COMBINING NEAREST NEIGHBOR PREDICTIONS AND MODEL-BASED PREDICTIONS OF REALIZED VARIANCE: DOES IT PAY?

Julián Andrada-Félix[†]
University of Las Palmas de Gran Canaria

Fernando Fernández-Rodríguez[††]
University of Las Palmas de Gran Canaria

Ana-Maria Fuertes[§]
Cass Business School
City University London

International Journal of Forecasting, forthcoming

October 10, 2015

[*] This research was funded by the Spanish Ministry of Education (project ECO2010-21318). We thank the editor (Esther Ruiz) and two anonymous referees for their constructive suggestions. We would also like to thank Alena Audzeyeva, Adrian Fernandez-Perez, Jerry Coakley, Jacques Pezier, Kavita Sirichand, and participants at the 2013 Annual Conference on Computational and Financial Econometrics (London), 5th Workshop on Risk Management and Insurance (Gran Canaria), 2013 Annual INFINITI Conference on International Finance (Aix en Provenze) and 2013 International Workshop on Market Microstructure and Nonlinear Dynamics for useful discussions and comments.

[†] Associate Professor in Quantitative Methods in Economics, Universidad de Las Palmas de Gran Canaria, Department. of Quantitative Methods in Economics and Management, Campus de Tafira 35017, Spain; e-mail: julian.andrada@ulpgc.es.

[††] Professor in Quantitative Methods in Economics, Universidad de Las Palmas de Gran Canaria, Department of Quantitative Methods in Economics and Management, Campus de Tafira 35017, Spain; e-mail: fernando.fernandez@ulpgc.es.

[§] Professor in Financial Econometrics, Cass Business School, City University London, 106 Bunhill Row, London ECIY 8TZ, UK; e-mail: a.fuertes@city.ac.uk. *Corresponding author*.

Combining Nearest Neighbor Predictions and Model-Based Predictions of Realized Variance: Does it Pay?

October 10, 2015

Abstract

The increasing availability of intraday financial data has led to improvements in daily volatility forecasting through long-memory models of realized volatility. This paper demonstrates the merit of the non-parametric Nearest Neighbor (NN) approach for S&P 100 realized variance forecasting. A priori the NN approach is appealing because it can reproduce complex dynamic dependencies while largely avoiding misspecification and parameter estimation uncertainty, unlike model-based methods. We evaluate the forecasts through straddle trading profitability metrics and using conventional statistical accuracy criteria. The ranking of individual forecasts confirms that statistical accuracy does not have a one-to-one mapping into profitability. In turbulent markets, the NN forecasts lead to higher risk-adjusted profitability even though the model-based forecasts are statistically superior. In both calm and turbulent market conditions, the directional combination of NN and model-based forecasts is more profitable than any of the individual forecasts.

Keywords: Realized Volatility; Volatility Forecasting; Non-parametric Forecasts; Nearest Neighbor; Long-Memory Models; Forecast Combination; Straddles; Options Trading.

JEL classification: C22; C53; G15.

1. Introduction

A decision that academic researchers and practitioners face when confronted with the task of financial time-series forecasting is whether to use a non-parametric or a model-based method. Non-parametric methods are attractive for variables with complex dynamics that would otherwise require heavily parameterized models. Combination of forecasts from non-parametric methods and time-series models provides a shield against model misspecification and parameter estimation uncertainty, given the distinct way in which both methods exploit the information set (Timmermann, 2006). The goal of this paper is to demonstrate the merit of the non-parametric Nearest Neighbor (NN) method in the novel context of *realized volatility* prediction. The motivation for choosing realized volatility as target variable is threefold.

First, in contrast with asset returns which are often portrayed as martingale difference series, volatility displays persistence and hence, it should be predictable. Second, volatility forecasts are key inputs in financial applications such as derivatives pricing, risk management and portfolio allocation. Third, since the seminal papers by French, Schwert and Stambaugh (1987) and Andersen and Bollerslev (1998), a paradigm shift has occurred in the volatility forecasting literature by which, instead of adopting daily GARCH or stochastic volatility models that treat volatility as latent, many studies construct forecasts from long memory models fitted to daily realized volatilities.[1] Recent studies have suggested extensions of standard long memory models of realized volatility to capture various nonlinearities. Little attention has been paid, however, to non-parametric forecasting methods for realized volatility.

The non-parametric NN method of interest in the paper is a "machine learning" tool that, since its inception several decades ago, has been successfully applied for pattern recognition in engineering and physic but it is far less well known in finance, except for foreign exchange prediction (Fernández-Rodríguez, Sosvilla-Rivero and Andrada-Félix, 1999; Arroyo and Mate, 2009). The merit of NN prediction has not been studied as yet in the context of realized market volatility. Our paper seeks to fill this gap. The most alluring aspect of the NN approach in this

[1] The theory behind realized volatility estimators establishes that by increasing the sampling frequency of asset returns it is possible to obtain arbitrarily precise estimates of daily volatility; essentially, the daily volatility becomes observable *ex post* (Andersen, Bollerslev, Diebold and Labys, 2003; Barndorff-Nielsen and Shephard, 2002). Various studies have shown empirically that volatility forecast improvements can be obtained via long-memory models of realized volatility; see, e.g., Pong, Shackleton, Taylor and Xu (2004); Koopman, Jungbacker and Hol (2005), and Fuertes and Olmo (2012).

context is that it does not require the specification of a functional form to describe the dynamics of realized volatility which is potentially complex. A stylized fact of realized volatility is long range dependence, but it may also exhibit other features such as news asymmetry, regime-switching dependence with the 'switch' dictated by either observable or latent variables, and other nonlinear dependencies induced by market microstructure effects or measurement errors.

In essence, the NN approach identifies the local histories within the available time-series of realized variances that are the most similar to the last observed history, and then pools together the subsequent observations to each history in order to build a prediction. The prediction can thus be interpreted as a 'projection' from sequential local non-parametric regressions. NN forecasts can thus parsimoniously capture complex nonlinear dependences.

This paper contributes to the realized volatility forecasting literature by investigating whether the NN scheme is a good competitor to long-memory models and, relatedly, whether it is worthwhile to combine the forecasts from both approaches. The evaluation tools include economic value measures and conventional statistical criteria. Our choice of economic scenario is a *straddle* trading strategy informed by stock market realized variance predictions. Performance is assessed by conventional risk-adjusted profitability metrics in the context of Markowitz mean-variance analysis (e.g., Sharpe ratios) and assorted non-normality robust metrics. The evaluation accommodates a range of trading scenarios through different noise-to-signal filtering rules, transaction cost levels, forecast horizons, and 'calm' versus 'turmoil' market conditions. As a byproduct, this comprehensive evaluation framework allows us to examine the degree of correspondence between the volatility forecast ranking implied by straddle-trading profitability measures and the ranking arising from purely statistical criteria.

The study is based on five-minute intraday data on the S&P 100 stock index from January 1997 until November 2012.[2] The forecast evaluation is carried out over two distinct holdout (or trading) periods of similar length; the 2003-2007 period that can be described as relatively 'calm' market conditions, and the 2008-2012 period that largely reflects 'turmoil'. The start of the second period is an important landmark of the late 2000s global financial crisis when various events shook the financial markets (*e.g.*, Lehman Brothers bankruptcy).

[2] Although S&P 100 options have been largely displaced by S&P 500 options in the empirical finance literature, in practice the S&P 100 stock index is still used for derivatives, e.g. OEX options.

Among the time-series models previously used for capturing long memory (and nonlinearities) in realized volatility, we focus on the most popular ones. The combination approaches are a Unanimity Rule or directional forecast combination by which a trading signal is triggered only if the direction of model-based forecasts and NN forecasts coincides, Equally-Weighted (EW) forecast combination, and OLS-weighted forecast combination.

The findings suggest that for the purpose of informing straddle trading, NN realized volatility forecasts are more effective than model-based forecasts during the post-2007 period. This confirms that the NN approach offers a 'shield' against heightened misspecification and parameter estimation uncertainty when markets are in turmoil. The comparison of individual and combined forecasts further reveals that the largest risk-adjusted profits over both calm and turbulent markets stem from the combination of NN and model-based predictions.

As a byproduct, the paper documents a mismatch between statistical and economic rankings of realized volatility forecasts. The evidence adds to a literature which contends that statistically accurate predictions of the first or second moment of the distribution of asset returns do not necessarily entail profitability; see, *e.g.* Satchell and Timmermann (1995), Gonçalves and Guidolin (2006), Cenesizoglu and Timmermann (2012), and Bernales and Guidolin (2014). The intuition is that the loss function implicit in a (possibly simple) trading strategy is not necessarily well captured by the Mean Squared Error and other statistical accuracy metrics.

The remaining of the paper is organized as follows. Section 2 briefly reviews the three strands of the literature that motivate our paper. Section 3 describes the dataset. Section 4 and Section 5 present the methodologies for forecast construction and forecast evaluation, respectively. Section 6 discusses the main empirical results. A final section concludes the paper.

2. Background Literature

A burgeoning empirical literature has emerged over the past two decades that derives volatility forecasts from long-memory models of realized volatilities. The motivation is that long-range dependence is a stylized property of realized volatilities whose autocorrelation function decays *hyperbolically* instead of exponentially as it is typical of short-memory processes. Realized volatilities are stationary but persistent. In particular, linear autoregressive Fractionally Integrated Moving Average (ARFIMA) models of realized volatility have become very popular; see *e.g.*, Andersen, Bollerslev, Diebold and Labys (2003), Li (2002), Pong, Shackleton, Taylor and Xu (2004), Koopman, Jungbacker and Hol (2005), Martens and Zein, (2004). Others have opted for the easier-to-handle Heterogenous AutoRegressive, HAR, model of Corsi (2009) which, by mixing volatility components over different horizons, can also capture long memory features; see, *e.g*. Brownlees and Gallo (2010), Andersen, Bollerslev and Huang (2011), and Becker, Clements and Hurn (2011).[3]

By focusing on the long-memory property of realized volatility, the aforementioned studies overlook the presence of nonlinear dynamic dependencies. Seeking to address this shortcoming, a recent strand of the literature has extended the linear long-memory framework. The ARFIMAX specification has been aimed at capturing also *leverage* or the stylized property that negative returns increase future volatility more than similar positive returns (*e.g.*, Oomen, 2004; Giot and Laurent, 2004; Martens, van Dijk and de Pooter, 2009; Fuertes and Olmo, 2012). Moreover, several studies have proposed models to capture *regime-switching* behavior over and above long memory features (Maheu and McCurdy, 2002; McAleer and Medeiros, 2008; Scharth and Medeiros, 2009; Martens, van Dijk and de Pooter, 2009).

In particular, Martens, van Dijk and de Pooter (2009) provide a comprehensive 'horserace' of ARFIMA models that accommodate leverage, level-dependent volatility persistence, structural breaks and day-of-week effects in S&P 500 realized volatility. Their out-of-sample analysis suggests that capturing leverage improves the forecasts but, by contrast, explicitly accounting for level shifts adds very little. Maheu and McCurdy (2002) show that a Markov-switching ARMA model fitted to DM/$ realized variance provides better in-sample fit than the linear ARMA, but the improvement in out-of-sample forecast performance is only marginal. McAleer and Medeiros (2008) argue that a pitfall of extant regime-switching models of realized volatility is that they only accommodate two regimes, and so they propose instead a multiple-regime Smooth Transition HAR model (called HARST). However, their analysis for 16 DJIA stocks does not produce convincing evidence on the superior forecast performance of the HARST model versus the baseline HAR model.

An overall aspect of this literature is that it has failed to produce convincing evidence on the merit of modeling nonlinear dependencies in realized volatility over and above its stylized long memory. It also reflects the challenge that researchers immediately face: the vast, if not

[3] The MIxed DAta Sampling (MIDAS) model of Ghysels, Santa-Clara and Valkanov (2004) and the *p*-spline Multiplicative Error Model (MEM) of Brownlees and Gallo (2010) can also capture long memory.

unlimited, number of candidate model specifications. In fact, many of the nonlinear time-series models put forward in recent decades could inspire nonlinear extensions of long memory models.[4] A practical issue is that, by trying to capture both long memory and nonlinearities, researchers may end up with heavily parameterized (and yet potentially misspecified) models. Hence, it is worthwhile to investigate the merit of combined realized volatility forecasts from parsimonious long-memory models and the non-parametric NN scheme.

Timmermann (2006) provides three reasons for using combined forecasts. They are more immune to misspecification risk than individual forecasts. They average across differences in the way in which individual forecasts are affected by structural breaks or parameter instability. They exploit jointly the information content of each individual forecast. Combined forecasts have been, however, barely used in the context of realized volatility. Patton and Sheppard (2009) show that the EW combination of forecasts from HAR models fitted to 32 distinct realized volatility measures is hard to beat. Liu and Maheu (2009) combine forecasts from (H)AR models fitted to realized variance and realized bipower variation measures.

Our paper speaks to a strand of the realized volatility forecasting literature that investigates the economic merit of the forecasts by considering financial applications; *e.g.*, for risk management (Giot and Laurent, 2004; Clements, Galvão and Kim, 2008; Martens, van Dijk and Pooter, 2009; Brownlees and Gallo, 2010; and Fuertes and Olmo, 2012) or trading purposes (Angelidis and Degiannakis, 2008; Fuertes, Kalotychou and Todorovic, 2015).

3. Data and summary statistics

The sample consists of five-minute quotes from 9:30am to 4:00pm for the S&P 100 stock index over the 16-year period from January 6, 1997 to November 16, 2012; T=3990 days.[5] The target variable is the *realized variance* (RV or realized 'volatility') measured daily as

$$RV_t = \sum_{j=1}^{78} r_{t,j}^2 \qquad (1)$$

[4] Well-known nonlinear time-series models include the bilinear model proposed by Granger and Andersen (1978), the threshold autoregression (Tong, 1978), the state-dependent model (Priestley, 1980), the Markov-switching model (Hamilton, 1989) and the smooth transition autoregression (Teräsvirta, 1994).

[5] The data source is DiskTrading, http://www.is99.com/disktrading. For stock indices, the 5-minute frequency has been shown to provide a good tradeoff between RV estimation accuracy and noise due to market microstructure effects such as asynchronous or infrequent trading, and the bid-ask bounce.

where $r_{t,j} = 100[\ln(p_{t,j}) - \ln(p_{t,j-1})]$ is the *j*th intraday return on day *t*.[6]

The out-of-sample forecasts are constructed sequentially using *rolling* estimation. The window length is T_0=1500 days which enables T_1=2490-h+1 out-of-sample forecasts where h is the horizon in days; *e.g.*, for h=1 the holdout period is December 30, 2002 to November 16, 2012 (T_1=2490 days) which lends itself as an interesting laboratory for forecast evaluation as it can be decomposed into two distinct periods: a 5-year period (from December 30, 2002 to December 31, 2007) when the stock market was relatively calm, and a 5-year period of turmoil (from January 2, 2008 to November 16, 2012) that captures the subprime mortgage crisis, the broader banking crisis, and Greek and Eurozone sovereign debt crises.

