Expert Systems With Applications 55 (2016) 222–242

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Combining high frequency data with non-linear models for forecasting energy market volatility☆

Jozef Baruník[a,b,*], Tomáš Křehlík[a,b]

[a] Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Opletalova 26, 110 00, Czech Republic
[b] Institute of Information Theory and Automation, The Czech Academy of Sciences, Pod Vodarenskou Vezi 4, 182 00, Prague, Czech Republic

ARTICLE INFO

JEL Classification:
C14
C53
G17

Keywords:
Artificial neural networks
Realized volatility
Multiple-step-ahead forecasts
Energy markets

ABSTRACT

The popularity of realized measures and various linear models for volatility forecasting has been the focus of attention in the literature addressing energy markets' price variability over the past decade. However, there are no studies to help practitioners achieve optimal forecasting accuracy by guiding them to a specific estimator and model. This paper contributes to this literature in two ways. First, to capture the complex patterns hidden in linear models commonly used to forecast realized volatility, we propose a novel framework that couples realized measures with generalized regression based on artificial neural networks. Our second contribution is to comprehensively evaluate multiple-step-ahead volatility forecasts of energy markets using several popular high frequency measures and forecasting models. We compare forecasting performance across models and across realized measures of crude oil, heating oil, and natural gas volatility during three qualitatively distinct periods: the pre-crisis period, the 2008 global financial crisis, and the post-crisis period. We conclude that the newly proposed approach yields both statistical and economic gains, while reducing the tendency to over-predict volatility uniformly during all the tested periods. In addition, the proposed methodology is robust to a substantial structural break induced by the recent financial crisis. Our analysis favors median realized volatility because it delivers the best performance and is a computationally simple alternative for practitioners.

1. Introduction

Predicting energy price variability has become one of the most significant issues faced by the natural gas industry and energy companies in recent decades. With their considerable volatility, the leading products of energy markets, i.e., crude oil, natural gas, and heating oil,[1] contributed to a climate of uncertainty and distrust of energy companies and investors, on one hand, and of consumers, regulators, and legislators, on the other. The high level of volatility in energy markets is likely due to supply uncertainty—such as from a variety of macroeconomic and political factors in the case of crude oil or simply storage constraints in the case of natural gas—and short-term inelasticity of demand, i.e., the difficulty of reducing consumption within a short period of time. The combination of these two factors makes it extremely difficult for both consumers and producers to forecast their costs and profits. The desire to protect market participants against the losses resulting from this unpredictability has led to immense interest in empirical research aiming to predict the variability in energy prices. In this paper, we contribute to this literature by proposing a novel framework to forecast energy commodity volatility that couples realized measures with generalized regression based on artificial neural networks. We demonstrate that our approach delivers precise forecasts even in the regime-switching moment of financial crisis.

Volatility research from previous decades is influenced mainly by the work of Engle (1982) and Bollerslev (1986, 1987) and has shown that price variability is much easier to understand than it is to forecast the direction of future price changes. However, the lion's share of previous research has focused on the financial markets, and the focus has only recently turned to the energy

☆ Support from the Czech Science Foundation under project no. P402/12/G097 DYME – "Dynamic Models in Economics" is gratefully acknowledged. Křehlík gratefully acknowledges financial support from the Grant Agency of Charles University under projects 588314 and 837413.

* Corresponding author at: Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Opletalova 26, 110 00, Czech Republic. Tel.: +420 776 259273.

E-mail addresses: barunik@fsv.cuni.cz (J. Baruník), tomas.krehlik@gmail.com (T. Křehlík).

[1] According to the CME Group Leading Products Resource, crude oil, natural gas, and heating oil futures are traded with the highest average volume among energy commodities (http://www.cmegroup.com/education/featured-reports/cme-group-leading-products.html).

http://dx.doi.org/10.1016/j.eswa.2016.02.008
0957-4174/

markets.[2] (Kang & Yoon, 2013; Kuper & van Soest, 2006; Linn & Zhu, 2004; Mohammadi & Su, 2010; Pindyck, 2004; Sévi, 2014; Wei, Wang, & Huang, 2010; Wilson, Aggarwal, & Inclan, 1996; Yang, Hwang, & Huang, 2002).

More recent advances in financial econometrics have led to the development of new estimators of volatility using high frequency data that make volatility observable. Although the pioneering studies in the realized volatility literature recognize the benefits of using high frequency data in terms of increased accuracy (Merton, 1980; Zhou, 1996), subsequent research[3] proposes several estimators that improve model efficiency, robustness to market microstructure effects, and the ability to separately estimate the variation due to the continuous part of the price process, on one hand, and the variation due to the jump part of the price process, on the other. For excellent reviews of the realized volatility literature, see Andersen, Bollerslev, Christoffersen, and Diebold (2006); McAleer and Medeiros (2008), or Barndorff-Nielsen and Shephard (2007). Moreover, recent studies utilize high frequency data in energy markets (Baum & Zerilli, 2016; Prokopczuk, Symeonidis, & Wese Simen, 2015).

However, estimating realized volatility is only the first step to more accurate predictions and using the appropriate model is the second step. Heterogeneous autoregressive (HAR) and autoregressive fractionally integrated (ARFIMA) models became widely used to forecast realized volatility because these models effectively capture the long memory of volatility (Andersen, Bollerslev, Diebold, & Labys, 2003; Corsi, 2009). In contrast to FIGARCH models that capture the long memory of volatility using daily returns data,[4] these approaches are more flexible and easier to estimate when high frequency data are available. Although both the HAR and the ARFIMA have been developed to capture the specific long memory feature of volatility, more complex patterns may be revealed and explored. Changes in market conditions and many types of noises induced by measurement error lead to non-linear patterns that cannot be captured by linear models that are based on restrictive distributional assumptions. Microstructure noise that can arise through the bid-ask bounce, asynchronous trading, infrequent trading or price discreteness are important examples of measurement error.

Artificial neural networks (ANN) may be understood as a generalization of these classical approaches that may help to uncover more complex volatility patterns. Concisely, neural networks are semi-parametric non-linear models that can approximate any reasonable function (Haykin, 2007; Hornik, Stinchcombe, & White, 1989). The number of models using machine learning is rapidly growing in the academic literature but applications that apply these models in energy market in energy markets are limited. Among the few that do, Fan, Liang, and Wei (2008) proposes a generalized pattern matching based on a genetic algorithm to predict crude oil prices on a multi-step-ahead basis. Xiong, Bao, and Hu (2013); Yu, Wang, and Lai (2008) proposes an empirical model that decomposes neural networks to forecast crude oil prices. Jammazi and Aloui (2012) uses a hybrid model for crude oil forecasting, Panella, Barcellona, and D'Ecclesia (2012) use a mixture of Gaussian neural network to forecast energy commodity prices, and Papadimitriou, Gogas, and Stathakis (2014) investigates the efficiency of support vector machines in forecasting next-day electricity prices. Although the focus has remained solely on forecasting prices, research using neural networks to forecast volatility continues to be developed.

This paper's primary contribution is that it proposes a model that couples measures of volatility from high frequency data with artificial neural networks to reliably forecast energy price volatility. Whereas researchers in financial econometrics have performed pioneering work using stock market index data (McAleer & Medeiros, 2011) or exchange rate data (Sermpinis, Theofilatos, Karathanasopoulos, Georgopoulos, & Dunis, 2013), we are the first to comprehensively test this strategy against competing models in the energy literature. Rather than choosing from among the plethora of advanced machine learning algorithms, we use the simplest and most popular feed-forward neural network as the first step in this field. Our main motivation is to show whether there are statistical and economic gains that can be realized by coupling high frequency data with easy-to-implement artificial neural networks.

This paper also contributes to the literature by comprehensively evaluating the most popular models and realized measures. These realized volatility measures rely on different assumptions, and there are no studies guiding practitioners to use a specific measure when working with the volatility forecasting of energy markets. To bridge this gap, we focus on the three most liquid energy commodities—crude oil, heating oil, and natural gas—during the period from January 5, 2004 to December 31, 2012 and put the models into a horse race through several discrete sub-periods to determine which model produces uniformly lower errors in multiple-step-ahead volatility forecasts. The period under study is especially interesting because it includes a sub-period of high and rapidly rising prices, a sub-period encapsulating the interruption of price increases in 2008 due to global turmoil in the financial markets, and the last sub-period that witnessed profound regime change over the most recent few years in which price variability became much calmer. In particular, the last period is particularly interesting from the forecaster's perspective, as it appears that demand for liquid transport fuels has peaked in the developed economies with car engines becoming more efficient and amid partial substitution by biofuels. On the supply side, high prices reversed the previous trend toward growing dependence on the conventional oil fields of the OPEC member states. Sophisticated modeling strategies should reflect these changes.

We test the ANN against widely used the HAR and ARFIMA long-memory models and a benchmark low frequency-based GARCH model. The tests are performed within the recently proposed frameworks of the Model Confidence Set (MCS) developed by Hansen, Lunde, and Nason (2011) and Superior Predictive Ability (SPA) developed by Hansen (2005) with several popular loss functions used in the literature. Moreover, we use realized variance (RV), realized kernel (RK), two-scale realized variance (TSRV), bipower variation (BV), median realized volatility (MedRV), and the recently proposed jump-adjusted wavelet two-scale realized variance (JWTSRV) as measures of volatility. Motivated by the possibility of reducing model uncertainty, we also experiment with the linear combination of forecasts from the popular HAR model and artificial neural network. This experiment yields the lowest error uniformly through all tested periods regardless of which realized measure is used. These low error levels also translate to economic benefits in terms of Value-at-Risk. One of the loss functions we use in the exercise allows us to assess whether the models tend to over-predict volatility as commonly found using GARCH-type models[5]. A uniform finding is that coupling neural networks with high frequency data results in substantial reductions in the

[2] For a complete review of GARCH-type models used in the energy literature, see Wang and Wu (2012).

[3] Andersen and Bollerslev (1998); Andersen, Bollerslev, Diebold, and Labys (2001, 2003); Bandi and Russell (2006); Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008); Hansen and Lunde (2006); Zhang, Mykland, and Aït-Sahalia (2005).

[4] Kang and Yoon (2013) recently investigate the ability of FIGARCH models to capture energy market volatility.

[5] For example, see Nomikos and Pouliasis (2011), who confirm the strong tendency of GARCH-type models to over-predict the volatility of crude oil, heating oil, and gasoline, which is further confirmed by Wang and Wu (2012), who find that multivariate GARCH-type models also suffer from over-predictions.

over-estimating tendency compared with previous studies. In addition, we find that MedRV delivers the best forecasts of the other measures. As a computationally simple alternative to other measures, we prefer the MedRV for forecasting energy volatility.

The remainder of this study is organized as follows. Section 2 describes the realized measure used in this study. Section 3 presents prediction models including the HAR, ARFIMA, and ANN models. Section 4 presents the data and discusses the research setup, including the methodology used for the statistical and economic forecast evaluations. Section 5 discusses the results, and, finally, Section 6 concludes. Notably, the number of results produced by this research setup is quite large, and the results that use different loss functions are overlapping; therefore, we relegate our auxiliary results to the online supplementary appendix that is available at 10.1016/j.eswa.2016.02.008.

2. Estimation of realized volatility

In this analysis, we assume that the latent logarithmic commodity price follows a standard jump-diffusion process contaminated by microstructure noise. Let $y_t = p_t + \epsilon_t$ be the observed logarithmic prices evolving over $0 \leq t \leq T$, which will have two components; the latent, the so-called *true log-price process* $dp_t = \mu_t dt + \sigma_t dW_t + \xi_t dq_t$, and zero mean *i.i.d.* microstructure noise, ϵ_t, with variance η^2. In a latent process, q_t is a Poisson process that is uncorrelated with W_t, and the magnitude of the jump, denoted as J_l, is controlled by factor $\xi_t \sim N(\bar{\xi}, \sigma_\xi^2)$.

The quadratic return variation over the interval $[t-h, t]$, for $0 \leq h \leq t \leq T$ that is associated with the price process y_t may be naturally decomposed into two parts: integrated variance and jump variation

$$QV_{t,h} = \underbrace{\int_{t-h}^{t} \sigma_s^2 ds}_{IV_{t,h}} + \underbrace{\sum_{t-h \leq l \leq t} J_l^2}_{JV_{t,h}} . \tag{1}$$

As detailed by Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002a), quadratic variation is a natural measure of variability in the logarithmic price process. A simple consistent estimator of the overall quadratic variation under the assumption that there is zero noise contamination in the price process is provided by the well-known realized variance developed by Andersen and Bollerslev (1998). The realized variance over $[t-h, t]$ may be estimated as

$$\widehat{QV}_{t,h}^{(RV)} = \sum_{k=1}^{N} (\Delta_k y_t)^2 , \tag{2}$$

where $\Delta_k y_t = y_{t-h+\left(\frac{k}{N}\right)h} - y_{t-h+\left(\frac{k-1}{N}\right)h}$ is the kth intraday return in the $[t-h, t]$ interval, and N is the number of intraday observations. The estimator in Eq. (2) converges in probability to $IV_{t,h} + JV_{t,h}$ as $N \to \infty$ (Andersen & Bollerslev, 1998; Andersen et al., 2001, 2003; Barndorff-Nielsen & Shephard, 2001, 2002a, 2002b).

Because observed price process y_t is contaminated with noise and jumps in real data, we must account for this feature as the main object of interest is the $IV_{t,h}$ part of quadratic variation. Zhang et al. (2005) propose a solution to noise contamination by introducing the two-scale realized volatility (TSRV) estimator. The authors adopt a methodology to estimate the quadratic variation utilizing all the available data with an idea of precise bias estimation. The two-scale realized variation over $[t-h, t]$ is measured by

$$\widehat{QV}_{t,h}^{(TSRV)} = \widehat{QV}_{t,h}^{(average)} - \frac{\bar{N}}{N} \widehat{QV}_{t,h}^{(all)} , \tag{3}$$

where $\widehat{QV}_{t,h}^{(all)}$ is computed as in Eq. (2) on all available data and $\widehat{QV}_{t,h}^{(average)}$ is constructed by averaging the estimators $\widehat{QV}_{t,h}^{(g)}$ obtained on G grids of average size $\bar{N} = N/G$ as $\widehat{QV}_{t,h}^{(average)} = \frac{1}{G}\sum_{g=1}^{G} \widehat{QV}_{t,h}^{(g)}$, where the original grid of observation times, $M = \{t_1, \ldots, t_N\}$ is subsampled to $M^{(g)}$, $g = 1, \ldots, G$, where $N/G \to \infty$ as $N \to \infty$. The estimator in Eq. (3) is the first consistent and asymptotic estimator of the quadratic variation of p_t. Zhang et al. (2005) also formulate the theory for the optimal choice of G grids, $G^* = cN^{2/3}$, where the constant c may be set to minimize total asymptotic variance.

A different approach to addressing noise developed by Barndorff-Nielsen et al. (2008) is realized kernels. The realized kernel variance estimator over $[t-h, t]$ is defined by

$$\widehat{QV}_{t,h}^{(RK)} = \gamma_0 + \sum_{\eta=1}^{H} K\left(\frac{\eta - 1}{H}\right)(\gamma_\eta + \gamma_{-\eta}), \tag{4}$$

with $\gamma_\eta = \sum_{k=1}^{N} \Delta_k y_t \Delta_{k-\eta} y_t$ denoting the ηth realized autocovariance with $\eta = -H, \ldots, -1, 0, 1, \ldots, H,$ and $K(.)$ denoting the kernel function. Notably, for $\eta = 0$, $\gamma_\eta = \gamma_0 = \widehat{QV}_{t,h}^{(RV)}$ is an estimate of the realized variance from Eq. (2). For the estimator to work, we must choose the kernel function $K(\,\cdot\,)$. In our study, we focus on the Parzen kernel because it satisfies the smoothness conditions, $K'(0) = K'(1) = 0$, and is guaranteed to produce a non-negative estimate. We should stress that the realized kernel estimator is computed without accounting for end effects, i.e., by replacing the first and last observations with local averages to eliminate the corresponding noise components (so-called *jittering*). Barndorff-Nielsen et al. (2008) argue that these effects are important theoretically but negligible practically.

When studying conditional volatility, it is important to separate the contributions of the two components of the quadratic variation, i.e., the continuous component from the jump component. Recent evidence from the volatility forecasting literature indicates that the two sources of variation in the price have substantially different time series properties and affect future volatility differently. Although we are mainly interested in forecasting integrated variance, we also estimate jumps in the data. Barndorff-Nielsen and Shephard (2004, 2006) develop a bipower variation estimator that may detect the presence of jumps in high frequency data. The main idea of the estimator is to compare two measures in the integrated variance, one containing the jump variation and the other that is robust to jumps and thus containing only the integrated variation component. In our study, we use the Andersen, Bollerslev, and Huang (2011) adjustment of the original (Barndorff-Nielsen & Shephard, 2004) estimator, which helps it become robust to certain types of microstructure noise. The bipower variation over $[t-h, t]$ is defined by

$$\widehat{IV}_{t,h}^{(BV)} = \mu_1^{-2} \frac{N}{N-2} \sum_{k=3}^{N} |\Delta_{k-2} y_t| \cdot |\Delta_k y_t|, \tag{5}$$

where $\mu_a = \pi/2 = E(|Z|^a)$, and $Z \sim N(0, 1)$, $a \geq 0$ and $\widehat{IV}_{t,h}^{(BV)} \to \int_{t-h}^{t} \sigma_s^2 ds$. Therefore, $\widehat{IV}_{t,h}^{(BV)}$ provides a consistent estimator of the integrated variance. Because $\widehat{QV}_{t,h}^{(RV)}$ provides a consistent estimator of the integrated variance plus the jump variation, the jump variation component may be estimated consistently as the difference between realized variance and realized bipower variation as $\text{plim}_{N\to\infty}(\widehat{QV}_{t,h}^{(RV)} - \widehat{IV}_{t,h}^{(BV)}) = JV_{t,h}$.

Under the assumption of no jump and certain other regularity conditions, Barndorff-Nielsen and Shephard (2006) provide the joint asymptotic distribution of the jump variation. This theory can be used to measure the contribution of the jump variation by

$$JV_{t,h} = I_{\{Z_{t,h} > \Phi_\alpha\}}\left(\widehat{QV}_{t,h}^{(RV)} - \widehat{IV}_{t,h}^{(BV)}\right), \tag{6}$$

where $I_{\{Z_{t,h}>\Phi_\alpha\}}$ denotes the indicator function and Φ_α refers to the chosen critical value from the standard normal distribution. The measure of integrated variance is defined as

$$\widehat{IV}^{(CBV)}_{t,h} = I_{\{Z_{t,h}\leq\Phi_\alpha\}}\widehat{QV}^{(RV)}_{t,h} + I_{\{Z_{t,h}>\Phi_\alpha\}}\widehat{IV}^{(BV)}_{t,h}, \quad (7)$$

which ensures that the jump measure and the continuous component add up to the estimated variance with no jumps.

To estimate the integrated volatility in the presence of jumps, we employ an additional estimator, the median realized volatility, developed by Andersen, Dobrev, and Schaumburg (2012):

$$\widehat{IV}^{(MedRV)}_{t,h} = \frac{\pi}{6-4\sqrt{3}+\pi}\left(\frac{N}{N-2}\right) \times \sum_{k=3}^{N} \text{med}(|\Delta_{k-2}y_t|, |\Delta_{k-1}y_t|, |\Delta_k y_t|)^2. \quad (8)$$

The median realized volatility offers a number of advantages over alternative measures of integrated variance in the presence of infrequent jumps. It is less sensitive to the presence of occasional zero intraday returns and enjoys smaller finite-sample bias induced by jumps, while it is also computationally simple to implement.

2.1. Estimation of quadratic variation using wavelets

Fan and Wang (2007) employ a different approach to measuring realized volatility by using wavelets to separate jump variation from the price process and to estimate the integrated variance on the jump-adjusted data. Although we use the wavelet-based estimator as one of our six realized measures, we do not discuss the details of the wavelet theory in this study and instead direct interested readers to the literature.

Assume that the sample path of the price process y_t has a finite number of jumps. Following the results of Wang (1995) regarding the wavelet jump detection of the deterministic functions with *i.i.d.* additive noise, we use a special form of a discrete wavelet transform, the maximal overlap discrete wavelet transform (MODWT) that is not restricted to a dyadic sample length, unlike the ordinary discrete wavelet transform. Jump locations are detected by the first-level wavelet coefficients obtained on the process y_t over $[t-h, t]$, $\mathcal{W}_{1,k}$. Because we use the MODWT, we have k wavelet coefficients at the first scale, which corresponds to the number of intraday observations, i.e., $k = 1, \dots, N$. If the value of the wavelet coefficient $\mathcal{W}_{1,k}$ is greater[6] than the universal threshold $d\sqrt{2\log N}$ (Donoho & Johnstone, 1994), then a jump of size $\Delta_k J_t$ is detected as

$$\Delta_k J_t = \left(y_{t-h+\left(\frac{k}{N}\right)h} - y_{t-h+\left(\frac{k-1}{N}\right)h}\right) I_{\left\{|\mathcal{W}_{1,k}|>d\sqrt{2\log N}\right\}} \qquad k\in[1,N], \quad (9)$$

where $d = \sqrt{2}\text{med}\{|\mathcal{W}_{1,k}|\}/0.6745$ for $k \in [1, N]$ denotes the intraday median absolute deviation estimator (Percival & Walden, 2000).

