International Journal of Forecasting 32 (2016) 1268–1283

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

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

Ioannis Psaradellis [a], Georgios Sermpinis [b,*]

[a] University of Liverpool, UK
[b] University of Glasgow, UK

ARTICLE INFO

Keywords:
Implied volatility indices
Heterogeneous autoregression
Heuristics
Volatility derivatives
Exchange traded notes

ABSTRACT

This paper concentrates on the modelling and trading of three daily market implied volatility indices issued on the Chicago Board Options Exchange (CBOE) using evolving combinations of prominent autoregressive and emerging heuristics models, with the aims of introducing an algorithm that provides a better approximation of the most popular U.S. volatility indices than those that have already been presented in the literature and determining whether there is the ability to produce 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 performances are benchmarked against the best forecasters on the VIX, VXN and VXD volatility indices. The trading performances of the forecasts are 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 the existence of strong nonlinearities in all indices examined, while the GASVR algorithm improves the statistical significance of the HAR processes. The trading performances of the hybrid models reveal the possibility of economically significant profits.

1. Introduction to the literature

The Chicago Board Options Exchange's (CBOE) implied volatility index (VIX), the so-called "investor fear gauge" (Whaley, 2000), has been used widely as a key measure of risk by both academics and practitioners, because it relies on the market expectations of volatility that are 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 other volatility indices as well, measuring the expectations conveyed by option prices traded in other markets; for example, 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 the expected future market volatility over the next 30 calendar days. They are all 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 or not the most liquid markets have reached an extreme level of sentiment. Thus, being able to predict these specific volatility indices accurately 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 using 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 the implied volatility, we find a consider-

* Corresponding author.
E-mail address: georgios.sermpinis@glasgow.ac.uk (G. Sermpinis).

http://dx.doi.org/10.1016/j.ijforecast.2016.05.004
0169-2070/

able degree of variation in the literature. Gonzalez Miranda and Burgess (1997) and Malliaris and Salchenberger (1996) apply non-parametric techniques successfully to the modelling of the Black–Scholes implied volatility of the S&P 100 ATM call options and Ibex35 index options, respectively. They find that neural networks (NNs) are able to express some characteristics of the data better than traditional models. Dumas, Fleming, and Whaley (1998) use a deterministic function (DVF) to capture the dynamic S&P 500 options' implied volatility, employing the asset prices, moneyness ratio and expiration date of options as inputs. The model that they examine does not show a significant stability across the implied volatility surface relative to a fully stochastic one. Following a similar methodology by Malliaris and Salchenberger (1996) and Refenes and Holt (2001) take a further step forward by not only applying a multi-layer perceptron network (MLP) for forecasting the implied volatility of Ibex35 options, but also using the Durbin–Watson test on (NN) residuals for purposes of misspecification analysis. Gonçalves and Guidolin (2006) express these dynamics using 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, Kim, Oh, and Kim (2012) follow a different and unique approach by forecasting the directional movements of the implied volatility of the KOPSI 200 options precisely as a function of Greeks using an artificial NN and a sliding window technique.

Other researchers, such as Blair, Poon, and Taylor (2001a,b), Fleming, Ostdiek, and Whaley (1995) and Harvey and Whaley (1992), have conducted noteworthy research on the predictability of the VXO implied volatility of the S&P 100 index. The first approach is an economic variables model under the Black–Scholes assumptions. The last three methodologies demonstrate that the movements in the VXO are explained by a first-order autocorrelation model that incorporates mean reversion and an ARCH model that consolidates leverage effects, index returns and VIX observations. Similarly, Brooks and Oozeer (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 that have investigated 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, Psychoyios, & Skiadopoulos, 2007). However, only a limited number of studies in the literature have addressed the question of forecasting the dynamics of the implied volatility indices directly. Ahoniemi (2006) uses a hybrid ARIMA-GARCH model to produce point forecasts of the VIX index, while Konstantinidi, Skiadopoulos, and Tzagkaraki (2008) examine the predictive ability of a mixture of methodologies, such as an economic variables model, a vector autoregressive (VAR) model and an autoregressive fractionally integrated moving average (ARFIMA) model for producing point and interval forecasts of several US and European implied volatility indices. Both studies indicate that the ARFIMA explains the US volatility indices better, and apply out-of-sample forecasts for trading purposes. Clements and Fuller's (2012) study focuses on the implementation of a long volatility hedge for an equity index, based on semi-parametric forecasts that capture increases in the VIX. Fernandes, Medeiros, and Scharth (2014) use a heterogeneous autoregressive (HAR) process (Corsi, 2009) to model the VIX, while considering numerous macro-finance variables from the US economy. The rationale behind the use of the HAR is the long memory which characterizes the implied and realized volatility of options (Bandi & Perron, 2006; Corsi, 2009; Koopman, Jungbacker, & Hol, 2005). They also develop a semi-parametric HAR model that includes a neural network (NN) term for capturing any nonlinearities of unknown form that define the index. Their stimulus lies in the fact that some macro-finance variables (e.g., the USD index) do not seem to have statistically significant effects on the VIX if one controls for nonlinear dependence; conversely, their effect on the index is significant in a linear structure.