Figure 1 shows time-series plots of daily S&P 100 realized variances in levels and logarithms, their histogram (with estimated density) and autocorrelation function for lags up to 100 days. Table 1 reports descriptive statistics for daily realized variances and open-to-close returns. The time-series plots illustrate the distinctive nature of the two holdout periods: low volatility (2003-2007) and high volatility (2008-2010), respectively. The highest volatility levels are observed during October 2008 following a stream of bad news.[7]

[Insert Figure 1 and Table 1 around here]

The histogram and correlogram together, and the descriptive statistics corroborate various stylized facts. The means of daily squared returns (r_t^2) and daily realized volatilities (RV_t) are quite close, as one would expect since both are unbiased estimators of *ex post* volatility. But the standard deviation of RV_t is about half the standard deviation of r_t^2 corroborating that the latter is far noisier. The unconditional distribution of logarithmic RV_t is essentially Gaussian. Daily returns are skewed and fat-tailed but the $\sqrt{RV_t}$ standardization brings the skewness and

[6] We have not employed an overnight-adjusted RV estimator because it would imply assuming that the squared overnight returns are part of the same process that generates the within-day return variation. In adopting a RV measure that ignores the overnight segment, we follow a strand of the literature (*e.g.* Oomen, 2004; Liu and Maheu, 2009; Brownlees and Gallo, 2010; Shephard and Sheppard, 2010). Bivariate models for the overnight variation and open-to-close RV have been proposed by Andersen, Bollerslev and Huang (2011) and Ahoniemi, Fuertes and Olmo (2015) to forecast the variation over the full-calendar day.

[7] On September 7, two large US mortgage lenders, Fannie Mae and Freddie Mac, were nationalized. On September 14, Bank of America acquired Merrill Lynch for $50 billion. On September 15, Lehman Brothers filed for Chapter 11 bankruptcy protection. On September 16, American International Group would fall victim to a liquidity crisis, as its shares lost 95% of their value and the company reported a $13.2 billion loss in just the first half of the year. On September 22, 2008, AIG was removed from the DJIA. After all these events, the Black Week began on October 6, 2008 and lasted five consecutive trading sessions.

kurtosis close to 0 and 3, respectively. Estimates of the fractional integration parameter d are largely downward biased when based on daily squared returns; see also Bollerslev and Wright (2000). Finally, the rejection of the unit root null hypothesis for realized variance in levels or logs and the slow autocorrelation decay (Figure 1) jointly indicate long memory behavior.

4. Forecasting approaches

This section outlines the distinct methods used in our study to obtain daily stock market volatility forecasts. We begin with the time-series models, and then turn to the NN method.

4.1. Long-memory models

We employ the ARFIMAX(p,d,q) model of Granger and Joyeux (1980) expressed as follows

$$\phi(L)(1-L)^d(\ln RV_t - \boldsymbol{\alpha}'\boldsymbol{X}) = \theta(L)\varepsilon_t \tag{2}$$

where ε_t is $NID(0, \sigma_\varepsilon^2)$, d is the fractional integration parameter, $\phi(L) \equiv 1 - \phi_1 L - \cdots - \phi_p L^p$ is an AR lag polynomial of order p, and $\theta(L) \equiv 1 + \theta_1 L + \cdots + \theta_q L^q$ is a MA lag polynomial of order q. This model allows for the inclusion of a $k \times 1$ vector $\boldsymbol{X}$ of exogenous variables. The linear ARFIMA specification, a particular case with $\boldsymbol{\alpha}'\boldsymbol{X}$ replaced by a constant α_0, is a widely-used realized volatility model; see *e.g.*, Andersen, Bollerslev, Diebold and Labys (2003), Pong, Shackleton, Taylor and Xu (2004), and Koopman, Jungbacker and Hol (2005).

Giot and Laurent (2004), Angelidis and Degiannakis (2008), and Martens, van Dijk and de Pooter (2009) among others, use the ARFIMAX model with $\boldsymbol{\alpha}'\boldsymbol{X} = \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 I_{t-1}^- r_{t-1}$ where $I_{t-1}^- = 1$ if $r_{t-1} < 0$ and $I_{t-1}^- = 0$ otherwise. Parameter values $\alpha_2 < 0$ suggest that negative returns have a larger impact on future volatility than positive returns of the same magnitude. We estimate this nonlinear long-memory model by Maximum Likelihood and select the lag orders p and q by the Bayesian Information Criterion (BIC).

Martens, van Dijk and de Pooter (2009) consider a Markov-switching ARFIMAX (hereafter, MSARFIMAX) model of realized volatility which allows for state-dependence in order to characterize sudden changes in financial market conditions. The regime that prevails at time t is dictated by a latent process $s_t=\{H, L\}$ where H denotes the 'crisis' or high volatility state and L the 'normal' or low volatility state; s_t follows a Markov chain of order one characterized by the following matrix of transition probabilities

$$\pi = \begin{pmatrix} \pi_{HH} & \pi_{LH} \\ \pi_{HL} & \pi_{LL} \end{pmatrix} \tag{3}$$

where π_{LH} is the probability that lnRV evolves from state H at t-1 to state L at t, so that $\pi_{HH} + \pi_{LH} = 1$ and $\pi_{LL} + \pi_{HL} = 1$. The MSARFIMAX function can be written as

$$\phi^{s_t}(L)(1-L)^{d^{s_t}}(\ln RV_t - \boldsymbol{\alpha}^{s_t\prime}\boldsymbol{X}) = \theta^{s_t}(L)\varepsilon_t \tag{4}$$

and estimation is by conditional Maximum Likelihood. The specification adopted allows for switching in interceptα_0, fractional integration d, error variance σ_ε^2 and ARMA parameters.

Various studies have shown that high-order AR models, and more parsimonious specifications such as the ARMA(2,1) model or the heterogeneous autoregressive (HAR) model proposed by Corsi (2009) can successfully approximate long memory. In a forecasting exercise for the realized variance of the S&P 500 index at horizons from 1 to 20 days ahead, Martens, van Dijk and de Pooter (2009) show that AR(22) forecasts yield a higher Mincer-Zarnowitz R^2 and a lower Mean Square Error than the ARFIMA(2,d,0). Using data on 16 DJIA stocks, Scharth and Medeiros (2009) highlight that the AR(10) model of realized volatility produces a lower Mean Absolute Error than the ARFIMA(0,d,0) for forecasting horizons of up to 20 days. For various FX rates, Pong, Shackleton, Taylor and Xu (2004) show that the ARMA(2,1) model can produce realized volatility forecasts of similar accuracy as ARFIMA models.

Martens, van Dijk and de Pooter (2009) and Scharth and Medeiros (2009) show that the HAR model of Corsi (2009) produces competitive realized variance forecasts vis-à-vis the conventional ARFIMA model. In our study, we consider the HAR parameterization

$$\ln RV_t = \beta_0 + \beta_d lnRV_{t-1} + \beta_w ln(RV_{t-1})_{t-5} + \beta_m ln(RV_{t-1})_{t-22} + \varepsilon_t \tag{5}$$

where $\ln(RV_{t-1})_{t-k} \equiv k^{-1}[\ln RV_{t-1} + \ldots + \ln RV_{t-k}]$ and $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$. Intuitively, this parsimonious HAR model can be obtained from a high-order AR model by imposing restrictions on the parameters so that the long lag length is implicit.[8]

As naïve benchmarks we consider the historical average, random walk and exponentially weighted moving average (EWMA) of realized variances. We deploy them sequentially over

[8] Throughout the analysis, log forecasts are transformed into level forecasts using the bias-corrected exponential mapping $\widehat{RV}_{t+h|t} = \exp(\widehat{\ln RV}_{t+h|t} + 0.5\sigma_\varepsilon^2)$; see Granger and Newbold (1976).

.

identical rolling windows as the long-memory models, and the smoothing parameter in the EWMA filter is identified by minimizing the Mean Square Error of in-sample predictions.

4.2. Nearest Neighbor (NN) method

The origins of the NN approach can be traced back to physics and engineering where pattern recognition has been one of the main applications. Farmer and Sidorowich (1987) and Yakowitz (1987) are the first to propose this non-parametric approach for time-series forecasting.

The fact that the NN approach is model-free implies that it sidesteps the need to assume a specific functional form to describe the behavior of the target variable; thus, the resulting forecasts are not affected by model misspecification risk. The NN tool is very flexible given its local nature which implies that it can easily adapt to structural breaks and regime-switching behavior. The process under study can be non-stationary. As an adaptive non-parametric method, NN can naturally account for time-variation in the underlying data generating process.

To outline the NN approach, let us state the problem as predicting the daily volatility process, denoted X_t here, which is persistent and may exhibit nonlinear dynamic dependencies of unknown form, $X_t = f_t(X_{t-1}, \dots, X_{t-d}) + \varepsilon_t$, where ε_t is white noise, and $f_t(\cdot)$ is not constrained to belong to the class of continuous functions. Let $\{x_t\}_{t=1}^{T}$ denote the daily time-series (length *T*) of realized variances available to the forecaster. The first stage of the NN approach is to subsample histories of equal length *m* from the time-series of realized variances

$$\mathbf{x}_t^m \equiv \left(x_t, x_{t-1}, \dots, x_{t-(m-1)}\right) \in \mathbb{R}^m, \quad m \leq t \leq T-1 \tag{6}$$

where *m* is the *embedding dimension*. These *m*-dimensional vectors are called *m*-histories and $\mathbb{R}^m$ is the *phase space*. Embedding is a well-known concept in the deterministic chaos literature; *e.g.*, see Hsieh (1991). The proximity of two *m*-histories allows the notion of ‘nearest neighbor’.

The NN approach begins by finding the *k* nearest neighbors defined as the *m*-histories $\mathbf{x}_{t_i}^m$, *i*=1,…,*k*, that represent the first *k* minima of the Euclidean distance function

$$||\mathbf{x}_{t_i}^m - \mathbf{x}_T^m|| \quad t_i = m, m+1, \dots, T-1 \tag{7}$$

where $\mathbf{x}_T^m$ represents the most recent *m*-history available, $\mathbf{x}_T^m \equiv \left(x_T, x_{T-1}, \dots, x_{T-(m-1)}\right)$. Let the scalar x_{t_i+1} denote the subsequent observation to the nearest neighbor $\boldsymbol{x}_{t_i}^m$. The NN prediction $\hat{x}_{T+1}^{NN}$ is defined as the combination of the subsequent observations to each of the *k* nearest neighbors, $x_{t_1+1}, x_{t_2+1}, \dots, x_{t_k+1}$, as follows

$$\hat{x}_{T+1}^{NN} \equiv F(x_{t_1+1}, x_{t_2+1}, \dots, x_{t_k+1}) \tag{8}$$

where the NN function $F(\cdot)$ denotes a measure of central tendency. In our application below, we employ the *median* so our NN predictor is then more precisely defined as

$$\hat{x}_{T+1}^{NN} \equiv med\left(x_{t_1+1}, x_{t_2+1}, \dots, x_{t_k+1}\right) \tag{9}$$

Other possible choices for $F(\cdot)$ are the *mean* which can be unweighted or weighted by a kernel function, and the *local linear autoregressive* function which defines $\hat{x}_{T+1}^{NN} \equiv \hat{a}_0 x_T + \hat{a}_1 x_{T-1} + \dots + \hat{a}_{m-1} x_{T-(m-1)} + \hat{a}_m$ where $(\hat{a}_0, \hat{a}_1, .., \hat{a}_{m-1})$ are the OLS parameters that minimize $\sum_{i=1}^{k}(x_{t_i+1} - a_0\, x_{t_i} - a_1 x_{t_i-1} - \dots - a_{m-1} x_{t_i-(m-1)} - a_m)^2$. The median-based approach is more robust to outliers than the mean-based approach, and less prone to ill-conditioning than the local OLS approach which can be contaminated by estimation noise due to collinearity (Casdagli, 1989). Figure 2 provides intuition on the *k*-NN forecasting approach.

[Insert Figure 2 around here]

The training vectors are disjoint sets in this illustrative graph, for the sake of clarity, but they can overlap in practice; *m* and *k* are, respectively, the training vector dimension and number of training vectors or smoothing parameter. For each horizon, we follow the classical identification approach for *m* and *k* that minimizes the Mean Squared Error of in-sample predictions (Härdle, 1990; Green and Silverman, 1994; Hastie, Tibshirani and Friedman, 2001).

5. Forecast evaluation

We begin by discussing the economic framework used to evaluate the relative merit of the competing out-of-sample forecasts. The statistical framework is presented in Section 5.2.

5.1. Economic criteria

The economic decision-making scenario for the volatility forecast evaluation is a *straddle* trading strategy on the S&P 100 stock index. Following Engle, Hong and Kane (1990), Xekalaki and Degiannakis (2005), and Angelidis and Degiannakis (2008), the analysis is based on theoretical option prices on the S&P 100 stock index derived from the Black-Scholes model.

More specifically, plugging the annualized VXO into the Black-Scholes equation – alongside the closing price of the S&P 100 index, annualized risk-free interest rate, and time-to-expiry in years – we derive daily at-the-money (ATM) put and call vanilla option prices.[9]

A *long* (*short*) straddle involves buying (selling) a call and put with the same strike price and expiration date. A straddle is a *delta neutral* option trading strategy – its symmetrical V-shape payoff profile implies identical profits associated with equal-size upward and downward changes in the stock price. Accordingly, the *straddle* holder can capitalize on accurate stock volatility predictions. Our hypothetical trader buys (sells) a straddle when she anticipates an increase (decrease) in the future stock market volatility using realized variance forecasts.

The S&P 100 index realized variance predictions are transformed into trading signals as follows. If the *h*-day-ahead realized variance prediction made at the close of day *t* is higher (lower) than the market realized variance on that day, a straddle is bought (sold). For concreteness, if $\widehat{RV}_{t+h|t} - RV_t > 0$, a long position is taken on the straddle at the open of day *t*+1; the position is held for *h* days until day *t*+*h*+1 open when the position is closed by selling the straddle.[10] If $\widehat{RV}_{t+h|t} - RV_t < 0$, a short position is taken on the straddle at the open of day *t*+1, and held for *h* days until the open of day *t*+*h*+1 when the position is closed by buying the straddle. The parameter *h* (forecast horizon) thus plays the role of *holding period*. A filter rule is adopted to discard the noisiest trading signals. A long position in straddles is taken on day

---

[9] Daily observations on the annualized implied volatility measure known as VXO and the annualized 3-month US T-Bill rate (the most liquid short term instrument emitted by the American Treasury) are obtained from *Datastream*. The VIX introduced in 1993 by the CBOE tracked the implied volatility of ATM 1-month synthetic options on the S&P 100; it was computed by averaging volatilities from puts and calls from the closest to ATM strikes in the nearest and next to nearest month. In 2003, the CBOE revamped the VIX as an average of out-of-money option price volatility across all available strikes on the S&P 500 index and introduced the new ticker VXO as an estimate of the 1-month ATM implied volatility on the S&P 100 index. The fact that VXO represents the implied volatility of ATM options on the S&P 100 index, permits us to derive the corresponding theoretical ATM call and put prices via the Black-Scholes model.

[10] In our trading simulation, ATM option contracts with maturity *M*=22 days are issued at the open of each day. Suppose that the realized volatility prediction made at the close of day *t* triggers a trading signal; the market price of the straddle traded, according to this signal, at the open of day *t*+1 is $P_{t+1} + C_{t+1}$ where the put and call option prices are derived, using the Black-Scholes formulae, as functions of the option's strike price, current stock price, stock price volatility per annum, annualized risk-free interest rate and option's time to expiry in years, i.e. $C_{t+1} \equiv f_{BS}^{call}\left(K, SP100_{t+1}, VXO_{t+1}, r_{f,t+1}, M/252\right)$ with $K = SP100_{t+1}$ for ATM options; likewise for P_{t+1}. The price of the same straddle contract at the open of day *t*+*h*+1 is given by $P_{t+h+1} + C_{t+h+1}$ with $C_{t+h+1} = f_{BS}^{call}\left(K, SP100_{t+h+1}, VXO_{t+h+1}, r_{f,t+h+1}, (M-h)/252\right)$; likewise for P_{t+h+1}.

t+1 only if the predicted h-step-ahead increase in daily realized variance as a proportion of the mean realized variance exceeds a pre-specified threshold, *i.e.* $\frac{\widehat{RV}_{t+h}-RV_t}{\overline{RV}_t} \times 100 > \delta\,\%$. Likewise, the trader takes a short position in straddles on day t only if $\frac{\widehat{RV}_{t+h}-RV_t}{\overline{RV}_t} \times 100 < -\delta\,\%$. The mean $\overline{RV}_t$ is computed iteratively over the same rolling window used to construct $\widehat{RV}_{t+h|t}$. On days with no trading signal, the risk-free interest rate is earned.