Following Fan and Wang (2007), the jump variation over $[t-h, t]$ in discrete time is estimated as the sum of the squares of all the estimated jump sizes, $\widehat{JV}_{t,h} = \sum_{k=1}^{N}(\Delta_k J_t)^2$.

Now that we have precisely detected the jumps, we proceed to the jump adjustment of the observed price process, y_t. We adjust the data for jumps by subtracting intraday jumps from the price process as follows:

$$\Delta_k y_t^{(J)} = \Delta_k y_t - \Delta_k J_t, \qquad k = 1, \dots, N, \quad (10)$$

where N is the number of intraday observations. Finally, volatility may be computed using the jump-adjusted wavelet two-scale realized variance (JWTSRV) estimator on the jump-adjusted data $\Delta_k y_t^{(J)}$. The JWTSRV is an estimator that can estimate integrated variance from the process under the assumption of data containing both noise and jumps. The estimator utilizes the TSRV approach of Zhang et al. (2005) as well as the wavelet jump detection method. Another of the estimator's advantages is that it decomposes the integrated variance into $J^m + 1$ components; therefore, we can study the dynamics of volatility at various investment horizons. Following Barunik, Krehlik, and Vacha (2016); Barunik and Vacha (2015), we define the JWTSRV estimator over $[t-h, t]$ on the jump-adjusted data as follows:

$$\widehat{IV}^{(JWTSRV)}_{t,h} = \sum_{j=1}^{J^m+1}\widehat{IV}^{(JWTSRV)}_{j,t,h} = \sum_{j=1}^{J^m+1}\left(\widehat{IV}^{(average)}_{j,t,h} - \frac{\bar{N}}{N}\widehat{IV}^{(all)}_{j,t,h}\right), \quad (11)$$

where $\widehat{IV}^{(average)}_{j,t,h} = \frac{1}{G}\sum_{g=1}^{G}\sum_{k=1}^{N}\left(\mathcal{W}^{(g)}_{j,k}\right)^2$ is obtained from wavelet coefficient estimates on a grid of size $\bar{N} = N/G$, and $\widehat{IV}^{(all)}_{j,t,h} = \sum_{k=1}^{N}\left(\mathcal{W}_{j,k}\right)^2$ is the wavelet realized variance estimator at a scale of j on all the jump-adjusted observed data, $\Delta_k y_t^{(J)}$. $\mathcal{W}_{j,k}$ denotes the MODWT wavelet coefficient at scale j with position k obtained on process y_t over $[t-h, t]$.

Barunik and Vacha (2015) show that the JWTSRV is a consistent estimator of the integrated variance because it converges in probability to the integrated variance of process p_t, and they assess the estimator's small sample performance in a large Monte Carlo study. The JWTSRV is found to be able to precisely recover true integrated variance from the noisy process with jumps. Moreover, the JWTSRV estimator is also tested in a forecasting exercise, which has been found to substantially improve the forecasting of the integrated variance (Barunik et al., 2016).

3. Prediction models

Well-documented evidence for the strong temporal dependence of realized volatility suggests that realized volatility should be modeled using an approach allowing for a slowly decaying autocorrelation function and possibly long memory. Arneodo, Muzy, and Sornette (1998); Müller et al. (1997) and Lynch and Zumbach (2003) show that volatility over long time intervals has a strong influence on volatility at shorter time intervals but that volatility over short time intervals has no effect on longer intervals. A possible economic interpretation is that long-term volatility matters to short-term traders, whereas short-term volatility has no effect on long-term trading strategies.

Standard, ARCH-type volatility models of Bollerslev (1986); 1987); Engle (1982) and one-factor stochastic volatility models treat volatility as a latent variable and do not capture long memory. In our study, we use realized volatility as ex-post observed variance, and we consider benchmark models for forecasting volatility by capturing its properties to assess the relative performance of the artificial neural network. We compare the forecasts from neural networks to the heterogeneous autoregressive (HAR) model of Corsi (2009) and an autoregressive fractionally integrated moving average (ARFIMA) model that is briefly described in this section.

3.1. The linear heterogeneous autoregressive (HAR) model

A simple and popular model for forecasting realized volatility is the heterogeneous autoregressive model (HAR) developed by Corsi (2009) that is based on heterogeneous realized volatility

[6] Using the MODWT filters, we must correct the position of the wavelet coefficients just slightly to obtain the precise jump position; see Percival and Mofjeld (1997).

components

$$\nu_{t+1} = \alpha + \beta_D \nu_t + \beta_W \nu_{t,t-5} + \beta_M \nu_{t,t-22} + \epsilon_{t+1}, \tag{12}$$

where $\nu_{t,t-k} = \frac{1}{k}\sum_{l=0}^{k-1} \nu_{t-j}$ is the average ν_t over the past k days; where $\nu_{t,\,h}$ is chosen from the estimated quadratic variation or its components, $\sqrt{\widehat{QV}^{(est)}_{t,h}}$, and $\sqrt{\widehat{IV}^{(est)}_{t,h}}$; and where (est) are the RV, RK, TRSV, CBV, MedRV, and JWTSRV measures.

3.2. Long-memory autoregressive fractionally integrated moving average (ARFIMA)

Although the HAR model is popular because of its simplicity, it is an approximate long-memory model and might not be able to capture the dynamics of long memory properties in volatility particularly well, as a result. Therefore, in our forecasting exercise, we follow Andersen et al. (2003) and adopt the autoregressive fractionally integrated moving average (ARFIMA) class of models.

If we assume that the volatility series belong to the class of ARFIMA processes of Granger and Joyeux (1980), then the dth difference of each series is a stationary and invertible ARMA process in which, to ensure stationarity and invertibility, parameter d may be any real number such that $-1/2 < d < 1/2$. More precisely, ν_t is an ARFIMA(p, d, q) process if it follows:

$$\alpha(L)(1-L)^d(\nu_t - \mu) = \beta(L)\nu_t, \tag{13}$$

where $\alpha(z) = 1 - \alpha_1 z - \cdots - \alpha_p z^p$ and $\beta(z) = 1 + \beta_1 z + \cdots + \beta_q z^q$ are polynomials of order p and q, respectively, in the lag operator L, which is rooted strictly outside the unit circle, ν_t is *iid* with zero mean and σ_ν^2 variance, and $(1-L)^d$ is defined by its binomial expansion. The model is estimated using a maximum likelihood method, and forecasting is performed by extrapolating the estimated model. Andersen et al. (2003); Deo, Hurvich, and Lu (2006) show that forecasting log realized volatility based on a simple ARFIMA(1, d, 0) specification performs well in comparison with other time-series methods of forecasting realized volatility. We estimate a simple ARFIMA(1, d, 0).

3.3. Artificial neural networks for predicting volatility

Both the HAR and ARFIMA models are developed to capture specific features of the time-series and are suitable to model volatility because they can capture long memory. However, these models require strict assumptions regarding distributions and simple linear structure that are never met by data. As a result of the substantial noise caused by changes in market conditions, volatility is a complex non-linear process, and more general methodologies are thus required to properly capture the dependence structures. Artificial neural networks may be viewed as a generalization of these classical approaches, which allows us to model other types of non-linearities in the data in addition to long memory. Specifically, neural networks are semi-parametric non-linear models that can approximate any reasonable function (Haykin, 2007; Hornik et al., 1989) that does not require strong distributional assumptions.

We use the neural network as a generalized nonlinear regression that can describe the complex patterns in volatility time series. As with linear or nonlinear methods, a neural network relates a set of input variables, such as lags of volatility to output, in the forecast. The only difference between network and other models is that the approximating function uses one or more so-called hidden layers, in which the input variables are squashed or transformed by a special function.

The most widely used artificial neural network in financial applications with one hidden layer (Hornik et al., 1989) is the feed-forward neural network. The general feed-forward or multilayered perception (MLP) network that we use for volatility ν_t forecasting may be described by as follows:

$$n_{k,t} = \omega_{k,\alpha} + \sum_{i=0}^{21} \omega_{k,i}\nu_{t-i} \tag{14}$$

$$\nu_{t+h} = \gamma_0 + \sum_{k=1}^{k^*} \gamma_k \Lambda(n_{k,t}) \tag{15}$$

where $\Lambda(n_{k,t}) = 1/(1+e^{-n_{k,t}})$ is a logistic function that introduces nonlinearity into the model. To make the model comparable to the HAR model, we use 22 lags of volatility ν_t as input variables and k^* neurons $n_{k,\,t}$. $\omega_{k,\,i}$ represents a coefficient vector to be found. The variable $n_{k,\,t}$ is squashed by the logistic function and becomes a "neuron" $\Lambda(n_{k,\,t})$. Next, the set of k^* neurons are combined linearly with the vector of coefficients $\{\gamma_k\}_{k=1}^{k^*}$ to form the final output, which is the volatility forecast ν_{t+h}. This model is the workhorse of the neural network modeling approach in finance, and almost all researchers begin with this network as their first alternative to linear models.

Notably, the HAR and ARFIMA are simple special cases within this framework if transformation $\Lambda(n_{k,\,t})$ is skipped and one neuron that contains a linear approximation function is used. Therefore, in addition to classical linear models, there are neurons that process inputs to improve the predictions.

To approximate the target function, the neural network must be able to learn. The process of learning is defined as the adjustment of weights using a learning algorithm. The main goal of the learning process is to minimize the sum of the prediction errors in all the training examples. The training phase is thus an unconstrained nonlinear optimization problem in which the goal is to find the optimal set of weights of the parameters by solving the minimization problem $\min\{\Psi(\omega) : \omega \in \mathbb{R}^n\}$, where $\Psi : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable error function. There are several ways to minimize $\Psi(\omega)$, but we are essentially searching for the gradient $G = \nabla\Psi(\omega)$ of function Ψ, which is the vector of the first partial derivatives of the error function $\Psi(\omega)$ with respect to the weight vector ω. Furthermore, the gradient specifies a direction that produces the steepest increase in Ψ. The negative of this vector thus reveals the direction of the steepest decrease.

Nevertheless, traditional gradient descent algorithms often fail to learn intricate patterns in the data efficiently because of the multitude of possible initial settings. An efficient method for learning the patterns in feed-forward neural networks, which we use, is the resilient propagation algorithm (Riedmiller & Braun, 1993). This algorithm differs from the previous one by concentrating solely on the sign of gradients rather than on the overall numerical estimate, which might be imprecise in many cases. This simple idea brings more stability and a higher convergence speed than plain backpropagation or quickpropagation algorithms. Here, the possibility of capturing the complex nature of the data and the possible nonlinearities comes with the cost of higher computational burden.

The best ANN model is chosen from a set of models with either 7 or 15 hidden neurons (to determine whether the amount of neurons in the hidden layer help to process the information better) and decay either at 0 (without decay) or $1e^{-10}$ (standard decay used in the literature). To prevent overfitting, we use cross-validation over time with a fixed window. The best model is chosen based on the cross-validation scheme.

4. Data description and research design

The data set consists of transaction prices for crude oil, heating oil, and natural gas traded on the New York Mercantile Exchange

Table 1
Descriptive statistics for the volatility estimates of crude oil, heating oil and natural gas for the sample period from July, 6, 2006 through December 31, 2012. Minimum, maximum, standard deviation and mean are multiplied by $\times 10^2$ for convenience. LB(l) is Ljung–Box statistics with l lag.

Asset	Estimator	N	Min.	Max.	Std.	Mean	Ex. Kurt.	Skew.	LB(5)	LB(20)
Crude oil	TSRV	1631	0.44	6.26	0.82	1.62	6.29	2.27	5038	18043
	RV	1631	0.48	6.29	0.81	1.62	6.14	2.24	4980	17879
	RK	1631	0.44	6.83	0.84	1.62	6.69	2.31	4790	17201
	JWTSRV	1631	0.45	6.07	0.79	1.58	6.03	2.24	5356	19070
	CBPV	1631	0.47	6.29	0.81	1.59	6.38	2.28	5047	18124
	medRV	1631	0.39	6.60	0.78	1.52	6.34	2.26	4991	17800
Heating oil	TSRV	1622	0.39	5.25	0.65	1.43	3.68	1.68	4625	16456
	RV	1622	0.42	5.53	0.65	1.43	3.97	1.72	4476	16005
	RK	1622	0.41	6.01	0.67	1.43	4.38	1.78	4301	15365
	JWTSRV	1622	0.38	4.75	0.62	1.39	3.42	1.67	5163	18127
	CBPV	1622	0.42	5.53	0.64	1.40	3.72	1.70	4642	16575
	medRV	1622	0.37	5.20	0.62	1.34	3.61	1.70	4500	15876
Natural gas	TSRV	1619	0.73	9.95	0.91	2.26	6.21	1.76	1999	5246
	RV	1619	0.75	9.67	0.95	2.29	6.54	1.88	1822	4676
	RK	1619	0.69	9.74	0.93	2.24	6.02	1.76	1718	4519
	JWTSRV	1619	0.73	7.30	0.77	2.11	3.72	1.36	3577	9682
	CBPV	1619	0.65	8.14	0.84	2.18	5.20	1.62	2822	7301
	medRV	1619	0.68	8.29	0.84	2.09	7.27	1.83	2629	6957

(NYMEX).[7] We use the most active rolling contracts from the pit (floor-traded) session during the main trading hours between 9:00–14:30 EST. From the raw and irregularly spaced prices, we extract 5-minute logarithmic returns using the last-tick method for the RV, RK, BV, and MedRV estimators and, in addition, one-second logarithmic returns for the TSRV and JWTSRV estimators. The 5-minute choice is guided by the volatility signature plot and the previous literature employing the same data. The sample period extends from January 5, 2004 to December 31, 2012, which covers the recent U.S. recession from December 2007–June 2009. We eliminate transactions executed on Saturdays and Sundays, U.S. federal holidays, December 24 to 26, and December 31 to January 2 because of the low activity on these days, which might lead to estimation bias.

4.1. Realized measures

We construct the following measures of the various components of quadratic variation: realized variance $\widehat{QV}_{t,h}^{(RV)}$, defined by Eq. (2), realized kernel $\widehat{QV}_{t,h}^{(RK)}$, defined by Eq. (4), two-scale realized variance $\widehat{QV}_{t,h}^{(TSRV)}$, defined by Eq. (3), the bipower variation $\widehat{IV}_{t,h}^{(CBV)}$, defined by Eq. (7), median realized volatility $\widehat{IV}_{t,h}^{(MedRV)}$, defined by Eq. (8), and jump-adjusted wavelet two-scale realized variance $\widehat{IV}_{t,h}^{(JWTSRV)}$, defined by Eq. (11). We work with forecasts of volatility, which is the square root of the component of quadratic variation. For ease of notation, we use only abbreviations in the analysis of results: RV, RK, TRSV, CBV, MedRV, and JWTSRV.

Our main motivation in using more realized measures in the forecasting is to determine the impact of noise and jumps on forecasting volatility. Although RV is simple to compute for a practitioner, RK and TSRV measure the volatility of the true price process contaminated by microstructure noise, and these three are measures of the quadratic variation. In addition, CBV, and MedRV measure integrated variance directly, whereas MedRV offers a number of advantages over alternative measures in the presence of infrequent jumps. This measure is less sensitive to the presence of occasional zero intraday returns and yields smaller finite-sample bias induced by jumps. Finally, the most complicated JWTSRV measure is robust to both microstructure noise and jumps.

Table 1 reports the summary statistics for the estimated realized measures. The price of natural gas shows the greatest degree of variability in comparison with crude oil and heating oil prices with averages that are twice as large as either. Ljung–Box statistics point to a substantial degree of dependence, as is commonly found in volatility time series. Daily prices, returns and volatility are plotted in Fig. 2.

4.2. Research design for forecast evaluation

The main interest of this work is in relative forecasting performance instead of the in-sample fit of various models. Although the literature describes the fits of particular models in detail, we are interested in comparing them in the forecasting exercise; therefore, the in-sample model fits are available upon request, and we aver that we have conducted all the necessary tests to conclude that all the models fit the data well. We focus on both statistical as well as economic evaluations of the applicable forecasts.

Our data sample covers the period from January 5, 2004 to December 31, 2012. The first 600 observations are used for the in-sample fit of the tested models, and we reserve the remaining 1631 observations to evaluate the out-of-sample forecasting performance. We compute and evaluate 1-step-ahead and cumulative 5- and 10-step-ahead forecasts of price volatility. The cumulative h-step-ahead forecasts are obtained from the usual multi-step-ahead forecast by adding together $\nu_{t+h}^2 = h^{-1}\sum_{j=1}^{h}\nu_{t+j}^2$. We focus on cumulative forecasts because they are more interesting in applications.

After obtaining the volatility forecasts for all 1631 observations from July 6, 2006 through December 31, 2012 on a rolling basis, we divide the forecasts into three periods. The main motivation for this division is the recent global financial crisis, which occurred in the middle of our forecast sample. As shown in Fig. 2, dividing the forecasts into these three equal periods allows us to evaluate the forecasting performance of all the models before the crisis with data from July 6, 2006 through August 31, 2008, during the crisis with data from September 1, 2008 through October 31, 2010, and after the crisis with data from November 1, 2010 through December 31, 2012.

[7] The data were obtained from Tick Data, Inc.

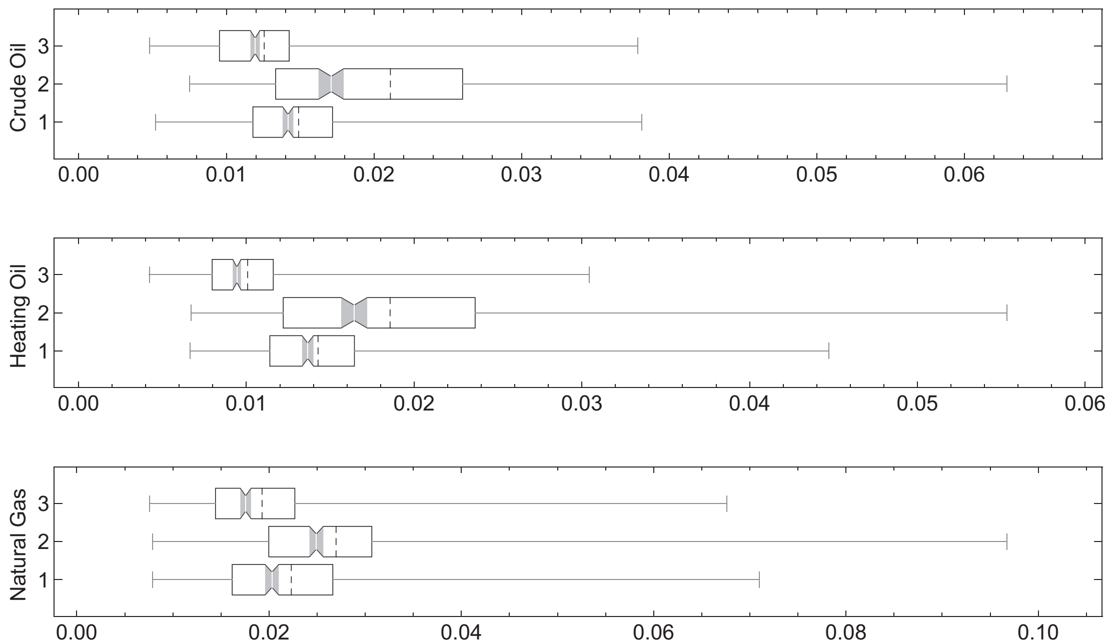

Fig. 1. Box and whisker charts showing distributions of volatilites in different sample periods from July, 6, 2006 through August 31, 2008 (period 1), from September 1, 2008 through October 31, 2010 (period 2) and from November 1, 2010 through December 31, 2012 (period 3). Note that each box starts at the 25% quantile and ends at the 75% quantile of volatility distribution. The left and right fences around the whiskers depict minimum and maximum, the gray notch around the median shows the confidence interval for the median, and the dashed line shows the average value.

To evaluate the choice of the sub-periods, we study the distributional properties of volatility with help of box whisker plots in Fig. 1. We can see that the first (pre-crisis) period is followed by much higher average volatility in the second period (during the crisis) and that the third period (post-crisis) is the period of lowest volatility. Fig. 1 shows that the sub-periods are significantly different. To strengthen the particular choice of the sub-periods, we use a standard procedure for testing multiple break points based on the F-test, and identify the break dates exactly at March 17, 2006, February 8, 2008, December 23, 2009, and December 21, 2010. The choice of the periods is mainly driven by the desire to obtain statistically comparable results, which requires that the three periods be of equal length. Although the break point dates are near the dates that separate our sub-periods, we follow the analysis of the three periods with equal sample lengths.

