Institute of Economic Studies, Faculty of Social Sciences
Charles University in Prague

Coupling High-Frequency Data with Nonlinear Models in Multiple-Step-ahead Forecasting of Energy Markets‘ Volatility

Jozef Baruník
Tomáš Křehlík

IES Working Paper: 30/2014

Institute of Economic Studies,
Faculty of Social Sciences,
Charles University in Prague

[UK FSV – IES]

Opletalova 26
CZ-110 00, Prague
E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Institut ekonomických studií
Fakulta sociálních věd
Univerzita Karlova v Praze

Opletalova 26
110 00 Praha 1

E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Bibliographic information:
Baruník J., Křehlík T. (2014). "Coupling high-frequency data with nonlinear models in multiple-step-ahead forecasting of energy markets' volatility" IES Working Paper 30/2014. IES FSV. Charles University.

Coupling High-Frequency Data with Nonlinear Models in Multiple-Step-ahead Forecasting of Energy Markets' Volatility

Jozef Baruník[a,b]
Tomáš Křehlík[a,b]

[a]Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Smetanovo nábreží 6, 111 01 Prague 1, Czech Republic
[b]Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod Vodarenskou Vezi 4, 182 00, Prague, Czech Republic
email: barunik@utia.cas.cz

September 2014

Abstract:

In the past decade, the popularity of realized measures and various linear models for volatility forecasting has attracted attention in the literature on the price variability of energy markets. However, results that would guide practitioners to a specifc estimator and model when aiming for the best forecasting accuracy are missing. This paper contributes to the ongoing debate with a comprehensive evaluation of multiple-step-ahead volatility forecasts of energy markets using several popular high-frequency measures and forecasting models. To capture the complex patterns hidden to linear models commonly used to forecast realized volatility, this paper also contributes to the literature by coupling realized measures with artificial neural networks as a forecasting tool. Forecasting performance is compared across models as well as realized measures of crude oil, heating oil, and natural gas volatility during three qualitatively distinct periods covering the precrisis period, recent global turmoil of markets in 2008, and the most recent post-crisis period. We conclude that coupling realized measures with arti_cial neural networks results in both statistical and economic gains, reducing the tendency to over-predict volatility uniformly during all tested periods. Our analysis favors the median realized volatility, as it delivers the best performance and is a computationally simple alternative for practitioners.

Keywords: artificial neural networks, realized volatility, multiple-step-ahead forecasts, energy markets
JEL: C14, C53, G17

Acknowledgements: Support from the Czech Science Foundation under the 13-32263S project is gratefully acknowledged. T. Krehlik gratefully acknowledges financial support from the Grant Agency of Charles University under the 588314 and 837413 projects.

1. Introduction

The prediction of 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, natural gas, crude oil, and heating oil, a leading product of energy markets,[1] contributed to a climate of uncertainty and distrust of energy companies and investors on the one side and consumers, regulators, and legislators on the other side. The high volatility in energy markets is likely due to supply uncertainty (depending on a variety of macroeconomic and political factors in crude oil or simply storage-level constraints in natural gas) and the short-term inelasticity of demand (difficulty of reducing consumption within a short period of time), which makes it extremely difficult for both consumers and producers to forecast their costs and profits. The desire to protect market participants against these losses has led to immense interest in empirical research focusing on the prediction of variability of energy prices.

Volatility research from previous decades, affected mainly by the work of Engle (1982) and Bollerslev (1986, 1987), showed that although it is difficult to forecast the direction of future price changes, price variability is much easier to understand. However, the vast majority of the research has focused on financial markets, with the focus only recently turning to the energy markets[2] (Wilson et al., 1996; Yang et al., 2002; Linn and Zhu, 2004; Pindyck, 2004; Kuper and van Soest, 2006; Mohammadi and Su, 2010; Wei et al., 2010; Kang and Yoon, 2013).

More recent advances in financial econometrics have led to the development of new estimators of volatility using high-frequency data, which made volatility observable. Whereas pioneering studies in the realized volatility literature recognized the benefits from using high-frequency data in terms of increased accuracy (Merton, 1980; Zhou, 1996), recent work[3] propose several estimators to improve the efficiency, robustness to market microstructure effects, and the ability to estimate the variation due to the continuous part of the price process separately from the variation due to the jump part of the price process. See Andersen et al. (2006), McAleer and Medeiros (2008), or Barndorff-Nielsen and Shephard (2007) for excellent reviews of the realized volatility literature.

Whereas the estimation of realized volatility is the first step to more accurate prediction, using an appropriate model is a second step. Popular heterogeneous autoregressive (HAR) models and autoregressive fractionally integrated (ARFIMA) models became widely used to forecast the realized volatility because they capture the long memory property of the volatility well (Corsi, 2009; Andersen et al., 2003). In contrast to FIGARCH models capturing the long memory of volatility using daily returns data,[4] these approaches are more flexible and easier to estimate once we have high-frequency data available. Whereas both HAR and ARFIMA are developed to capture a specific long memory feature of the volatility, there may be more complex patterns to be explored.

