 1591–1635.
- Gonzalez Miranda, F., & Burgess, N., 1997. Modelling market volatilities: the neural network perspective. *The European Journal of Finance*, 3 (2), 137–157.

- Hansen, P. R., 2005. A Test for Superior Predictive Ability. *Journal of Business & Economic Statistics*, 23(4), 365–380.
- Hansen, P. R., Lunde, A., Nason, J. M., 2011. The Model Confidence Set. *Econometrica*, 79(2), 453–497.
- Harvey, C. R., Whaley, R.E., 1992. Market volatility prediction and the efficiency of the S&P 100 index option market. *Journal of Financial Economics* 31, 43–73.
- Holland J., 1995. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. MIT Press: Cambridge, MA.
- Ince, H., Trafalis, T.B., 2006. Kernel methods for short-term portfolio management. *Expert Systems with Applications* 30, 535–542..
- Ince, H., Trafalis, T.B., 2008. Short term forecasting with support vector machines and application to stock price prediction. *International Journal of General Systems* 37, 677–687.
- Jiang, G., Tian, Y., 2005. Model-free implied volatility and its information content. *Review of Financial Studies* 18, 1305–1342.
- Konstantinidi, E., Skiadopoulos, G., & Tzagkaraki, E., 2008. Can the evolution of implied volatility be forecasted? Evidence from European and US implied volatility indices. *Journal of Banking & Finance*, 32 (11), 2401–2411.
- Konstantinidi, E., Skiadopoulos, G., 2011. Are VIX futures prices predictable? An empirical investigation. *International Journal of Forecasting* 27, 543–560.
- 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. *Journal of Empirical Finance* 12, 445–475.
- Koza, J. R., 1992. *Genetic programming: on the programming of computers by means of natural selection* (Vol. 1). MIT press.
- LeBaron, B. 2000, The stability of moving average technical trading rules on the Dow Jones Index. *Derivatives Use, Trading and Regulation*, 5 (4), 324-338.
- Lee, D., Schmidt, P., 1996. On the power of the KPSS test of stationarity against fractionally-integrated alternatives. *Journal of Econometrics* 73, 285-302.
- Leland, H.E., 1999. Beyond mean-variance: Performance measurement in a nonsymmetrical world. *Financial Analysts Journal* 55, 27–35.
- Malliaris, M., Salchenberger, L., 1996. Using neural networks to forecast the S&P 100 implied volatility. *Neurocomputing*, 10 (2), 183–195.

- McAleer, M., Medeiros, M. C., 2008. A multiple regime smooth transition Heterogeneous Autoregressive model for long memory and asymmetries. *Journal of Econometrics*, 147(1), 104–119.
- Müller, U., Dacorogna, M., Dav, R., Olsen, R., Pictet, O., von Weizsacker, J., 1997. Volatilities of different time resolutions: analysing the dynamics of market components. *Journal of Empirical Finance* 4, 213–239.
- Müller, U., Dacorogna, M., Dav, R., Olsen, R., Pictet, O., Ward, J., 1993. Fractals and intrinsic time: a challenge to econometricians. In: Proceedings of the XXXIX International AEA Conference on Real Time Econometrics.
- Pai, P. F., Lin, C. S., Hong, W. C. & Chen, C. T., 2006. A Hybrid Support Vector Machine Regression for Exchange Rate Prediction. *International Journal of Information and Management Sciences*, 17 (2), 19- 32.
- Psychoyios, D., Skiadopoulos, G., 2006. Volatility Options: Hedging Effectiveness, Pricing and Model Error. *Journal of Futures Markets*, vol. 26, no. 1, 1-31.
- Qi, M., Wu, Y., 2006. Technical trading-rule profitability, data snooping, and reality check: Evidence from the foreign exchange market. *Journal of Money, Credit and Banking*, 38(8), 2135–2158.
- Refenes, A. N., & Holt, W. T., 2001. Forecasting Volatility with Neural Regression: A Contribution to Model Adequacy. *IEEE Transactions on Neural Networks*, 12, no. 4: 850–864.
- Scholkopf B., Bartlett P., Smola A., Williamson R., 1999. Shrinking the tube: a new support vector regression algorithm. In *Advances in Neural Information Processing Systems 11*, Kearns MJ (ed.). MIT Press: Cambridge MA; 330–336.
- Scholkopf B., Smola A., 2002. *Learning with Kernels*. MIT Press: Cambridge, MA.
- Sermpinis G., Laws J., Karathanasopoulos A., Dunis C.L., 2012. Forecasting and trading the EUR/USD exchange rate with gene expression and psi sigma neural networks. *Expert Systems with Applications* 39: 8865–8877.
- Sermpinis, G., Stasinakis, C., Theofilatos, K., Karathanasopoulos, A., 2014. Inflation and Unemployment Forecasting with Genetic Support Vector Regression. *Journal of Forecasting*, 33 (6), 471-487.
- Shapiro A.F., 2000. A Hitchhiker's guide to the techniques of adaptive nonlinear models. *Insurance: Mathematics and Economics* 26:119–132.
- Shu, J., Zhang, J., E., 2011. Causality in the VIX Futures Market. *The Journal of Futures Markets*, vol. 32, no. 1, 24-46.
- Siedlecki W, Sklansky J., 1989. A note on genetic algorithms for large-scale feature selection. *Pattern Recognition Letters* 10:335–347.

Sun Z, Bebis G, Miller R., 2004. Object detection using feature subset selection. *Pattern Recognition* 37: 2165–2176.

Suykens, J.A.K., Brabanter, J.D., Lukas, L., Vandewalle, L., 2002. Weighted least squares support vector machines: robustness and sparse approximation. *Neurocomputing* 48, 85–105.

Tenti P., 1996. Forecasting foreign exchange rates using recurrent neural networks. *Applied Artificial Intelligence* 10: 567–582.

Trafalis TB, Ince H., 2000. Support vector machine for regression and applications to financial forecasting. In Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks, IJCNN 2000, Vol. 1. IEEE Press: New York; 348–353.

Vapnik VN., 1995. The Nature of Statistical Learning Theory. Springer: Berlin.

Whaley, R. E., 2000. The investor fear gauge. *Journal of Portfolio Management* 26, 12-17.

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

Yuang, F.C., 2012. Parameters Optimization Using Genetic Algorithms in Support Vector Regression for Sales Volume Forecasting. *Applied Mathematics*, 3 (1), pp. 1480 - 1486.