5. Empirical results

The following section compares the performance of neural networks with the performances of the competing ARFIMA and HAR models in volatility forecasting. Each model is estimated using all six realized measures: RV, TSRV, RK, CBV, MedRV, and JWTSRV. In addition, we experiment with equally weighted combinations of the popular HAR and neural network model, as model averaging may help reduce model uncertainty. Although these two alternatives offer the best forecasts, the linear combination is a good candidate for offering the best forecasting framework for a practitioner in any situation. High frequency data-based forecasts are also benchmarked against a low frequency-based GARCH(1,1) model to study the contribution of high frequency data in the volatility forecasting.

We begin the discussion with a statistical evaluation of the forecasting models and move to the economic implications later on. As discussed above, we aim to assess the forecasting performance of all the models over three separate periods: before, during, and after the 2008 financial crisis. We thus discuss the results in this logical sequence. A substantial number of tables have been produced by this research setup, and the results using different loss functions are overlapping; therefore, we report part of the results in the online supplementary appendix available at 10.1016/j.eswa.2016.02.008.

5.1. Statistical evaluation of forecasts

To statistically compare the accuracy of the volatility forecasts from different models, we employ two common loss functions, i.e., the root mean square error (RMSE) and the mean absolute error (MAE) defined as $RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\nu}_{t+h} - \nu_{t+h}\right)^2}$, and $MAE = \frac{1}{T}\sum_{t=1}^{T}|\hat{\nu}_{t+h} - \nu_{t+h}|$ respectively. As discussed by Nomikos and Pouliasis (2011), these metrics do not provide information about the asymmetry of the errors commonly found in the literature, particularly for the parametric GARCH models. Nonetheless, the asymmetry of forecast error is important for practitioners because it alerts us to whether the modeling strategy tends to over- or under-predict the volatility. Testing energy commodities forecasts, Nomikos and Pouliasis (2011) confirm the strong tendency of GARCH type models to over-predict the volatility of crude oil, heating oil, and natural gas. This finding was further confirmed by Wang and Wu (2012), who find that multivariate GARCH-type models suffer from over-predictions as well.

This bias then translates to direct economic losses. Hence, as suggested by Nomikos and Pouliasis (2011), we employ two additional mean mixed error (MME) loss functions (Brailsford & Faff, 1996) to assess the forecasts. These functions use a mixture of positive and negative forecast errors with different weights that reveal the cases when the model tends to over- or under-predict. Statistics are defined as

$$MME(O) = \frac{1}{T}\left(\sum_{t\in U}|\hat{\nu}_{t+h} - \nu_{t+h}| + \sum_{t\in O}\sqrt{|\hat{\nu}_{t+h} - \nu_{t+h}|}\right) \quad (16)$$

$$MME(U) = \frac{1}{T}\left(\sum_{t\in U}\sqrt{|\hat{\nu}_{t+h} - \nu_{t+h}|} + \sum_{t\in O}|\hat{\nu}_{t+h} - \nu_{t+h}|\right), \quad (17)$$

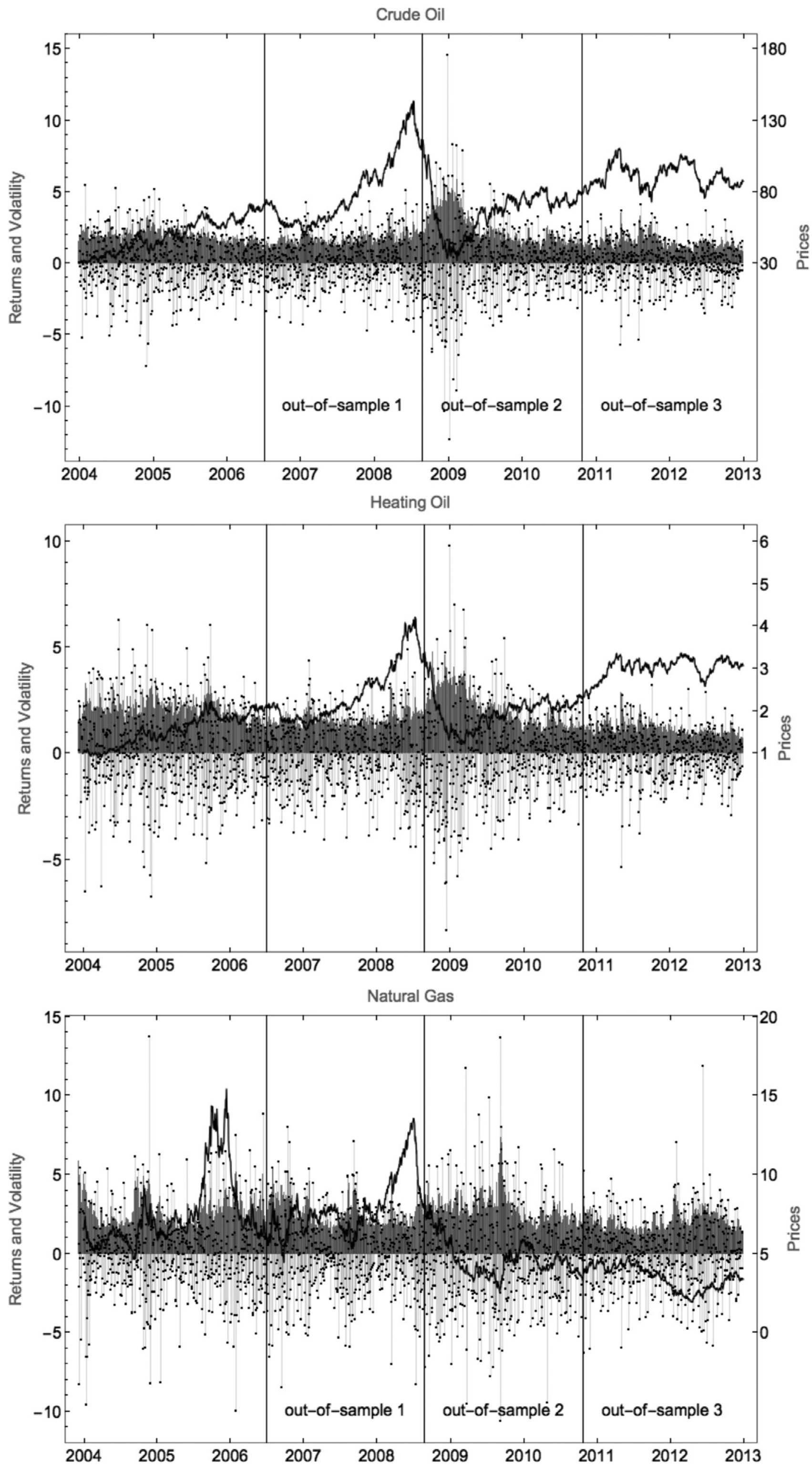

Fig. 2. Realized volatility, returns, and prices of crude oil, heating oil and natural gas. The forecast period is divided into three equal sample periods from July, 6, 2006 through August 31, 2008, from September 1, 2008 through October 31, 2010 and from November 1, 2010 through December 31, 2012.

where U is the set containing under-predictions and O is the set containing over-predictions.

To test the significant differences among competing models, we use the Model Confidence Set (MSC) methodology of Hansen et al. (2011). Given a set of forecasting models, $\mathcal{M}_0$, we identify the model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha} \subset \mathcal{M}_0$, which is the set of models that contain the "best" forecasting model at a given level of confidence α. For a given model $i \in \mathcal{M}_0$, the p-value is the threshold confidence level. Model i belongs to the MCS only if $\widehat{p}_i \geq \alpha$. MSC methodology repeatedly tests the null hypothesis of equal forecasting accuracy

$$H_{0,\mathcal{M}} : E[L_{i,t} - L_{j,t}] = 0, \qquad \text{for all } i, j \in \mathcal{M}$$

where $L_{i,\,t}$ is an appropriate loss function of the i-th model. Beginning with the full set of models, $\mathcal{M} = \mathcal{M}_0$, this procedure sequentially eliminates the worst-performing model from $\mathcal{M}$ when the null is rejected. The surviving set of models then belong to the model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha}$. Following Hansen et al. (2011), we implement the MCS using a stationary bootstrap with an average block length of 20 days.[8]

In addition, we employ the superior predictive ability (SPA) test developed by Hansen (2005) to identify the best performing model. The null hypothesis of the SPA methodology is that the chosen benchmark model is the best forecasting model among its competitors, indicating that the benchmark model produces the smallest loss. Again, we use bootstrapped p-values and follow Hansen (2005) in implementing the test.

We determine the set of the statistically best models in three steps.

1. Determine the MCS $\widehat{\mathcal{M}}^*_{1-\alpha}$ across the forecasting models: the ARFIMA, HAR, Neural Networks (ANN) and HAR-ANN combination.
2. Determine the best forecasting model based on the SPA by benchmarking all the models against the rest.
3. Determine the MCS $\widehat{\mathcal{M}}^*_{1-\alpha}$ across the following realized measures: RV, TSRV, RK, CBV, MedRV, and JWTSRV.

As a result, the best forecasting model is the one we are unable to reject using the SPA and which belongs to the MCS across forecasting models and across realized measures. We repeat the procedure for all the chosen loss functions, the MAE, RMSE, MME (O), and MME (U)[9].

We present the results for the RMSE, MAE, MME(O), and MME(U) in separate tables for each period. Each table contains results for all three commodities (crude oil, heating oil, and natural gas) and several forecasting horizons: 1-step-ahead, 5-step-ahead and 10-step-ahead. The statistical significance of the differences in performance is evaluated across forecasting models (a row-wise comparison) and across volatility estimators (a column-wise comparison) using MCS. (^{b}) and (^{a}) denote the model and estimators that belong to the corresponding 10% model confidence sets, respectively. In addition, a bold entry signifies a model that cannot be rejected as the best forecasting model against its competitors using the SPA test.

Although we do not know the true process generating the data, we must make a decision about a volatility proxy in the testing procedure. When testing model performance, we use the realized measure that is being forecast by the model as a volatility proxy. When testing the performance across measures, we choose a simple proxy of the absolute value of open-close returns, which is common in the literature. This approach lets us identify which realized measures perform best. We also experimented with all different measures as a proxy for volatility, but the results do not change; therefore, we offer these results upon request from the authors.

5.1.1. Forecasting performance before the crisis

We begin by studying the forecasting performance of the models in the pre-crisis period (July, 6, 2006–August 31, 2008). Table 2 presents the results for the RMSE and MAE.

To assist with interpreting the tables, consider the results in Table 2, the first column of which shows the RMSE of the models forecasting the volatility of crude oil. Beginning with 1-step-ahead forecasts and holding the realized measure, such as the TSRV (first column), fixed, ANN and HAR-ANN models produce the lowest RMSE at 0.357×10^{-2}, whereas all high frequency-based models belong to the model confidence set, as they are depicted by (^{b}). GARCH(1,1) produces the largest RMSE and is statistically outperformed by other models. Moreover, the HAR, ANN, and HAR-ANN combinations are set forth in bold, indicating that they are not rejected as the best benchmark forecasting model by the SPA test, whereas the ARFIMA model is rejected because it has the largest RMSE of 0.365×10^{-2}. This result holds for all columns (all realized measures) except for JWTSRV and CBV, which forecast only the integrated variation component. For the JWTSRV and CBV, the ANN and the HAR-ANN combinations are the only two models in the model confidence set. This approach indicates that if we are interested in forecasting the entire quadratic variation, the HAR and ANN models are both in the model confidence set and produce statistically indistinguishable results, whereas the ANN model produces the lowest RMSE. If we are interested in forecasting only the integrated variation component, the ANN is superior to other models. Holding the model and comparing the RMSE column-wise, the MedRV is the only measure belonging to the confidence set. Note that the RMSE and MAE values for comparison across measures are different from those reported in the table, as we use a single volatility proxy for the absolute value of the open-close returns to conduct the MCS.

For the 5-step-ahead and 10-step-ahead forecasts, all realized measures belong to the model confidence set, and the HAR, the ANN, and the combination HAR-ANN produce statistically identical forecasts, whereas the ARFIMA model is rejected, and the ANN models induce the lowest RMSE. Turning to the results found for the MAE, they lead to similar conclusions, but the ARFIMA model is not rejected. Nonetheless, a higher forecasting horizon h implies a lower RMSE for the ANN than the competing models. The GARCH(1,1) model produces uniformly the largest errors, confirming large statistical gains by using high frequency data.

The remaining results reported in Table 2 for heating oil and natural gas show similar—although more mixed—results. In conclusion, a larger forecasting horizon h implies less error from the ANN or a combination of the HAR-ANN model than the HAR and the ARFIMA models (with the exception of heating oil). Whereas on many occasions, the HAR or even the ARFIMA belong to the model confidence set, note that the HAR-ANN combination always belongs to the model confidence set and is never rejected by the SPA test (again, except for a few occasions concerning heating oil). As for the comparison across realized measures, MedRV belongs to the MCS in all cases, whereas other estimators of integrated variance, i.e., the CBV and JWTSRV, belong to the MCS more often than in the case of crude oil. This fact points us to the conclusion that the MedRV is the best measure for forecasting volatility. One may argue that the results are not robust, as these are measures of integrated variance, excluding jumps. However, the results are strong, as the volatility proxy used is the absolute value of open-close returns, which also includes jumps. The large statistical gains from using high frequency data are visible from the largest errors from the GARCH(1,1) model, which is rejected by all the competing models.

Turning our attention to the over- and under-predictions reported in the online appendix, the main conclusions remain unchanged.[10] Notably, the models yield similar results for both the MME(O) and MME(U) in terms of significance but also in terms of % predicted. We may conclude that for all the tested futures, crude oil, heating oil, and natural gas, the models tend to over-predict slightly, but only by approximately 55% on average (with the maxi-

[8] We also used different block lengths, including those that depended on forecasting horizons, to assess the robustness of the results and witnessed no change in the final results. These results are available from the authors upon request.

[9] The results from the statistical testing in the case of the MME(O) and MME(U) merit a cautionary note, as the results on the two loss functions should not be interpreted separately because severe underprediction might lead to favorable results in the case of MME(U) and disastrous results in the case of the MME(O).

[10] To conserve space, we report the actual MME(U), and MME(O) values together with the percentages of the over- and under-predictions in the online supplementary appendix, available at 10.1016/j.eswa.2016.02.008.

Table 2

Statistical comparison of forecasts: pre-crisis period. The table reports average RMSE/MAE loss functions. The Model Confidence Set (MSC) is used to compare the errors row-wise (across forecasting models) as well as column-wise (across realized measures). We use (^{a}) to denote the volatility measures that belong to the $\widehat{\mathcal{M}}^*_{10\%}$ and (^{b}) to denote the forecasting models that belong to the $\widehat{\mathcal{M}}^*_{10\%}$. Moreover, each forecasting model is benchmarked to the rest of the competing models using the Superior Predictive Ability (SPA) test. Cases in which the null hypothesis that the benchmark model is the best forecasting model cannot be rejected are set in bold. Note that numbers are multiplied by $\times\ 10^2$.