[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`).

[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 et al. (2001, 2003); Zhang et al. (2005); Bandi and Russell (2006); Hansen and Lunde (2006); Barndorff-Nielsen et al. (2008).

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

Artificial neural networks (ANN) may be viewed as a generalization of these classical approaches that may help to uncover the more complex patterns in volatility. Concisely, neural networks are semi-parametric non-linear models, which are able to approximate any reasonable function Haykin (2007); Hornik et al. (1989). Whereas the number of models using machine learning is rapidly growing in the academic literature, applications to forecasting in energy markets are very limited. Among the few, Fan et al. (2008) proposes a generalized pattern matching based on genetic algorithm for multi-step-ahead prediction of crude oil prices. Yu et al. (2008); Xiong et al. (2013) proposes an empirical mode based on the decomposition of neural networks to forecast crude oil prices. Jammazi and Aloui (2012) uses a hybrid model for crude oil forecasting, Panella et al. (2012) use a mixture of gaussian neural network to forecast energy commodity prices, and Papadimitriou et al. (2014) investigates the efficiency of a support vector machines in forecasting next day electricity prices. Moreover, focus is placed solely on the forecasting of prices, whereas research using neural networks to forecast volatility is still being developed.

The main contribution of this paper is in coupling measures of volatility from high-frequency data with artificial neural networks to deliver a reliable forecast of energy price volatility. Whereas researchers in financial econometrics have done the pioneering work using stock market index data (McAleer and Medeiros, 2011), we are the first to comprehensively test the strategy against competing models in the energy literature. Instead of choosing from a plethora of advanced machine learning algorithms, we use the simplest and most popular feed-forward neural network, as a first step in this field. The main motivation is to show whether there are statistical and economic gains from coupling high-frequency data and easy-to-implement artificial neural networks.

Another contribution of this work is a comprehensive evaluation of the most popular models and realized measures. The realized volatility measures rely on different assumptions, and results guiding practitioners to use a specific one when working with the volatility forecasting of energy markets are still missing from the literature. 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 distinct periods to see which model is able to produce uniformly lower errors in a multiple-step-ahead forecasts of volatility. The period under study is especially interesting, as it covers the period of high and rapidly rising prices, an interruption of the price increase in 2008 due to turmoil in financial markets, and profound regime change in past few years when price variability became much calmer. In particular, the last period is interesting from the forecaster's perspective, as it appears that from the demand side, the consumption of liquid transport fuels has peaked in the developed economies as car engines become more efficient and amid partial substitution by biofuels. On the supply side, high prices reversed the previous trend towards growing dependence on conventional oil fields in the OPEC member states. Proper modeling strategies should be able to reflect these changes.

We test the ANN against widely used HAR and ARFIMA models using the recently proposed frameworks of Model Confidence Set (MCS) from Hansen et al. (2011) and Superior Predictive Ability (SPA) from 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) measures of volatility. Motivated by the possible reduction in model uncertainty, we also experiment with the linear combination of forecasts from the popular HAR model and artificial neural network. This experiment proves to yield the lowest error uniformly through all tested periods regardless of the realized measure used. These 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 the volatility, as commonly found by the GARCH-type models (e.g., 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, further confirmed by Wang and Wu (2012), who find multivariate GARCH-type models to suffer from over-predictions as well). A uniform finding is that coupling neural networks with high-frequency data brings large reductions in over-estimation tendency compared with the previous literature. In addition, we find that MedRV delivers the best forecasts when compared to other measures. As a computationally simple alternative to other measures, we prefer the MedRV for forecasting energy volatility.

The rest of the study is organized as follows. Section 2 describes the realized measure used in this study. Section 3 presents prediction models including HAR, ARFIMA, and ANN. Section 4 presents the data and discusses research setup, including methodology for statistical and economic evaluation of forecasts. Section 5 discusses the results, and, finally, Section 6 concludes. Note that the number of results produced by this research setup is large, and results using different loss functions are overlapping; therefore, we relegate auxiliary results to the online supplementary appendix available at `http://ies.fsv.cuni.cz/sci/publication/show/id/5062/lang/en`.

2. Estimation of realized volatility

In this analysis, we assume that the latent logarithmic commodity price follows a standard jump-diffusion process contaminated with microstructure noise. Let y_t be the observed logarithmic prices evolving over $0 \leq t \leq T$, which will have two components; the latent, 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,

$$y_t = p_t + \epsilon_t. \tag{1}$$

In a latent process, q_t is a Poisson process 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 $0 \leq h \leq t \leq T$ associated with the price process y_t may be naturally decomposed into two parts: integrated variance of the latent price process, $IV_{t,h}$ and jump variation $JV_{t,h}$

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

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 of zero noise contamination in the price process is provided by the well-known realized variance, introduced by Andersen and Bollerslev (1998). The realized variance over $[t-h,t]$ may be estimated as

$$\widehat{QV}_{t,h}^{(RV)} = \sum_{k=1}^{N} (\Delta_k y_t)^2 \ , \tag{3}$$

where $\Delta_k y_t = y_{t-h+(\frac{k}{N})h} - y_{t-h+(\frac{k-1}{N})h}$ is the k-th intraday return in the $[t-h,t]$ and N is the number of intraday observations. The estimator in Eq (3) converges in probability to $IV_{t,h} + JV_{t,h}$ as $N \to \infty$ (Andersen and Bollerslev, 1998; Andersen et al., 2001, 2003; Barndorff-Nielsen and Shephard, 2001, 2002a,b).