Let TC denote the transaction costs incurred by buying or selling straddles on the S&P 100 index.[11] The return associated with a buy (sell) signal triggered by the forecast $\widehat{RV}_{t+h|t}$ is

$$\begin{aligned} r_{t+1,t+h+1}^{buy} &= \frac{C_{t+h+1}+P_{t+h+1}-(C_{t+1}+P_{t+1})-(C_{t+1}+P_{t+1})\cdot TC-(C_{t+h+1}+P_{t+h+1})\cdot TC}{C_{t+1}+P_{t+1}} \\ r_{t+1,t+h+1}^{sell} &= \frac{C_{t+1}+P_{t+1}-(C_{t+h+1}+P_{t+h+1})-(C_{t+1}+P_{t+1})\cdot TC-(C_{t+h+1}+P_{t+h+1})\cdot TC}{C_{t+1}+P_{t+1}} \end{aligned} \qquad (10)$$

where $(C_{t+1}, P_{t+1})'$ and $(C_{t+h+1}, P_{t+h+1})'$ are the call and put option prices at the open of day t+1 and t+h+1, respectively. The return of the trading rule over the holdout period (T_1 days) is

$$\begin{aligned} R_h = \textstyle\sum_{\#b}\left(\frac{C_{t+h+1}+P_{t+h+1}}{C_{t+1}+P_{t+1}} - 1\right) + \sum_{\#s}\left(1 - \frac{C_{t+h+1}+P_{t+h+1}}{C_{t+1}+P_{t+1}}\right) \\ - \textstyle\sum_{\#bs}\left(1 + \frac{C_{t+h+1}+P_{t+h+1}}{C_{t+1}+P_{t+1}}\right) TC + \sum_{T_1-\#bs} \frac{r_{f,t}}{252} \end{aligned} \qquad (11)$$

where $r_{f,t}/252$ is the daily risk-free interest rate, $\#b$ and $\#s$ are the total number of buy and sell signals, respectively, over the entire holdout period, and $\#bs$ is the sum of the two. Equation (11) implies an almost linear relation between TC and profitability. We accommodate diverse trading scenarios by varying the filter rule or threshold parameter δ (from 0% to 50% in steps of 1%), transaction costs TC (from 0% to 2.5% of the straddle price, in steps of 0.1%), forecast horizon h (from 1 to 10 days), and type of market conditions ('calm' versus 'turmoil'). The economic criteria are the Sharpe Ratio, alpha and non-normality robust profitability metrics.

[11] Option traders face explicit or direct costs such as commissions and other fees, and implicit or indirect costs such as the bid-ask spread that market makers charge for shouldering the trader's undesired inventory position, and information costs. All these cost elements are typically proxied by a fixed transaction cost; e.g., Noh, Engle and Kane (1994) consider a transaction cost of $0.25 per straddle. We assign transaction costs as a percentage of the straddle price. Further discussion of transaction costs in the context of various option trading strategies can be found in Santa-Clara and Saretto (2009) and Broadie, Chernov and Johannes (2009).

5.2. Statistical criteria

In this section we outline the statistical volatility forecast evaluation that we carry out, for completeness, that relies on standard error measures for *point* forecasts, a scoring rule for *density* forecasts, and directional forecast measures of the accuracy of *sign* predictions.

First, defining the out-of-sample forecast error as $e_{t+h} \equiv RV_{t+h} - \widehat{RV}_{t+h|t}$, we compute various point forecast error measures: Mean Square Error $MSE = \frac{1}{T_1}\sum_{t=1}^{T_1}(e_{t+h})^2$, Mean Absolute Error (percentage) $MAE = \frac{1}{T_1}\sum_{t=1}^{T_1}|e_{t+h}|/(RV_{t+h} + \widehat{RV}_{t+h|t})$, Heteroskedasticity-adjusted Mean Square Error $HMSE = \frac{1}{T_1}\sum_{t=1}^{T_1}(1 - RV_{t+h}^{-1}\,\widehat{RV}_{t+h|t})^2$, Logarithm Loss $LL = \frac{1}{T_1}\sum_{t=1}^{T_1}(\ln RV_{t+h} - \ln\widehat{RV}_{t+h|t})^2$, and Quasi-Gaussian log-likelihood $QLIKE = \frac{1}{T_1}\sum_{t=1}^{T_1}\frac{RV_{t+1}}{\widehat{RV}_{t+h|t}} - \ln(\frac{RV_{t+1}}{\widehat{RV}_{t+h|t}}) - 1$. Patton and Sheppard (2009) advocate the loss functions implicit in MSE and QLIKE as the most robust to noise in the volatility proxy (here, RV_t).

Second, we also assess density forecast accuracy using the continuous ranked probability score $CRPS(F,h) = \frac{1}{T_1}\sum_{t=1}^{T_1}\int_{-\infty}^{+\infty}(F_{t+h}(y) - I\{y \geq RV_{t+h}\})^2 dy$ where F_{t+h} is the cumulative density function of the forecast distribution, RV_{t+h} is the verifying observation, and $I\{\cdot\}$ is the *Heaviside* function. A density forecast is more informative that a point forecast because it also incorporates a measure of the uncertainty that ought to be associated with the point forecast.

Third, we estimate Mincer-Zarnowitz style regressions in which the realized variances are regressed on the corresponding out-of-sample forecasts, $RV_{t+h} = b_0 + b_1\widehat{RV}_{t+h|t}^{j} + v_t$, where *j* denotes the predictive method at hand. The joint restrictions H_0: $b_0 = 0 \cup b_1 = 1$ are tested using a Wald test statistic; test rejection suggests systematic bias (under- or over-prediction). Volatility forecast accuracy is further measured by the coefficient of determination *MZ*-R^2 which reflects the variance but not the bias-squared component of the MSE.

Finally, we consider two *hit ratios* which are aimed at measuring the ability of the forecasts to anticipate directional-change. Let T_1^+ and T_1^- denote the total number of actual volatility increases and decreases, respectively, over the out-of-sample period. The positive hit ratio is defined as $Hit^+ = \frac{1}{T_1^+}\sum_{t=1}^{T_1^+} I_{\{(\widehat{RV}_{t+h|t} - RV_t)\cdot(RV_{t+h} - RV_t) > 0\}} \cdot I_{\{(RV_{t+h} - RV_t) > 0\}}$, where the indicator

$I_{\{\cdot\}}$ takes value 1 if the condition expressed inside the curly brackets is met. The negative hit ratio is similarly defined as $Hit^{-} = \frac{1}{T_1^-}\sum_{t=1}^{T_1^-} I_{\{(\widehat{RV}_{t+h|t}-RV_t)\cdot(RV_{t+h}\text{-}RV_t)>0\}} \cdot I_{\{(RV_{t+h}\text{-}RV_t)<0\}}$.

6. Empirical Results

In this section we begin by presenting the model estimation results before discussing the economic and statistical evaluation of model-based and non-parametric NN forecasts.

6.1. Model estimation and diagnostics

Among the long-memory ARFIMA(*p*,*d*,*q*) parameterizations resulting from combinations of lag orders $p, q \in \{0,1,2\}$, the lowest BIC is attained by the ARFIMA(1,*d*,0) which we subsequently adopt as our baseline long-memory model. This specification has been advocated in other studies as providing a parsimonious and effective description of the dynamics of realized variance; see, *e.g.* Koopman, Jungbacker and Hol (2005), and Pong, Shackleton, Taylor and Xu (2004) among others. We extend the ARFIMA(1,*d*,0) model in order to capture also leverage (ARFIMAX), and regime-switching (MSARFIMAX) as discussed in Section 4.1. The ARMAX(2,1) model is included as a reasonable competitor. Table 2 reports model parameter estimates and diagnostics for the in-sample period from January 6, 1997 until December 27, 2002 (T_0=1500 days) and full-sample period ending in November 16, 2012 (T=3990 days).

[Insert Table 2 around here]

The estimate of the fractional integration parameter of the ARFIMA model, $\hat{d} = 0.481$, is very near the estimate reported in Koopman, Jungbacker and Hol (2005) for the S&P 100 volatility, notwithstanding some differences in the realized variance measure adopted and time span.

Several observations can be made from the diagnostics reported in Table 2. First, the linear ARFIMA(1,*d*,0) versus nonlinear ARFIMAX(1,*d*,0) comparison suggests that by parameterizing the *leverage* effect through α_1 and α_2 there are notable in-sample fit improvements as borne out by a reduction in the AIC and BIC. Second, over both sample periods, the ARMAX(2,1) model achieves lower AIC and BIC values than the ARFIMAX(1,*d*,0), but the persistence of the realized variance materializes as a significant unit root in the autoregressive component of the ARMAX (*i.e.*, $\hat{\phi}_1$ is very close to 1). The estimated vector $(\hat{\alpha}_1, \hat{\alpha}_2)'$ is very close in the ARFIMAX and ARMAX models, and the parameter signs and magnitude are well aligned with the *leverage* interpretation: past negative returns significantly increase current volatility but past positive returns have a mute effect.[12]

Overall, the best diagnostics correspond to the nonlinear MSARFIMAX whose parameter estimates are aligned with those reported in previous studies (*e.g.*, Maheu and McCurdy, 2002). Volatility shocks have permanent-like effects in the high-volatility or crisis regime ($\hat{d}_H > 0.5,$) but are transitory in the normal regime ($\hat{d}_L < 0.5$) . The staying probabilities $\hat{\pi}_{HH}$ and $\hat{\pi}_{LL}$ are much lower in the in-sample than full-sample estimation. This is plausible because, as Figure 1 shows, the in-sample period 1997-2002 (first rolling estimation window) broadly reflects a single volatility state; a 'high' and 'low' volatility regimes are therefore difficult to identify. Over the full-sample period, by contrast, there is a clear low volatility state over the period 2003-2007 and a subsequent jittery period 2008-2012 with various high volatility clusters.

6.2. Out-of-sample forecast evaluation

We begin by plotting the realized variances and corresponding forecasts derived from a long-memory model and the non-parametric NN approach (*h*=5 days ahead). Figure 3 focuses on the MSARFIMAX model for brevity; additional graphs for all the long-memory models entertained are shown in the on-line Appendix. The graphs provide evidence to suggest, *prima facie*, that the NN forecasts track relatively well the ups and downs of daily realized variance.

[Insert Figure 3 around here]

In order to assess the economic value of the forecasts, we plot in Figure 4 the Sharpe ratio of the straddle trading strategy discussed in Section 5 for various filter rules, transaction costs and predictive horizons.[13] The top graphs (Panel A) correspond to the calm period 2003-2007, and the bottom graphs (Panel B) to the jittery period 2008-2012.[14]

[Insert Figure 4 around here]

[12] Reassuringly, estimation of an ARMAX(10,0) model yields log-likelihood, AIC and BIC values of -3121.11, 6272.22 and 6366.59, respectively, which are similar to those of the parsimonious ARMAX(2,1).

[13] The Sharpe ratio for a straddle held during h days is $SR(\tilde{r}_{t,t+h}) = \sqrt{h}SR(\tilde{r}_{t,t+1})$ where $\tilde{r}_{t,t+h} = r_{t,t+h} - h * r_{f,t}/252$ and $r_{t,t+h}$ is the return of the straddle position (long or short) over an *h*-day period, as defined in (10). We annualize the Sharpe ratio by multiplying it by $\sqrt{252/h}$. Likewise with the Sortino ratio.

[14] The ARFIMA forecasts are not considered in Figure 4 to avoid too dense and cumbersome graphs. Instead, detailed summary profitability statistics of the ARFIMA model are provided in subsequent tables.

Trading profitability improves as the noise-filtering rule becomes tougher and the horizon increases; the horizon effect is more noticeable in the calm period. Trading profitability is adversely affected by transaction costs, almost linearly, as suggested by equation (10).

The comparison of Sharpe ratios across forecasts is quite revealing. It can be seen that NN forecasts are excellent competitors to model-based forecasts, particularly, during the turbulent post-2007 trading period (Panel B) when they result in the most attractive Sharpe ratios. This evidence supports the non-parametric NN approach for forecasting realized variance.

Figure 4 further shows that the directional (Unanimity Rule) combination of model-based and NN predictions is beneficial, namely, the combined forecasts entail higher Sharpe ratios than any of the individual forecasts. In the pre-2007 period, the ARFIMAX+NN and MSARFIMA+NN combinations are generally more effective than the HAR+NN combination. In the post-2007 combination, the three combinations are quite close. However, in some trading scenarios (*i.e.*, mild filter rule and low transaction costs) the NN predictions outperform all other volatility predictions, individual or combined, during the turbulent post-2007 period.

In addition to the directional (Unanimity Rule) combination, we consider two well-known combination schemes for point forecasts: Equal Weights (EW) and Ordinary Least Squares Weights (OLS). To avoid look-ahead bias in the latter scheme, we obtain the weights iteratively by regressing over each rolling estimation windows the realized variances on the in-sample (NN and model) forecasts. Figure 5 plots the Sharpe ratios of the straddles informed by the three types of combined forecasts; for simplicity, we focus on the ARFIMAX and HAR models.

[Insert Figure 5 around here]

The graphs suggest that directional-forecast combination is more effective for straddle trading than point-forecast combination. Among the two point-forecast combination schemes, the simple EW scheme produces higher Sharpe ratios than the OLS scheme. This is not surprising, given that the OLS weights are contaminated by estimation noise due to multicollinearity. Similar evidence stems from the (unreported) graphs of Sharpe ratios resulting from the combination of NN forecasts with either ARFIMA, ARMAX or MSARFIMAX forecasts.

A battery of trading performance statistics for naïve forecasts, model-based forecasts, non-parametric NN forecasts and combinations of model-based and NN forecasts are shown in Table 3; for brevity, the table focuses on δ=35%, *TC*=2% and *h*=5 days.[15]

Several observations can be made. The annualized mean excess returns of our options trading strategy are broadly in line with those reported in Santa-Clara and Saretto (2009) and Broadie, Chernov and Johannes (2009) for similar transaction cost levels. The returns of buy (sell) signals are negative (positive) which suggests that, the profitability of the straddle trading strategy hinges especially on anticipating correctly the directional change of volatility for future volatility falls. Consistent with this observation, among all individual forecasts the NN forecasts excel versus naïve and model-based forecasts because they earn the highest returns in the *sells*.

[Insert Table 3 around here]

Examining the trading activity associated with the different forecasting approaches, we observe an interesting pattern. In the relatively calm market period 2003-2007, all forecasts trigger more sells than buys, but the sell/buy ratio is notably greater for the combined forecasts. In the turbulent market period 2008-2012, the ratio of sell to buy signals generally decreases (that is, for NN forecasts and model-based forecasts) but less so for the NN forecasts which now produce the largest sell/buy ratio. This provides a rationale for our previous finding that during turbulent markets the NN forecasts excel among all individual forecasts. The main findings from this economic analysis of predictability is that: i) in the comparison of individual forecasts, the non-parametric NN forecasts stand out as the most effective for *straddle* trading during turbulent market conditions possibly because they mitigate the noise-to-signal ratio relative to model-based predictions, ii) both in calm and turbulent market conditions the largest profitability is obtained by directional combination of NN forecasts and model-based forecasts.