	Crude oil						Heating oil						Natural gas					
	TSRV	RV	RK	JWTSRV	CBV	MedRV	TSRV	RV	RK	JWTSRV	CBV	MedRV	TSRV	RV	RK	JWTSRV	CBV	MedRV
RMSE	$h = 1$						$h = 1$						$h = 1$					
GARCH	0.390	0.379	0.403	0.336	0.358	0.355	0.414	0.423	0.429	0.375	0.417	0.452	0.655	0.695	0.702	0.619	0.638	0.670
ARFIMA	0.365^{b}	0.357^{b}	**0.379**^{b}	0.315	0.338	**0.326**^{a,b}	**0.352**^{b}	**0.373**^{b}	**0.367**^{b}	**0.283**^{b}	**0.353**^{a,b}	**0.364**^{a,b}	0.584	0.630	**0.636**^{b}	**0.509**^{b}	0.545^{a,b}	0.556^{a,b}
HAR	**0.360**^{b}	**0.351**^{b}	**0.375**^{b}	0.309	**0.332**	**0.323**^{a,b}	**0.354**^{b}	**0.373**^{b}	**0.369**^{a,b}	**0.286**^{a,b}	**0.353**^{a,b}	**0.365**^{a,b}	**0.575**^{b}	**0.622**^{b}	**0.629**^{b}	**0.504**^{b}	**0.539**^{a,b}	**0.550**^{a,b}
ANN	**0.357**^{b}	**0.349**^{b}	**0.373**^{b}	**0.305**^{b}	**0.329**^{b}	**0.321**^{a,b}	**0.354**^{b}	0.375^{a,b}	**0.371**^{a,b}	0.288^{a,b}	**0.355**^{a,b}	**0.367**^{a,b}	**0.575**^{b}	**0.621**^{b}	**0.631**^{b}	**0.501**^{b}	**0.540**^{a,b}	**0.550**^{a,b}
HAR-ANN	**0.357**^{b}	**0.349**^{b}	**0.373**^{b}	**0.306**^{b}	**0.329**^{b}	**0.321**^{a,b}	**0.353**^{b}	0.372^{a,b}	**0.369**^{a,b}	**0.287**^{a,b}	**0.352**^{a,b}	**0.365**^{a,b}	**0.573**^{b}	**0.620**^{b}	**0.628**^{b}	**0.501**^{b}	**0.538**^{a,b}	**0.548**^{a,b}
	$h = 5$						$h = 5$						$h = 5$					
GARCH	0.654	0.632	0.660	0.559	0.588	0.587	0.709	0.703	0.720	0.693	0.713	0.803	1.128	1.189	1.177	1.136	1.129	1.184
ARFIMA	0.606	0.582	0.619	0.519^{a}	0.542^{a}	0.529^{a}	**0.538**^{a}	**0.567**^{a}	**0.553**^{a}	**0.444**^{a}	**0.531**^{a}	**0.569**^{a}	0.974	1.032	1.038	0.881^{a,b}	0.915^{a}	0.893^{a}
HAR	**0.571**^{b}	**0.546**^{b}	**0.581**^{b}	**0.480**^{b}	**0.504**^{b}	**0.497**^{a,b}	0.550^{a}	**0.574**^{a,b}	**0.558**^{a,b}	0.458^{a}	**0.537**^{a,b}	**0.580**^{a}	0.920	**0.975**^{b}	**0.979**^{b}	**0.848**^{b}	0.872^{a}	0.859^{a}
ANN	**0.568**^{b}	**0.545**^{b}	**0.581**^{b}	**0.479**^{b}	**0.502**^{b}	**0.500**^{a,b}	0.562^{a}	0.591^{a,b}	0.577^{a,b}	0.468^{a}	0.554^{a,b}	0.596^{a,b}	**0.906**^{b}	**0.969**^{b}	**0.975**^{b}	**0.853**^{b}	**0.858**^{a,b}	**0.859**^{a,b}
HAR-ANN	**0.567**^{b}	**0.543**^{b}	**0.579**^{b}	**0.478**^{b}	**0.501**^{b}	**0.495**^{a,b}	0.554^{a,b}	**0.579**^{a,b}	**0.564**^{a,b}	0.461^{a,b}	**0.543**^{a,b}	**0.586**^{a,b}	**0.906**^{b}	**0.963**^{b}	**0.970**^{b}	**0.845**^{b}	**0.858**^{a,b}	**0.850**^{a,b}
	$h = 10$						$h = 10$						$h = 10$					
GARCH	0.895	0.871	0.900	0.773	0.814	0.805	0.969	0.938	0.982	0.971	0.969	1.097	1.600	1.651	1.647	1.635	1.616	1.700
ARFIMA	0.853^{a}	0.830^{a}	0.870^{a}	0.746^{a}	0.774^{a}	0.753^{a}	**0.693**	**0.721**	**0.712**	**0.591**	**0.677**^{a}	**0.735**^{a}	1.370	1.415	1.432	1.240^{a,b}	1.294^{a}	1.247^{a}
HAR	**0.769**^{b}	**0.751**^{b}	**0.782**^{b}	**0.658**^{b}	**0.696**^{a,b}	**0.689**^{a,b}	0.721	**0.739**^{b}	**0.725**^{b}	0.625	**0.698**^{a,b}	0.766^{a}	1.273	**1.305**	**1.328**^{b}	**1.177**^{a,b}	**1.216**^{a}	**1.181**^{a}
ANN	**0.761**^{b}	**0.745**^{a,b}	**0.778**^{a,b}	**0.652**^{a,b}	**0.692**^{a,b}	**0.694**^{a,b}	0.728	0.755^{b}	0.743^{b}	0.636	**0.713**^{a,b}	0.786^{a}	**1.238**^{b}	**1.277**^{b}	**1.315**^{a,b}	**1.163**^{a,b}	**1.188**^{a,b}	**1.162**^{a,b}
HAR-ANN	**0.762**^{b}	**0.744**^{b}	**0.777**^{a,b}	**0.652**^{a,b}	**0.691**^{a,b}	**0.686**^{a,b}	0.721^{b}	**0.742**^{b}	**0.730**^{b}	0.628^{b}	**0.702**^{a,b}	0.773^{a,b}	**1.244**^{b}	**1.276**^{b}	**1.309**^{b}	**1.161**^{a,b}	**1.190**^{a,b}	**1.158**^{a,b}
MAE	$h = 1$						$h = 1$						$h = 1$					
GARCH	0.282	0.285	0.297	0.257	0.272	0.278	0.328	0.329	0.339	0.309	0.332	0.358	0.515	0.531	0.550	0.503	0.509	0.546
ARFIMA	**0.260**^{b}	**0.262**^{b}	**0.274**^{b}	**0.233**^{b}	**0.247**^{b}	**0.241**^{a,b}	**0.264**^{b}	**0.271**	**0.278**^{b}	**0.221**	**0.259**	**0.251**^{a}	0.429^{b}	0.449^{b}	**0.457**^{b}	**0.378**^{b}	**0.405**^{a,b}	**0.409**^{a,b}
HAR	**0.262**^{b}	**0.265**^{b}	**0.276**^{b}	**0.231**^{b}	**0.249**^{b}	**0.243**^{a,b}	**0.262**^{b}	**0.270**^{b}	**0.276**^{b}	**0.221**^{b}	**0.258**^{b}	**0.249**^{a,b}	**0.421**^{b}	**0.442**^{b}	**0.452**^{b}	**0.374**^{b}	**0.400**^{a,b}	**0.403**^{a,b}
ANN	**0.260**^{b}	**0.266**^{b}	**0.275**^{b}	**0.231**^{b}	**0.250**^{b}	**0.243**^{a,b}	**0.264**^{b}	0.275^{b}	**0.278**^{b}	0.225^{b}	0.262^{b}	0.253^{a,b}	**0.419**^{b}	**0.441**^{b}	**0.453**^{b}	**0.372**^{b}	0.403^{a,b}	0.406^{a,b}
HAR-ANN	**0.260**^{b}	**0.265**^{b}	**0.275**^{b}	**0.230**^{b}	**0.248**^{b}	**0.242**^{a,b}	**0.262**^{b}	**0.272**^{b}	**0.276**^{b}	**0.222**^{b}	**0.259**^{b}	**0.250**^{a,b}	**0.418**^{b}	**0.439**^{b}	**0.451**^{b}	**0.372**^{b}	**0.400**^{a,b}	**0.403**^{a,b}
	$h = 5$						$h = 5$						$h = 5$					
GARCH	0.467	0.466	0.479	0.434	0.459	0.468	0.580	0.573	0.582	0.589	0.605	0.696	0.951	0.970	0.987	0.977	0.965	1.016
ARFIMA	**0.428**^{b}	0.430	**0.443**^{b}	**0.381**^{b}	**0.404**^{b}	**0.389**^{a,b}	**0.386**	**0.406**	**0.393**	**0.333**	**0.389**^{a}	**0.392**^{a}	0.770	0.794	0.793	0.698	0.726^{a}	0.709^{a}
HAR	**0.406**^{b}	**0.404**^{b}	**0.415**^{b}	**0.359**^{b}	**0.386**^{b}	**0.372**^{a,b}	**0.396**	**0.412**	**0.395**	**0.339**^{a}	**0.393**^{a}	**0.398**^{a}	0.714	**0.736**^{b}	**0.748**^{b}	**0.666**^{b}	0.688^{a}	0.669^{a}
ANN	**0.406**^{b}	**0.405**^{b}	**0.417**^{b}	**0.361**^{b}	**0.386**^{b}	**0.374**^{a,b}	0.409^{b}	0.433^{b}	0.416^{b}	0.354^{a,b}	0.413^{a,b}	0.416^{a,b}	**0.703**^{b}	**0.733**^{b}	**0.750**^{b}	**0.669**^{b}	**0.672**^{a,b}	**0.668**^{a,b}
HAR-ANN	**0.404**^{b}	**0.402**^{b}	**0.414**^{b}	**0.358**^{b}	**0.384**^{b}	**0.370**^{a,b}	0.401^{b}	0.420^{b}	0.404^{b}	0.345^{a,b}	0.402^{a,b}	0.405^{a,b}	**0.701**^{b}	**0.725**^{b}	**0.741**^{b}	**0.663**^{b}	**0.673**^{a,b}	**0.663**^{a,b}
	$h = 10$						$h = 10$						$h = 10$					
GARCH	0.662	0.658	0.674	0.611	0.641	0.657	0.816	0.802	0.825	0.853	0.855	0.985	1.388	1.427	1.424	1.440	1.417	1.489
ARFIMA	**0.612**^{b}	**0.614**^{b}	**0.628**^{a,b}	**0.559**^{a,b}	**0.580**^{a,b}	**0.556**^{a,b}	**0.509**	**0.537**	**0.517**	**0.448**	**0.516**^{a}	**0.540**^{a}	1.135	1.159	1.162^{b}	1.044^{a}	1.080^{a}	1.041^{a}
HAR	**0.567**^{b}	**0.570**^{b}	**0.583**^{b}	**0.507**^{b}	**0.538**^{b}	**0.526**^{a,b}	**0.528**^{b}	**0.543**^{b}	**0.520**^{b}	0.474^{a}	**0.525**^{a,b}	0.561^{a}	1.043	1.056	**1.094**^{b}	**0.982**^{b}	1.019^{a}	0.973^{a}
ANN	**0.568**^{b}	**0.571**^{b}	**0.583**^{b}	**0.509**^{b}	**0.540**^{b}	**0.539**^{a,b}	0.544^{b}	0.565^{b}	0.543^{b}	0.494^{a}	0.547^{a,b}	0.587^{a}	**0.996**^{b}	**1.027**^{b}	**1.069**^{b}	**0.956**^{b}	**0.979**^{a,b}	**0.940**^{a,b}
HAR-ANN	**0.566**^{b}	**0.568**^{b}	**0.582**^{b}	**0.506**^{b}	**0.537**^{b}	**0.529**^{a,b}	**0.533**^{b}	**0.550**^{b}	**0.527**^{b}	0.481^{a,b}	**0.533**^{a,b}	0.571^{a,b}	**1.011**^{b}	**1.028**^{b}	**1.071**^{b}	**0.963**^{b}	**0.991**^{a,b}	**0.947**^{a,b}

mum levels of over-predictions for natural gas under 60%), whereas on many occasions, models yield an equal number of over- and under-predictions. This is an important finding because, in comparison with the GARCH-type models that strongly over-predict volatility (Nomikos & Pouliasis, 2011; Wang & Wu, 2012), high frequency data appear to yield substantial improvement in this respect. We confirm this result as the GARCH(1,1) models are found to over-predict the volatility much more strongly than high frequency data-based models.

5.1.2. Forecasting performance during the crisis

The forecasting performance of the models during the crisis, i.e., during the September 1, 2008–October 31, 2010 period, follows in terms of RMSE and MAE, as reported in Table 3. A general overview of the results from the pre-crisis period hold, whereas all the RMSE and MAE are larger than in the pre-crisis period. The ANN and HAR-ANN combination of models produce the fewest errors, whereas in most cases, the HAR, the ANN, and their combination belong to the model confidence set. The ARFIMA is rejected as a best-performing model several times, whereas the combination of the ANN and HAR models is never rejected and always belongs to the model confidence set. The GARCH(1,1) model using daily data produces even larger errors than the previous period. This is an important result, i.e., the GARCH model when it was fit onto the pre-crisis data cannot forecast the volatility after large structural break, whereas the ANN models are much more robust.

When comparing the results across realized measures, we observe that the MedRV again belongs to the MCS across all commodities and forecasting horizons. In addition, when forecasting crude oil 1 and 5 steps ahead, it does not matter which measure is used. Therefore (and logically), the simplest realized volatility is preferred in this case. In many cases, the CBV and JWTSRV belong to the model confidence set as well.

The CBV and MedRV belong to the model confidence set most often together with the JWTSRV. From the remaining estimators, the RK appears to perform best.

The comparison using over- and under-prediction loss functions reported in the online appendix provides even more support for the ANN models. The ANN or the HAR-ANN combination belong to the model confidence set, whereas HAR may not be rejected as the best forecasting model more often. Generally, models tend to over-predict volatility during the crisis little bit more on average, but again, the degree of over-prediction is not greater than 60%. The GARCH models generally over-predict the volatility, but to a lesser extent.

To conclude, the results from forecasting volatility during the recent crisis produce larger errors than before the crisis. Generally, ANNs frequently offer worse performance against the remaining models because of greater uncertainty in forecasts. When combined with the HAR, the ANNs prove to be the uniformly best forecasting vehicle. In terms of realized measure, the MedRV is decisively the best choice. Notably, the rate of over-prediction is not much higher, which proves the models' general ability to correctly forecast volatility. In comparison with the low frequency-based GARCH model, coupling neural networks with high frequency data yields reliable forecasts even with large structural breaks, when the models that were fit on pre-crisis data are producing sound out-of-sample forecasts during the crisis.

5.1.3. Forecasting performance after the crisis

Next, we compare the models' performance on the data following the crisis, November 1, 2010–December 31, 2012. Table 4 presents the results for the RMSE and MAE. Although the reported loss functions are lower than in both previous periods, the statistical tests tend to reject more models. The ANN tends to deliver larger errors than competing models, but its combination with the HAR produces the fewest errors. After turmoil of the 2008, the HAR-ANN combination again always belongs to the model confidence set, although it is the only model in the model confidence set in many occasions. Interestingly, ARFIMA produces lowest errors in many cases as well. The results of column-wise comparison favor the MedRV, and the GARCH alternative is again rejected by all the competing models.

Comparing the errors from a volatility forecast through the lens of over- and under-prediction yields similar conclusions. The HAR-ANN combination again belongs to the model confidence set in all cases. This time, all the models tend to over-predict the volatility to a greater extent—up to 70%. This result is attributed to the fact that the model parameters are estimated during the high volatility times of 2008, whereas the predictions are made during a calmer period. In this respect, the models all perform well in terms of statistical criteria. Looking at the statistics for the GARCH models, we find an even larger degree of over-predictions showing that the daily-based model in which volatility is latent can hardly compete with the high frequency data-based strategies

5.1.4. Forecasting performance over the entire period.

As a robustness check, we also compute the statistics for all 1631 forecasts that we obtained. The RMSE/MAE are reported in Table 5, whereas the results for the over- and under-prediction statistics are reported in the online appendix. The combination of the HAR and the ANN always belongs to the model confidence set, and generally produces the best forecast, with few exceptions when forecasting heating oil. A longer forecasting period improves the errors produced by the ANN or HAR-ANN when compared with those produced by competing models. When we compare the errors through the realized measures, the MedRV again belongs to the model confidence set in most cases. In addition, the CBV and JWTSRV belong to the model confidence set in many cases as well. The forecasts based on the low frequency GARCH model decisively yields the largest forecast errors.

In comparison with more the complicated TSRV and JWTSRV measures, the MedRV is a simple alternative and provides the best performance. Therefore, the MedRV is a preferred measure in forecasting the variability of energy prices.

5.2. Comparison of forecasts across realized measures

In addition, we analyze the forecasting efficiency and information content of different volatility estimators and models with the help of simple (Mincer & Zarnowitz, 1969) regressions. Although we do not know which is the most accurate measure of true process underlying the volatility, we simply test the efficiency of all estimators against the rest and expect that if there is an estimator to be chosen among the others, it should also be predicted by all the others. This approach allows us to avoid making decisions about choosing a volatility proxy, as all measures become a proxy. In other words, we seek to describe the information content of the measures and the forecasting models. The regression takes the following form:

$$\hat{\nu}^{RM_1}_{t+h} = \alpha + \beta \hat{\nu}^{(RM_2,f)}_{t+h} + \epsilon_t, \tag{18}$$

where $\hat{\nu}_{t+h}$ is the volatility estimated with RM measures, i.e., the TSRV, RV, RK, JWTSRV, CBV, and MedRV volatility, and $\hat{\nu}^{(RM,f)}_{t+h}$ is its forecast using the ARFIMA, HAR, ANN, HAR-ANN and GARCH models. For example, we first consider $RM = TSRV$ as a true process underlying the data; therefore, we use forecasts from all four models using all six measures to determine which measure and model combination carries over the most information for forecasting TSRV. In this manner, we test all the remaining realized measures, which results in 144 final regressions for one commodity.

Table 3

Statistical comparison of forecasts: crisis period. The table reports average RMSE/MAE loss functions. The Model Confidence Set (MSC) is used to compare the errors row-wise (across forecasting models) as well as column-wise (across realized measures). We use (^{a}) to denote the volatility measures that belong to the $\widehat{\mathcal{M}}^{*}_{10\%}$ and (^{b}) to denote the forecasting models that belong to the $\widehat{\mathcal{M}}^{*}_{10\%}$. Moreover, each forecasting model is benchmarked to the rest of the competing models using the Superior Predictive Ability (SPA) test. Cases in which the null hypothesis that the benchmark model is the best forecasting model cannot be rejected are set in bold. Note that numbers are multiplied by $\times\ 10^{2}$.

	Crude oil						Heating oil						Natural gas					
	TSRV	RV	RK	JWTSRV	CBV	MedRV	TSRV	RV	RK	JWTSRV	CBV	MedRV	TSRV	RV	RK	JWTSRV	CBV	MedRV
RMSE	$h = 1$						$h = 1$						$h = 1$					
GARCH	0.704	0.711	0.731	0.691	0.708	0.725	0.537	0.542	0.571	0.520	0.535	0.537	0.994	1.028	1.028	0.860	0.899	0.954
ARFIMA	**0.536**$^{a, b}$	**0.541**$^{a, b}$	**0.579**$^{a, b}$	**0.498**$^{a, b}$	**0.541**$^{a, b}$	**0.531**$^{a, b}$	**0.420**	**0.421**	**0.466**	**0.378**	**0.411**	**0.395**^{a}	0.897	0.971	0.931	**0.545**^{b}	**0.724**^{b}	**0.766**^{b}
HAR	**0.522**^{b}	**0.525**^{b}	**0.569**^{b}	**0.489**^{b}	**0.526**$^{a, b}$	**0.517**$^{a, b}$	**0.416**^{b}	**0.412**	**0.464**^{b}	**0.376**^{b}	**0.403**^{b}	**0.388**$^{a, b}$	0.872	0.936	0.907	**0.548**^{b}	**0.716**^{b}	**0.759**^{b}
ANN	**0.524**^{b}	**0.530**$^{a, b}$	**0.573**$^{a, b}$	**0.533**$^{a, b}$	**0.530**$^{a, b}$	0.524$^{a, b}$	0.420^{b}	0.419^{b}	0.471^{b}	0.381^{b}	0.410^{b}	0.395$^{a, b}$	**0.856**$^{a, b}$	**0.912**$^{a, b}$	**0.893**$^{a, b}$	**0.551**$^{a, b}$	**0.719**$^{a, b}$	**0.759**$^{a, b}$
HAR-ANN	**0.519**^{b}	**0.524**$^{a, b}$	**0.566**$^{a, b}$	**0.498**$^{a, b}$	**0.525**$^{a, b}$	**0.517**$^{a, b}$	**0.416**^{b}	**0.413**^{b}	**0.464**^{b}	**0.377**^{b}	**0.404**^{b}	**0.389**$^{a, b}$	**0.859**$^{a, b}$	**0.918**$^{a, b}$	**0.894**$^{a, b}$	**0.548**$^{a, b}$	**0.716**$^{a, b}$	**0.757**$^{a, b}$
	$h = 5$						$h = 5$						$h = 5$					
GARCH	1.358	1.372	1.362	1.365	1.353	1.384	1.031	1.050	1.057	1.040	1.043	1.051	1.677	1.638	1.680	1.794	1.712	1.802
ARFIMA	**0.877**^{a}	0.879$^{a, b}$	**0.942**^{a}	**0.852**^{a}	0.882$^{a, b}$	**0.874**^{a}	**0.684**	**0.691**^{a}	**0.751**^{a}	**0.650**^{a}	**0.677**^{a}	**0.653**^{a}	**1.339**^{b}	**1.403**^{b}	**1.342**^{b}	**1.030**$^{a, b}$	**1.225**$^{a, b}$	**1.277**$^{a, b}$
HAR	**0.779**^{a}	**0.776**$^{a, b}$	**0.848**^{a}	**0.764**^{a}	**0.772**$^{a, b}$	**0.780**^{a}	**0.643**^{b}	**0.639**^{b}	**0.715**^{b}	**0.630**^{b}	**0.630**$^{a, b}$	**0.614**$^{a, b}$	**1.327**^{b}	**1.387**^{b}	**1.325**^{b}	**1.061**^{b}	**1.250**$^{a, b}$	**1.312**$^{a, b}$
ANN	0.796$^{a, b}$	0.787$^{a, b}$	0.875$^{a, b}$	0.779$^{a, b}$	0.786$^{a, b}$	0.816$^{a, b}$	0.680$^{a, b}$	0.667$^{a, b}$	0.762$^{a, b}$	0.658$^{a, b}$	0.664$^{a, b}$	0.651$^{a, b}$	**1.355**^{b}	**1.409**^{b}	**1.349**^{b}	**1.082**^{b}	**1.308**$^{a, b}$	**1.369**$^{a, b}$
HAR-ANN	**0.776**$^{a, b}$	**0.773**$^{a, b}$	**0.849**$^{a, b}$	**0.762**$^{a, b}$	**0.771**$^{a, b}$	**0.786**$^{a, b}$	**0.650**$^{a, b}$	**0.643**$^{a, b}$	**0.726**$^{a, b}$	**0.636**$^{a, b}$	**0.638**$^{a, b}$	**0.621**$^{a, b}$	**1.335**^{b}	**1.390**^{b}	**1.329**^{b}	**1.065**^{b}	**1.271**$^{a, b}$	**1.329**$^{a, b}$
	$h = 10$						$h = 10$						$h = 10$					
GARCH	1.961	1.987	1.957	1.967	1.965	1.987	1.464	1.503	1.483	1.477	1.494	1.493	2.391	2.349	2.388	2.579	2.459	2.566
ARFIMA	**1.345**$^{a, b}$	**1.336**$^{a, b}$	**1.420**$^{a, b}$	**1.312**$^{a, b}$	**1.344**$^{a, b}$	**1.313**^{a}	**0.997**	**1.025**	**1.075**	**0.957**^{a}	**1.003**^{a}	**0.964**^{a}	**1.864**^{b}	**1.979**^{b}	**1.862**^{b}	**1.501**^{b}	**1.761**$^{a, b}$	**1.793**$^{a, b}$
HAR	**1.114**^{b}	**1.108**^{b}	**1.215**$^{a, b}$	**1.088**$^{a, b}$	**1.105**$^{a, b}$	**1.096**^{a}	**0.904**^{b}	**0.918**	**0.988**^{b}	**0.903**$^{a, b}$	**0.905**^{a}	**0.880**$^{a, b}$	**1.861**^{b}	**1.975**^{b}	**1.845**^{b}	**1.538**^{b}	**1.790**^{b}	**1.826**^{b}
ANN	1.135^{b}	**1.115**^{b}	**1.234**^{b}	**1.086**$^{a, b}$	**1.113**$^{a, b}$	1.134$^{a, b}$	0.953^{b}	0.963^{b}	1.064^{b}	0.945$^{a, b}$	0.942$^{a, b}$	0.962$^{a, b}$	**1.882**^{b}	**2.010**^{b}	**1.831**^{b}	**1.549**^{b}	**1.821**^{b}	**1.805**^{b}
HAR-ANN	**1.110**^{b}	**1.099**^{b}	**1.204**^{b}	**1.075**$^{a, b}$	**1.098**$^{a, b}$	**1.097**$^{a, b}$	**0.911**^{b}	**0.926**^{b}	**1.005**^{b}	**0.911**$^{a, b}$	**0.910**$^{a, b}$	**0.899**$^{a, b}$	**1.851**^{b}	**1.968**^{b}	**1.810**^{b}	**1.531**^{b}	**1.775**^{b}	**1.798**^{b}
MAE	$h = 1$						$h = 1$						$h = 1$					
GARCH	0.491	0.498	0.515	0.485	0.495	0.517	0.396	0.398	0.421	0.394	0.402	0.413	0.757	0.783	0.790	0.708	0.720	0.767
ARFIMA	**0.364**	**0.365**^{b}	**0.400**	**0.343**^{b}	**0.357**^{b}	**0.354**^{a}	**0.304**^{b}	0.302	**0.341**	**0.278**^{b}	0.295	**0.291**^{a}	0.607	0.653	0.642^{b}	**0.409**^{b}	**0.509**^{b}	**0.523**^{b}
HAR	**0.359**	**0.364**^{b}	**0.398**^{b}	**0.339**^{b}	**0.359**^{b}	**0.357**$^{a, b}$	**0.304**^{b}	**0.294**	**0.340**^{b}	**0.278**^{b}	**0.288**	**0.288**$^{a, b}$	0.597	0.641	0.632^{b}	**0.410**^{b}	**0.509**^{b}	**0.525**^{b}
ANN	0.362^{b}	**0.365**^{b}	0.404^{b}	**0.351**^{b}	**0.361**^{b}	0.364$^{a, b}$	**0.305**^{b}	0.299^{b}	0.345^{b}	**0.279**^{b}	0.293^{b}	0.294$^{a, b}$	**0.582**^{b}	**0.625**^{b}	**0.623**^{b}	**0.414**^{b}	**0.512**^{b}	**0.530**^{b}
HAR-ANN	**0.356**^{b}	**0.362**^{b}	**0.396**^{b}	**0.343**^{b}	**0.358**$^{a, b}$	**0.358**$^{a, b}$	**0.302**^{b}	**0.294**^{b}	**0.340**^{b}	**0.277**^{b}	**0.288**^{b}	**0.289**$^{a, b}$	**0.585**^{b}	**0.629**^{b}	**0.623**^{b}	**0.411**^{b}	**0.509**^{b}	**0.526**^{b}
	$h = 5$						$h = 5$						$h = 5$					
GARCH	0.954	0.972	0.946	0.980	0.977	1.026	0.752	0.782	0.761	0.802	0.805	0.842	1.288	1.246	1.280	1.460	1.387	1.463
ARFIMA	**0.593**	**0.600**	**0.636**	**0.584**	**0.600**	**0.593**^{a}	**0.495**	**0.501**	**0.536**	**0.487**	**0.498**	**0.486**^{a}	**0.940**^{b}	**0.982**^{b}	**0.947**^{b}	**0.718**	**0.841**^{a}	**0.852**^{a}
HAR	**0.559**	**0.563**	**0.612**^{b}	**0.547**	**0.560**	**0.558**$^{a, b}$	**0.469**^{b}	**0.466**^{b}	**0.523**^{b}	**0.465**^{b}	**0.461**^{b}	**0.450**$^{a, b}$	**0.942**^{b}	**0.996**^{b}	**0.956**^{b}	**0.744**^{b}	**0.877**^{a}	**0.894**^{a}
ANN	0.580^{b}	0.576^{b}	0.637^{b}	0.568^{b}	0.574^{b}	0.583$^{a, b}$	0.498^{b}	0.490^{b}	0.555^{b}	0.491^{b}	0.491^{b}	0.486$^{a, b}$	**0.957**^{b}	**1.001**^{b}	**0.966**^{b}	0.757^{b}	0.893^{a}	0.915^{a}
HAR-ANN	**0.562**^{b}	**0.563**^{b}	**0.616**^{b}	**0.549**^{b}	**0.561**^{b}	**0.561**$^{a, b}$	**0.475**^{b}	**0.470**^{b}	**0.530**^{b}	**0.473**^{b}	**0.467**^{b}	**0.457**$^{a, b}$	**0.944**^{b}	**0.994**^{b}	**0.955**^{b}	**0.745**^{b}	**0.880**$^{a, b}$	**0.897**$^{a, b}$
	$h = 10$						$h = 10$						$h = 10$					
GARCH	1.409	1.444	1.400	1.453	1.454	1.513	1.090	1.134	1.085	1.163	1.169	1.210	1.899	1.792	1.868	2.121	1.995	2.112
ARFIMA	**0.916**^{b}	**0.920**^{b}	**0.964**^{b}	**0.902**^{b}	**0.928**^{b}	**0.923**$^{a, b}$	**0.736**	**0.759**	**0.782**	**0.727**	**0.752**	**0.734**^{a}	**1.319**^{b}	**1.404**^{b}	**1.309**^{b}	**1.013**	**1.195**$^{a, b}$	**1.180**$^{a, b}$
HAR	**0.813**^{b}	**0.807**^{b}	**0.884**^{b}	**0.792**^{b}	**0.804**^{b}	**0.813**$^{a, b}$	**0.681**^{b}	**0.681**	**0.740**^{b}	**0.680**^{b}	**0.679**	**0.669**$^{a, b}$	**1.378**^{b}	**1.445**^{b}	**1.331**^{b}	**1.078**	**1.265**^{b}	**1.253**^{b}
ANN	**0.830**^{b}	0.812^{b}	**0.896**^{b}	**0.788**$^{a, b}$	**0.814**$^{a, b}$	**0.831**$^{a, b}$	0.712^{b}	0.722^{b}	0.801^{b}	0.719^{b}	0.717^{b}	0.731$^{a, b}$	**1.358**^{b}	**1.442**^{b}	**1.289**^{b}	1.079	**1.254**^{b}	**1.237**^{b}
HAR-ANN	**0.810**^{b}	**0.795**^{b}	**0.877**^{b}	**0.779**^{b}	**0.798**$^{a, b}$	**0.809**$^{a, b}$	**0.682**^{b}	**0.688**^{b}	**0.751**^{b}	**0.685**^{b}	**0.686**^{b}	**0.682**$^{a, b}$	**1.352**^{b}	**1.424**^{b}	**1.291**^{b}	**1.071**^{b}	**1.244**^{b}	**1.231**^{b}