Due to the fact that the 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 the noise contamination by introducing the so-called two-scale realized volatility (TSRV) estimator. The authors adopt a methodology for the estimation of the quadratic variation utilizing all of the available data using an idea of precise bias estimation. The two-scale realized variation over $[t-h,t]$ is measured by

$$\widehat{QV}_{t,h}^{(TSRV)} = \widehat{QV}_{t,h}^{(average)} - \frac{\bar{N}}{N}\widehat{QV}_{t,h}^{(all)}, \tag{4}$$

where $\widehat{QV}_{t,h}^{(all)}$ is computed as in Eq (3) 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

$$\widehat{QV}_{t,h}^{(average)} = \frac{1}{G}\sum_{g=1}^{G} \widehat{QV}_{t,h}^{(g)}. \tag{5}$$

where the original grid of observation times, $M = \{t_1, \dots, t_N\}$ is subsampled to $M^{(g)}$, $g = 1, \dots, G$, where $N/G \to \infty$ as $N \to \infty$. The estimator in Eq (4) provides the first consistent and asymptotic estimator of the quadratic variation of p_t. Zhang et al. (2005) also provide the theory for the optimal choice of G grids, $G^* = cN^{2/3}$, where the constant c may be set to minimize the 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

$$\widehat{QV}_{t,h}^{(RK)} = \gamma_0 + \sum_{\eta=1}^{H} K\left(\frac{\eta-1}{H}\right)(\gamma_\eta + \gamma_{-\eta}), \tag{6}$$

with $\gamma_\eta = \sum_{k=1}^{N} \Delta_k y_t \Delta_{k-\eta} y_t$ denoting the η-th realized autocovariance with $\eta = -H, \dots, -1, 0, 1, \dots, H$, and $K(.)$ denotes the kernel function. Note that for $\eta = 0$, $\gamma_\eta = \gamma_0 = \widehat{QV}_{t,h}^{(RV)}$ is an estimate of the realized variance from Eq (3). For the estimator to work, we must choose the kernel function $K(.)$. In our study, we will 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.[5] We should note that the realized kernel estimator is computed without accounting for end effects, i.e., replacing the first and the last observation by local averages to eliminate the corresponding noise components (so-called "jittering"). Barndorff-Nielsen et al. (2008) argue that these effects are important theoretically but are negligible practically.

When studying conditional volatility, it is important to separate the contribution of the two components of the quadratic variation process, i.e., the continuous part from the jump part. Recent evidence from the volatility forecasting literature indicates that the two sources of variation in the price possess substantially different time series properties and impact future volatility in different ways. Despite being mainly interested in forecasting integrated variance, we also disentangle jumps from the data. Barndorff-Nielsen and Shephard (2004, 2006) first develop a bipower variation estimator, which may detect the presence of jumps in high-frequency data. The main idea of the estimator is to compare two measures of the integrated variance, one containing the jump variation and the other being robust to jumps and thus containing only the integrated variation part. In our work, we use the Andersen et al. (2011) adjustment of the original Barndorff-Nielsen and Shephard (2004) estimator, which helps to render it robust to certain types of microstructure noise. The bipower variation over $[t-h, t]$ is defined by

$$\widehat{IV}_{t,h}^{(BV)} = \mu_1^{-2} \frac{N}{N-2} \sum_{k=3}^{N} |\Delta_{k-2} y_t|.|\Delta_k y_t|, \tag{7}$$

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

$$\text{plim}_{N\rightarrow\infty}(\widehat{QV}_{t,h}^{(RV)} - \widehat{IV}_{t,h}^{(BV)}) = JV_{t,h}. \tag{8}$$

Under the assumption of no jump and some other regularity conditions, Barndorff-Nielsen and Shephard (2006) provide the joint asymptotic distribution of the jump variation.[6] Using this theory, the contribution of the jump variation to the quadratic variation of

---

[5]The Parzen kernel function is given by $K(x) = \begin{cases} 1 - 6x^2 + 6x^3 & 0 \leq x \leq 1/2 \\ 2(1-x)^3 & 1/2 \leq x \leq 1. \\ 0 & x > 1 \end{cases}$