These findings are not challenged when we consider instead the alpha and non-normality robust performance measures (*e.g.*, Leland's alpha, Sortino and Omega ratio) as shown in Table 3 for δ=35%, *TC*=2% and *h*=5 days. The alpha and Sortino ratios plotted in Figures 6 and 7,

[15] Becker, Clements and Hurn (2011) propose a semi-parametric approach where the volatility forecast is a *weighted* average of the most-recent realized volatilities. The weight function is a multivariate Gaussian kernel that exploits the available time-series $\{RV_t\}$ and *N*-1 additional time-series to account for short-term trends. We deployed this semi-parametric approach by defining the additional series, as in their paper, as λ-day moving averages of RV_t for λ={2,3,5,10}. Results in the on-line Appendix suggest that the risk-adjusted profitability afforded by the NN forecasts is at least as good as that of the kernel-based forecasts.

respectively, for a range of transaction costs, filter rules and horizons confirm the main findings. Additional graphs shown in the on-line Appendix (Figures A3 and A4) confirm the superiority of the directional forecast combination vis-à-vis the EW and OLS point forecast combinations.

[Insert Figures 6 and 7 around here]

Finally, we examine the statistical ranking of forecasts according to several conventional criteria as outlined in Section 5.2. Table 4 shows the results for horizon *h*=5 days, for brevity.

[Insert Table 4 around here]

Panel A of Table 4 reports forecast accuracy metrics for the *level* of daily realized volatility. As suggested by the Mincer-Zarnowitz regressions, the predictive power of the NN scheme is not better than that of long-memory models; *i.e.*, $1-\frac{R^2_{NN}}{R^2_j}>0$.[16] Likewise, relative average point forecast error measures, such as $1-MSE_{NN}/MSE_j$ in the context of quadratic losses, and the relative continuous ranking probability score $1-CRPS_{NN}/CRPS_j$ that takes into account the entire predictive density, are negative and therefore do not favor the NN forecasting approach either. Overall, the best forecasting approach according to the statistical criteria is the HAR model. The statistical ranking of forecasts thus stands in sharp contrast with the economic ranking associated with the straddle trading analysis. This evidence adds to a growing literature that documents a mismatch between statistical accuracy and profitability of forecasts (*e.g.*, Leitch and Tanner, 1991; Satchell and Timmermann, 1995; Abhyankar, Sarno and Valente, 2005; Cenesizoglu and Timmermann, 2012; and Bernales and Guidolin, 2014).

We next assess the relative directional accuracy of NN forecasts using the hit ratios for volatility rises and falls, $\Delta Hit^+ \equiv 1-\frac{Hit^+_{NN}}{Hit^+_j}$ and $\Delta Hit^- \equiv 1-Hit^-_{NN}/Hit^-_j$, respectively, as defined in Section 5.2. Panel B of Table 4 shows that while ΔHit^+ is generally positive, ΔHit^- is negative suggesting that the NN forecasts excel in anticipating future market volatility falls. This contrast is reminiscent of our prior findings that: i) the profitability of the trading strategy is mainly driven by the straddle sell signals, and ii) the frequency of sells relative to buys during the turbulent market period is clearly largest for the NN scheme. This explains why the NN forecasts are the most profitable, among all individual forecasts, when the market is in turmoil.

[16] The last column of Panel A in Table 4 also reports Wald test statistics for the null hypothesis of forecast unbiasedness. Common across approaches, the hypothesis is refuted in the turbulent (but not the calm) period.

7. Conclusions

This paper investigates the merit of the non-parametric Nearest Neighbor (NN) approach for daily realized stock market volatility prediction relative to widely-used long memory models, and their combination. The main appeal of the NN approach in this context is that it lends itself as a flexible forecasting tool that can mitigate misspecification risk and parameter uncertainty.

The analysis is based on daily realized variances of the S&P 100 index over the time period from January 6, 1997 to November 16, 2012. We focus on the widely-used class of Autoregressive Fractionally Integrated Moving Average Models (ARFIMA) models, the convenient Heterogeneous Autoregressive (HAR) approximation to long memory, and nonlinear extensions to capture leverage (ARFIMAX), and leverage together with Markov-switching behavior (MSARFIMAX). The evaluation of forecast ability is conducted over the relatively ‘calm’ 2003-2007 episode and the ‘turbulent’ 2008-2012 episode, using both standard statistical criteria and economic profitability measures associated with straddle trading.

Our findings suggest that in turbulent markets, when the noise-to-signal ratio is relatively large, the NN predictions emerge as more profitable than the model-based forecasts. More pervasively, in both calm and turbulent markets, the combination of NN forecasts and model-based forecasts outperforms all the individual forecasts in terms of risk-adjusted profitability. As a byproduct, our analysis verifies a previously documented mismatch between the ranking of forecasting approaches according to economic value and the ranking dictated by statistical accuracy. Further research can explore ways to enhance the NN forecasts by calibrating the number of training vectors and embedding dimension using a broader measure of ‘nearest neighborhood’ than the Euclidean distance like, for instance, the Kullback-Leibler metric.

The paper cannot and does not advocate our simple straddle trading rule based on realized variance forecasts as an optimal trading strategy. In fact, modeling the dynamics of the implied volatility surface for forecasting purposes is a very active area of research with implications for the design of option trading strategies; see *e.g.*, Bernales and Guidolin (2014). There is room for extending the present analysis to other trading strategies such as S&P 500 straddles and variance swaps, assessing whether the combination of implied and intraday-based realized volatility forecasts is beneficial, and to other economic problems informed by volatility forecasts such as Value-at-Risk where periods of crisis and high volatility are of most concern.

References

Abhyankar, A., Sarno, L. & Valente, G. (2005). Exchange rates and fundamentals: Evidence on the economic value of predictability. *Journal of International Economics* 66, 325-48.

Ahoniemi, K., Fuertes, A. -M. & Olmo, J. (2015). Overnight News and Daily Equity Trading Limits. *Journal of Financial Econometrics* 13, 1-27.

Andersen, T. & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39, 885-905.

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

Andersen, T., Bollerslev, T. & Huang, X. (2011). A reduced form framework for modeling volatility of speculative prices based on realized variation measures. *Journal of Econometrics* 160, 176-189.

Angelidis, T. & Degiannakis, S. (2008). Volatility forecasting: intra-day versus inter-day models. *Journal of International Financial Markets, Institutions and Money* 18, 449-465.

Arroyo, J. & Mate, C. (2009). Forecasting histogram time series with *k*-nearest neighbors methods. *International Journal of Forecasting* 25, 192-207.

Barndorff-Nielsen, O. E., & Shephard, N. (2002). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society*, Series B, 64, 253–280.

Becker, R., Clements, A. & Hurn, S. (2011). Semi-parametric forecasting of realized volatility. *Studies in Nonlinear Dynamics and Econometrics* 15, Article 1.

Bernales, A. & Guidolin, M. (2014). Can we forecast the implied volatility dynamics of equity options. *Journal of Banking and Finance* 46, 326-342.

Bollerslev, T. & Wright, J. H. (2000). Semiparametric estimation of long memory volatility dependencies: The role of high-frequency data. *Journal of Econometrics* 98, 81-106.

Broadie, M., Chernov, M. & Johannes, M. (2009). Understanding index option returns. *Review of Financial Studies* 22, 4493-4529.

Brownlees, C. T. & Gallo, G. M. (2010). Comparison of volatility measures: A risk management perspective. *Journal of Financial Econometrics* 8, 29-56.

Casdagli, M. (1989). Nonlinear prediction of chaotic time series. *Physica D* 35, 335-356.

Cenesizoglu, T. & Timmermann, A. (2012). Do return prediction models add economic value? *Journal of Banking & Finance* 36, 2974-2987.

Clements, M. P, Galvão, A. B. & Kim, J. H. (2008). Quantile forecasts of daily exchange returns from forecasts of realized volatility. *Journal of Empirical Finance* 15, 729-750.

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

Engle, R., Hong, C. & Kane, A. (1990). Valuation of variance forecasts with simulated option markets. NBER Working Paper 3350, National Bureau of Economic Research, Inc.

Farmer, D. & Sidorowich, J. (1987). Predicting chaotic time series. *Physical Review Letters* 59, 845-848.

Fernández-Rodríguez, F., Sosvilla-Rivero, S. & Andrada-Félix, J. (1999). Exchange rate forecast with simultaneous nearest-neighbour methods: Evidence from the EMS. *International Journal of Forecasting* 15, 383-392.

French, K. R., Schwert, G. W. & Stambaugh, R. E. (1987). Expected stock returns and volatility. *Journal of Financial Economics* 19, 3-29.

Fuertes, A.-M., & Olmo, J. (2012). Optimally harnessing inter-day and intra-day information for Value-at-Risk prediction. *International Journal of Forecasting* 29, 28-42.

Fuertes, A. -M., Kalotychou, E. & Todorovic, N. (2015). Daily volume, intraday and overnight returns for volatility prediction: Profitability or accuracy? *Review of Quantitative Finance and Accounting* 45, 251-278.

Ghysels, E., Santa-Clara, P. & Valkanov, R. (2004). The MIDAS touch: mixed data sampling regression models. Chapel Hill, mimeo.

Giot, L. & Laurent, S. (2004). Modeling daily value-at-risk using realized volatility and ARCH-type models. *Journal of Empirical Finance* 11, 379-398.

Gonçalves, S., G. & Guidolin, M. (2006). Predictable dynamics in the S&P 500 index options implied volatility surface. *Journal of Business* 79, 1591-1635.

Granger, C. W. J. & Andersen, A.P. (1978). *An Introduction to Bilinear Time Series Models*. Göttingen: Vandenhoeck & Ruprecht.

Granger, C. W. J. & Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* 1, 15-29.

Granger, C. W. J. & Teräsvirta, T. (1993). *Modeling Nonlinear Economic Relationships*. Oxford University Press.

Granger, C. W. J. & Newbold, P. (1976). Forecasting transformed series. *Journal of the Royal Statistical Society B* 38, 189-203.

Green, P. J. & Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*. Chapman & Hall: London.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57, 357-384.

Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press: Cambridge.

Hastie, T., Tibshirani, R. & Friedman, J. (2001). *The Elements of Statistical Learning*. Springer: New York.

Hsieh, D. (1991). Chaos and nonlinear dynamics: Applications to financial markets. *Journal of Finance* 47, 1145-1189.

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.

Leitch, G. & Tanner, J. E. (1991). Economic forecast evaluation: profits versus the conventional error measures. *American Economic Review* 81, 580-90.

Li, K. (2002), Long memory versus option-implied volatility predictions. *Journal of Derivatives* 9, 9-25.

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

Maheu, J. M. & McCurdy, T. H. (2002). Nonlinear features of realized FX volatility. *Review of Economics & Statistics* 84, 668-681.

Martens, M. & Zein, J. (2004), Forecasting financial volatility: High-frequency time series forecasts vis-vis implied volatility. *Journal of Futures Markets* 24(11), 1005-1028.

Martens, M., van Dijk, D. & de Pooter, M. (2009). Forecasting S&P 500 volatility: Long memory, level shifts, leverage effects, day-of-the-week seasonality and macroeconomic announcements. *International Journal of Forecasting* 25, 282-303.

McAleer, M. & Medeiros, M. C. (2008). A multiple regime smooth transition Heterogeneous AutoRegressive model for long memory and asymmetries. *Journal of Econometrics* 147, 104-119.

Noh, J., Engle, R. F. & Kane, A. (1994). Forecasting volatility and option prices of the S&P 500 index. *Journal of Derivatives* 2(1), 17-30.

Oomen, R. C. A. (2004). Modeling realized variance when returns are serially correlated. Leibniz Information Centre for Economics 2004-11. http://hdl.handle.net/10419/51076.

Patton, A. & Sheppard, K. (2009). Optimal combinations of realized volatility estimators. *International Journal of Forecasting* 25, 218-238.

Pong, S., Shackleton, M., Taylor, S. & Xu, X. (2004). Forecasting sterling/dollar volatility: A comparison of implied volatilities and AR(FI)MA models. *Journal of Banking and Finance* 28, 2541-2563.

Priestley, M. B. (1980). State-dependent models: A general approach to nonlinear time series analysis. *Journal of Time Series Analysis* 1, 47-71.

Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *Annals of Statistics* 23, 1630-1661.

Santa-Clara, P. & Saretto, A. (2009). Option strategies: Good deals and margin calls. *Journal of Financial Markets*, 12, 391-417.

Satchell, S. & Timmermann, A. (1995). An assessment of the economic value of non-linear foreign exchange rate forecasts. *Journal of Forecasting* 14, 477-497.

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

Shephard N. & Sheppard K. K. (2010). Realising the future: Forecasting with high-frequency based volatility (HEAVY) models. *Journal of Applied Econometrics* 25, 197-231.

Teräsvirta, T. (1994). Specification, estimation and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association* 89, 208-218.

Tong, H. (1978). On a threshold model. In *Pattern Recognition and Signal Processing*, eds. C. H. Chen, Sijhoff & Noordhoff: Amsterdam.

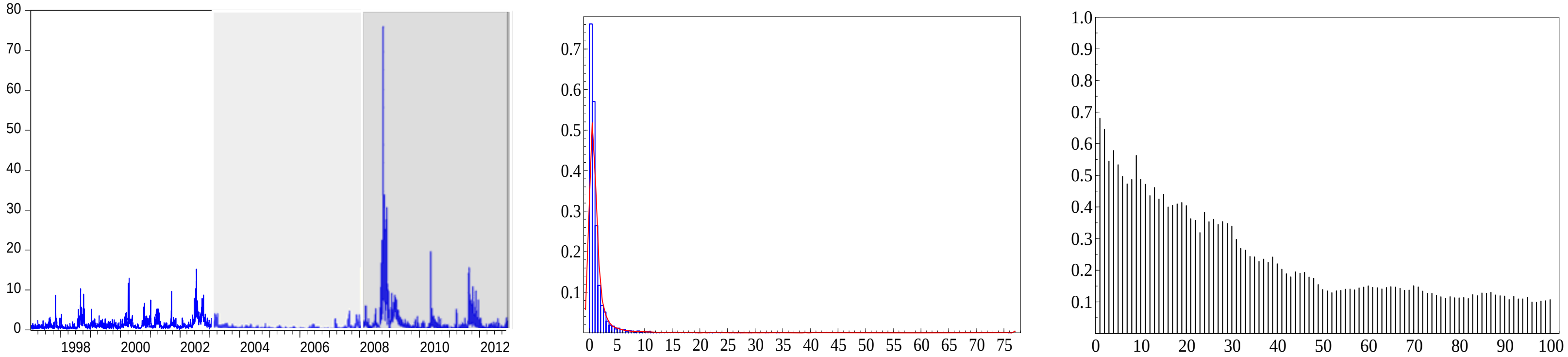
Timmermann, A. (2006) *Forecast combinations*. In G. Elliott, C. W. J. Granger & A. G. Timmermann (Eds.) Handbook of economic forecasting. Amsterdam: North Holland Press.

Xekalaki, E. & Degiannakis, S. (2005). Evaluating volatility forecasts in option pricing in the context of a simulated options market. *Computational Statistics and Data Analysis* 49, 611-629.