This study employs a heterogeneous autoregressive process (Corsi, 2009; Müller et al., 1997) for predicting 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 that is free of the data snooping effect and parameterization bias, and has had only a small number of applications in the field of forecasting (Dunis, Likothanassis, Karathanasopoulos, Sermpnis, & Theofilatos, 2013; Pai, Lin, Hong, & Chen, 2006; Sermpinis, Stasinakis, Theofilatos, & Karathanasopoulos, 2014; Yuang, 2012). This is the first application of the GASVR to the modelling of option volatilities.

Financial series (particularly tradable series such as the ones under study) are vulnerable to both behavioural (Froot, Schaferstein, & Stein, 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 (like those that dominate the relevant literature) will be only partially successful at capturing the relevant underlying trend. They seem to be unable to help traders to generate profitable series, and have low forecasting power and a volatile behaviour over time (LeBaron, 2000; Qi & 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, when considering our proposed semiparametric approach, we find that the daily component of the HAR specification is no longer statistically significant. This indicates that the series under study exhibits nonlinear characteristics. By combining the best linear performers for forecasting the US volatility indices with one of the most up-to-date and promising non-linear heuristic approaches, this research aims to create a superior hybrid forecaster that will surpass the statistical and trading performances of the models presented previously 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 performances of the hybrid models indicate whether there are non-linear

elements that HAR is unable to capture and whether the evolutionary concept of GASVR can actually mimic the market dynamics and be capable of producing profitable forecasts. Their performances are benchmarked against a hybrid non-linear heuristic model (incorporating a HAR process and a recurrent neural network (RNN)), a simple HAR process, an ARFIMA model and a hybrid ARFIMA algorithm. In this study, we do not consider macro-finance variables models because their forecasting performances for predicting the VIX, VXN and VXD are poor relative to those of stochastic processes (see Fernandes et al., 2014; and Konstantinidi et al., 2008).[1] We verify the robustness of our proposed methodology by examining two out-of-sample datasets. The first one covers the period between the Lehmann Brothers' collapse (mid-September 2008) and the end of 2009, the period of the financial crisis. The second one is 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, with the aim of extracting any other unknown form of nonlinearity that is not captured by the residuals of the HAR and ARFIMA specifications. The aim of the paper is not to map the series under study, as this would be impossible to achieve for any financial tradable series when using a mathematical model. Instead, this study aims to introduce an algorithm that approximates the examined indices better than those already presented in the literature.

The forecasting performances of the models under study are examined using three different predictive ability tests: the superior predictive ability (SPA) and model confidence set (MCS) tests of Hansen (2005) and Hansen, Lunde, and Nason (2011) respectively, and the Giacomini and White (2006) test. Finally, we perform a realistic out-of-sample 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), for the iPath S&P 500 VIX mid-term futures index (VXZ). ETNs linked with volatility indices are strongly preferred by investors as a good diversification hedge, and are available with tiny investor fee rates. Our results indicate that a HAR and GASVR residual hybrid model is the only algorithm that has a statistically significant trading performance, when taking futures contracts into account. When considering trading performances for the VXZ ETN, where the transaction costs are substantially lower, all HAR specifications are capable of producing statistically significant profits. To the best of our knowledge, this is the first time that a HAR process has been employed for trading purposes other than modelling.

The remainder of the paper is structured as follows. Section 2 provides a detailed description of the implied volatility indices, the VIX and VXN futures contracts and the VXZ ETN. Section 3 presents a synopsis of the benchmark models, the semiparametric architectures applied and the combination methods implemented. The statistical forecasting and trading performances are discussed in Sections 4 and 5, respectively. Finally, the last section presents our 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 mix together different types of stock index options from S&P 500, Nasdaq 100 and DJIA, respectively. As has been 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 that is expected by the market in the short term. All indices apply the VIX algorithm (a model-free implied volatility estimator; see Jiang & Tian, 2005) to the calculation of index values (see, Chicago Board Options Exchange, 2015). Thus, they do not depend on any particular option pricing structure such as the Black–Scholes model (Britten-Jones & Neuberger, 2000).

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

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

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

The time series (in logs) were also tested for stationarity, a unit root and long memory through a variety of testing techniques (Table 3), including the augmented Dickey–Fuller (ADF) and Phillips–Perron (PP) unit root tests. In addition, the KPPS test statistic for the null hypothesis of stationarity and the long memory rescaled variance test statistic (V/S) (see Giraitis, Kokoszka, Leipus, & Teyssière, 2003) were employed to confirm 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 using the quadratic spectral kernel with bandwidth choice (Andrews, 1991).

Table 3 reports that the null hypothesis of a unit root is rejected at the 99% statistical level for the full sample according to the 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,[2] which

[1] Experimentation has shown that the use of explanatory variables, such as the continuously compounded return on the 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, does not improve the performance of our hybrid algorithm.

[2] The small size of our sample does not allow us to distinguish reliably between long and short memory processes (Lee & Schmidt, 1996).

Table 1
The VIX, VXN and VXD datasets: the training datasets of the neural network model and the GASVR algorithm.

	Name of period	Trading days	Start date	End date
Dataset 1	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
Dataset 2	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

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

	VIX	VXN	VXD
Summary statistics in levels			
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
Summary statistics in logs			
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 period examined is 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.