[6]Under the null hypothesis of no within-day jumps,

$$Z_{t,h} = \frac{\frac{\widehat{QV}_{t,h}^{(RV)} - \widehat{IV}_{t,h}^{(BV)}}{\widehat{QV}_{t,h}^{(RV)}}}{\sqrt{\left(\left(\frac{\pi}{2}\right)^2 + \pi - 5\right) \frac{1}{N} \max\left(1, \frac{\widehat{TQ}_{t,h}}{\left(\widehat{IV}_{t,h}^{(BV)}\right)^2}\right)}},$$

where $\widehat{TQ}_{t,h} = N\mu_{4/3}^{-3} \left(\frac{N}{N-4}\right) \sum_{k=5}^{N} |\Delta_{k-4} y_t|^{4/3} |\Delta_{k-3} y_t|^{4/3} |\Delta_{k-2} y_t|^{4/3}$ is asymptotically standard normally distributed.

the price process is measured by

$$JV_{t,h} = I_{Z_{t,h}>\Phi_\alpha} \left(\widehat{QV}_{t,h}^{(RV)} - \widehat{IV}_{t,h}^{(BV)} \right), \tag{9}$$

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

$$\widehat{IV}_{t,h}^{(CBV)} = I_{Z_{t,h}\leq\Phi_\alpha} \widehat{QV}_{t,h}^{(RV)} + I_{Z_{t,h}>\Phi_\alpha} \widehat{IV}_{t,h}^{(BV)}, \tag{10}$$

ensuring that the jump measure and the continuous part add up to the estimated variance without jumps.

To estimate the integrated volatility in the presence of jumps, we employ an additional estimator, the median realized volatility, introduced by Andersen et al. (2012):

$$\widehat{IV}_{t,h}^{(MedRV)} = \frac{\pi}{6-4\sqrt{3}+\pi} \left(\frac{N}{N-2} \right) \sum_{k=3}^{N} \text{med} \left(|\Delta_{k-2} y_t|, |\Delta_{k-1} y_t|, |\Delta_k y_t| \right)^2. \tag{11}$$

2.1. Estimation of quadratic variation using Wavelets

Fan and Wang (2007) use a different approach to measuring realized volatility. They use wavelets to separate jump variation from the price process as well as for the estimation of the integrated variance on the jump–adjusted data. Although we use the wavelet-based estimator as one of the six realized measures, we do not discuss the details of the wavelet theory here and direct the reader 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) on 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), which, unlike the ordinary discrete wavelet transform, is not restricted to a dyadic sample length. 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., $k = 1, \ldots, N$. In case the value of the wavelet coefficient $\mathcal{W}_{1,k}$ is greater[7] than the universal threshold $d\sqrt{2 \log N}$ (Donoho and Johnstone, 1994), then a jump with size $\Delta_k J_t$ is detected as

$$\Delta_k J_t = \left(y_{t-h+\left(\frac{k}{N}\right)h} - y_{t-h+\left(\frac{k-1}{N}\right)h} \right) \mathbb{1}_{\left\{|\mathcal{W}_{1,k}|>d\sqrt{2\log N}\right\}} \quad k \in [1, N], \tag{12}$$

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 and Walden, 2000).

Following Fan and Wang (2007), the jump variation over $[t-h, t]$ in discrete time is estimated as the sum of squares of all the estimated jump sizes,

$$\widehat{JV}_{t,h} = \sum_{k=1}^{N} \left(\Delta_k J_t \right)^2. \tag{13}$$

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

Fan and Wang (2007) prove that using Eq (13), we are able to estimate the jump variation from the process consistently.

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

$$\Delta_k y_t^{(J)} = \Delta_k y_t - \Delta_k J_t, \qquad k = 1, \ldots N, \tag{14}$$

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 is able to estimate integrated variance from the process under the assumption of data containing noise as well as jumps. The estimator utilizes the TSRV approach of Zhang et al. (2005) as well as the wavelet jump detection method. Another advantage of the estimator is that it decomposes the integrated variance into $J^m + 1$ components; therefore, we are able to study the dynamics of volatility at various investment horizons. Following Barunik and Vacha (2014), we define the JWTSRV estimator over $[t-h, t]$ on the jump-adjusted data as follows:

$$\widehat{IV}_{t,h}^{(JWTSRV)} = \sum_{j=1}^{J^m+1} \widehat{IV}_{j,t,h}^{(JWTSRV)} = \sum_{j=1}^{J^m+1} \left(\widehat{IV}_{j,t,h}^{(average)} - \frac{\bar{N}}{N} \widehat{IV}_{j,t,h}^{(all)} \right), \tag{15}$$

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

Barunik and Vacha (2014) show that the JWTSRV is a consistent estimator of the integrated variance, as it converges in probability to the integrated variance of the process p_t, and they assess the small sample performance of the estimator in a large Monte Carlo study. The JWTSRV is found to be able to recover true integrated variance from the noisy process with jumps very precisely. Moreover, the JWTSRV estimator is also tested in a forecasting exercise, which has been found to improve substantially the forecasting of the integrated variance.

3. Prediction models

Well documented evidence for the strong temporal dependence of realized volatility suggests that realized volatility should be modeled by an approach allowing for a slowly decaying autocorrelation function and possibly long memory. Müller et al. (1997), Arneodo et al. (1998) 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 does not have an effect on longer intervals. A possible economic interpretation is that long-term volatility matters for short-term traders, whereas short-term volatility does not affect long-term trading strategies.