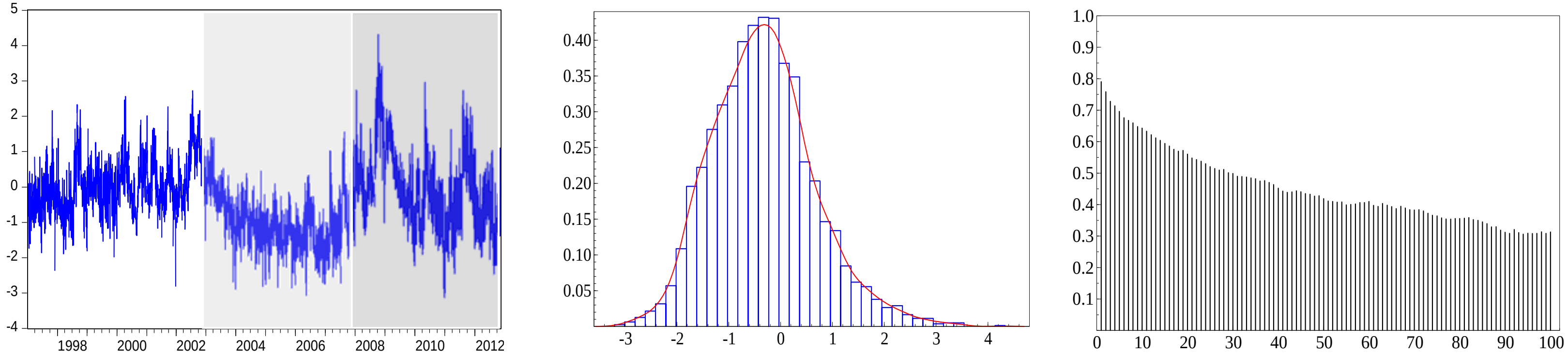
Yakowitz, S. (1987). Nearest-neighbour methods for time series analysis. *Journal of Time Series Analysis* 8, 235-247.

Figure 1. Daily Realized Volatility of S&P 100 Index.

(a) Realized Variance

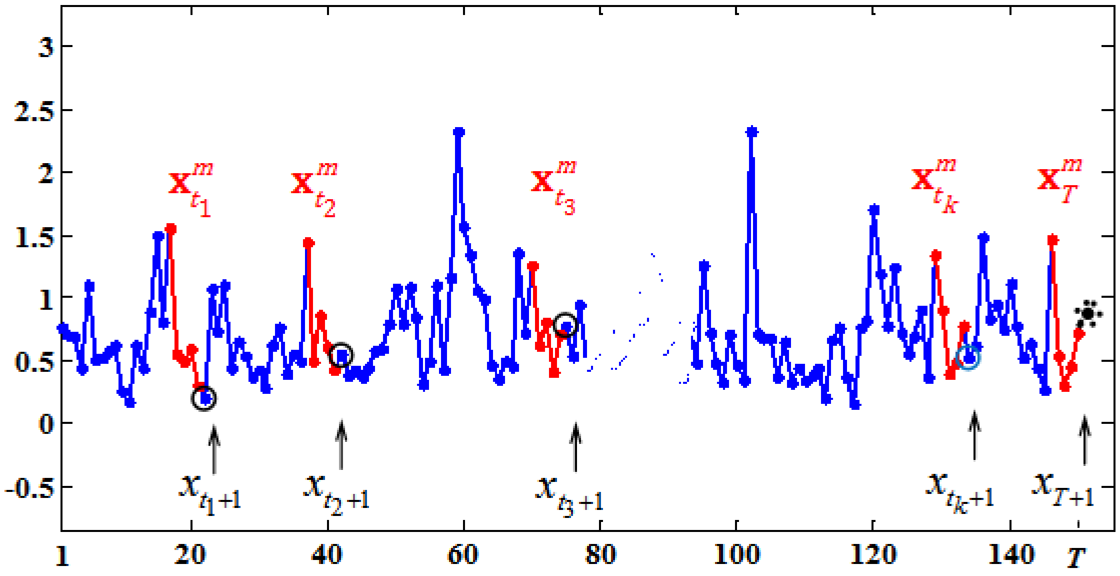


(b) Logarithmic Realized Variance



Notes: Time-series (first column), histogram (second column) and correlogram (third column) of S&P 100 daily realized variance in levels and logarithms constructed at a sampling frequency of 5 minutes over the period from January 6, 1997 until November 16, 2012 (3990 observations). The shaded area is the out-of-sample or holdout period (2490 days) for the forecast evaluation that comprises a low-volatility period from December 30, 2002 to December 31, 2007 (1260 days; light shade) and a high-volatility period from January 2, 2008 until the sample end (1230 days; dark shade).

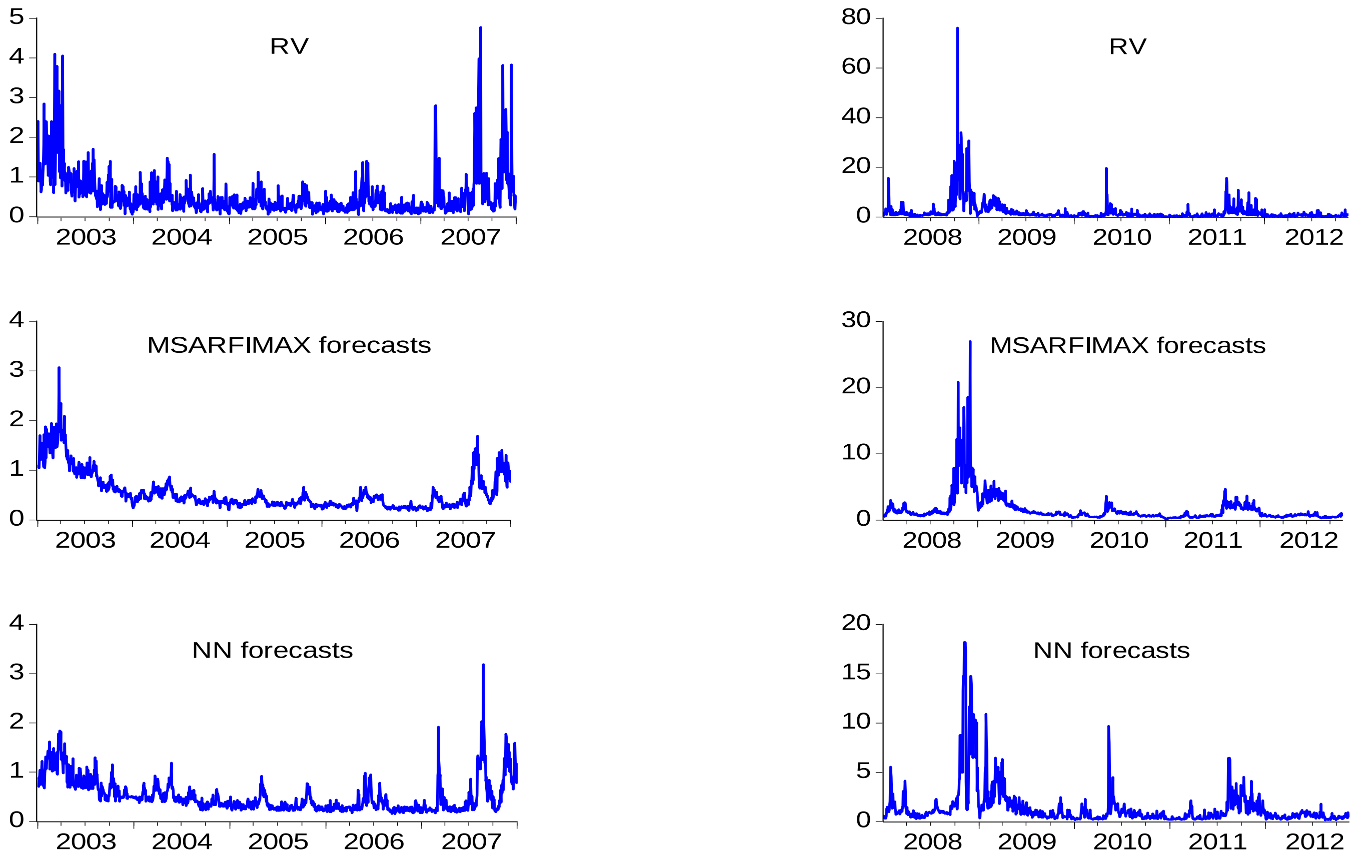
Figure 2. Nearest Neighbor Prediction Technique



Notes: The figure illustrates the nearest neighbor (NN) forecasting approach. The entire sequence of blue/red points is the available time-series $\{x_1, x_2, \ldots, x_T\}$, the last point in black is the 1-day-ahead NN prediction, $\hat{x}_{T+1}$. The first k sequences of red points, denoted $\mathbf{x}_{t_j}^m$, j=1,…k, are the $m \times 1$ nearest neighbor vectors or histories of m consecutive observations that are the closest to the most recent m-history denoted $\mathbf{x}_T^m$ (*i.e.*, the final sequence of red points). The NN prediction $\hat{x}_{T+1}$ is the median of $\{x_{t_1+1}, x_{t_2+1}, \ldots, x_{t_k+1}\}$, the subsequent observations to the m-histories (*i.e.*, the k circled points).

Figure 3. Daily S&P 100 Realized Variance and Out-of-Sample Forecasts

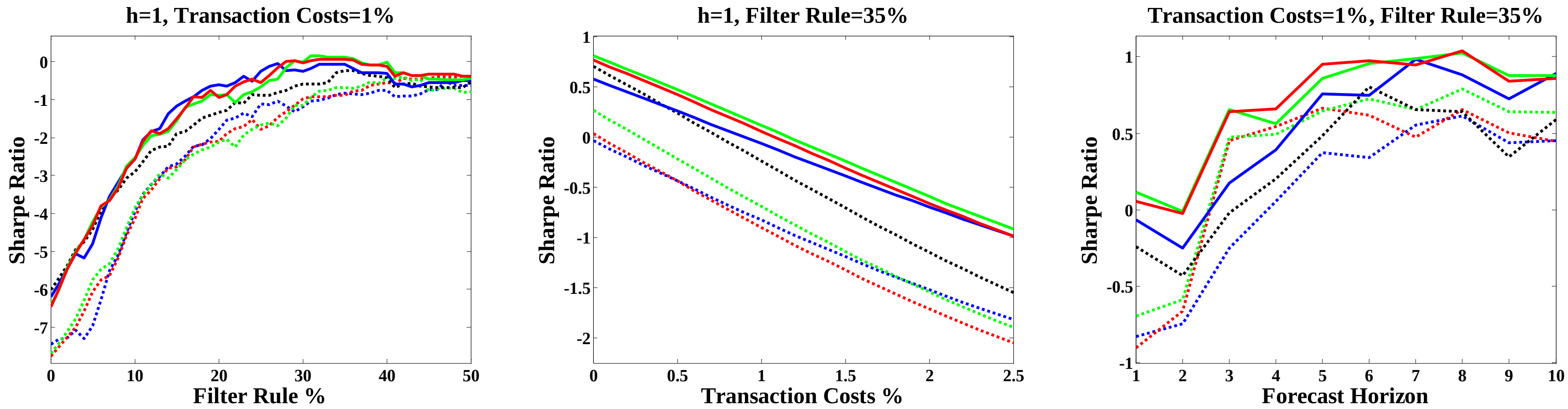
(a) Calm period: December 2002-December 2007 (b) Turbulent period: January 2008-November 2012



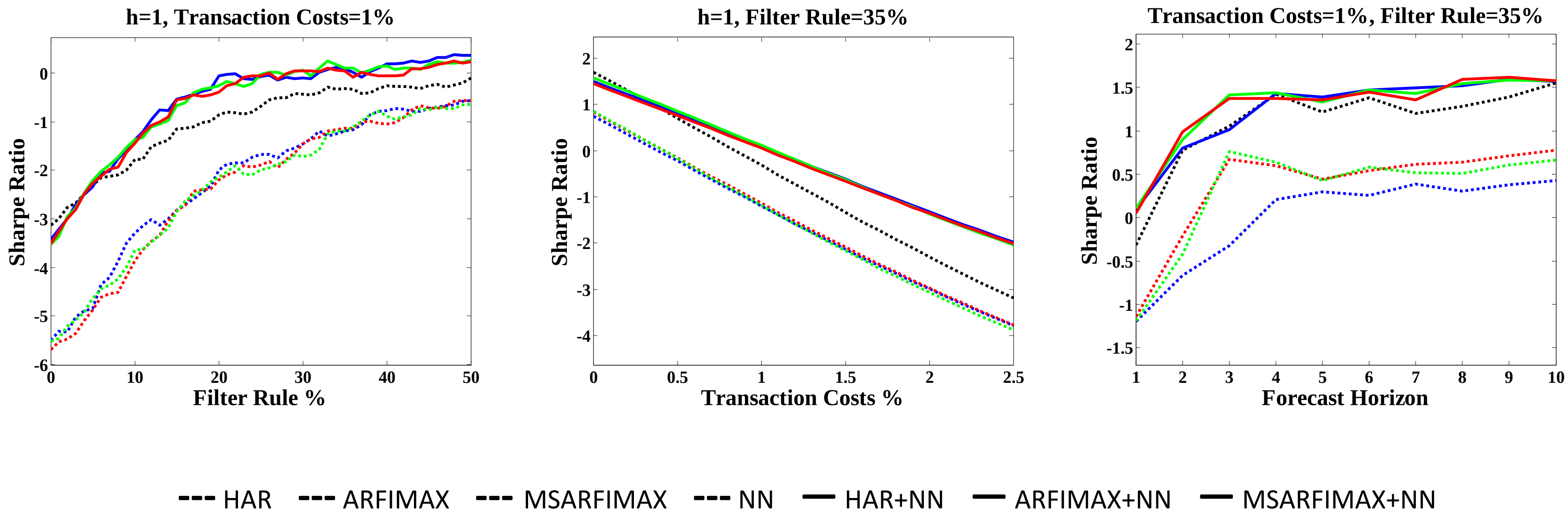
Notes: The figure plots actual S&P 100 realized variances and corresponding out-of-sample forecasts obtained from the MSARFIMAX model and the non-parametric nearest neighbor (NN) approach. The forecast horizon is h=5 days ahead and the forecasts are generated sequentially using a rolling window scheme.

Figure 4. Sharpe Ratio of Straddle Trading Informed by Model-based Forecasts, Non-parametric NN Forecasts and Combined Forecasts.