Table 3
Unit root, stationarity and long memory tests for the logarithms of VIX, VXN and VXD.

Tests	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 the KPSS test statistic for the stationarity property, the critical values of which are 0.119, 0.146 and 0.216 at the 10%, 5% and 1% significance levels, respectively. Finally, the values of the V/S test for long memory are reported, with the critical values being 1.36 and 1.63 at the 5% and 1% levels, respectively.

confirms the stationarity property. The V/S test null hypothesis for short memory is rejected for both levels of significance, indicating that our sample is characterized by long memory.

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

The VIX and 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 Nasdaq 100 index options expire. The contract multiplier for each future is $1000. Our application examines seven different futures contracts traded in the second out-of-sample data set, which expire in 2013 and 2014.[5] We trade the contracts much closer to their expiration dates, when the futures price is almost equal to the spot price, in order to minimize the basis risk. Finally, we minimize the effect of noisy data by rolling from every future contract series to the following one, five days before each matures (see Dotsis et al., 2007). Table 4 presents the characteristics of the VIX and VXN futures contracts considered.

Table 4
Volatility indices (VIX and VXN) futures contracts.

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

VXZ offers investors a cheap alternative relative to the more expensive (in terms of transaction costs) futures. In addition, trading with ETN does not require a margin account. The VXZ ETN is designed to provide 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 rate for the VXZ ETN is 0.89% per annum. The VXZ ETN is the second biggest CBOE volatility index ETF in terms of total assets. The biggest is the iPath S&P 500 short-term VIX features ETN (VXX), which seeks to replicate the daily rolling long position in the immediately first and second month VIX futures contracts. We choose the VXZ ETN because it is subject to less contango effect arising from the volatility forward curve[6] and has a lower basis risk than the VXX.[7]

3. Forecasting models

3.1. ARFIMA model

An ARFIMA (1, *d*, 1) model is employed as a benchmark for capturing the short and long memory properties of the

[3] We do not take VXD futures contracts into account because they were delisted from the CBOE Futures Exchange in 2009.

[4] The futures contract and ETN specifications and settlement processes were retrieved from the CBOE and Barclays websites, respectively.

[5] VXN futures and the VXZ ETN were developed after the period of the first dataset, unlike the VIX futures. Contracts with trading volumes of settlement prices that are less than five are excluded.

[6] Volatility ETFs are subject to a contango effect arising from the volatility forward curve, which is upward sloping, because they track VIX futures, not the VIX index itself.

[7] We have computed the basis risks for the two ETNs in the in-sample subperiods. VXZ demonstrates less basis risk than the VXX in both subperiods.

implied volatility index. ARFIMA $(1, d, 1)$ performs better than VAR models and other simple linear models based on economic variables for forecasting the US implied volatility indices (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 are most likely to follow a nonlinear pattern. The GASVR algorithm attempts to extract these non-linear elements from the residuals and to combine them with the ARFIMA forecasts in order to present a superior forecasting model.

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

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

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

3.2. HAR model

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

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

where $lriv^{(h)} = \frac{1}{h}\sum_{j=1}^{h} lriv_{t-j+1}$, 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 forecasting ability on implied and realised volatilities (see Busch, Christensen, & Nielsen, 2011; and McAleer & Medeiros, 2008, amongst others). In addition, we also 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 involving the HAR process and a recurrent neural network (RNN). As was discussed above, several researchers have applied NNs successfully 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, and the number of nodes corresponds to the number of explanatory variables. The last layer is called the output layer, and the number of nodes corresponds to the number of response variables. An intermediate layer of nodes, called the hidden layer, separates the input and output layers. The number of nodes in this layer controls the amount of complexity that the model is capable of fitting. In addition, the input and hidden layers each 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 in a given layer is connected to all of the nodes in the next layer. The training of the network (which involves the adjustment of its weights such that the network maps the input values of the training data to the corresponding output values) starts with randomly chosen weights and proceeds by applying a learning algorithm called the back-propagation 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. While Tenti (1996) notes that RNNs need more connections and memory than standard back-propagation networks, these additional memory inputs allow RNNs to yield better results than simple MLPs. For more information on RNNs, see Sermpinis, Laws, Karathanasopoulos, and Dunis (2012). A similar hybrid HAR process and a simple NN model (NNHARX) performed equally well for forecasting the VIX relative to different types of HAR processes (see Fernandes et al., 2014). Straightforward modelling of the implied volatility indices using only RNNs[8] does not seem sufficient.