Standard, ARCH-type volatility models of (Engle, 1982; Bollerslev, 1986, 1987) and one-factor stochastic volatility models treat volatility as a latent variable and do not capture the long memory. In this study, we use the realized volatility as ex-post observed variance, and to assess the relative performance of the artificial neural network, we consider benchmark models for forecasting volatility capturing its properties. 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 briefly described in this section.

3.1. The linear heterogeneous autoregressive (HAR) model

A simple, popular model for forecasting realized volatility is the heterogeneous autoregressive model (HAR) of Corsi (2009) based on the heterogeneous realized volatility components

$$\nu_{t+1} = \alpha + \beta_D \nu_t + \beta_W \nu_{t,t-5} + \beta_M \nu_{t,t-22} + \epsilon_{t+1}, \tag{16}$$

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, $\sqrt{\widehat{QV}_{t,h}^{(est)}}$, and $\sqrt{\widehat{IV}_{t,h}^{(est)}}$, where (est) are RV, RK, TRSV, CBV, MedRV, and JWTSRV measures.

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

Although the HAR model is popular due to its simplicity, it is an approximate long-memory model and as a result might not be able to capture the dynamics of long memory properties in volatility well. 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 where parameter d may be any real number such that $-1/2 < d < 1/2$ to ensure stationarity and invertibility. More precisely, ν_t is an ARFIMA(p, d, q) process if it follows:

$$\alpha(L)(1 - L)^d(\nu_t - \mu) = \beta(L)v_t, \tag{17}$$

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 $L(L\nu_t = \nu_{t-1})$, which is rooted strictly outside the unit circle, v_t is iid with zero mean and σ_v^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. Deo et al. (2006), Andersen et al. (2003) show that forecasting log realized volatility based on a simple ARFIMA$(1, d, 0)$ specification is a very good competitor to 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 HAR and ARFIMA models are developed to capture a specific features of the time series, and they are suitable for volatility modeling due to its ability to capture long memory. Artificial neural networks (ANN) may be viewed as a generalization of these classical approaches, which allows us to model another type of nonlinearities in the data in addition to long memory. Concisely, neural networks are semi-parametric non-linear models, which are able to approximate any reasonable function Haykin (2007); Hornik et al. (1989).

Although neural networks are primarily associated with biological systems and successfully applied in numerous fields, such as pattern recognition, medical diagnostics, many econometricians argue that neural networks are a black box. A standard ANN imitates neural processing in brain activation. Together with the fact that one must make arbitrary decisions about the implementation of the network, i.e., the number of hidden layers, the choice of transformation functions, the number of neurons, etc., neural networks are still not commonly used for financial time series modeling.

Abandoning these concerns, we use neural network as a generalized nonlinear regression, being able to describe the complex patterns in volatility time series. Like linear or nonlinear methods, a neural network relates a set of input variables, say lags of volatility to output, 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 we use for volatility ν_t forecasting may be described by the following equations:

$$n_{k,t} = \omega_{k,0} + \sum_{i=0}^{21} \omega_{k,i} \nu_{t-i} \tag{18}$$

$$\Lambda(n_{k,t}) = \frac{1}{1 + e^{-n_{k,t}}} \tag{19}$$

$$\nu_{t+h} = \gamma_0 + \sum_{k=1}^{k^*} \gamma_k \Lambda(n_{k,t}) \tag{20}$$

where $\Lambda(n_{k,t})$ is the log-sigmoid activation function. To make the model comparable to 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 or *weights* vector to be found. The variable $n_{k,t}$ is squashed by the logsigmoid 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, as almost all researchers begin with this network as the first alternative to linear models.

Note that 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 the inputs to improve the predictions.

To be able 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 for all training examples. The training phase is thus an unconstrained nonlinear optimization problem, where the goal is to find the optimal set of weights of the parameters by solving the minimization problem:

$$\min\{\Psi(\omega) : \omega \in \mathbb{R}^n\}, \tag{21}$$

where $\Psi : \mathbb{R}^n \to \mathbb{R}^n$ is a continuously differentiable error function. There are several ways of minimizing $\Psi(\omega)$, but basically we are 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 provides us the direction of steepest decrease.

Nevertheless, the traditional gradient descent algorithms often fail in learning intricate patterns in the data efficiently due to many possible initial settings. One of the efficient methods for learning the patterns in feed-forward neural networks, which we use, is the resilient propagation algorithm Riedmiller and Braun (1993). The 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 speed of convergence than in the case of plain backpropagation or quickpropagation algorithms.