Table 4

Statistical comparison of forecasts: after-crisis period. The table reports average RMSE/MAE loss functions. The Model Confidence Set (MSC) is used to compare the errors row-wise (across forecasting models) as well as column-wise (across realized measures). We use (^{a}) to denote the volatility measures that belong to the $\widehat{\mathcal{M}}^*_{10\%}$ and (^{b}) to denote the forecasting models that belong to the $\widehat{\mathcal{M}}^*_{10\%}$. Moreover, each forecasting model is benchmarked to the rest of the competing models using the Superior Predictive Ability (SPA) test. Cases in which the null hypothesis that the benchmark model is the best forecasting model cannot be rejected are set in bold. Note that numbers are multiplied by $\times 10^2$.

	Crude oil					
	TSRV	RV	RK	JWTSRV	CBV	MedRV
RMSE	$h=1$					
GARCH	0.394	0.392	0.393	0.379	0.388	0.403
ARFIMA	**0.321**	**0.330**	**0.332**	**0.288**	**0.318**	**0.309**^{a}
HAR	**0.333**	**0.340**^{b}	**0.342**^{b}	**0.299**	**0.328**	**0.318**^{a,b}
ANN	0.335	**0.341**^{b}	0.343^{b}	0.299	0.328	**0.318**^{a,b}
HAR-ANN	**0.333**^{b}	**0.339**^{b}	**0.341**^{b}	**0.298**^{b}	**0.327**^{b}	**0.317**^{a,b}
	$h=5$					
GARCH	0.760	0.746	0.731	0.772	0.755	0.805
ARFIMA	**0.612**^{b}	**0.616**^{b}	**0.592**^{b}	**0.578**^{b}	**0.600**^{b}	**0.588**^{a,b}
HAR	**0.629**^{b}	**0.631**^{b}	**0.609**^{b}	**0.598**^{b}	**0.622**^{b}	**0.610**^{a,b}
ANN	**0.635**^{b}	**0.636**^{b}	**0.613**^{b}	**0.601**^{b}	**0.627**^{b}	**0.616**^{a,b}
HAR-ANN	**0.628**^{b}	**0.630**^{b}	**0.606**^{b}	**0.595**^{b}	**0.621**^{b}	**0.608**^{a,b}
	$h=10$					
GARCH	1.038	1.014	0.994	1.069	1.039	1.116
ARFIMA	**0.860**^{b}	**0.857**^{b}	**0.825**^{b}	**0.827**^{b}	**0.848**^{b}	**0.833**^{a,b}
HAR	**0.881**^{b}	**0.873**^{b}	**0.842**^{b}	**0.857**^{b}	**0.877**^{b}	**0.865**^{a,b}
ANN	**0.887**^{b}	**0.879**^{b}	**0.855**^{b}	**0.858**^{b}	**0.886**^{b}	**0.869**^{a,b}
HAR-ANN	**0.878**^{b}	**0.871**^{b}	**0.841**^{b}	**0.852**^{b}	**0.876**^{b}	**0.861**^{a,b}
MAE	$h=1$					
GARCH	0.300	0.304	0.309	0.300	0.305	0.328
ARFIMA	**0.231**	**0.236**	**0.250**	**0.210**	**0.231**	**0.221**^{a}
HAR	0.241	0.247	0.259	0.218	0.241	0.229^{a}
ANN	0.243	0.248	0.261	0.218	0.241	0.228^{a}
HAR-ANN	0.241^{b}	0.246^{b}	0.259^{b}	0.217^{b}	0.239^{b}	0.228^{a,b}
	$h=5$					
GARCH	0.578	0.587	0.571	0.616	0.604	0.666
ARFIMA	**0.405**	**0.414**	**0.405**	**0.386**	**0.405**	**0.383**^{a}
HAR	0.435	0.454	0.436	0.425	0.445	0.427^{a}
ANN	0.449	0.465	0.452	0.433	0.455	0.437^{a}
HAR-ANN	0.439^{b}	0.458^{b}	0.440^{b}	0.427^{b}	0.448^{b}	0.429^{a,b}
	$h=10$					
GARCH	0.796	0.801	0.776	0.864	0.831	0.923
ARFIMA	**0.559**^{b}	**0.569**	**0.558**	**0.544**	**0.559**	**0.533**^{a}
HAR	**0.612**^{b}	**0.629**^{b}	**0.600**^{b}	0.610	0.626	0.616^{a}
ANN	**0.626**^{b}	0.643^{b}	0.623^{b}	0.618	0.640	0.625^{a}
HAR-ANN	**0.613**^{b}	**0.632**^{b}	**0.605**^{b}	**0.609**^{b}	0.629^{b}	0.615^{a,b}

	Heating oil					
	TSRV	RV	RK	JWTSRV	CBV	MedRV
RMSE	$h=1$					
GARCH	0.306	0.314	0.313	0.300	0.311	0.333
ARFIMA	**0.257**	**0.270**	**0.268**	**0.228**	**0.254**	**0.254**^{a}
HAR	**0.262**	**0.275**	**0.272**	**0.232**^{b}	**0.257**	**0.259**^{a}
ANN	0.265	0.280	0.277^{b}	0.234^{b}	0.261^{a,b}	0.261^{a}
HAR-ANN	0.263^{b}	**0.276**^{b}	0.273^{b}	**0.232**^{b}	**0.259**^{b}	0.259^{a,b}
	$h=5$					
GARCH	0.559	0.566	0.553	0.585	0.576	0.635
ARFIMA	**0.440**	**0.458**	**0.437**	**0.412**	**0.436**^{a}	**0.438**^{a}
HAR	0.478	0.492	0.471	0.446	0.472^{a}	0.473^{a}
ANN	0.497	0.507	0.491	0.457	0.484^{a}	0.484^{a}
HAR-ANN	0.484^{b}	0.496^{b}	0.476^{b}	0.448^{b}	0.475^{a,b}	0.475^{a,b}
	$h=10$					
GARCH	0.765	0.767	0.750	0.816	0.797	0.880
ARFIMA	**0.596**	**0.605**	**0.578**	**0.566**^{a}	**0.593**^{a}	**0.591**^{a}
HAR	0.669	0.669	0.640^{a}	0.640^{a}	0.665^{a}	0.659^{a}
ANN	0.710	0.702	0.690^{a}	0.666^{a}	0.693^{a}	0.687^{a}
HAR-ANN	0.682^{b}	0.679^{b}	0.656^{b}	0.646^{a,b}	0.673^{a,b}	0.667^{a,b}
MAE	$h=1$					
GARCH	0.251	0.250	0.256	0.250	0.255	0.277
ARFIMA	**0.191**	**0.193**	**0.204**	**0.169**	**0.185**	**0.183**^{a}
HAR	**0.191**^{b}	**0.194**^{b}	**0.205**^{b}	**0.169**^{b}	**0.186**^{b}	**0.185**^{a,b}
ANN	0.196^{b}	0.198^{b}	0.211^{b}	0.174^{b}	0.190^{b}	0.188^{a,b}
HAR-ANN	**0.193**^{b}	**0.195**^{b}	**0.207**^{b}	0.171^{b}	**0.187**^{b}	**0.186**^{a,b}
	$h=5$					
GARCH	0.458	0.465	0.453	0.496	0.486	0.550
ARFIMA	**0.324**	**0.337**	**0.330**	**0.304**	**0.326**^{a}	**0.322**^{a}
HAR	0.361	0.372	0.359	0.337	0.360^{a}	0.353^{a}
ANN	0.378	0.387	0.381	0.350	0.373^{a}	0.367^{a}
HAR-ANN	0.367^{b}	0.377^{b}	0.368^{b}	0.340^{b}	0.366^{a,b}	0.358^{a,b}
	$h=10$					
GARCH	0.625	0.638	0.609	0.688	0.665	0.763
ARFIMA	**0.449**	**0.467**	**0.444**	**0.424**^{a}	**0.456**^{a}	**0.450**^{a}
HAR	0.520	0.529	0.501	0.501^{a}	0.520^{a}	0.514^{a}
ANN	0.555	0.558	0.547	0.526^{a}	0.544^{a}	0.539^{a}
HAR-ANN	0.532^{b}	0.539^{b}	0.517^{b}	0.508^{a,b}	0.527^{a,b}	0.522^{a,b}

	Natural gas					
	TSRV	RV	RK	JWTSRV	CBV	MedRV
RMSE	$h=1$					
GARCH	0.670	0.690	0.690	0.599	0.623	0.663
ARFIMA	0.620	0.645	0.636	**0.365**	**0.485**^{a,b}	**0.454**^{a,b}
HAR	0.605	0.629	0.625	0.371	**0.481**^{a,b}	**0.455**^{a,b}
ANN	**0.585**^{b}	**0.599**^{b}	**0.610**^{b}	0.370	**0.477**^{a,b}	**0.455**^{a,b}
HAR-ANN	**0.590**^{b}	**0.606**^{b}	**0.612**^{b}	0.370^{b}	**0.478**^{a,b}	**0.453**^{a,b}
	$h=5$					
GARCH	0.956	0.976	0.970	1.207	1.098	1.247
ARFIMA	**0.792**	**0.829**^{b}	**0.822**	**0.622**	**0.697**^{a}	**0.689**^{a}
HAR	**0.826**	0.863^{b}	0.857	0.660	0.737^{a}	0.736^{a}
ANN	0.823	**0.860**^{b}	0.859	0.649	0.722^{a}	0.722^{a}
HAR-ANN	0.820^{b}	**0.858**^{b}	**0.852**^{b}	0.651^{b}	0.725^{a,b}	0.724^{a,b}
	$h=10$					
GARCH	1.293	1.313	1.285	1.721	1.558	1.759
ARFIMA	**1.079**^{b}	**1.120**^{b}	**1.093**^{b}	**0.909**	**1.009**^{a}	**0.980**^{a}
HAR	**1.130**^{b}	**1.188**^{b}	**1.148**^{b}	0.967	1.081^{a}	1.056^{a}
ANN	**1.135**^{b}	**1.195**^{b}	**1.157**^{b}	0.950	**1.056**^{a,b}	**1.030**^{a,b}
HAR-ANN	**1.124**^{b}	**1.183**^{b}	**1.144**^{b}	0.953^{b}	1.062^{a,b}	1.035^{a,b}
MAE	$h=1$					
GARCH	0.542	0.549	0.558	0.513	0.515	0.570
ARFIMA	0.468	0.487	0.480	**0.286**^{b}	0.366^{a}	0.345^{a}
HAR	0.447	0.462	0.465	0.290^{b}	**0.360**^{a,b}	**0.344**^{a,b}
ANN	**0.433**^{b}	**0.443**^{b}	**0.460**^{b}	0.290^{b}	**0.360**^{a,b}	**0.345**^{a,b}
HAR-ANN	**0.436**^{b}	**0.444**^{b}	**0.458**^{b}	0.290^{b}	**0.359**^{a,b}	**0.344**^{a,b}
	$h=5$					
GARCH	0.762	0.756	0.776	1.062	0.937	1.090
ARFIMA	**0.596**	**0.624**	**0.621**	**0.474**	**0.524**^{a}	**0.526**^{a}
HAR	**0.622**	0.656	0.655	0.500	0.552^{a}	0.560^{a}
ANN	0.627	0.662	0.665	0.501	0.548^{a}	0.557^{a}
HAR-ANN	**0.620**^{b}	**0.655**^{b}	0.657^{b}	0.498^{b}	0.547^{a,b}	0.555^{a,b}
	$h=10$					
GARCH	1.048	1.052	1.043	1.537	1.360	1.567
ARFIMA	**0.820**^{b}	**0.849**^{b}	**0.840**^{b}	**0.684**	**0.746**^{a}	**0.725**^{a}
HAR	**0.868**^{b}	**0.915**^{b}	**0.896**^{b}	0.727	0.818^{a}	0.796^{a}
ANN	**0.877**^{b}	**0.926**^{b}	0.909^{b}	0.732	0.797^{a}	0.777^{a}
HAR-ANN	**0.865**^{b}	**0.916**^{b}	**0.896**^{b}	0.725^{b}	0.801^{a,b}	0.781^{a,b}

Table 5

Statistical comparison of forecasts: whole period. The table reports average RMSE/MAE loss functions. The Model Confidence Set (MSC) is used to compare the errors row-wise (across forecasting models) as well as column-wise (across realized measures). We use (^{a}) to denote the volatility measures that belong to the $\widehat{\mathcal{M}}^{*}_{10\%}$ and (^{b}) to denote the forecasting models that belong to the $\widehat{\mathcal{M}}^{*}_{10\%}$. Moreover, each forecasting model is benchmarked to the rest of the competing models using the Superior Predictive Ability (SPA) test. Cases in which the null hypothesis that the benchmark model is the best forecasting model cannot be rejected are set in bold. Note that numbers are multiplied by $\times 10^{2}$.