A) Calm period: December 2002-December 2007

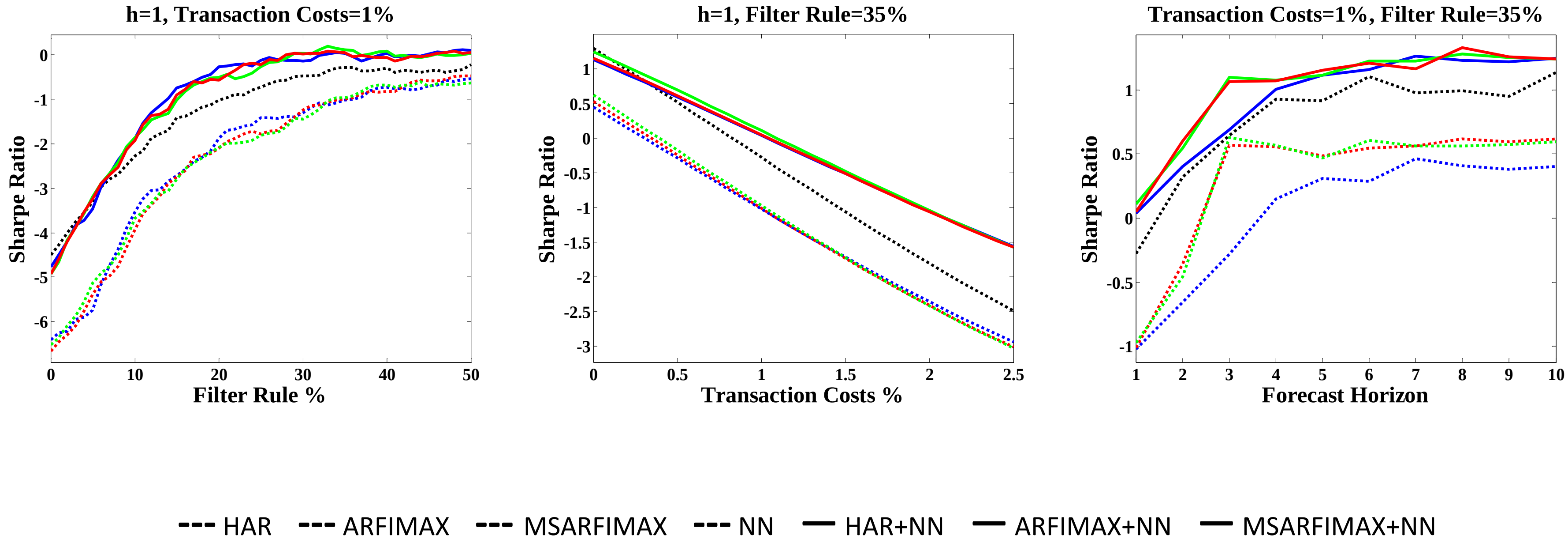


B) Turbulent period: January 2008-November 2012



(*cont.*)

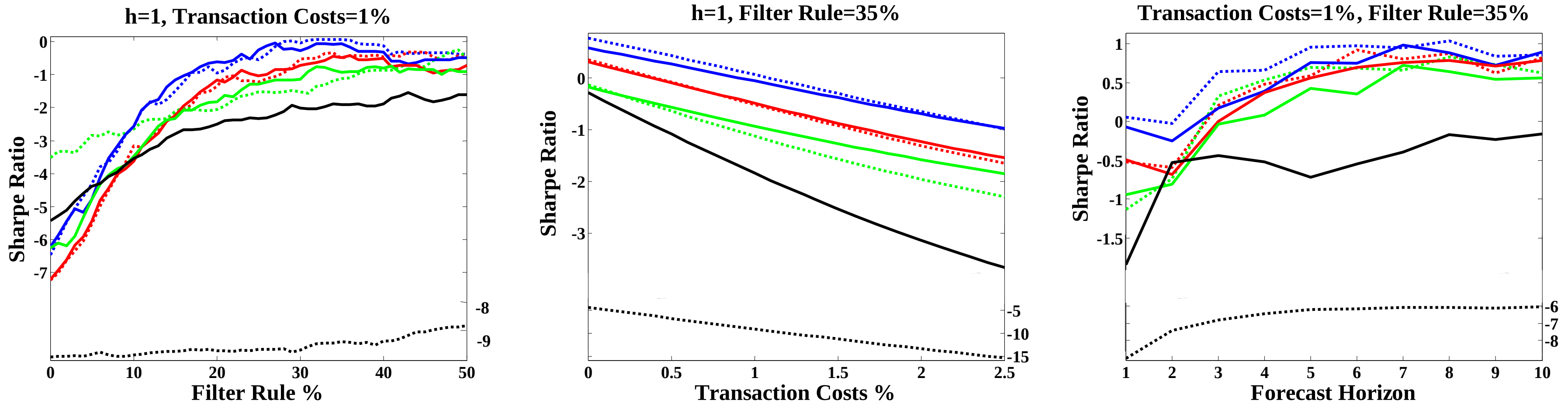
C) Full period: December 2002-November 2012



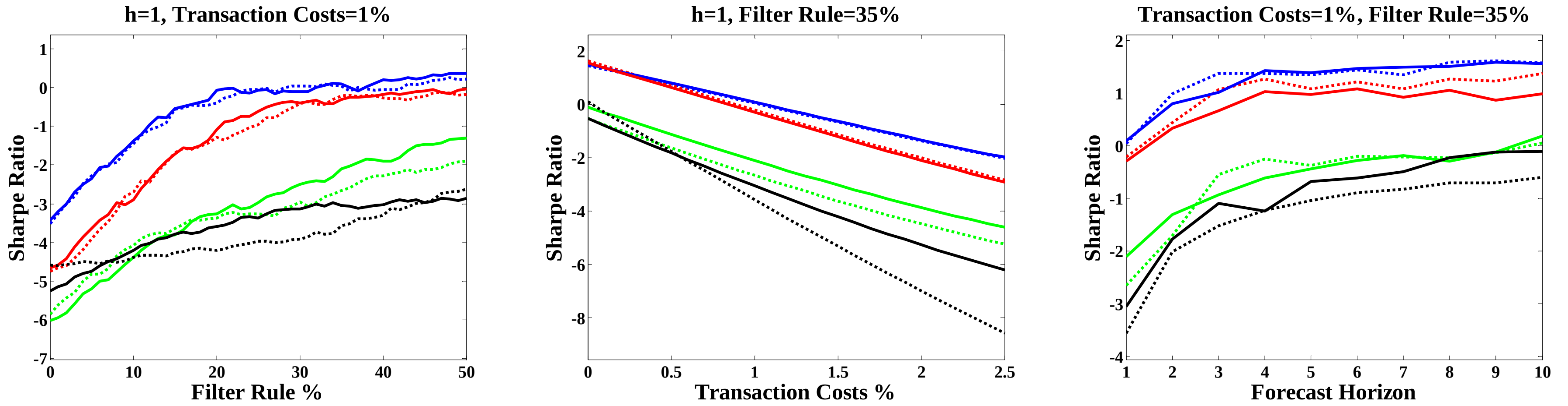
Notes: The graphs show the profitability of a straddle trading strategy informed by HAR, ARFIMAX, and MSARFIMAX model forecasts, non-parametric NN forecasts and combinations thereof according to a unanimity rule (directional forecast combination) by which a buy/sell signal is triggered only if the individual buy/sell signals agree.

Figure 5. Sharpe Ratio of Straddle Trading Informed by Directional Forecast and Point Forecast Combinations.

A) Calm period: December 2002-December 2007

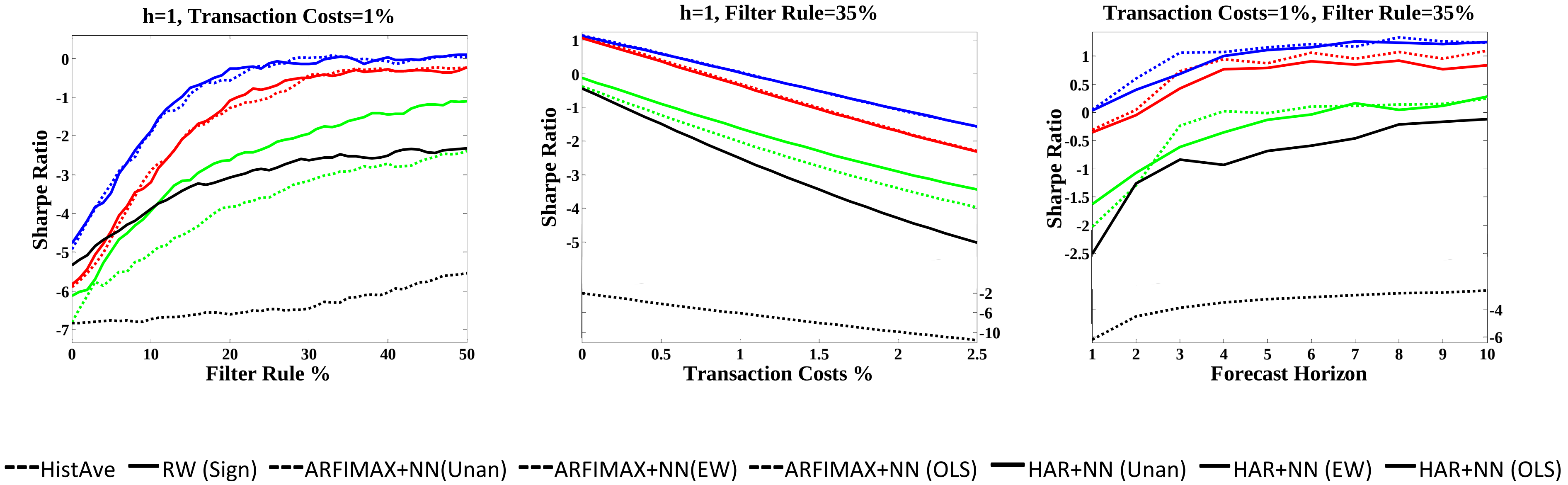


B) Turbulent period: January 2008-November 2012



HistAve — RW (Sign) — ARFIMAX+NN(Unan) — ARFIMAX+NN(EW) — ARFIMAX+NN (OLS) — HAR+NN (Unan) — HAR+NN (EW) — HAR+NN (OLS)

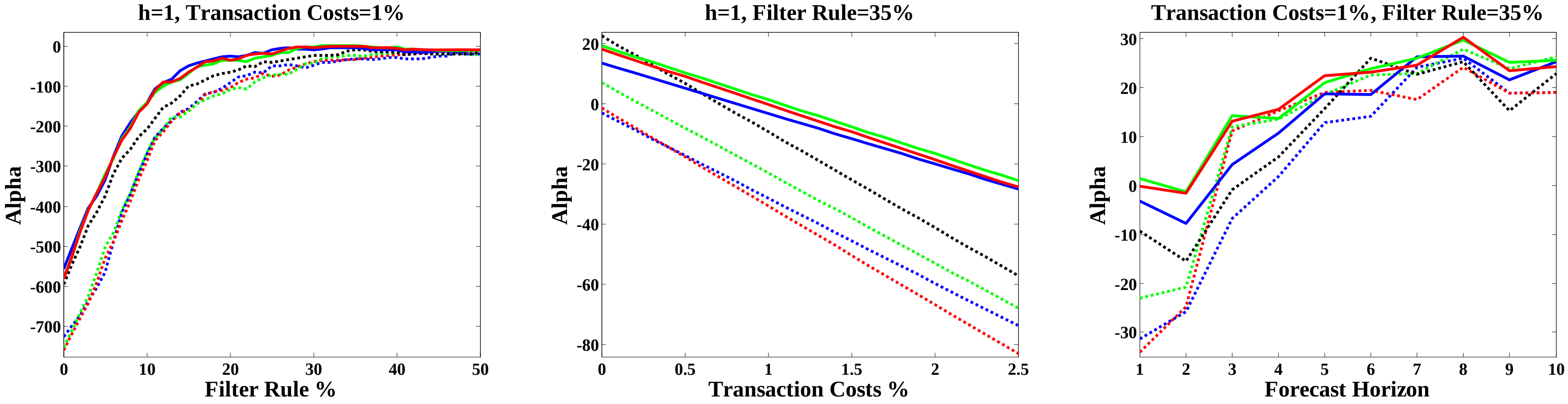
C) Full period: December 2002-November 2012



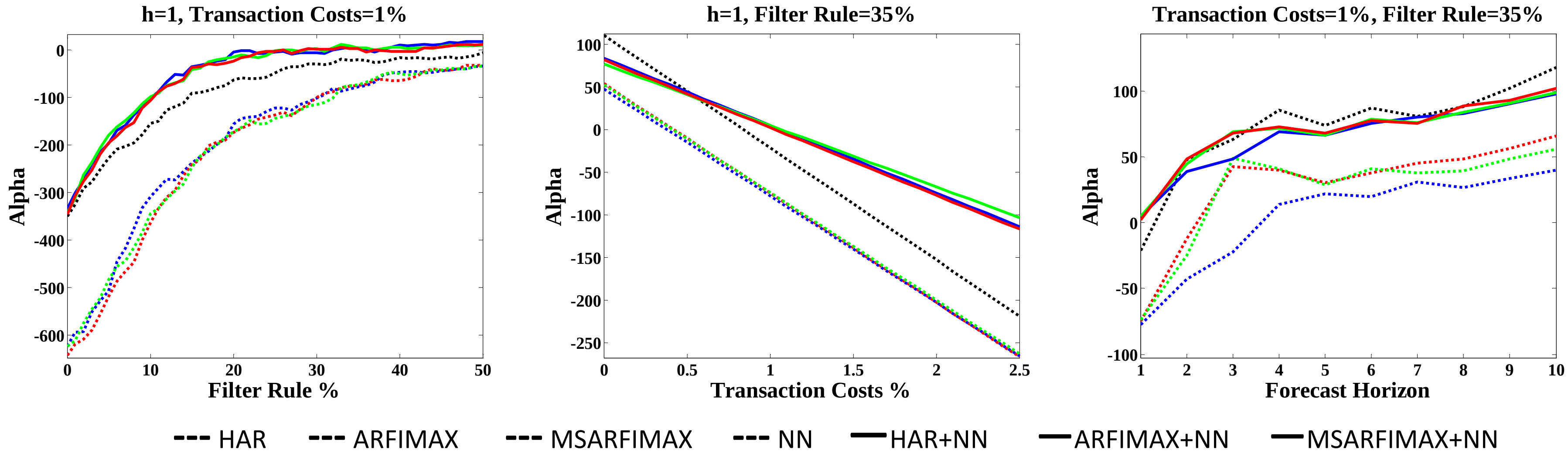
Notes: The graphs show the profitability of a straddle trading strategy informed by individual naïve forecasts (historical average of realized volatility or sign-of-change in volatility prediction according to the random walk principle), and by combinations of model-based (HAR, ARFIMAX) forecasts and nearest neighbor (NN) forecasts of realized volatility. The combination method is either the unanimity rule (directional forecast combination scheme by which a buy/sell signal is triggered only if the individual buy/sell signals agree), the equal-weight (EW) combination scheme or the OLS weight combination scheme.

Figure 6. Alpha of Straddle Trading Informed by Model-based Forecasts, Non-parametric NN Forecasts and Combined Forecasts.

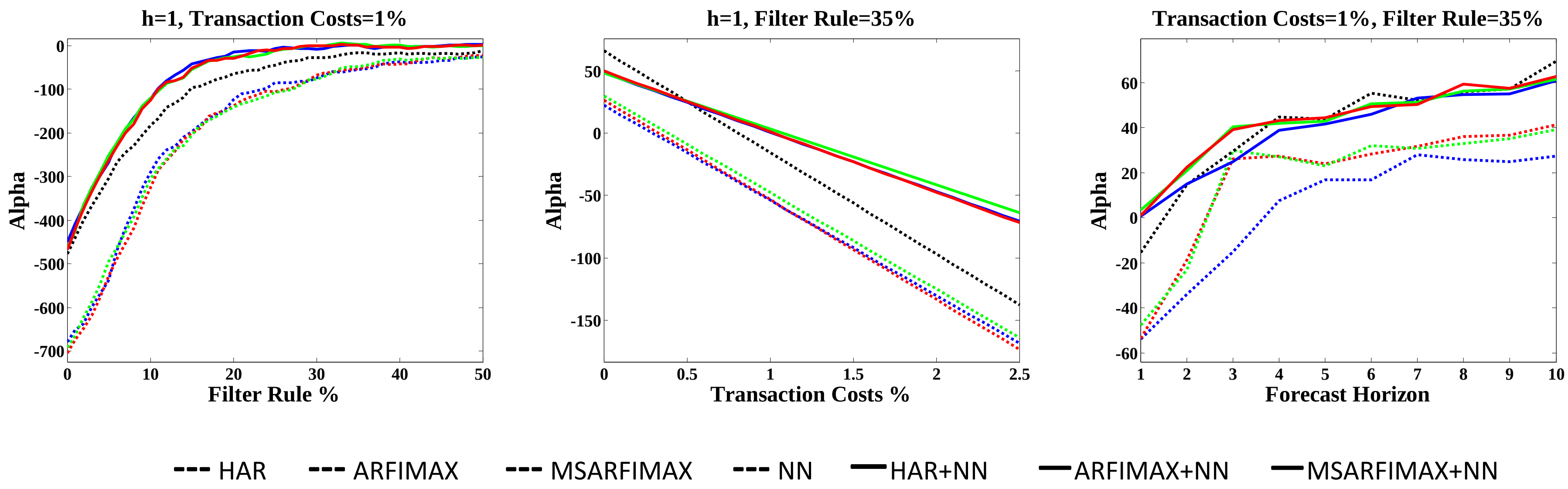
A) Calm period: December 2002-December 2007