The best ANN model is chosen from a set of models having 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 0 (without decay) or 1e-10 (standard decay used in the literature). To prevent overfitting, we use cross-validation over time with 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 (NYMEX).[8] 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 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 TSRV and JWTSRV estimators. The 5-minute choice is guided by the volatility signature plot, and previous literature employing the same data. The sample period extends from January 5, 2004 to December 31, 2012 covering the recent U.S. recession from Dec 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, due to the low activity on these days, which could 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 (3), realized kernel $\widehat{QV}_{t,h}^{(RK)}$, defined by Eq (6), two-scale realized variance $\widehat{QV}_{t,h}^{(TSRV)}$, defined by Eq (4), the "continuous part" of the

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

bipower variation $\widehat{IV}_{t,h}^{(CBV)}$, defined by Eq (10), the median realized volatility $\widehat{IV}_{t,h}^{(MedRV)}$, defined by Eq (11), and jump-adjusted wavelet two-scale realized variance $\widehat{IV}_{t,h}^{(JWTSRV)}$, defined by Eq (15). We work with forecasts of volatility, which is the square root of the component of quadratic variation. For ease of notation, we will 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 very simple to compute for a practitioner, RK and TSRV measure the volatility of the true price process contaminated with microstructure noise, and these three are measures for the quadratic variation. In addition, CBV, and MedRV measure integrated variance directly, whereas MedRV offers a number of advantages over the 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 offers robustness to both microstructure noise and jumps.

Table 1 reports the summary statistics for the estimated realized measures. The natural gas price shows greatest degree of variability in comparison to crude oil and heating oil with twice larger averages. Ljung-Box statistics point to large degree of dependence, as commonly found in the volatility time series. The daily prices, returns and volatility are plotted in Figure 1.

4.2. Research design for forecast evaluation

The main interest of this work lies in relative forecasting performance rather than in 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, we have made the in-sample model fits available upon request and wish to state that we have conducted all the necessary tests to conclude that all the models fit the data well.

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, whereas we reserve the remaining 1631 observations for evaluating 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 cumulating $\nu_{t,t+h} = h^{-1} \sum_{j=1}^{h} \nu_{t+j}$ daily volatilities. We focus on cumulative forecasts, as they are more interesting in applications.

After estimating the volatility forecasts for all 1631 observations from July, 6, 2006 until December 31, 2012 on a rolling basis, we divide the forecasts into three periods. The main motivation is the recent crisis, which occurred in the middle of our forecasted sample. Therefore, evaluating the forecasts in the three equal periods will allow us to evaluate the forecasting performance of all models before the crisis on the data from July 6, 2006 until August 31, 2008, during the crisis, using the period from September 1, 2008 until October 31, 2010, and after the crisis, covering the period from November 1, 2010 until December 31, 2012. The periods are visually depicted by Figure 1. We observe a highly distinct evolution of prices in all three studied assets through the three periods.

4.3. Statistical evaluation of the forecasts

We split the evaluation analysis into two parts to evaluate the out-of sample forecasts statistically as well as economically. To statistically compare the accuracy of the volatility forecasts from different models, we employ two common loss functions, namely the root mean square error (RMSE) and the mean absolute error (MAE). The measures are calculated for the $t = 1, \ldots, T$ forecasts as

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{T} (\hat{\nu}_{t+i} - \nu_{t+i})^2} \tag{22}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{T} |\hat{\nu}_{t+i} - \nu_{t+i}| \tag{23}$$

As discussed by Nomikos and Pouliasis (2011), these metrics do not provide information about the asymmetry of the errors commonly found by the literature, especially for the parametric GARCH models. Nonetheless, the asymmetry of forecast error is important for the practitioners, as it alerts us to whether the modeling strategy tends to over-predict or under-predict the volatility. Testing the forecasts of energy commodities, 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 multivariate GARCH-type models to 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 and Faff, 1996) to assess the forecasts. These functions use a mixture of positive and negative forecast errors with different weights allowing us to discover the cases if the model tends to over- or under-predict

$$MME(O) = \frac{1}{N} \left(\sum_{i \in U} |\hat{\nu}_{t+i} - \nu_{t+i}| + \sum_{i \in O} \sqrt{|\hat{\nu}_{t+i} - \nu_{t+i}|} \right), \tag{24}$$

$$MME(U) = \frac{1}{N} \left(\sum_{i \in U} \sqrt{|\hat{\nu}_{t+i} - \nu_{t+i}|} + \sum_{i \in O} |\hat{\nu}_{t+i} - \nu_{t+i}| \right), \tag{25}$$

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

To test significant differences of 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 given a 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

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

with $L_{i,t}$ being an appropriate loss function of the i-th model. Starting 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.[9]

In addition, we employ the superior predictive ability (SPA) test of 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 to its competitors, indicating that the benchmark model produces the smallest loss. Again, we use the bootstrapped p-values and follow Hansen (2005) in the implementation of the test.

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

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

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

4.4. Economic Evaluation of the forecasts

A modelÕs statistical superiority does not necessarily translate to economic benefits; therefore, in addition to performing statistical evaluation, we evaluate the forecasts economically. Quantile forecasts are central to risk management decisions due to 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%, 5% VaR for both long positions and short positions.