The hybrid HAR-RNN method is defined as follows:

$$lriv = \beta_0 + \beta_{(d)} lriv_{t-1}^{(d)} + \beta_{(w)} lriv_{t-1}^{(w)} + \beta_{(m)} lriv_{t-1}^{(m)} + \sum_{m=1}^{M} \frac{\lambda_m}{1+e^{-\alpha-\beta_{(m)} lriv_{t-1}^{(d)}-\beta_{(m)} lriv_{t-1}^{(w)}-\beta_{(m)} lriv_{t-1}^{(m)}}} + \varepsilon_t, \qquad (3)$$

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

[8] 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 the in-sample and out-of-sample datasets. The problem is probably that a simple NN model cannot capture the long memory of implied volatilities efficiently, although it is very capable of capturing nonlinearities.

transfer sigmoid function of the neural network. The neural network architecture is trained through the back-propagation method, and the regularization parameter is optimized based on a cross-validation algorithm. The number of hidden units M is set through a trial-and-error procedure on the in-sample dataset, which reveals the optimal results. In this study, the optimal number of hidden units is three. 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 have been 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 that are used to solve classification problems in supervised learning frameworks. SVM processes belong to the general category of kernel methods (Scholkopf & Smola, 2002). Their development involves, first, sound theory, then implementation and experiments, in contrast to the development of other heuristics that are purely atheoretic, such as NNs. Their main advantage is that, while they can generate nonlinear decision boundaries through linear classifiers, they still have a simple geometric interpretation. In addition, the solution to an SVM is global and unique; in other words, it does not suffer from multiple local minima, 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 ε-sensitive loss function of Vapnik (1995). This function established SVRs as a robust technique for the construction of 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, among others, Dunis et al., 2013; Pai et al., 2006; Sermpinis et al., 2014; and Yuang, 2012). Finally, they also seem able to cope well with high-dimensional, noisy and complex feature problems (Suykens, Brabanter, Lukas, & Vandewalle, 2002). A theoretical framework for SVRs is provided in Appendix A.

Although SVR has emerged as a highly effective technique for solving nonlinear regression problems, the design of such a model can be impeded by the complexity and sensitivity of parameter selection. The performances of SVRs depend on all parameters being set optimally. Numerous different approaches to this optimization have been presented in the literature, such as setting ε to a non-negative constant for the sake of convenience (Trafalis & Ince, 2000), using data-driven approaches (Cherkassky & Ma, 2004), applying cross-validation techniques (Cao, Chua, & Guan, 2003; Duan, Keerthi, & Poo, 2003), and controlling ε with v-SVR (Scholkopf, Bartlett, Smola, & Williamson, 1999).

In this study, SVR parametrization is conducted via a genetic algorithm (GA).[9] The resulting algorithm (GASVR) searches genetically over a feature space and then provides a single optimized SVR forecast for each series under study. We perform this process using 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. For our hybrid approach, we implement an RBF v-SVR kernel, which is specified generally as

$$K(x_i, x) = \exp\left(-\gamma \|x_i - x\|^2\right), \quad \gamma > 0, \tag{4}$$

where γ represents the variance of the kernel function. Thus, the parameters optimized by the GA are C, v and γ. RBF kernels are the most common in similar SVR applications (see Ince & Trafalis, 2006, 2008, amongst others), because they overcome overfitting efficiently 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 v-SVR. In this hybrid model, the v-SVR parameters (C, v and γ) are optimized through a genetic algorithm. This HAR-type genetic support vector (HAR-GASVR) model is specified as follows:

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

where

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

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

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, in the hope of creating better new chromosomes from the good parts of the old ones. The mutation probability is set to 0.1 in order 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 the use of training and test subsets to validate the goodness-of-fit of each chromosome. The population of chromosomes is initialized in the training sub-period, and 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 set to 200. However, the algorithm may terminate the evolution earlier if the population is deemed to have converged. The population is deemed to have 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,

[9] For a description of the GA algorithm, see Appendix B.

the diversity of the population is very low, and more generations of evolution are unlikely to produce different and better individuals than the existing ones or those examined by the algorithm in previous generations.

3.5. Modelling the residuals

Adding to the previous models, we now proceed to a residual analysis of the implied volatility indices estimation approaches in order to express potential asymmetric effects that show up among the residuals. The GASVR regression method is applied to the residuals generated from our two linear benchmarks (ARFIMA and HAR). The idea behind these HAR and ARFIMA-type genetic support vector regression residual models (ARFIMA-GASVR(res) and HAR-GASVR (res), respectively) is to perform a heuristic analysis on the ARFIMA and HAR residuals and capture the nonlinear elements that are hidden in their noise. This specification should be able to forecast VIX, VXN and VXD more accurately than its linear counterparts.

We follow a two-step approach. The first step is to feed and train the GASVR algorithm using 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 goals are to genetically optimize the v-SVR parameters and to minimize the mean squared error (MSE) between the residuals and those that emerge from the SVR regression by employing the same fitness function. For this purpose, the optimization problem describing the v-SVR is transformed into

$$f(\varepsilon) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) \exp\left(-\gamma \left\|\varepsilon_i - \varepsilon\right\|^2\right) + b,$$

$$0 \le \alpha_i,\ \alpha_i^* \le \frac{C}{n},\ \gamma > 0, \tag{7}$$

where ε_i are the lagged values of the residuals for each benchmark model and ε are the actual ones. In the absence of any formal theory for 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 the VIX, VXN and VXD estimation residuals, representing a weekly time interval.[10] In addition, we keep the population and generation levels and the crossover and mutation probabilities the same as in the previous approach.