	Crude oil						Heating oil						Natural gas					
	TSRV	RV	RK	JWTSRV	CBV	MedRV	TSRV	RV	RK	JWTSRV	CBV	MedRV	TSRV	RV	RK	JWTSRV	CBV	MedRV
RMSE	$h = 1$						$h = 1$						$h = 1$					
GARCH	0.517	0.516	0.532	0.494	0.509	0.521	0.430	0.437	0.451	0.409	0.432	0.449	0.789	0.820	0.822	0.703	0.731	0.774
ARFIMA	**0.417**^{b}	**0.419**^{b}	**0.443**	**0.378**^{b}	**0.411**^{b}	**0.401**^{a}	**0.350**	**0.361**	**0.376**	**0.303**	**0.346**	**0.344**^{a}	0.714	0.765	0.747	**0.479**^{b}	0.594^{a}	0.606^{a}
HAR	**0.413**^{b}	**0.414**^{b}	**0.440**	**0.375**^{b}	**0.406**^{b}	**0.397**^{a}	**0.351**^{b}	**0.359**^{b}	**0.378**^{b}	**0.304**^{b}	**0.344**^{b}	**0.342**$^{a,\ b}$	0.697	0.744	0.732	**0.480**^{b}	**0.587**$^{a,\ b}$	**0.602**$^{a,\ b}$
ANN	**0.414**^{b}	**0.415**^{b}	0.441^{b}	**0.394**^{b}	**0.407**$^{a,\ b}$	0.399$^{a,\ b}$	0.353^{b}	0.363^{b}	0.382^{b}	0.308^{b}	0.348^{b}	0.347$^{a,\ b}$	**0.684**^{b}	**0.725**^{b}	**0.723**^{b}	**0.480**^{b}	**0.588**$^{a,\ b}$	**0.601**$^{a,\ b}$
HAR-ANN	**0.411**^{b}	**0.412**^{b}	**0.438**^{b}	**0.378**^{b}	**0.404**^{b}	**0.396**$^{a,\ b}$	**0.350**^{b}	**0.359**^{b}	**0.378**^{b}	**0.305**^{b}	**0.344**^{b}	**0.343**$^{a,\ b}$	**0.686**^{b}	**0.729**^{b}	**0.723**^{b}	**0.479**^{b}	**0.586**$^{a,\ b}$	**0.600**$^{a,\ b}$
	$h = 5$						$h = 5$						$h = 5$					
GARCH	0.974	0.972	0.970	0.962	0.957	0.985	0.793	0.801	0.806	0.798	0.804	0.849	1.291	1.297	1.310	1.410	1.344	1.438
ARFIMA	0.709^{a}	0.705^{a}	**0.735**^{a}	**0.665**^{a}	0.691^{a}	**0.680**^{a}	**0.564**	**0.581**	**0.596**	**0.514**	**0.558**^{a}	**0.561**^{a}	**1.060**^{b}	**1.114**^{b}	**1.088**^{b}	**0.861**^{b}	**0.970**$^{a,\ b}$	**0.984**$^{a,\ b}$
HAR	**0.666**	**0.658**	**0.690**^{a}	**0.625**^{a}	**0.643**^{a}	**0.640**^{a}	**0.562**^{b}	**0.572**^{b}	**0.591**^{b}	**0.519**^{b}	**0.551**$^{a,\ b}$	**0.559**$^{a,\ b}$	**1.047**^{b}	**1.099**^{b}	**1.072**^{b}	**0.872**^{b}	**0.978**$^{a,\ b}$	**1.000**$^{a,\ b}$
ANN	0.673^{b}	0.664^{b}	0.702$^{a,\ b}$	0.632$^{a,\ b}$	0.649$^{a,\ b}$	0.657$^{a,\ b}$	0.585^{b}	0.593^{b}	0.621^{b}	0.537^{b}	0.573$^{a,\ b}$	0.582$^{a,\ b}$	**1.054**^{b}	**1.106**^{b}	**1.081**^{b}	**0.879**^{b}	**0.995**$^{a,\ b}$	**1.022**$^{a,\ b}$
HAR-ANN	**0.663**^{b}	**0.656**^{b}	**0.689**$^{a,\ b}$	**0.623**$^{a,\ b}$	**0.641**$^{a,\ b}$	**0.642**$^{a,\ b}$	**0.567**^{b}	**0.576**^{b}	**0.599**^{b}	**0.523**^{b}	**0.557**$^{a,\ b}$	**0.565**$^{a,\ b}$	**1.045**^{b}	**1.095**^{b}	**1.070**^{b}	**0.870**^{b}	**0.980**$^{a,\ b}$	**1.002**$^{a,\ b}$
	$h = 10$						$h = 10$						$h = 10$					
GARCH	1.383	1.384	1.370	1.369	1.368	1.397	1.108	1.118	1.117	1.127	1.129	1.187	1.822	1.824	1.833	2.025	1.923	2.048
ARFIMA	**1.045**^{b}	**1.035**$^{a,\ b}$	**1.073**^{a}	**0.994**$^{a,\ b}$	**1.021**$^{a,\ b}$	**0.998**^{a}	**0.783**	**0.805**	**0.818**	**0.729**^{a}	**0.780**^{a}	**0.780**^{a}	**1.474**^{b}	**1.547**^{b}	**1.496**^{b}	**1.240**^{b}	**1.390**$^{a,\ b}$	**1.383**$^{a,\ b}$
HAR	**0.933**^{b}	**0.924**^{b}	**0.966**	**0.887**^{b}	**0.909**^{b}	**0.900**^{a}	**0.772**$^{a,\ b}$	**0.783**$^{a,\ b}$	**0.800**^{a}	**0.735**$^{a,\ b}$	**0.764**$^{a,\ b}$	**0.775**$^{a,\ b}$	**1.457**^{b}	**1.530**^{b}	**1.471**^{b}	**1.250**^{b}	**1.398**^{b}	**1.397**^{b}
ANN	0.941^{b}	**0.927**^{b}	0.976^{b}	**0.885**$^{a,\ b}$	**0.915**$^{a,\ b}$	0.918$^{a,\ b}$	0.806^{b}	0.816^{b}	0.850	0.763$^{a,\ b}$	0.792$^{a,\ b}$	0.821$^{a,\ b}$	**1.457**^{b}	**1.539**^{b}	**1.464**^{b}	**1.246**^{b}	**1.397**$^{a,\ b}$	**1.375**$^{a,\ b}$
HAR-ANN	**0.929**^{b}	**0.917**^{b}	**0.959**^{b}	**0.878**^{b}	**0.905**^{b}	**0.898**$^{a,\ b}$	**0.779**^{b}	**0.790**^{b}	**0.813**^{b}	**0.741**$^{a,\ b}$	**0.770**$^{a,\ b}$	**0.786**$^{a,\ b}$	**1.443**^{b}	**1.517**^{b}	**1.450**^{b}	**1.239**^{b}	**1.378**$^{a,\ b}$	**1.372**$^{a,\ b}$
MAE	$h = 1$						$h = 1$						$h = 1$					
GARCH	0.358	0.362	0.373	0.347	0.358	0.374	0.326	0.326	0.339	0.318	0.330	0.350	0.605	0.621	0.632	0.574	0.581	0.628
ARFIMA	**0.285**	**0.288**^{b}	**0.307**	**0.262**^{b}	**0.278**^{b}	**0.272**^{a}	**0.253**	**0.256**	**0.275**	**0.223**	**0.247**	**0.242**^{a}	0.501	0.530	0.526	**0.358**^{b}	**0.427**^{a}	**0.426**^{a}
HAR	**0.287**^{b}	**0.292**^{b}	**0.311**^{b}	**0.263**^{b}	**0.282**^{b}	**0.276**$^{a,\ b}$	**0.253**^{b}	**0.253**^{b}	**0.274**^{b}	**0.223**^{b}	**0.244**^{b}	**0.241**$^{a,\ b}$	0.488	0.515	0.517	**0.358**^{b}	**0.423**^{a}	**0.424**^{a}
ANN	0.288^{b}	0.293^{b}	0.313^{b}	**0.266**^{b}	0.284^{b}	0.278$^{a,\ b}$	0.255^{b}	0.258^{b}	0.279^{b}	0.226^{b}	0.249^{b}	0.245$^{a,\ b}$	**0.478**^{b}	**0.503**^{b}	**0.512**^{b}	**0.359**^{b}	0.425$^{a,\ b}$	0.427$^{a,\ b}$
HAR-ANN	**0.286**^{b}	**0.291**^{b}	**0.310**^{b}	**0.263**^{b}	**0.282**^{b}	**0.276**$^{a,\ b}$	**0.253**^{b}	**0.254**^{b}	**0.275**^{b}	**0.224**^{b}	**0.245**^{b}	**0.242**$^{a,\ b}$	**0.480**^{b}	**0.504**^{b}	**0.511**^{b}	**0.357**^{b}	**0.423**$^{a,\ b}$	**0.424**$^{a,\ b}$
	$h = 5$						$h = 5$						$h = 5$					
GARCH	0.667	0.675	0.666	0.677	0.680	0.721	0.598	0.608	0.600	0.630	0.633	0.697	1.000	0.991	1.014	1.167	1.096	1.190
ARFIMA	**0.475**	**0.481**	**0.494**	**0.450**	**0.469**	**0.455**^{a}	**0.402**	**0.415**	**0.421**	**0.375**	**0.405**	**0.401**^{a}	**0.769**^{b}	**0.800**^{b}	**0.787**	**0.630**	**0.697**$^{a,\ b}$	**0.696**$^{a,\ b}$
HAR	**0.467**^{b}	**0.474**^{b}	**0.488**^{b}	**0.444**^{b}	**0.464**^{b}	**0.452**$^{a,\ b}$	**0.409**	**0.417**	**0.426**	**0.381**	**0.405**	**0.401**^{a}	**0.760**^{b}	**0.796**^{b}	**0.786**^{b}	**0.637**^{b}	**0.706**$^{a,\ b}$	**0.707**$^{a,\ b}$
ANN	0.478^{b}	0.482^{b}	0.502^{b}	0.454^{b}	0.472^{b}	0.465$^{a,\ b}$	0.429^{b}	0.437^{b}	0.451^{b}	0.399^{b}	0.426^{b}	0.423$^{a,\ b}$	**0.762**^{b}	**0.799**^{b}	0.794^{b}	0.642^{b}	**0.705**$^{a,\ b}$	**0.713**$^{a,\ b}$
HAR-ANN	**0.468**^{b}	**0.475**^{b}	**0.490**^{b}	**0.445**^{b}	**0.465**^{b}	**0.454**$^{a,\ b}$	0.415^{b}	0.423^{b}	0.434^{b}	0.387^{b}	0.412^{b}	0.407$^{a,\ b}$	**0.755**^{b}	**0.791**^{b}	**0.784**^{b}	**0.635**^{b}	**0.700**$^{a,\ b}$	**0.705**$^{a,\ b}$
	$h = 10$						$h = 10$						$h = 10$					
GARCH	0.957	0.969	0.951	0.978	0.977	1.033	0.845	0.860	0.842	0.903	0.898	0.988	1.445	1.424	1.445	1.701	1.592	1.724
ARFIMA	**0.695**	**0.701**	**0.716**	**0.668**^{b}	**0.689**	**0.671**^{a}	**0.566**	**0.589**	**0.582**	**0.534**	**0.576**	**0.576**^{a}	**1.091**	**1.137**^{b}	**1.103**	**0.913**^{b}	**1.007**^{a}	**0.982**^{a}
HAR	**0.665**	**0.669**	**0.689**	**0.637**^{b}	**0.657**	**0.652**$^{a,\ b}$	**0.577**^{b}	**0.585**^{b}	**0.588**	**0.552**^{b}	**0.576**^{b}	**0.582**$^{a,\ b}$	1.096^{b}	**1.139**^{b}	1.107^{b}	**0.929**^{b}	1.034$^{a,\ b}$	1.007$^{a,\ b}$
ANN	0.675^{b}	0.676^{b}	0.701^{b}	**0.639**^{b}	0.665^{b}	0.666$^{a,\ b}$	0.604^{b}	0.616^{b}	0.631^{b}	0.581^{b}	0.603^{b}	0.620$^{a,\ b}$	**1.078**^{b}	**1.132**^{b}	**1.089**^{b}	**0.922**^{b}	**1.010**$^{a,\ b}$	**0.985**$^{a,\ b}$
HAR-ANN	**0.663**^{b}	**0.665**^{b}	**0.688**^{b}	**0.632**^{b}	**0.655**^{b}	**0.652**$^{a,\ b}$	**0.583**^{b}	**0.593**^{b}	0.600^{b}	**0.559**^{b}	0.583^{b}	**0.592**$^{a,\ b}$	**1.076**^{b}	**1.123**^{b}	**1.086**^{b}	**0.919**^{b}	**1.012**$^{a,\ b}$	**0.987**$^{a,\ b}$

Table 6

Economic comparison of forecasts: whole period. The table reports unconditional coverage for VaR (long). In addition, models are compared through loss function using the Model Confidence Set (MSC) is used to compare the errors row-wise (across forecasting models) as well as column-wise (across realized measures). We use (^{a}) to denote the volatility measures that belong to the $\widehat{\mathcal{M}}^*_{10\%}$ and (^{b}) to denote the forecasting models that belong to the $\widehat{\mathcal{M}}^*_{10\%}$. Moreover, each of the forecasting models is benchmarked to the rest of the competing models using the Superior Predictive Ability (SPA) test. Cases where the null hypothesis that the benchmark model is the best forecasting model cannot be rejected are set in bold.

	Crude oil						Heating oil						Natural gas					
	TSRV	RV	RK	JWTSRV	CBV	MedRV	TSRV	RV	RK	JWTSRV	CBV	MedRV	TSRV	RV	RK	JWTSRV	CBV	MedRV
1% VaR	$h=1$						$h=1$						$h=1$					
GARCH	1.349	1.349	1.349	1.349	1.349	1.349	1.850	1.850	1.850	1.850	1.850	1.850	1.174	1.174	1.174	1.174	1.174	1.174
ARFIMA	1.901	**1.901**	2.023	2.452^{a}	2.330^{a}	**2.820**^{a,b}	1.726	1.603	1.726	2.035	1.911	2.035^{a}	**1.853**^{b}	**1.606**^{b}	**1.853**	**2.409**^{b}	**1.915**^{a,b}	**2.162**^{a,b}
HAR	1.594	**1.533**^{b}	1.594	2.085	**1.901**^{b}	**2.269**^{a,b}	1.850	1.726	1.726	2.096	2.096	2.281^{a}	**1.915**^{b}	**1.729**^{b}	2.038	**2.471**^{b}	**1.977**^{a,b}	**2.285**^{a,b}
ANN	1.655	1.717^{b}	1.717	2.207^{a}	2.085^{a,b}	**2.391**^{a,b}	1.788	1.788	1.665	2.096	2.035	2.281^{a}	**1.482**^{b}	**1.606**^{b}	**1.668**^{b}	**2.409**^{a,b}	**1.977**^{a,b}	**2.162**^{a,b}
HAR-ANN	**1.655**^{b}	**1.655**^{b}	**1.533**^{b}	**2.146**^{b}	**1.901**^{b}	**2.330**^{a,b}	**1.850**^{b}	**1.788**^{b}	**1.665**^{b}	**2.035**^{b}	**2.035**^{b}	**2.281**^{a,b}	**1.668**^{b}	**1.606**^{b}	**1.791**^{b}	**2.532**^{a,b}	**1.977**^{a,b}	**2.224**^{a,b}
	$h=5$						$h=5$						$h=5$					
GARCH	1.659	1.659	1.659	1.659	1.659	1.659	1.669	1.669	1.669	1.669	1.669	1.669	1.053	1.053	1.053	1.053	1.053	1.053
ARFIMA	1.659	1.598^{a}	1.782^{a}	1.844^{a}	1.782^{a}	2.151^{a}	1.483	**1.483**^{b}	1.422	**1.607**^{b}	**1.545**^{a,b}	**1.792**^{a,b}	1.362	1.176	1.362	2.167	2.043^{a}	2.291^{a}
HAR	1.291	1.291	1.291	1.414	1.352	1.598^{a}	1.669	**1.545**^{b}	1.545	**1.669**^{b}	**1.669**^{b}	**1.916**^{a,b}	0.867	0.867	0.867	1.981	1.548^{a}	1.796^{a}
ANN	1.352	1.229	1.414	1.352	1.291	1.659^{a}	1.483	**1.422**^{b}	1.483	**1.669**^{b}	**1.607**^{a,b}	**1.978**^{a,b}	1.053	0.805	0.991	1.981	1.548	1.610
HAR-ANN	**1.352**^{b}	**1.229**^{b}	**1.291**^{b}	**1.414**^{b}	**1.352**^{b}	**1.721**^{a,b}	**1.607**^{b}	**1.545**^{b}	**1.545**^{b}	**1.731**^{b}	**1.669**^{a,b}	**1.916**^{a,b}	**0.929**^{b}	**0.867**^{b}	**0.991**^{b}	**2.043**^{b}	**1.424**^{a,b}	**1.858**^{a,b}
	$h=10$						$h=10$						$h=10$					
GARCH	1.850	1.850	1.850	1.850	1.850	1.850	2.294^{b}	**2.294**	**2.294**^{b}	2.294	2.294	2.294	1.180	1.180	1.180	1.180	1.180	1.180
ARFIMA	2.466	2.466	2.589	2.589	2.589	2.959^{a}	**2.294**^{b}	**2.356**^{b}	**2.170**^{b}	**2.728**^{b}	**2.728**^{a,b}	**3.100**^{a,b}	1.491	**1.429**	**1.491**^{b}	**2.236**^{b}	**1.615**^{a,b}	**2.050**^{a,b}
HAR	1.480	1.418	1.480	1.788	1.665	2.035^{a}	**2.108**^{b}	**2.294**^{b}	**2.108**^{b}	**2.666**^{b}	**2.542**^{a,b}	**2.728**^{a,b}	1.242	1.242	**1.304**^{b}	**1.863**^{b}	**1.553**^{b}	**1.801**^{b}
ANN	1.480	1.541	1.418	1.726	1.541^{a}	2.158^{a}	**2.294**^{b}	**2.232**^{b}	**2.108**^{b}	**2.666**^{a,b}	**2.604**^{a,b}	**3.224**^{a,b}	**1.366**^{b}	**1.366**	**1.366**^{b}	**1.988**^{b}	**1.615**^{b}	**1.863**^{b}
HAR-ANN	**1.480**^{b}	**1.480**^{b}	**1.480**^{b}	**1.726**^{b}	**1.541**^{b}	**1.973**^{a,b}	**2.232**^{b}	**2.170**^{b}	**2.108**^{b}	**2.604**^{a,b}	**2.604**^{a,b}	**2.976**^{a,b}	**1.242**^{b}	**1.242**^{b}	**1.366**^{b}	**1.863**^{b}	**1.615**^{b}	**1.739**^{b}
5% VaR	$h=1$						$h=1$						$h=1$					
GARCH	5.457	5.457	5.457	5.457	5.457	5.457	4.932	4.932	4.932	4.932	4.932	4.932	5.250	5.250	5.250	5.250	5.250	5.250
ARFIMA	6.193	6.131	6.254	6.560	6.499	6.99^{a}	6.165	6.165	6.104	6.782	6.535	7.398^{a}	**6.547**^{b}	**6.177**	**6.733**	**7.350**^{b}	**7.103**^{a,b}	**7.844**^{a,b}
HAR	5.763	5.886	5.947	6.070	6.009	6.683^{a}	5.980	6.042	6.042	6.720	6.597	7.583^{a}	**6.362**^{b}	**6.362**^{b}	6.733	**7.597**^{b}	**7.165**^{a,b}	**7.721**^{a,b}
ANN	6.009	6.070	5.886	6.193	6.254	6.560^{a}	6.104	6.042	5.980	6.843	6.289	7.707^{a}	**6.177**^{b}	**6.238**^{b}	**6.733**^{b}	**7.659**^{b}	**7.165**^{a,b}	**7.906**^{a,b}
HAR-ANN	**5.886**^{b}	**5.886**^{b}	**5.886**^{b}	**6.131**^{b}	**6.193**^{b}	**6.622**^{a,b}	**5.795**^{b}	**6.104**^{b}	**5.919**^{b}	**6.658**^{b}	**6.473**^{b}	**7.645**^{a,b}	**6.424**^{b}	**6.300**^{b}	**6.609**^{b}	**7.474**^{b}	**6.980**^{a,b}	**7.783**^{a,b}
	$h=5$						$h=5$						$h=5$					
GARCH	4.548	4.548	4.548	4.548	4.548	4.548	6.428	6.428	6.428	6.428	6.428	6.428	4.706	4.706	4.706	4.706	4.706	4.706
ARFIMA	5.839	5.839^{a}	6.023^{a}	**6.515**^{a}	**6.392**^{a}	**7.068**^{a,b}	**7.293**^{b}	**7.293**^{b}	**7.231**^{b}	**7.911**^{b}	**7.726**^{a,b}	**8.653**^{a,b}	6.378	6.130	6.625	**8.111**^{b}	**7.678**^{a}	**8.421**^{a}
HAR	4.733	4.917	4.794	5.224	5.163^{b}	**6.146**^{a,b}	**6.737**^{b}	**6.922**^{b}	**6.860**^{b}	**7.540**^{b}	**7.355**^{a,b}	**8.405**^{a,b}	6.006	5.573	6.068	**8.050**^{b}	7.492^{a,b}	8.297^{a,b}
ANN	4.610	4.917	4.733	5.163	**5.224**^{b}	**6.146**^{a,b}	**6.984**^{b}	**6.922**^{b}	**6.984**^{b}	**7.726**^{b}	**7.540**^{a,b}	**8.653**^{a,b}	6.006	5.820	6.130	**7.926**^{b}	**7.368**^{a,b}	**8.297**^{a,b}
HAR-ANN	**4.733**^{b}	**4.978**^{b}	**4.856**^{b}	**5.286**^{b}	**5.163**^{b}	**6.146**^{a,b}	**6.860**^{b}	**6.984**^{b}	**6.737**^{b}	**7.540**^{b}	**7.540**^{a,b}	**8.529**^{a,b}	**6.068**^{b}	**5.573**^{b}	**6.192**^{b}	**7.926**^{b}	**7.307**^{a,b}	**8.235**^{a,b}
	$h=10$						$h=10$						$h=10$					
GARCH	5.487	5.487	5.487	5.487	5.487	5.487	8.122	8.122	8.122	8.122	8.122	8.122	4.410	4.410	4.410	4.410	4.410	4.410
ARFIMA	6.782	6.658	6.720	**7.213**^{a,b}	**6.905**^{a}	**8.138**^{a,b}	**8.865**^{b}	**9.175**^{b}	**8.927**^{b}	**9.857**^{a,b}	**9.733**^{a,b}	**10.353**^{a,b}	4.907	4.720	5.155	**6.584**^{b}	**5.839**^{a,b}	**6.894**^{a,b}
HAR	5.179	5.425	5.055	**5.734**^{b}	5.734	**6.412**^{a,b}	**8.617**^{b}	**8.555**^{b}	**8.493**^{a,b}	**9.361**^{a,b}	**9.237**^{a,b}	**9.857**^{a,b}	4.658^{b}	4.348	4.534	**6.149**^{b}	5.404^{a,b}	6.460^{a,b}
ANN	5.425	5.302	5.364	**5.795**^{b}	5.610^{a}	**6.473**^{a,b}	**8.617**^{b}	**8.803**^{b}	**8.679**^{b}	**9.547**^{a,b}	**9.361**^{a,b}	**10.167**^{a,b}	**4.845**^{b}	4.472	4.783	**6.460**^{b}	**5.528**^{a,b}	**6.584**^{a,b}
HAR-ANN	**5.179**^{b}	**5.302**^{b}	**5.302**^{b}	**5.919**^{b}	**5.610**^{b}	**6.289**^{a,b}	**8.679**^{b}	**8.555**^{b}	**8.493**^{a,b}	**9.423**^{a,b}	**9.361**^{a,b}	**9.857**^{a,b}	**4.720**^{b}	**4.410**^{b}	**4.720**^{b}	**6.398**^{b}	**5.528**^{a,b}	**6.584**^{a,b}

Table 7
Economic comparison of forecasts: whole period. The table reports unconditional coverage for VaR (short). In addition, models are compared through loss function using the Model Confidence Set (MSC) is used to compare the errors row-wise (across forecasting models) as well as column-wise (across realized measures). We use (^{a}) to denote the volatility measures that belong to the $\widehat{\mathcal{M}}^{*}_{10\%}$, and (^{b}) to denote the forecasting models that belong to the $\widehat{\mathcal{M}}^{*}_{10\%}$. Moreover, each of the forecasting models is benchmarked to the rest of the competing models using the Superior Predictive Ability (SPA) test. Cases where the null hypothesis that the benchmark model is the best forecasting model cannot be rejected are set in bold.