Although quantile forecasts may be readily evaluated by comparing their actual (estimated) coverage $\hat{C}_\alpha = 1/n \sum_{n=1}^{T} 1(y_{t+h} < \hat{q}^{\alpha}_{t+h})$, against their nominal coverage rate, $C_\alpha = E[1(y_{t,t+1} < q^{\alpha}_{t,t+1})]$, 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 a forecaster m as follows:

$$\widehat{L}_{\alpha,m} = E\left[\alpha - 1\left(y_{t,t+1} < \widehat{q}^{\alpha,m}_{t,t+1}\right)\right]\left[y_{t,t+1} - \widehat{q}^{\alpha,m}_{t,t+1}\right], \tag{26}$$

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

[9]We have used different block lengths, including the ones depending on the forecasting horizons, to assess the robustness of the results, without any change in the final results. These results are available from the authors upon request.

5. Empirical results

The following section compares the performance of neural networks against the competing ARFIMA and HAR models in volatility forecasting. Each of the models 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, its linear combination is a good candidate for offering the best forecasting framework for a practitioner in any situation.

We begin the discussion with the statistical evaluation of the forecasting models and move to the economic implications later on. As mentioned earlier, we aim to assess the forecasting performance of all models in three separate periods: before, during, and after the recent financial crisis of 2008. We thus discuss the results in this logical sequence. The number of tables produced by this research setup is large, and results using different loss functions are overlapping; therefore, we report part of the results in the online supplementary appendix available at `http://ies.fsv.cuni.cz/sci/publication/show/id/5062/lang/en`.

5.1. Statistical evaluation of forecasts

We present the results for RMSE, MAE, MME(O), and MME(U) in separate tables for each of the periods. Each table contains results for all three commodities; crude oil, heating oil, and natural gas, and several forecasting horizons: 1-step-ahead, 5- and 10-step-ahead. The statistical significance of the difference in the performance is evaluated across forecasting models (row-wise comparison) and across volatility estimators (column-wise comparison) using MCS. We use (^{b}), and (^{a}) to denote the model and estimators that belong to the corresponding 10% model confidence sets, respectively. In addition, a bold entry depicts 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 volatility proxy in the testing procedure. When testing model performance, we use the realized measure being forecasted by the model as a volatility proxy. When testing the performance across measures, we choose a simple proxy of absolute value of open-close returns, as common in the literature. This approach allows us to identify which realized measures perform best. We have also experimented with all different measures as a proxy for volatility, but the results do not change; therefore, we leave these results upon request from authors.

5.1.1. Forecasting performance before the crisis

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

To assist the interpretation of the tables, let us consider the results in Table 2, the first column of which shows the RMSE of the models forecasting volatility of the crude oil. Starting with 1-step-ahead forecasts and holding the realized measure, say, TSRV (first column), fixed, ANN and HAR-ANN models produce the lowest error of 0.357×10^{-2}, whereas all models belong to the model confidence set, as they are depicted by (^{b}). Moreover, HAR, ANN, and HAR-ANN combinations are set in bold, indicating that they are not rejected as the best benchmark forecasting model by the SPA test, whereas ARFIMA is rejected due to its largest error of 0.365×10^{-2}. This result holds for all columns (all realized measures) except JWTSRV and CBV, which forecast only integrated variation part. For the JWTSRV and CBV, ANN and HAR-ANN combination are the only two models in the model confidence set. This approach indicates that if we are interested in forecasting the whole quadratic variation, the HAR and ANN models are both in the model confidence set and produce statistically indistinguishable results, whereas ANN produces lowest error. If we are interested in forecasting only the integrated variation part, ANN is superior to other models. Holding the model and comparing the errors column-wise, MedRV is the only measure belonging to the confidence set. Note that 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 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 HAR, ANN, and the combination HAR-ANN produce statistically same forecasts, whereas ARFIMA is rejected, and the ANN models bring the lowest error. Turning to the results found for MAE, they yield similar conclusions, but the ARFIMA model is not rejected. Nonetheless, a higher forecasting horizon h implies a lower error for ANN compared to other competing models.

When examining the rest of the results reported by Table 2 for heating oil and natural gas, we observe a similar picture, although the results are more mixed. In conclusion, a larger forecasting horizon h implies less error from ANN or a combination of the HAR-ANN model in comparison to the HAR and ARFIMA (with exception of heating oil). Whereas on many occasions HAR or even 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, while other estimators of integrated variance; namely, CBV and JWTSRV belong to the MCS more often than in the crude oil case. This fact points us to the conclusion that MeRV is the best measure for forecasting volatility. One may oppose that the results are not robust, as these are measures of integrated variance, excluding jumps. However, the result is strong, as the volatility proxy used is the absolute value of open-close returns, which also includes jumps.

Turning our attention to the over- and under- predictions reported in the online appendix, the main conclusions remain unchanged.[10] An important result emerges: the models yield similar results for both 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 maximum levels of over-predictions for natural gas under 60%), whereas in many occasions, models yield an equal number of over- and under-predictions. This is an important finding, as, in comparison to GARCH-type models that strongly over-predict volatility (Nomikos and Pouliasis, 2011; Wang and Wu, 2012), high-frequency data appear to yield substantial improvement in this respect.