4. Statistical performance

Tables 5 and 6 present the out-of-sample statistical performances[11] of each of the models inspected, for the two periods considered. We report the root mean squared error (RMSE) and the mean absolute error (MAE) criteria for the statistical evaluation of our one-day-ahead forecasts, where a lower output value for a model indicates that it has a better forecasting accuracy. Apart from the models mentioned above, we also consider the predictive ability of a random walk without drift[12] and an AR(1) model.[13]

Concerning the statistical performance for the global financial crisis out-of-sample period, we observe that the HAR-GASVR(res) displays the best statistical results according to both measures for all of the implied volatility indices. In the case of the VIX index in particular, the HAR-GASVR(res) and HAR-GASVR models outperform their competitors considerably. For instance, the former clearly outperforms the ARFIMA model in forecasting accuracy, and achieves even better results than the HAR and HAR-RNN methods, which were established recently as the most accurate techniques for forecasting the VIX (Fernandes et al., 2014). The second-best predictive ability is achieved by the HAR-GASVR method, which presents a considerably better performance than either the HAR or HAR-RNN. Only in the case of VXN index does the HAR-GAVSR method perform equally well with the above two processes. The results undoubtedly reveal the existence of nonlinearities and asymmetric effects on the implied volatility index, though from long memory and persistence. Strong evidence for this is provided by 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.

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

The above findings confirm the relative success of the HAR method, bearing out the findings of Fernandes et al. (2014). Indeed, it is obvious that every HAR-type specification outperforms the ARFIMA ones. This advantage might be attributable to the special ability of the HAR method to capture strong persistence in our dataset. A

[10] We experimented with various different numbers of lags in the in-sample period (orders three to fifteen). In all cases, we obtained the best forecast performance when using the first five autoregressive lags. The performance of GA-SVR is highly sensitive to the selection of the inputs (see Dunis et al., 2013; Pai et al., 2006; and Sermpinis et al., 2014, among others).

[11] The in-sample statistical performances for both periods considered are available upon request.

[12] Note that we also computed a random walk model with a drift, but found that incorporating the drift had a negative impact on the forecasting performance.

[13] In addition to the proposed models, we also explored forecast combinations of the best three and all 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, Burnham, & Augustin, 1997), and a weighted average technique (Aiolfi & Timmermann, 2006). In all cases, the performances of the forecast combinations were inferior to that of our best model (HAR-GASVR(res)).

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

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

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

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

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 there are also strong nonlinearities in the above indices, which makes our hybrid models perform better.

We authenticate the results above by computing the unconditional Giacomini and White (2006) test for out-of-sample predictive ability testing and forecast selection, when the model may be misspecified. The null hypothesis of the test is the equivalence of forecasting accuracy between two forecasting models. The sign of the test statistic indicates which model has a superior forecasting performance. A positive GW test statistic indicates that the second model is more accurate than the first one, which produces larger losses, whereas a negative statistic 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 performance of the model in the column is equivalent to that of the model in the row, for every index separately.

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

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

Table 7
Giacomini–White test for the mean squared error: the VIX index.

VIX	ARFIMA	ARFIMA-GASVR	HAR	HAR-RNN	HAR-GASVR
12/09/2008–06/ 11/2009					
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***
12/02/2013–09/04/2014					
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 indicate agreement with the null hypothesis that the performance of the model in the column is equivalent to that of the model in the row in terms of mean squared errors.
* Denotes a rejection of the null hypothesis at the 10% level of significance.
** Denotes a rejection of the null hypothesis at the 5% level of significance.
*** Denotes a rejection of the null hypothesis at the 1% level of significance.

Table 8
Giacomini–White test for the mean squared error: the VXN index.

VXN	ARFIMA	ARFIMA-GASVR	HAR	HAR-RNN	HAR-GASVR
12/09/2008–06/ 11/2009					
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***
12/02/2013–09/04/2014					
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 indicate agreement with the null hypothesis that the performance of the model in the column is equivalent to that of the model in the row in terms of mean squared errors.
* Denotes a rejection of the null hypothesis at the 10% level of significance.
** Denotes a rejection of the null hypothesis at the 5% level of significance.
*** Denotes a rejection of the null hypothesis at the 1% level of significance.

Table 9
Giacomini–White test for the mean squared error: the VXD index.

VXD	ARFIMA	ARFIMA-GASVR	HAR	HAR-RNN	HAR-GASVR
12/09/2008–06/ 11/2009					
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***
12/02/2013–09/04/2014					
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 indicate agreement with the null hypothesis that the performance of the model in the column is equivalent to that of the model in the row in terms of mean squared errors.
* Denotes a rejection of the null hypothesis at the 10% level of significance.
** Denotes a rejection of the null hypothesis at the 5% level of significance.
*** Denotes a rejection of the null hypothesis at the 1% level of significance.

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

Tables 10 and 11 exhibit some descriptive results from Hansen's (2005) SPA test and Hansen et al.'s (2011) MCS procedure in order to allow an equal comparison of various methodologies considered under the mean squared error (MSE) and mean absolute error (MAE) criteria. The SPA test focuses on a comparison of the relative forecasting performances of 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 in turn 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 (rejecting 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, while 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 that includes 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 that the model will belong to the set of the 'best' models. Therefore, p-values that exceed the usual levels of significance are preferable.