	Crude oil						Heating oil						Natural gas					
	TSRV	RV	RK	JWTSRV	CBV	MedRV	TSRV	RV	RK	JWTSRV	CBV	MedRV	TSRV	RV	RK	JWTSRV	CBV	MedRV
99% VaR	$h = 1$						$h = 1$						$h = 1$					
GARCH	**98.774**	**98.774**^{b}	**98.774**^{b}	**98.774**^{b}	98.774	98.774^{b}	99.260	99.260	99.260	99.260	99.260	99.260	99.259	99.259	99.259	99.259	99.259	99.259
ARFIMA	**98.651**	**98.590**^{b}	**98.406**^{b}	**98.345**^{a,b}	**98.406**^{a}	**98.099**^{a,b}	**98.767**^{b}	**98.644**^{b}	**98.767**^{b}	**98.705**^{b}	**98.582**^{b}	**98.274**^{a,b}	**98.703**^{b}	**98.826**^{b}	**98.888**^{b}	**98.023**^{b}	**98.394**^{a,b}	**98.147**^{a,b}
HAR	98.712^{a,b}	**98.651**^{a,b}	**98.712**^{a,b}	**98.467**^{a,b}	98.651^{a,b}	**98.283**^{a,b}	**98.767**^{b}	**98.520**^{b}	**98.890**^{b}	**98.582**^{b}	**98.582**^{b}	**98.335**^{a,b}	**98.765**^{b}	**98.950**^{b}	**98.888**^{b}	**97.962**^{a,b}	**98.456**^{a,b}	**98.085**^{a,b}
ANN	**98.712**^{b}	**98.774**^{a,b}	**98.651**^{a,b}	**98.651**^{a,b}	**98.651**^{a,b}	**98.099**^{a,b}	**98.829**^{b}	**98.582**^{b}	**98.829**^{b}	**98.644**^{b}	**98.459**^{b}	**98.335**^{a,b}	**98.641**^{b}	**98.888**^{b}	**98.888**^{a,b}	**97.900**^{a,b}	**98.394**^{a,b}	**98.147**^{a,b}
HAR-ANN	98.835^{b}	**98.774**^{a,b}	**98.835**^{a,b}	**98.590**^{a,b}	**98.590**^{a,b}	**98.345**^{a,b}	**98.705**^{b}	**98.582**^{b}	**98.890**^{b}	**98.582**^{b}	**98.582**^{b}	**98.274**^{a,b}	**98.765**^{b}	**98.950**^{b}	**98.888**^{a,b}	**97.962**^{a,b}	**98.394**^{a,b}	**97.962**^{a,b}
	$h = 5$						$h = 5$						$h = 5$					
GARCH	99.385	**99.385**	99.385	**99.385**^{b}	**99.385**^{b}	99.385^{b}	99.506	99.506	99.506	99.506	99.506	99.506	99.443	99.443	99.443	99.443	99.443	99.443
ARFIMA	99.262	99.262	99.262	**99.017**^{b}	**99.017**^{a}	**98.648**^{a,b}	99.320	99.320	99.320	99.258	99.258	**99.073**^{a}	99.257	99.319	99.257	98.576	98.824^{a}	98.638^{a}
HAR	99.385	99.508	99.447	**99.262**^{b}	**99.447**^{b}	**99.078**^{a,b}	99.444	99.506	99.567	99.197	99.382	**99.073**^{a}	99.257	99.319	99.257	98.638	98.700^{a}	98.700^{a}
ANN	99.570	99.508	99.508	**99.447**^{b}	**99.447**^{b}	**99.201**^{a,b}	99.506	99.444	99.506	99.258	99.382	99.197^{a}	99.195	99.381	99.195	98.638	98.824^{a}	98.700^{a}
HAR-ANN	**99.570**^{b}	**99.508**^{b}	**99.508**^{b}	**99.385**^{b}	**99.447**^{b}	**99.201**^{a,b}	**99.567**^{b}	**99.444**^{b}	**99.506**^{b}	**99.197**^{b}	**99.320**^{b}	**99.135**^{a,b}	**99.257**^{b}	**99.443**^{b}	**99.257**^{b}	**98.638**^{b}	**98.824**^{a,b}	**98.700**^{a,b}
	$h = 10$						$h = 10$						$h = 10$					
GARCH	99.630	99.630	99.630	99.630	99.630	99.630	99.752	99.752	99.752	99.752	99.752	99.752	99.814	99.814	99.814	99.814	99.814	99.814
ARFIMA	99.260	99.260	99.260	99.075	99.260	98.890^{a}	99.814	99.876	99.814	99.628	99.876	99.442^{a}	99.689	99.627	99.627	99.317	99.565^{a}	99.255^{a}
HAR	99.753	99.753	99.692	99.630	99.630	99.383^{a}	99.814	99.752	99.814	99.628	99.752	99.442^{a}	99.627	99.627	99.565	99.379	99.441^{a}	99.379^{a}
ANN	99.630	99.630	99.568	99.445	99.630	99.322^{a}	99.690	99.752	99.752	99.566	99.690	99.504^{a}	99.565	99.379	99.503	99.317	99.441^{a}	99.193^{a}
HAR-ANN	**99.630**^{b}	**99.692**^{b}	**99.630**^{b}	**99.507**^{b}	**99.630**^{b}	**99.322**^{a,b}	**99.752**^{b}	**99.752**^{b}	**99.814**^{b}	**99.628**^{b}	**99.752**^{b}	**99.442**^{a,b}	**99.565**^{b}	**99.441**^{b}	**99.503**^{b}	**99.379**^{b}	**99.503**^{a,b}	**99.317**^{a,b}
95% VaR	$h = 1$						$h = 1$						$h = 1$					
GARCH	94.421	94.421	94.421	94.421	94.421	94.421	95.561	95.561	95.561	95.561	95.561	95.561	95.738	95.738	95.738	95.738	95.738	95.738
ARFIMA	**94.482**^{b}	**94.298**^{b}	**94.237**^{a,b}	**93.746**^{a}	**93.930**^{a}	**93.256**^{a,b}	**95.253**^{b}	**95.191**^{b}	**95.253**^{b}	**94.636**^{b}	**94.698**^{b}	**93.896**^{a,b}	**94.441**^{b}	**94.565**^{b}	**94.194**^{b}	**93.206**^{b}	**93.638**^{a,b}	**92.959**^{a,b}
HAR	94.543^{b}	94.359^{b}	94.421^{b}	94.237	94.237	**93.624**^{a,b}	**95.068**^{b}	**95.068**^{b}	**95.006**^{b}	**94.451**^{b}	**94.451**^{b}	**93.527**^{a,b}	**94.194**^{b}	**94.194**^{b}	**93.885**^{b}	**93.144**^{b}	**93.391**^{a,b}	**92.588**^{a,b}
ANN	**94.850**^{b}	94.359^{b}	94.605^{b}	94.543	94.482	**93.930**^{a,b}	**95.438**^{b}	**95.314**^{b}	**95.314**^{b}	**94.760**^{b}	**94.575**^{b}	**93.773**^{a,b}	**94.132**^{b}	**94.194**^{b}	**94.256**^{b}	**92.835**^{b}	**93.391**^{a,b}	**92.712**^{a,b}
HAR-ANN	**94.788**^{b}	**94.421**^{b}	**94.727**^{b}	**94.298**^{b}	**94.359**^{b}	**93.746**^{a,b}	**95.191**^{b}	**95.129**^{b}	**95.191**^{b}	**94.513**^{b}	**94.513**^{b}	**93.835**^{a,b}	**94.009**^{b}	**94.132**^{b}	**94.132**^{b}	**92.959**^{b}	**93.515**^{a,b}	**92.773**^{a,b}
	$h = 5$						$h = 5$						$h = 5$					
GARCH	95.206	95.206	95.206	**95.206**	95.206	95.206	96.601	96.601	96.601	96.601	96.601	96.601	97.028	97.028	97.028	97.028	97.028	97.028
ARFIMA	94.776	94.530	94.345	94.161	94.038^{a}	93.239^{a}	95.797	95.859	95.921	95.426	95.612	95.179^{a}	95.604	96.037	95.789	94.737	95.480^{a}	94.923^{a}
HAR	95.698	95.882	95.698	95.022	95.452	94.407^{a}	95.921	95.921	95.983	95.488	95.612	95.056^{a}	96.223	96.285	96.223	94.923	95.604^{a}	95.046^{a}
ANN	95.636	95.698	95.636	95.022	95.267	94.345^{a}	95.859	95.921	95.921	95.488	95.797	95.117^{a}	96.037	96.285	95.975	94.737	95.356^{a}	95.046^{a}
HAR-ANN	**95.698**^{b}	**95.882**^{b}	**95.698**^{b}	**95.083**^{b}	**95.452**^{b}	**94.407**^{a,b}	**95.797**^{b}	**95.921**^{b}	**95.921**^{b}	**95.426**^{b}	**95.797**^{b}	**95.117**^{a,b}	**96.223**^{b}	**96.285**^{b}	**96.161**^{b}	**94.799**^{b}	**95.418**^{a,b}	**95.046**^{a,b}
	$h = 10$						$h = 10$						$h = 10$					
GARCH	95.623	95.623	95.623	95.623	95.623	95.623	96.714	96.714	96.714	96.714	96.714	96.714	98.137	98.137	98.137	98.137	98.137	98.137
ARFIMA	95.191	95.376	95.006	94.698	94.575	93.711^{a}	95.970	96.156	96.032	95.598	95.970	95.040^{a}	97.267	97.267	f96.957	96.025	96.584^{a}	95.963^{a}
HAR	95.993	95.993	96.054	95.746	95.869	95.314^{a}	95.846	95.908	95.970	95.350	95.474	95.102^{a}	97.516	97.516	97.143	96.211	96.646^{a}	96.087^{a}
ANN	95.993	95.931	96.239	95.869	95.869	95.253^{a}	95.908	96.156	96.156	95.598	96.032	95.226^{a}	97.205	97.267	96.957	96.087	96.584^{a}	96.149^{a}
HAR-ANN	**95.993**^{b}	**96.054**^{b}	**96.054**^{b}	**95.808**^{b}	**95.993**^{b}	**95.191**^{a,b}	**95.846**^{b}	**96.094**^{b}	**95.970**^{b}	**95.350**^{b}	**95.784**^{b}	**95.102**^{a,b}	**97.267**^{b}	**97.453**^{b}	**97.081**^{b}	**96.087**^{b}	**96.646**^{a,b}	**96.211**^{a,b}

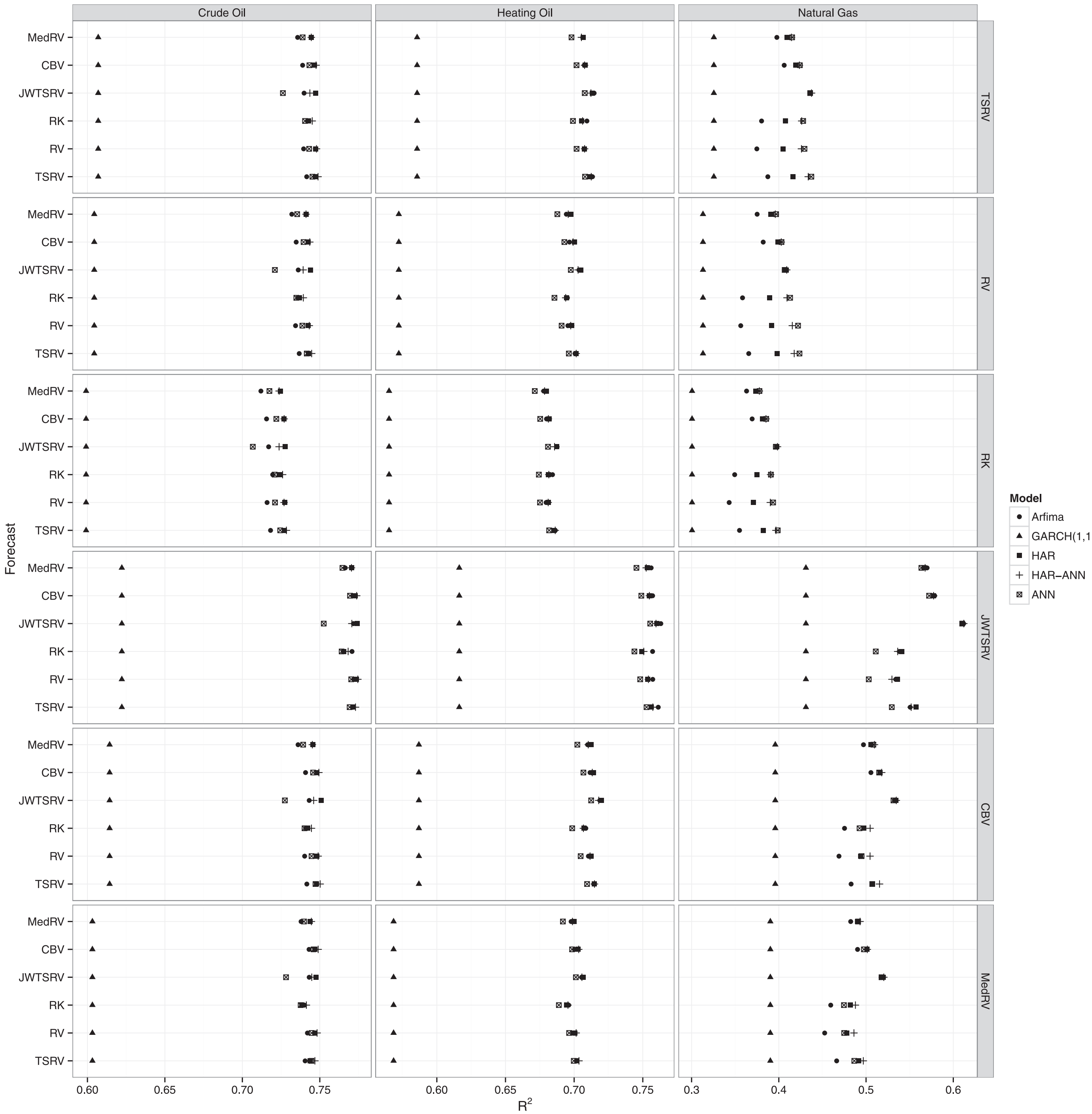

Fig. 3. Whole period: R^2 from the Mincer–Zarnowitz regressions for the 1-day forecast horizon.

Following Patton and Sheppard (2009), we estimate the Mincer–Zarnowitz (MZ) regression using Generalized Least Squares (GLS), employing the form $\hat{\nu}_{t+h}^{RM}/\hat{\nu}_{t+h}^{(RM,f)} = \alpha/\hat{\nu}_{t+h}^{(RM,f)} + \beta + \epsilon_t^*$. In cases in which the forecast is unbiased, we expect $\alpha = 0$ and $\beta = 1$ jointly.

The results from the MZ regressions are reported in the online appendix for all periods. Testing the joint null hypothesis that $(\alpha, \beta) = (0, 1)$ shows us that after November 2010, we never reject the hypothesis that the parameters are significantly different for the high frequency data-based models—except for the heating oil for the last period. This finding leads us to the conclusion that all the forecasts are uniformly unbiased. As for the daily based GARCH model, the joint hypothesis is frequently rejected, leading us to the result that forecasts from the GARCH model are frequently biased.

Finally, we study R^2 from the regressions because it will tell us what portion of variance is explained by forecasts. The results from the MZ regressions for the entire period are incorporated into Figs. 3–5 for all forecasting horizons. We also include the R^2 results for all three periods in the online supplementary appendix. We observe from the figures that all the models perform well in all the forecasting horizons, with R^2 over 70% in all cases except for natural gas, which is forecasted 1-step-ahead for the high frequency data-based models. This is the expected result, as natural

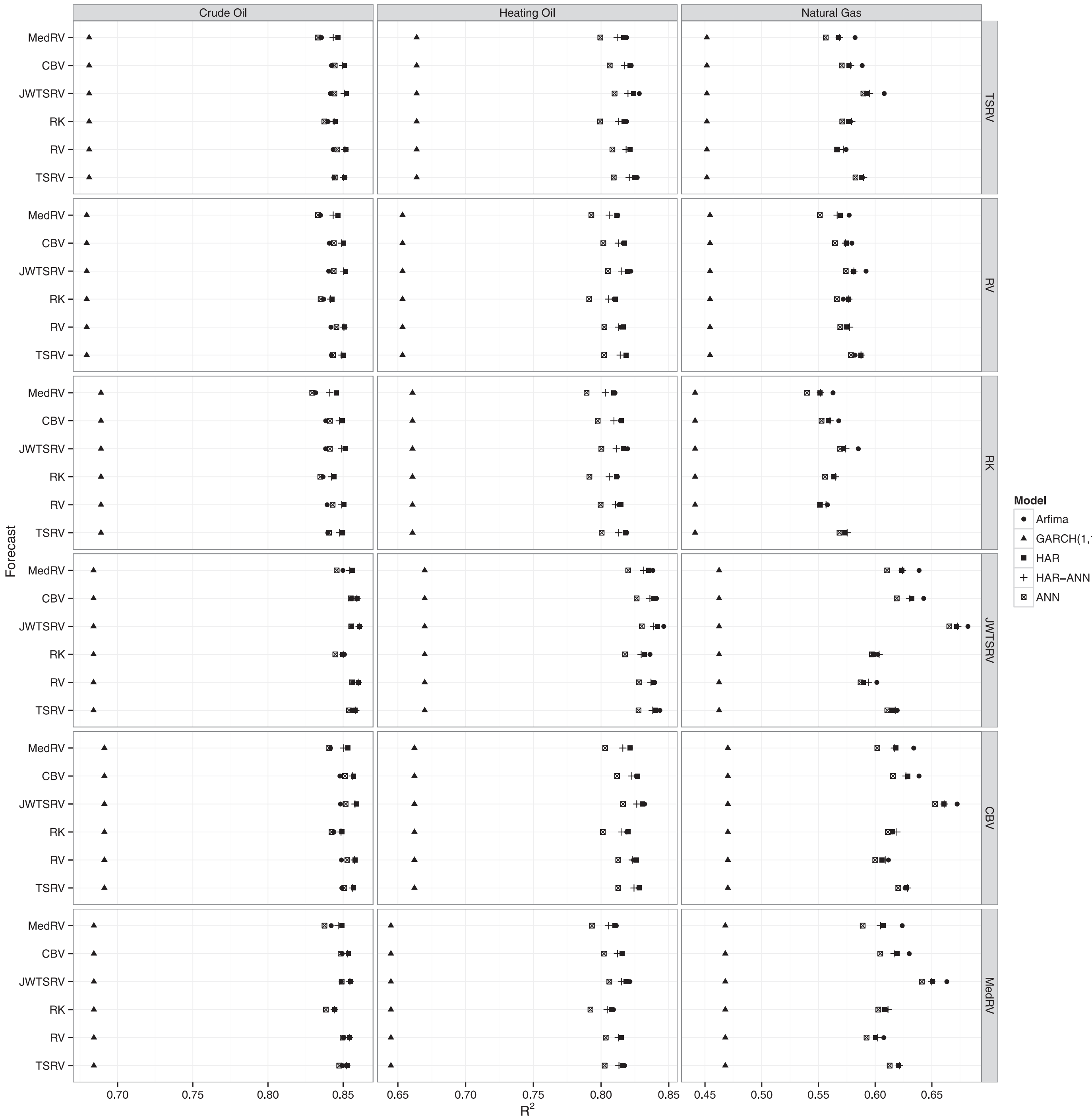

Fig. 4. Whole period: R^2 from the Mincer–Zarnowitz regressions for the 5-day forecast horizon.

gas shows the greatest degree of price variability, leading the models to be able to explain less variance. When comparing performance across models, we may conclude that all the models deliver similar accuracy of the explained variance, and the results are consistent with previous analyzes. Comparing the high frequency-based models to the low frequency GARCH model confirms the result from the previous analysis, as R^2 is more than 10% lower for all the GARCH forecasts in comparison with the competing models.