B) Turbulent period: January 2008-November 2012



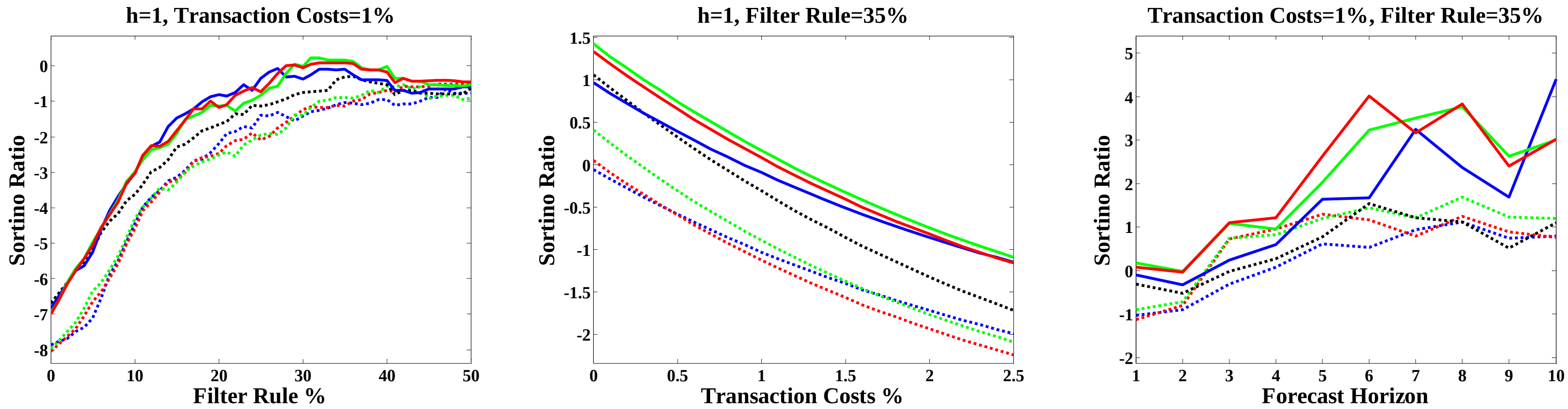
C) Full period: December 2002-November 2012



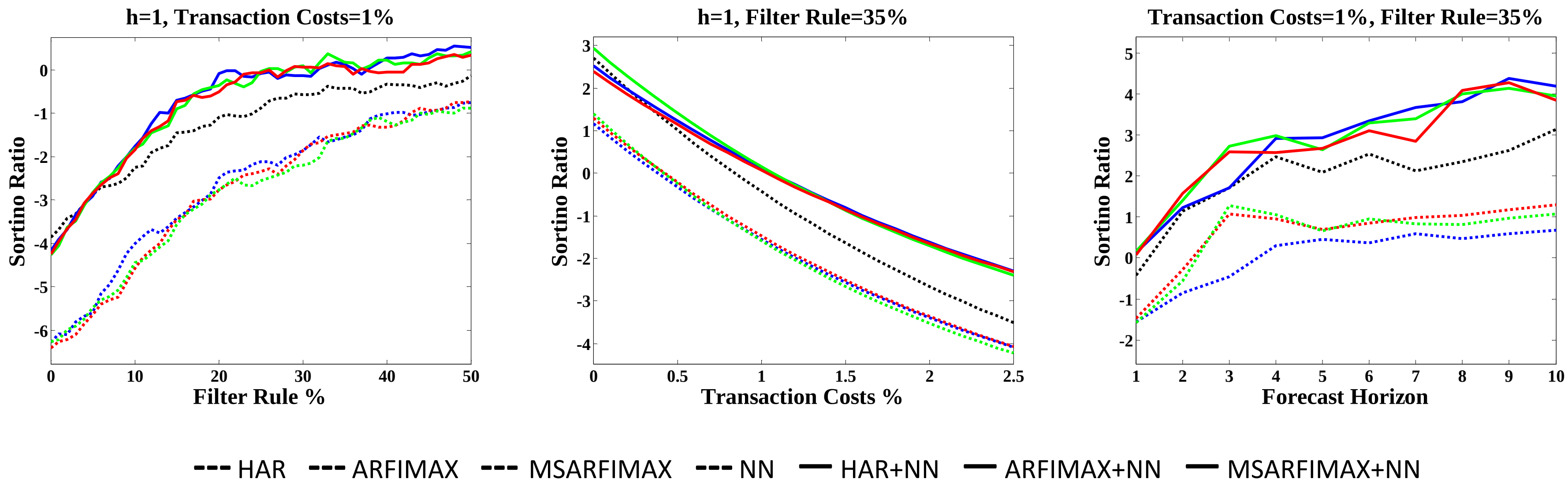
Notes: The graphs show the profitability of a straddle trading strategy informed by HAR, ARFIMAX, and MSARFIMAX model forecasts, non-parametric NN forecasts and combinations thereof according to a unanimity rule (directional forecast combination) by which a buy/sell signal is triggered only if the individual buy/sell signals agree.

Figure 7. Sortino Ratio of Straddle Trading Informed by Model-based Forecasts, Non-parametric NN Forecasts and Combined Forecasts.

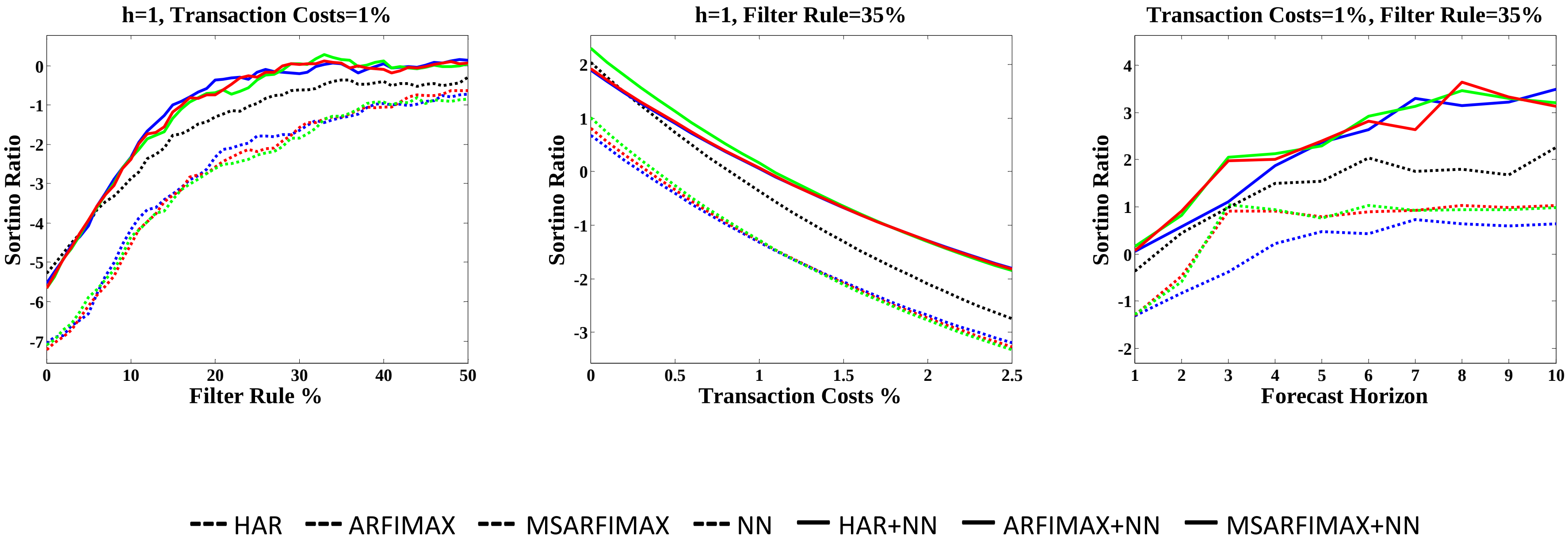
A) Calm period: December 2002-December 2007



B) Turbulent period: January 2008-November 2012



C) Full period: December 2002-November 2012



Notes: The graphs show the profitability of a straddle trading strategy informed by HAR, ARFIMAX, and MSARFIMAX model forecasts, non-parametric NN forecasts and combinations thereof according to a unanimity rule (directional forecast combination) by which a buy/sell signal is triggered only if the individual buy/sell signals agree.

Table 1. Descriptive Statistics for Daily S&P 100 Index

	Mean	Max	Min	StDev	Skew	Kurt	ADF(30)	Robinson *d*
Returns	0.010	9.102	-8.432	1.246	-0.185	9.095	-49.20	0.057
Standardized returns	0.073	3.262	-3.080	1.080	0.039	2.627	-65.71	0.057
Squared returns	1.553	82.837	0.000	4.416	9.345	125.8	-7.976	0.181
Realized variance	1.285	76.188	0.044	2.503	11.361	242.6	-5.999	0.388
Log realized variance	-0.329	4.333	-3.118	0.983	0.464	3.559	-7.024	0.426

Notes: The table contains summary statistics for daily logarithmic returns from open-to-close and realized variance over the period from January 6, 1997 until November 16, 2012 (3990 days). Standardized returns are the raw returns divided by the square root of the realized variance. All returns are expressed in percentage terms. ADF(30) is the Augmented Dickey Fuller test statistic for the null hypothesis of a unit root with lag order selected according to the BIC (max lag set at 30); the simulated 5% and 10% critical values are -2.8 and -2.5, respectively. Robinson *d* is the fractional integration estimate obtained via the Gaussian semi-parametric approach proposed by Robinson (1995).

Table 2. Estimated Models for Daily S&P Logarithmic Realized Variance

HAR	A: In-sample	B: Full sample
β_d	0.301	0.261
	(0.028)	(0.017)
β_w	0.562	0.642
	(0.052)	(0.030)
β_m	0.286	0.286
	(0.060)	(0.032)
β_0	0.006	-0.023
	(0.019)	(0.013)
σ_ε	0.504	0.543
s	5	5
ln$L$	-1083.43	-3203.90
AIC	2176.86	6417.80
BIC	2203.35	6449.23

	A: In-sample			B: Full sample		
	ARFIMA $(1, d, 0)$	ARFIMAX $(1, d, 0)$	ARMAX (2, 1)	ARFIMA $(1, d, 0)$	ARFIMAX $(1, d, 0)$	ARMAX (2, 1)
ϕ_1	-0.071	-0.163	1.016	-0.114	-0.164	1.035
	(0.032)	(0.031)	(0.052)	(0.016)	(0.017)	(0.030)
ϕ_2			-0.053			-0.054
			(0.046)			(0.028)
d	0.481	0.489		0.498	0.498	
	(0.019)	(0.014)		(0.003)	(0.003)	
α_0	0.005	-0.073	-0.074	-0.353	-0.382	-0.382
	(1.221)	(1.449)	(0.104)	(3.844)	(4.091)	(0.131)
α_1		-0.004	-0.004		-0.040	-0.044
		(0.017)	(0.017)		(0.012)	(0.012)
α_2		-0.155	-0.157		-0.069	-0.064
		(0.029)	(0.029)		(0.020)	(0.020)
θ_1			-0.696			-0.698
			(0.042)			(0.023)
σ_ε	0.504	0.488	0.487	0.541	0.531	0.531
s	4	6	7	4	6	7
ln$L$	-1101.58	-1052.18	-1048.07	-3211.09	-3137.02	-3133.70
AIC	2211.16	2116.36	2110.13	6430.18	6286.05	6281.40
BIC	2232.41	2148.24	2147.33	6455.30	6323.80	6325.44

MSARFIMAX $(1, d, 0)$	A: In-sample	B: Full sample
$\phi_{1,H}$	-0.215	-0.197
	(0.062)	(0.047)
$\phi_{1,L}$	-0.113	-0.226
	(0.080)	(0.029)
d_H	0.557	0.603
	(0.046)	(0.035)
d_L	0.394	0.472
	(0.049)	(0.021)
$\alpha_{0,U}$	2.440	6.065
	(1.318)	(4.708)
$\alpha_{0,L}$	-1.361	-0.939
	(0.350)	(0.319)
α_1	-0.002	-0.045
	(0.016)	(0.012)
α_2	-0.145	-0.055
	(0.033)	(0.023)
$\sigma_{\varepsilon,H}$	0.571	0.706
$\sigma_{\varepsilon,L}$	0.328	0.434
π_{HH}	0.779	0.921
π_{LL}	0.731	0.968
s	12	12
ln$L$	**-1029.72**	***-3036.48***
AIC	**2083.44**	***6096.96***
BIC	**2147.20**	***6172.46***

Notes: The table shows estimation results for the AR(FI)MAX, MSARFIMAX and HAR models, equations (2), (4) and (5), respectively, over the first rolling window from January 6, 1997 to December 27, 2002 (1500 days) and over the full sample from January 6, 1997 to November 16, 2012 (3990 days). Standard errors from the robust HAC covariance matrix are in parentheses. The Akaike Information Criterion is computed as $-2lnL + 2s$ where s is the number of estimated parameters including the residual variance; the Bayesian Information Criterion is computed as $-2lnL + slnT$ where T is the number of effective observations. A negative α_2 indicates that past negative returns have a stronger effect on future realized volatility than past positive returns. Bold (bold italics) denotes the largest log-likelihood, smallest AIC and BIC values attained on the first estimation window (full period).

Table 3. Performance Statistics of Straddle Trading (Filter Rule = 35%, TC = 2%, *h* = 5)

I. Calm period: December 2002-December 2007

	A. Individual forecasts									B. Combined model forecasts and NN forecasts (Unanimity rule)				
	Naïve			Model-based					Nearest neighbor					
	Hist. Ave.	RW (Sign)	EWMA	HAR	ARFIMA	ARFIMAX	ARMAX	MSARFIMAX	NN	HAR + NN	ARFIMA + NN	ARFIMAX + NN	ARMAX + NN	MSARFIMAX + NN
Percentage of buy signals	79.78	10.43	2.95	3.11	2.07	2.39	3.03	2.31	2.39	0.88	0.48	0.48	1.19	0.80
Percentage of sell signals	5.65	9.87	3.42	5.18	5.73	6.05	4.94	6.05	6.53	4.14	4.78	4.94	4.62	4.94
Frequency sells/buys ratio	0.07	0.95	1.16	1.67	2.77	2.53	1.63	2.62	2.73	4.73	10.00	10.33	3.87	6.20
Frequency ratio buys/sells	14.11	1.06	0.86	0.60	0.36	0.39	0.61	0.38	0.37	0.21	0.10	0.10	0.26	0.16
Excess returns														
All trading signals	-588.46	-58.70	-10.67	4.53	10.34	12.68	-4.12	11.89	7.55	14.48	17.89	18.64	9.13	16.85
	(-35.31)	(-5.35)	(-1.61)	(0.66)	(1.67)	(2.04)	(-0.63)	(1.94)	(1.13)	(2.95)	(3.77)	(3.93)	(1.69)	(3.36)
Buy signals	-616.41	-92.16	-27.08	-20.65	-15.69	-15.64	-25.13	-16.07	-18.00	-5.15	-2.29	-2.29	-10.44	-5.23
Sell signals	27.94	33.46	16.41	25.19	26.03	28.33	21.01	27.96	25.56	19.63	20.18	20.93	19.57	22.08
Risk measures														
Annualized volatility	83.03	54.67	33.03	34.14	30.87	30.97	32.39	30.61	33.42	24.49	23.61	23.64	26.98	24.98
Annual. downside volatility (0%)	50.93	66.44	76.07	55.93	52.99	49.99	61.41	52.13	69.30	59.71	48.08	45.88	63.59	48.02
Skewness	0.95	-1.72	-2.39	0.67	1.16	1.50	-1.04	1.11	-0.63	2.37	3.99	3.92	0.31	2.95
Kurtosis	5.16	10.57	37.10	31.29	27.53	26.29	28.94	26.14	29.55	44.34	44.93	44.29	41.69	39.25
99% VaR (Cornish-Fisher)	14.61	15.14	20.70	17.76	12.95	11.26	17.89	12.34	18.56	12.42	4.61	4.75	18.80	8.26
Risk-adjusted performance														
Sharpe ratio	-7.09	-1.07	-0.32	0.13	0.33	**0.41**	-0.13	0.39	0.23	0.59	0.76	**0.79**	0.34	0.67
Annualized alpha	-575.87	-57.56	-10.97	4.45	8.93	**10.36**	-4.94	10.07	6.61	13.66	16.17	**16.87**	8.15	15.15
	(-36.54)	(-5.24)	(-1.65)	(0.65)	(1.44)	**(1.69)**	(-0.76)	(1.65)	(0.98)	(2.78)	(3.45)	**(3.60)**	(1.51)	(3.05)
Sortino ratio (0%)	-5.23	-1.20	-0.39	0.20	0.53	**0.68**	-0.17	0.62	0.32	1.09	1.69	**1.76**	0.52	1.32
Omega ratio (0%)	0.08	0.44	0.65	1.16	1.49	**1.60**	0.86	1.56	1.29	2.52	3.39	**3.50**	1.60	2.69
Annualized Leland's alpha	-577.20	-58.30	-11.60	3.82	8.42	**9.98**	-5.53	9.63	6.07	13.14	15.74	**16.45**	7.62	14.73