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

The MCS findings reveal the same picture. HAR-GASVR and HAR-GASVR(res) are the only specifications that belong to the 'best' set for the VIX and VXN indices in the first out-of-sample period, whereas HAR-GASVR(res) is the only superior model for the rest of the cases considered.[15] This allows us to conclude that the data examined are indeed informative.

5. Economic significance (out-of-sample trading simulation)

In this section, we apply a trading strategy to assess the economic significance of our models by employing the time series of VIX[16] and VXN futures and the VXZ ETN[17] for the 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. The transaction costs are estimated at \$0.5 per transaction (see the CBOE specifications) for future contracts and 0.89% per annum for the VXZ ETN (see Barclay's specifications).

We evaluate the trading efficiency of our forecasts and compare our results with those of previous studies by following a simple trading rule. The investor goes long (short) in the volatility futures and the ETN when 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 performance measures. We calculate the Sharpe ratio and Leland's alpha using the continuously compounded annual US Libor rate as the risk-free rate. Moreover, we bootstrap the 95% confidence intervals of the SRs and A_P values for each forecasting model in order 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.[18] It is specified as

$$A_p = E\left(r_P\right) - B_p\left[E\left(r_{mkt}\right) - r_f\right] - r_f, \tag{8}$$

where r_p is the return on the trading strategy, r_f is the risk free rate, r_{mkt} is the return on the market portfolio, $B_p = \frac{cov(r_P, -(1+r_{mkt})^{-\gamma})}{cov(r_{mkt}, -(1+r_{mkt})^{-\gamma})}$ is a measure of risk that is similar to the CAPM's beta, and $\gamma = \frac{\ln[E(1+r_{mkt})] - \ln(1+r_f)}{var[\ln(1+r_{mkt})]}$ is a risk aversion criterion (see Konstantinidi & Skiadopoulos, 2011). In addition, the continuously compounded annual return of the S&P 500 and the 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 performances of our models are presented in Table 12, while Appendix C presents the cumulative returns of the two best models over time. It is

[14] Applying the SPA test without considering the HAR-GASVR(res) approach, we find that all of the models are beaten by the second-best algorithm, the HAR-GASVR approach.

[15] The results remain the same whether we set the confidence level in our application to 10%, 5% or 1%, with the number of replications set to 10,000. Only when we exclude the HAR-GASVR(res) model and apply the procedure do we obtain a larger 'best' set.

[16] The regular VIX calculation uses the mid-point in the bid–ask spread of out-of-the money SPX options. The VIX futures settlement price is based on actually traded prices of SPX options. This difference can lead the 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. However, Shu and Zhang (2011) have found that spot VIX and VIX futures generally react to information synchronously.

[17] 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 (Alexander & Korovilas, 2013; Engle & Figlewski, 2015; Psychoyios & Skiadopoulos, 2006).

[18] A statistical analysis shows that the distributions of the returns of the individual models are non-normal and far from Gaussian.

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

	VIX				VXN			
	SPA		MSC		SPA		MCS	
12/09/2008–06/ 11/2009	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
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*
12/02/2013–09/04/2014								
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 table reports the *p*-values of the SPA (Hansen, 2005) and MCS (Hansen et al., 2011) tests in terms of the MSE and MAE criteria. Low *p*-values indicate either that the respective benchmark model is inferior to at least one alternative (SPA) or that it is unlikely that the model will belong to the set of the 'best' models (MCS).
* Denotes that the model examined belongs to the set of 'best' models at the 95% confidence level.

Table 11
SPA and MCS tests for the out-of-sample periods: the VXD index.

	VXD			
	SPA		MSC	
12/09/2008–06/ 11/2009	MSE	MAE	MSE	MAE
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*
12/02/2013–09/04/2014				
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 table reports the *p*-values of the SPA (Hansen, 2005) and MCS (Hansen et al., 2011) tests in terms of the MSE and MAE criteria. Low *p*-values indicate either that the respective benchmark model is inferior to at least one alternative (SPA) or that it is unlikely for the model to belong to the set of the 'best' models (MCS).
* Denotes that the model examined belongs to the set of 'best' models at the 95% confidence level.

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

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

Table 12
Trading performances of the VIX, the 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
ARFIMA			
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
ARFIMA-GASVR(res)			
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
HAR			
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
HAR-RNN			
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
HAR-GASVR			
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
HAR-GASVR(res)			
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

* Denotes the rejection of the null hypothesis of a zero return at the 5% significance level.

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

6. Conclusions

This paper examines the existence of nonlinearities in the evolution of the implied volatility. In particular, it provides evidence concerning the daily settlement of three market volatility indices, the VIX, VXN and VXD. Fernandes et al. (2014) recently showed that a HAR process seems to be very promising for forecasting the VIX, due to its long-range dependence and persistent nature. Two semiparametric methodologies are introduced as a combination of the HAR specification and one of the most promising heuristic techniques, a hybrid genetic algorithm–support vector regression (GASVR) model. The first semiparametric approach includes an extra optimization term in the HAR model. Specifically, the GASVR algorithm is fed the three volatility components (daily, weekly and monthly) of the HAR specification as inputs. The second specification performs a residual analysis for expressing the potential asymmetric effects that may be 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 (1) an ARFIMA model, which predicts the US implied volatility indices well according to the literature (see Konstantinidi et al., 2008), (2) a semiparametric approach similar to our first, but using a recurrent neural network (RNN) instead of the GASVR algorithm, and (3) a semiparametric technique that is focused on the residual analysis of the ARFIMA model.