More interestingly, a distinction may be made when comparing realized measures. The JWTSRV, together with the CBV and MedRV, may be forecast with the highest degree of success on all horizons. Although a longer forecasting horizon implies less difference, we may conclude that measures of integrated volatility are the best choice when a forecaster requires an accurate forecast of a 'true' volatility process underlying the data. In addition to previous results that have indicated that the MedRV performs the best statistically, the results of this analysis find this simple measure to outperform the others.

5.3. Economic evaluation of forecasts

A model's statistical superiority does not necessarily translate to economic benefits; therefore, in addition to performing a statistical evaluation, we evaluate the forecasts economically. Quantile

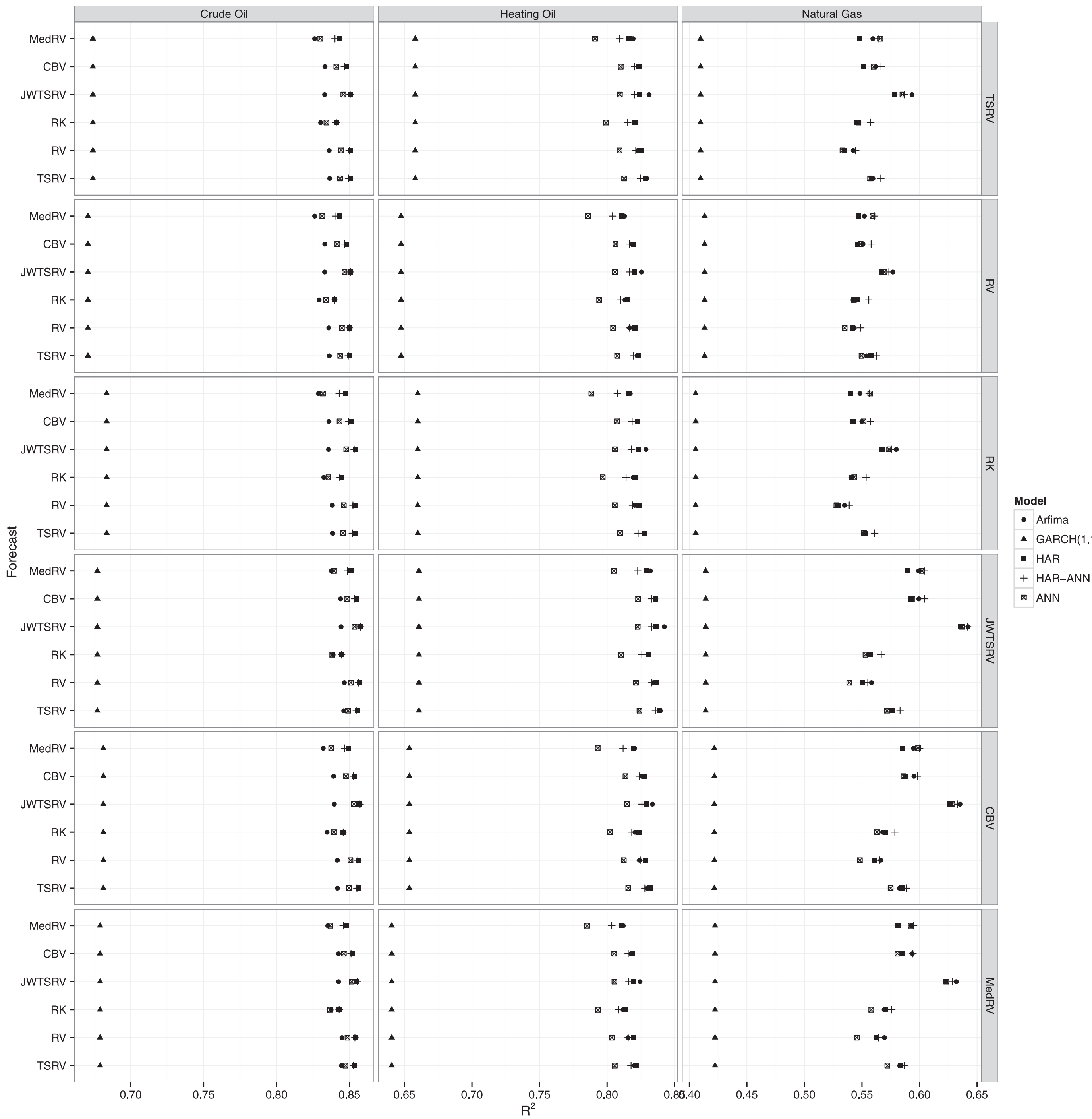

Fig. 5. Whole period: R^2 from the Mincer–Zarnowitz regressions for the 10-day forecast horizon.

forecasts are central to risk management decisions because of a widespread Value-at-Risk (VaR); therefore, we use VaR metrics for the economic evaluation of the forecasts. From the volatility forecasts, we compute 1% and 5% VaR for both long positions and short positions.

Although quantile forecasts may be readily evaluated by comparing their actual (estimated) coverage, $\widehat{C}_\alpha = 1/T\sum_{t=1}^{T} I_{\{y_{t+h}<\hat{q}^{\alpha}_{t+h}\}}$, against their nominal coverage rate, $C_\alpha = E[I_{\{y_{t+h}<q^{\alpha}_{t+h}\}}]$, with $\hat{q}^{\alpha}_{t+h}$ being h-step-ahead forecast of VaR at α, this approach reduces to the simple comparison of unconditional coverage rates. Therefore, we evaluate the accuracy of VaR forecasts statistically by defining the expected loss of VaR forecasts of Giacomini and Komunjer (2005) made by forecaster m as follows:

$$\widehat{L}_{\alpha,m} = E\left[\left(\alpha - I_{\{y_{t+h}<\widehat{q}^{\alpha,m}_{t+h}\}}\right)\left(y_{t+h} - \widehat{q}^{\alpha,m}_{t+h}\right)\right], \quad (19)$$

and VaR forecasts are tested using the same methodology as employed in the previous section, i.e., using MSC and SPA procedures. Again, we test the performance across both forecasting models and realized measures.

To conserve space, we discuss the economic evaluation of the results for the entire forecasted period, although the results from the three periods studied previously are the same, and the comparison of the forecasting performance does not change over time.

Tables 6 and7 report conditional coverage as well as a statistical comparison by means of the loss function of Giacomini and Komunjer (2005) that was described in previous sections for the long and short positions, at 1%, 5%, 95%, and 99% forecasts of return distribution.

Examining the model confidence set and the SPA results, the HAR-ANN model combination belongs to the model confidence set uniformly yielding the statistically best results. Notably, the ARFIMA model belongs to the model confidence set in many occasions. Forecasts from the realized volatility tend to overestimate VaR, forcing a forecaster to hold more capital than required. VaRs of 1%, 5%, 95% and 99% are forecasted on average at approximately 2%, 6%, 94%, and 98%. However, the results are much better than expected, as this is a well-documented feature of realized volatility forecasts.

Turning to the comparison of the VaR forecasts through realized measures used, it appears that although MedRV again provides the best statistical performance, it also yields greater bias in the unconditional coverage. This feature is common for measures of integrated variance, and it is expected, as they do not include jumps, although the forecasts are compared with the original returns containing jumps. Therefore, to use these measures, it is recommended to also include the jump variation. This approach is nevertheless beyond the scope of this study. We conduct an economic evaluation as a robustness check for the results from the statistical evaluation. The general conclusion is that the results from the statistical evaluation materialized into economic benefits.

Finally, we compare the economic value of forecasts from high frequency data-based models to the GARCH model. In most situations, the statistical results translate also to significant economic gains, as the GARCH is outperformed almost in all the situations except $h = 1$ in 99% VaR for crude oil, in which its performance can not be statistically distinguished. Thus, high frequency data do contribute to better VaR forecasts in most of the situations.

6. Conclusion

Predicting energy price variability is of immense interest to both practitioners and the academic literature. Nonetheless, most relevant studies focus on the usage of daily data and rely on the popular GARCH-type models when predicting the volatility of energy prices. Many recent studies in expert and intelligent systems implemented neural networks for forecasting volatility with intention to improve the volatility forecasts (Cheng & Wei, 2009; Hajizadeh, Seifi, Zarandi, & Turksen, 2012; Kristjanpoller, Fadic, & Minutolo, 2014; Kristjanpoller & Minutolo, 2015; Roh, 2007). Although all the studies focus on the daily data, a few recent works utilize high frequency data in expert systems for trading (Araújo, Oliveira, & Meira, 2015; Kotkatvuori-Örnberg, 2016).

In this paper, we contribute to this literature by combining the information included in high frequency data with popular artificial neural networks to improve volatility forecasts. Precise volatility forecasting is the core issue in risk management, as portfolio pricing, hedging, and option strategies rely on it heavily. Hence our results not only contribute to the academic literature but also are of great importance for market participants and practitioners as precise forecasts of volatility translate directly to precise forecasts of risk. The results are also important because of increased interest in volatility trading and hedging.

Examining the most liquid energy commodity markets of crude oil, heating oil, and natural gas, we comprehensively evaluate the most popular models for realized volatility forecasting. We test the widely used HAR and ARFIMA models against the simple ANN using the Model Confidence Set and Superior Predictive Ability. Moreover, we use realized variance, realized kernel, two-scale realized variance, bipower variation, median realized volatility, and the recently proposed jump-adjusted wavelet two-scale realized variance measures of volatility. Driven by the possible reduction of model uncertainty, we also experiment with the linear combination of forecasts from the popular HAR model and the ANN, which yields the lowest error uniformly through all tested periods. These errors also translate to economic benefits in terms of VaR. In addition, we find that high frequency data-based forecasting strategies substantially outperform the benchmark GARCH model.

Our main finding is that coupling realized measures with artificial neural networks results in both statistical and economic gains. Although the proposed methodology delivers less precise short-term forecasts during the crisis period, the forecasts remain economically valuable. Importantly, the methodology reduced the tendency to over-predict the volatility confirmed by previous research. Even in those cases in which the model is fit on the data during a period of high uncertainty and forecasts a period of reduced uncertainty, the results hold. Therefore, the findings hold uniformly throughout the tested periods, and the methodology yields substantial advances to previously used methodologies, which tend to over-predict the volatility. Another important finding is that the median realized volatility is preferred as the best approximation of volatility when we are interested in forecasting. This result holds across all studied periods, and models with median realized volatility deliver the best forecasts both statistically and economically.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.eswa.2016.02.008.

References

Andersen, T., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review, 39*(4), 885–905.

Andersen, T., Bollerslev, T., Diebold, F., & Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association, 96*(453), 42–55.

Andersen, T., Bollerslev, T., Diebold, F., & 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*(1), 176–189.

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., & Diebold, F. X. (2006). Volatility and correlation forecasting. *Handbook of economic forecasting, 1*, 777–878.

Andersen, T. G., Dobrev, D., & Schaumburg, E. (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics, 169*(1), 75–93.

Araújo, R. d. A., Oliveira, A. L., & Meira, S. (2015). A hybrid model for high-frequency stock market forecasting. *Expert Systems with Applications, 42*(8), 4081–4096.

Arneodo, A., Muzy, J., & Sornette, D. (1998). Casual cascade in stock market from the "infrared"to the "ultraviolet". *European Physical Journal B*, (2), 277–282.

Bandi, F., & Russell, J. (2006). Separating microstructure noise from volatility. *Journal of Financial Economics*, (79), 655–692.

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

Barndorff-Nielsen, O., & Shephard, N. (2001). Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society, Series B, 63*(2), 167–241.

Barndorff-Nielsen, O., & Shephard, N. (2002a). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B, 64*(2), 253–280.

Barndorff-Nielsen, O., & Shephard, N. (2002b). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics, 17*(5), 457–477.

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

Barndorff-Nielsen, O., & Shephard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics, 4*(1), 1–30.

Barndorff-Nielsen, O. E., & Shephard, N. (2007). Variation, jumps, market frictions and high frequency data in financial econometrics. In *Proceedings of the theory and applications, ninth world congress advances in economics and econometrics.: Vol. 3* (pp. 328–372). http://www.forskningsdatabasen.dk/en/catalog/2185870445

Barunik, J., Krehlik, T., & Vacha, L. (2016). Modeling and forecasting exchange rate volatility in time-frequency domain. *European Journal of Operational Research, 251*(1), 329–340.
Barunik, J., & Vacha, L. (2015). Realized wavelet-based estimation of integrated variance and jumps in the presence of noise. *Quantitative Finance, 15*(8), 1347–1364.
Baum, C. F., & Zerilli, P. (2016). Jumps and stochasticvolatility in crude oil futures prices using conditional moments of integrated volatility. *Energy Economics, 53*, 175–181.
Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*(3), 307–327.
Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economic and Statistics, 69*, 542–547.
Brailsford, T. J., & Faff, R. W. (1996). An evaluation of volatility forecasting techniques. *Journal of Banking & Finance, 20*(3), 419–438.
Cheng, C.-H., & Wei, L.-Y. (2009). Volatility model based on multi-stock index for taiex forecasting. *Expert Systems with Applications, 36*(3), 6187–6191.
Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics, 7*(2), 174–196.
Deo, R. S., Hurvich, C. M., & Lu, Y. (2006). Forecasting realized volatility using a long memory stochastic volatility model: Estimation, prediction and seasonal adjustment. *Journal of Econometrics, 131*(1–2), 29–58.
Donoho, D. L., & Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrica, 81*(3), 425–455.
Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica, 50*(4), 987–1007.
Fan, J., & Wang, Y. (2007). Multi-scale jump and volatility analysis for high-frequency financial data. *Journal of the American Statistical Association, 102*(480), 1349–1362.
Fan, Y., Liang, Q., & Wei, Y.-M. (2008). A generalized pattern matching approach for multi-step prediction of crude oil price. *Energy Economics, 30*(3), 889–904.
Giacomini, R., & Komunjer, I. (2005). Evaluation and combination of conditional quantile forecasts. *Journal of Business & Economic Statistics, 23*(4), 416–431.
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.
Hajizadeh, E., Seifi, A., Zarandi, M. F., & Turksen, I. (2012). A hybrid modeling approach for forecasting the volatility of S&P 500 index return. *Expert Systems with Applications, 39*(1), 431–436.
Hansen, P., & Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business and Economic Statistics, 24*(2), 127–161.
Hansen, P. R. (2005). A test forsuperior predictive ability. *Journal of Business & Economic Statistics, 23*(4), 365–380.
Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica, 79*(2), 453–497.
Haykin, S. (2007). *Neural networks: a comprehensive foundation*. Prentice Hall Englewood Cliffs, NJ.
Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks, 2*(5), 359–366.
Jammazi, R., & Aloui, C. (2012). Crude oil price forecasting: Experimental evidence from wavelet decomposition and neural network modeling. *Energy Economics, 34*(3), 828–841.
Kang, S. H., & Yoon, S.-M. (2013). Modeling and forecasting the volatility of petroleum futures prices. *Energy Economics, 36*, 354–362.
Kotkatvuori-Örnberg, J. (2016). Measuring actual daily volatility from high frequency intraday returns of the S&P futures and index observations. *Expert Systems with Applications, 43*, 213–222.
Kristjanpoller, W., Fadic, A., & Minutolo, M. C. (2014). Volatility forecast using hybrid neural network models. *Expert Systems with Applications, 41*(5), 2437–2442.
Kristjanpoller, W., & Minutolo, M. C. (2015). Gold price volatility: A forecasting approach using the artificial neural network–Garch model. *Expert Systems with Applications, 42*(20), 7245–7251.
Kuper, G. H., & van Soest, D. P. (2006). Does oil priceuncertainty affect energy use? *Energy Journal, 27*(1), 55–78.
Linn, S. C., & Zhu, Z. (2004). Natural gas prices and the gas storage report: Public news and volatility in energy futures markets. *Journal of futures markets, 24*(3), 283–313.
Lynch, P., & Zumbach, G. (2003). Market heterogeneities and the causal structure of volatility. *Quantitative Finance*, (3), 320–331.
McAleer, M., & Medeiros, M. (2008). Realized volatility: A review. *Econometric Reviews*, (27), 10–45.
McAleer, M., & Medeiros, M. C. (2011). Forecasting realized volatility with linear and nonlinear univariate models. *Journal of Economic Surveys, 25*(1), 6–18.
Merton, R. C. (1980). On estimating the expeted return on the market. *Journal of Financial Econometrics, 8*, 323–361.
Mincer, J., & Zarnowitz, V. (1969). *The evaluation of economic forecasts*. New York: National Bureau of Economic Research.
Mohammadi, H., & Su, L. (2010). International evidence on crude oil price dynamics: Applications of Arima–Garch models. *Energy Economics, 32*(5), 1001–1008.
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.
Nomikos, N. K., & Pouliasis, P. K. (2011). Forecasting petroleum futures markets volatility: The role of regimes and market conditions. *Energy Economics, 33*(2), 321–337.
Panella, M., Barcellona, F., & D'Ecclesia, R. L. (2012). Forecasting energy commodity prices using neural networks. *Advances in Decision Sciences, 2012*. Article ID 289810, 26 pages http://www.hindawi.com/journals/ads/2012/289810/
Papadimitriou, T., Gogas, P., & Stathakis, E. (2014). Forecasting energy markets using support vector machines. *Energy Economics, 44*, 135–142.
Patton, A. J., & Sheppard, K. (2009). Evaluating volatility and correlation forecasts. In *Handbook of financial time series* (pp. 801–838). Springer.
Percival, D. B., & Mofjeld, H. (1997). Analysis of subtidal coastal sea level fluctuations using wavelets. *Journal of the American Statistical Association, 92*(439), 868–880.
Percival, D. B., & Walden, A. T. (2000). *Wavelet methods for time series analysis*. Cambridge University Press.
Pindyck, R. S. (2004). Volatility and commodity price dynamics. *Journal of Futures Markets, 24*(11), 1029–1047.
Prokopczuk, M., Symeonidis, L., & Wese Simen, C. (2015). Do jumps matter for volatility forecasting? evidence from energy markets. *Journal of Futures Markets*. doi:10.1002/fut.21759.
Riedmiller, M., & Braun, H. (1993). A direct adaptive method for faster backpropagation learning: The RPROP algorithm. In *IEEE international conference on neural networks* (pp. 586–591). IEEE.
Roh, T. H. (2007). Forecasting the volatility of stock price index. *Expert Systems with Applications, 33*(4), 916–922.
Sermpinis, G., Theofilatos, K., Karathanasopoulos, A., Georgopoulos, E. F., & Dunis, C. (2013). Forecasting foreign exchange rates with adaptive neural networks using radial-basis functions and particle swarm optimization. *European Journal of Operational Research, 225*(3), 528–540.
Sévi, B. (2014). Forecasting the volatility of crude oil futures using intraday data. *European Journal of Operational Research, 235*(3), 643–659.
Wang, Y. (1995). Jump and sharp cusp detection via wavelets. *Biometrika, 82*(2), 385–397.
Wang, Y., & Wu, C. (2012). Forecasting energy market volatility using Garch models: Can multivariate models beat univariate models? *Energy Economics, 34*(6), 2167–2181.
Wei, Y., Wang, Y., & Huang, D. (2010). Forecasting crude oil market volatility: Further evidence using Garch-class models. *Energy Economics, 32*(6), 1477–1484.
Wilson, B., Aggarwal, R., & Inclan, C. (1996). Detecting volatility changes across the oil sector. *Journal of Futures Markets, 16*(3), 313–330.
Xiong, T., Bao, Y., & Hu, Z. (2013). Beyond one-step-ahead forecasting: Evaluation of alternative multi-step-ahead forecasting models for crude oil prices. *Energy Economics, 40*, 405–415.
Yang, C., Hwang, M.-J., & Huang, B.-N. (2002). An analysis of factors affecting price volatility of the us oil market. *Energy Economics, 24*(2), 107–119.
Yu, L., Wang, S., & Lai, K. K. (2008). Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Economics, 30*(5), 2623–2635.
Zhang, L., Mykland, P., & Aït-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high frequency data. *Journal of the American Statistical Association, 100*(472), 1394–1411.
Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates. *Journal of Business & Economic Statistics, 14*(1), 45–52.