(*cont.*)

II. Turbulent period: January 2008 –November 2012

	A. Individual forecasts									B. Combined model forecasts and NN forecasts (Unanimity rule)				
	Naïve			Model-based					Nearest neighbor					
	Hist. Ave.	RW (Sign)	EWMA	HAR	ARFIMA	ARFIMAX	ARMAX	MSARFIMAX	NN	HAR + NN	ARFIMA + NN	ARFIMAX + NN	ARMAX + NN	MSARFIMAX + NN
Percentage of buy signals	40.81	21.71	15.93	16.50	10.98	13.01	10.98	13.17	4.96	2.52	2.36	2.44	3.09	2.60
Percentage of sell signals	28.86	22.20	13.41	17.80	19.59	18.86	16.34	17.72	21.54	13.33	15.93	15.28	13.98	14.47
Frequency sells/buys ratio	0.71	1.02	0.84	1.08	1.79	1.45	1.49	1.35	4.34	5.29	6.76	6.27	4.53	5.56
Frequency buys/sells ratio	1.41	0.98	1.19	0.93	0.56	0.69	0.67	0.74	0.23	0.19	0.15	0.16	0.22	0.18
Excess returns														
All trading signals	-169.28	-98.97	-55.58	-13.71	4.88	-2.88	-0.55	-3.26	46.25	49.58	48.46	49.38	40.66	48.35
	(-8.73)	(-5.97)	(-4.07)	(-0.96)	(0.37)	(-0.21)	(-0.04)	(-0.24)	(3.90)	(5.41)	(5.04)	(5.09)	(4.19)	(5.05)
Buy signals	-300.66	-184.75	-123.18	-115.79	-86.18	-96.33	-81.21	-92.00	-46.87	-23.92	-19.69	-20.25	-26.14	-20.56
Sell signals	131.39	85.78	67.61	102.08	91.06	93.45	80.66	88.74	93.11	73.50	68.14	69.64	66.80	68.91
Risk measures														
Annualized volatility	95.59	81.67	67.25	70.12	64.69	66.31	62.17	65.71	58.39	45.16	47.44	47.80	47.85	47.20
Annual. downside volatility (0%)	60.27	73.28	69.98	63.84	62.01	62.25	67.46	61.25	76.57	66.21	73.87	76.91	77.99	76.74
Skewness	-0.01	-0.69	-0.79	-0.07	-0.13	-0.16	-0.40	-0.02	-0.38	0.94	0.24	0.20	-0.05	0.24
Kurtosis	3.30	5.80	8.54	6.56	8.72	8.22	10.01	8.89	11.93	13.51	20.04	19.74	20.19	20.45
99% VaR (Cornish-Fisher)	15.97	17.91	17.31	14.31	15.25	15.30	16.46	15.41	16.64	10.27	17.83	17.87	18.87	18.04
Risk-adjusted performance														
Sharpe ratio	-1.77	-1.21	-0.83	-0.20	0.08	-0.04	-0.01	-0.05	**0.79**	**1.10**	1.02	1.03	0.85	1.02
Annualized alpha	-169.26	-265.20	-55.55	-12.75	5.75	-1.82	0.34	-2.21	**47.03**	**50.35**	49.30	50.32	41.46	49.23
	(-8.73)	(-6.01)	(-4.07)	(-0.91)	(0.44)	(-0.14)	(0.03)	(-0.17)	**(4.01)**	**(5.59)**	(5.22)	(5.31)	(4.34)	(5.25)
Sortino ratio (0%)	-2.06	-1.40	-1.00	-0.27	0.11	-0.06	-0.01	-0.07	**1.21**	**2.00**	1.77	1.78	1.38	1.77
Omega ratio (0%)	0.50	0.53	0.60	0.90	1.05	0.98	0.99	0.97	**1.70**	**2.61**	2.35	2.40	2.05	2.41
Annualized Leland's alpha	-169.36	-99.55	-55.63	-12.80	5.70	-1.86	0.29	-2.25	**46.98**	**50.29**	49.25	50.27	41.40	49.18

(***cont.***)

III. Full period: January 2008 –November 2012

	A. Individual forecasts									B. Combined model forecasts and NN forecasts (Unanimity rule)				
	Naïve			Model-based					Nearest neighbor					
	Hist. Ave.	RW (Sign)	EWMA	HAR	ARFIMA	ARFIMAX	ARMAX	MSARFIMAX	NN	HAR + NN	ARFIMA + NN	ARFIMAX + NN	ARMAX + NN	MSARFIMAX + NN
Percentage of buy signals	60.50	16.05	9.45	9.73	6.52	7.68	7.00	7.72	3.70	1.69	1.41	1.45	2.13	1.69
Percentage of sell signals	17.14	16.09	8.37	11.46	12.59	12.39	10.58	11.83	13.96	8.69	10.30	10.06	9.25	9.65
Frequency sells/buys ratio	0.28	1.00	0.89	1.18	1.93	1.61	1.51	1.53	3.77	5.14	7.31	6.94	4.34	5.71
Frequency buys/sells ratio	3.53	1.00	1.13	0.85	0.52	0.62	0.66	0.65	0.27	0.19	0.14	0.14	0.23	0.18
Excess returns														
All trading signals	-380.30	-78.85	-33.56	-4.33	7.60	4.94	-2.38	4.36	26.81	31.78	32.94	33.78	24.68	32.37
	(-28.35)	(-7.95)	(-4.44)	(-0.55)	(1.06)	(0.67)	(-0.34)	(0.60)	(3.96)	(6.15)	(6.19)	(6.30)	(4.47)	(6.03)
Buy signals	-459.27	-138.40	-75.22	-67.59	-50.49	-55.48	-52.80	-53.56	-32.05	-14.41	-10.88	-11.16	-18.17	-12.79
Sell signals	78.96	59.55	41.66	63.26	58.09	60.42	50.42	57.91	58.86	46.19	43.82	44.94	42.85	45.16
Risk measures														
Annualized volatility	94.13	69.55	53.07	54.93	50.46	51.52	49.36	51.03	47.48	36.26	37.37	37.60	38.75	37.67
Annual. downside volatility (0%)	54.56	71.24	71.24	62.52	60.61	60.49	66.19	59.91	74.70	64.33	69.63	71.91	74.58	71.00
Skewness	0.44	-1.03	-1.23	-0.04	-0.01	-0.05	-0.54	0.06	-0.37	1.37	0.86	0.80	0.10	0.76
Kurtosis	3.44	7.56	13.99	10.97	13.58	12.88	15.13	13.77	17.08	20.85	29.25	28.87	28.08	28.66
99% VaR (Cornish-Fisher)	15.42	17.13	17.78	14.63	15.22	15.12	16.96	15.39	17.22	10.64	17.49	17.58	19.59	17.62
Risk-adjusted performance														
Sharpe ratio	-4.04	-1.13	-0.63	-0.08	0.15	0.10	-0.05	0.09	**0.56**	0.88	0.88	**0.90**	0.64	0.86
Annualized alpha	-379.65	-78.11	-33.70	-4.71	6.61	3.78	-3.17	3.38	**25.87**	31.16	32.00	**32.77**	23.85	31.43
	(-28.36)	(-7.92)	(-4.46)	(-0.60)	(0.94)	(0.53)	(-0.46)	(0.47)	**(3.90)**	(6.12)	(6.22)	**(6.36)**	(4.43)	(6.05)
Sortino ratio (0%)	-3.86	-1.30	-0.77	-0.11	0.22	0.14	-0.07	0.12	**0.85**	1.60	1.60	**1.62**	1.02	1.53
Omega ratio (0%)	0.22	0.50	0.60	0.95	1.12	1.07	0.96	1.07	**1.58**	2.59	2.53	**2.59**	1.92	2.48
Annualized Leland's alpha	-380.06	-78.61	-34.01	-4.97	6.47	3.67	-3.34	3.25	**25.71**	30.95	31.84	**32.63**	23.68	31.28

Notes: The table shows performance statistics for a straddle trading strategy informed by various out-of-sample forecasts of realized variance: naïve forecasts (historical average, sign-of-volatility change according to the random walk principle, EWMA), model-based forecasts (HAR, ARFIMA, ARFIMAX, ARMAX and MSARFIMAX) and nearest neighbor (NN) forecasts. All forecasts are obtained recursively using 1500-day length rolling windows. The combined forecasts are based on the (directional) unanimity rule. Returns are annualized in percentage. Excess returns are the returns of the strategy in excess of the risk-free rate which is proxied by the 3-month Treasury Bill rate. Shaded areas in Panel A denote the forecasting approaches that yield inferior risk-adjusted profitability than the NN approach individually. Bold areas in Panel A (Panel B) denote the individual (combined) forecasting approach that yields superior performance.

Table 4. Statistical Evaluation of Out of-Sample Realized Variance Forecasts

	A. Volatility-level prediction accuracy								B. Sign-of-volatility-change prediction	
	$1-\frac{MSE_{NN}}{MSE_j}$	$1-\frac{MAPE_{NN}}{MAPE_j}$	$1-\frac{HMSE_{NN}}{HMSE_j}$	$1-\frac{LL_{NN}}{LL_j}$	$1-\frac{QLIKE_{NN}}{QLIKE_j}$	$1-\frac{CRPS_{NN}}{CRPS_j}$	$1-\frac{R^2_{NN}}{R^2_j}$	$\chi^2_{(2)}$	$1-\frac{Hit^+_{NN}}{Hit^+_j}$	$1-\frac{Hit^-_{NN}}{Hit^-_j}$
I. Calm period: December 2002-December 2007										
Historical average	0.719	0.533	0.931	0.775	0.638	0.606	-70.70	170.4 ***	**0.205**	-3.281
"Random walk"	0.205	0.099	0.138	0.218	0.331	0.134	0.007	47.23 ***	-1.244	-0.488
EWMA	-0.054	-0.020	-0.028	-0.023	0.172	-0.050	0.173	11.591 ***	-0.034	0.021
HAR	**-0.311**	**-0.103**	**-0.465**	**-0.288**	**-0.379**	**-0.212**	**0.311**	2.203	0.119	-0.089
ARFIMA	-0.084	-0.001	-0.066	-0.035	-0.072	-0.082	0.135	1.438	0.128	-0.241
ARFIMAX	-0.117	-0.024	*-0.153*	-0.092	-0.137	-0.110	0.169	0.651	0.134	-0.194
ARMAX	-0.151	-0.002	0.041	-0.033	-0.108	-0.101	0.212	3.731	*0.143*	-0.300
MSARFIMAX	*-0.191*	*-0.034*	-0.139	*-0.114*	*-0.182*	*-0.141*	*0.232*	0.651	0.134	-0.176
NN								2.123		
II. Turbulent period: January 2008-December 2012										
Historical average	0.186	0.309	0.807	0.545	0.582	0.369	-0.563	10.53 ***	0.213	-0.395
"Random walk"	0.090	0.023	0.550	0.046	-0.133	0.125	0.417	39.04 ***	-0.366	-0.829
EWMA	-0.165	-0.044	0.300	-0.166	-0.581	-0.088	0.532	14.57 ***	0.303	-0.264
HAR	**-0.647**	**-0.200**	**-0.509**	**-0.657**	**-1.310**	**-0.332**	**0.673**	7.287 **	**0.361**	-0.205
ARFIMA	-0.220	-0.064	-0.033	-0.222	-0.491	-0.114	0.538	10.04 ***	0.313	-0.337
ARFIMAX	-0.286	-0.080	-0.101	-0.270	-0.546	-0.138	0.588	11.11 ***	0.320	-0.307
ARMAX	-0.418	*-0.098*	*-0.079*	-0.315	-0.677	-0.182	0.599	9.241 **	0.317	-0.249
MSARFIMAX	*-0.472*	-0.093	-0.003	*-0.316*	*-0.756*	*-0.213*	*0.642*	10.55 ***	*0.331*	-0.267
NN								19.26 ***		
III. Full period: January 2002-December 2012										
Historical average	0.206	0.432	0.889	0.667	0.599	0.426	-2.78	27.44 ***	0.208	-1.057
"Random walk"	0.092	0.058	0.440	0.116	0.046	0.127	0.355	46.35 ***	-0.761	-0.649
EWMA	-0.163	-0.033	0.172	-0.110	-0.357	-0.082	0.477	17.535 ***	0.143	-0.110
HAR	**-0.641**	**-0.154**	**-0.491**	**-0.502**	**-0.954**	**-0.310**	**0.621**	4.979 *	**0.242**	-0.147
ARFIMA	-0.218	-0.035	-0.046	-0.148	-0.349	-0.109	0.485	7.972 *	0.218	-0.289
ARFIMAX	-0.284	-0.054	-0.122	-0.200	-0.410	-0.133	0.534	8.756 *	0.225	-0.251
ARMAX	-0.414	-0.052	-0.026	-0.197	-0.473	-0.167	0.545	6.058 *	0.227	-0.273
MSARFIMAX	*-0.467*	*-0.065*	*-0.055*	*-0.235*	*-0.553*	*-0.200*	*0.589*	8.323 *	*0.231*	-0.222
NN								16.33 ***		

Notes: The table shows statistical accuracy ratios for NN predictions versus alternative predictions (denoted *j*) of realized variance: mean square error (*MSE*), adjusted mean absolute percentage error (*MAPE*) or mean of the absolute ratio of forecast error to the sum of observation and forecast, heteroskedasticity-adjusted mean square error (*HMSE*), logarithmic loss (*LL*), quasi Gaussian log-likelihood (*QLIKE*), continuous rank probability score (*CRSP*), R^2 of Mincer-Zarnowitz regressions $RV_{t+h} = b_0 + b_1\widehat{RV}^m_{t+h|t} + v_t$, and the Wald $\chi^2_{(2)}$ test statistic for the forecast unbiasedness hypothesis H_0: $b_0 = 0 \cup b_1 = 1$; ***, ** and * denotes significance at the 1%, 5% and 10% levels. Hit^+ is the ratio of correct positive directional change predictions to actual positive directional changes, Hit^- is defined similarly for negative changes. All forecasts are computed 5-days-ahead recursively using 1500-day rolling estimation windows. “Random walk” forecasts are defined as the last observed daily realized variance (in Panel A) and directional volatility forecasts as the sign of the last observed 5-day change in realized variance (in Panel B). Shaded area means that the NN forecasts outperform the competing forecasts; bold (italic) font denotes the (second) best forecasting approach among those that outperform the NN scheme.