The HAR-GASVR(res) approach produces predictions that are more accurate than those of the other models by a significant margin. The second-best performance is achieved by the HAR-GASVR model. We authenticate the above results by applying the SPA test (Hansen, 2005), the MCS procedure (Hansen et al., 2011) and the Giacomini and White (2006) test. However, all of the HAR processes have better predictive abilities than the benchmark model. This justifies Fernandes et al.'s (2014) finding that this process cannot be beaten for forecasting the VIX, because of its persistent feature. The forecasting superiority of hybrid models confirms that the VIX, VXN and VXD indices exhibit nonlinear characteristics.

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 evaluated economically for the first time by using futures and ETNs. The results indicate that the HAR specifications, and particularly those optimized using the GASVR algorithm, are capable of producing statistically significant profits in normal conditions to some extent, 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 gracious support of this work from the EPSRC and ESRC Centre for Doctoral Training on Quantification and Management of Risk and Uncertainty in Complex Systems and Environment (EP/L015927/1). We would like also to thank Fundboiler for their support.

Appendix A. SVR theoretical framework

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

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

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 ε-sensitive loss function L_ε finds that the predicted points lie within the tube created by two slack variables, ξ_i, ξ_i^*:

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

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

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

$$\textit{Minimize} \quad C\left[v\varepsilon + \frac{1}{n}\sum_{i=1}^{n}(\xi_i + \xi_i^*)\right] + \frac{1}{2}\|w\|^2, \tag{A.3}$$

$$\textit{subject to} \quad \left\{\begin{matrix} \xi_i \ge 0 \\ \xi_i^* \ge 0 \\ C \ge 0 \end{matrix}\right\} \text{ and}$$

$$\left\{\begin{matrix} y_i - w^T\varphi(x) - b \le +\varepsilon + \xi_i \\ w^T\varphi(x) + b - y_i \le +\varepsilon + \xi_i^* \end{matrix}\right\}.$$

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

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)K(x_i, x) + b,$$

$$\text{where } 0 \le \alpha_i,\ \alpha_i^* \le \frac{C}{n}. \tag{A.4}$$

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 the Karush–Kuhn–Tucker conditions. A detailed mathematical explanation of the above solution is provided by Vapnik (1995). Support vectors (SVs; x_i in Eq. (7)) lie outside the ε-tube,[19] whereas non-SVs lie within the ε-tube. Increasing ε leads to less SV selection, whereas decreasing it results in 'flatter' estimates. The norm term $\|W\|^2$ characterizes the complexity (flatness) of the model. 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 'v-trick', as presented by Scholkopf et al. (1999), increasing ε leads to a proportional increase in the first term (training error) in Eq. (7), whereas its second term decreases proportionally to the fraction of points outside the ε-tube. Hence, v can be considered the upper bound on the fraction of errors. Conversely, decreasing ε again leads to a proportional change in the first term, but the change in the second term is also proportional to the fraction of SVs. In other words, ε will shrink as long as the fraction of SVs is smaller than v, meaning that v is also the lower bound in the fraction of SVs. Consequently, the introduction of the parameter C satisfies the need to trade model complexity for training error, and vice versa (Cherkassky & Ma, 2004). In general, the two terms cannot both be minimal or close to zero at the same time. The SVR algorithm estimates the w and b of the linear function of Eq. (4) with the predefined ε and C for the resulting regression function so as to achieve a good generalization ability. This result should not be too complex, while at the same time avoiding many training errors. If this balance is achieved, the SVR offers a solution to the overfitting problem.

Appendix B. GA theoretical framework

GAs, introduced by Holland (1995), are search algorithms that are inspired by the principle of natural selection. They are useful and efficient if the search space is large and complicated or if there is no mathematical analysis of the problem available. A population of candidate solutions, called chromosomes, is optimized via a number of evolutionary cycles and genetic operations, such as crossovers or mutations.[20] Chromosomes consist of genes, which are the optimizing parameters. At each iteration (generation),

[19] 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 $\xi_i,\ \xi_i^* > 0$).

[20] The specifications of the GAs were based on the guidelines of Koza (1992).

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 certain termination criteria are met. In general, GAs can address large search spaces and do not become trapped in local optimal solutions like other search algorithms.

The GA uses the *one-point crossover* and the *mutation operator*. The one-point crossover creates two offspring from each pair of 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 the *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, called the mutation probability. This operator is very important for avoiding local optima and ensuring the exploration of 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 of being selected. Usually, elitism is used to raise the evolutionary pressure on better solutions and to accelerate the evolution. Thus, we ensure that the best solution is copied to the new population without changes, so that the best solution found in a generation can survive at the end of that generation.

Appendix C. Cumulative return

Figs. C.1–C.3 present the cumulative returns of the two best models in terms of their profitability over time for the VIX futures, VXN futures and VXZ ETN.

These figures show that all of the model strategies present relatively stable performances in terms of profitability, with no large drawdowns.

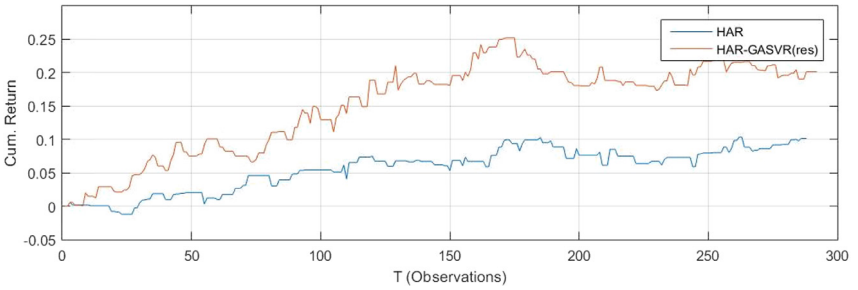

Fig. C.1. Cumulative returns of HAR and HAR-GASVR(res) in the out-of-sample period for VIX futures.

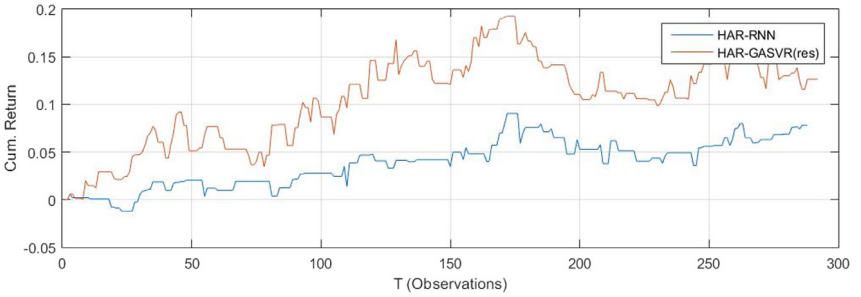

Fig. C.2. Cumulative returns of HAR-RNN and HAR-GASVR(res) in the out-of-sample period for VXN futures.

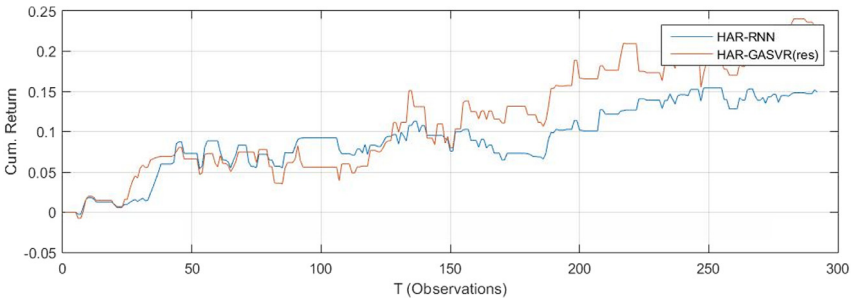

Fig. C.3. Cumulative returns of HAR and HAR-GASVR(res) in the out-of-sample period for VXZ ETN.

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*, *16*(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. (2001a). 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. (2001b). Modelling S&P 100 volatility: the information content of stock returns. *Journal of Banking and 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, L. J., Chua, K. S., & Guan, L. K. (2003). C-ascending support vector machines for financial time series forecasting. In *2003 IEEE international conference on computational intelligence for financial engineering* (pp. 317–323). New York: IEEE.

Cherkassky, V., & Ma, Y. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks*, *17*, 113–126.

(2015). *Chicago Board Options Exchange. The CBOE Volatility Index – VIX*. CBOE White paper, Revised. http://www.cboe.com/micro/vix/vixwhite.pdf.

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, S. S., & Poo, A. N. (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. *The 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*, *14*, 52–71.

Engle, R., & Figlewski, S. (2015). Modelling the dynamics of correlations among implied volatilities. *Review of Finance*, *19*, 991–1018.

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

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

Frisman, R. (2001). Estimating the value of political connections. *The American Economic Review*, *91*(4), 1095–1102.

Froot, K., Schaferstein, D., & Stein, J. (1992). Herd on the street: informational inefficiencies in a market with short-term speculation. *The 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 and 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. *The 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*. Cambridge, MA: MIT Press.

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. (2011). Are VIX futures prices predictable? An empirical investigation. *International Journal of Forecasting*, *27*, 543–560.

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

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*, *26*(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*(4), 850–864.

Scholkopf, B., Bartlett, P., Smola, A., & Williamson, R. (1999). Shrinking the tube: a new support vector regression algorithm. In M. J. Kearns (Ed.), *Advances in neural information processing systems, Vol. 11* (pp. 330–336). Cambridge MA: MIT Press.

Scholkopf, B., & Smola, A. (2002). *Learning with kernels*. Cambridge, MA: MIT Press.

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 & Economics*, *26*, 119–132.
Shu, J., & Zhang, J. E. (2011). Causality in the VIX futures market. *The Journal of Futures Markets*, *32*(1), 24–46.
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. *Applications of Artificial Intelligence*, *10*, 567–582.
Trafalis, T. B., & 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*: *Vol. 1* (pp. 348–353). IJCNN 2000, New York: IEEE Press.
Vapnik, V. N. (1995). *The nature of statistical learning theory*. Berlin: Springer.
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), 1480–1486.