[10] To conserve space, we report the actual MME(U), and MME(O) values together with percentages of over- and under-predictions in the online supplementary appendix, available at `http://ies.fsv.cuni.cz/sci/publication/show/id/5062/lang/en`.

5.1.2. Forecasting performance during the crisis

Forecasting performance of the models during the crisis, covering the September 1, 2008 – October 31, 2010 period follows in terms of RMSE and MAE is reported in Table 3. A general overview of results from the pre-crisis period hold, whereas all of the RMSE and MAE are larger in comparison with the pre-crisis period. ANN and the HAR-ANN combination of models produce the lowest errors, whereas in most cases, HAR, ANN, and their combination belong to the model confidence set. ARFIMA is rejected as a best-performing model several times, whereas the combination of ANN and HAR models is never rejected and always belongs to the model confidence set.

When comparing the results across realized measures, we observe that 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, logically, the simplest realized volatility is preferred in this case. In many cases, CBV and JWTSRV belong to the model confidence set as well.

CBV and MedRV belong to the model confidence set most often together with JWTSRV. From the remaining estimators, 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. 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 larger than 60%.

To conclude, the results from forecasting volatility during the recent crisis produce larger errors than before the crisis, but generally, ANNs again prove to be uniformly best forecasting vehicle. In terms of realized measure, MedRV is decisively the best choice. Interestingly, the rate of over-prediction is not much higher, which proves the modelsÕ general ability to forecast the volatility correctly.

5.1.3. Forecasting performance after the crisis

Next, we compare the model performance on the data after the crisis, November 1, 2010 – December 31, 2012. Table 4 presents the results for the RMSE, and MAE. Whereas the reported loss functions are the lowest in comparison to the previous periods, the statistical tests tend to reject more models. ANN tends to deliver larger errors than competing models, but its combination with HAR produces lowest errors. After turmoil of the 2008, the HAR-ANN combination again always belongs to the model confidence set, while 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 MedRV.

The comparison of 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 of the models tend to over-predict the volatility to a larger extent, up to 70%. This result is attributed to the fact that the model parameters are estimated in times of high volatility during 2008, whereas the predictions are made during a calmer period. In this respect, the models all perform very well in terms of statistical criteria.

5.1.4. Forecasting performance over whole period.

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

In comparison to more complicated TSRV, JWTSRV measures, MedRV is a simple alternative, and it provides the best performance. Therefore, MedRV would be 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 and Zarnowitz (1969) regressions. Although we do not know which measure 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 be predicted by all others as well. 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 forecasting models. The regression takes the following form:

$$\hat{\nu}_{t+h}^{RM} = \alpha + \beta \hat{\nu}_{t+h}^{(RM,f)} + \epsilon_t, \tag{27}$$

with $\hat{\nu}_{t+h}$ being the volatility estimated with RM measures, namely the TSRV, RV, RK, JWTSRV, CBV, and MedRV volatility, and $\hat{\nu}_{t+h}^{(RM,f)}$ its forecast using ARFIMA, HAR, ANN, HAR-ANN 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 way, we test all the remaining realized measures, resulting in 144 final regressions for one commodity. 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 where the forecast is optimal, 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 except for the heating oil for the last period – after November 2010 – we never reject the hypothesis that the parameters are significantly different. This finding leads us to the conclusion that all the forecasts are uniformly optimal.

Finally, we study R^2 from the regressions, as it will tell us what portion of variance is explained by forecasts. The results from the MZ regressions for the entire period are reported by Figures 2, 3, and 4 for all forecasting horizons. We also include the R^2 results for all three periods in an online supplementary appendix. We observe from the figures that all of the models perform well in all forecasting horizons, with R^2 being over 70% in all cases except for natural gas forecasted 1-step-ahead. This is the expected result, as natural gas shows the greatest degree of price variability, leading the models to be able to explain lower amounts of variance. When comparing performance across models, we may conclude that all the models deliver very similar accuracy of explained variance, and the results are in line with previous analyses.

More interestingly, a distinction may be made when comparing realized measures. JWTSRV, together with CBV and MedRV, may be forecasted with the highest degree of success at all horizons. Although a longer forecasting horizon implies a lower 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 MedRV performs the best statistically, the results of this analysis find this simple measure to outperform the others.

5.3. Economic evaluation of forecasts

Most practitioners are interested in the economic evaluation of the performance of the models. Therefore, we evaluate the models using VaR forecasts. To conserve space, we discuss the economic evaluation of 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.

Table 6 and Table 7 report conditional coverage as well as statistical comparison through the loss function of Giacomini and Komunjer (2005) 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 SPA results, the HAR-ANN model combination belongs to the model confidence set uniformly yielding the statistically best results. Interestingly, ARFIMA 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 needed. 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 to the original returns containing jumps. Therefore, to use these measures, it would be recommended to include the jump variation as well. This approach is nevertheless beyond the scope of this study. We conduct an economic evaluation as a robustness check for the results from statistical evaluation. The general conclusion is that the results from statistical evaluation materialized into economic benefits.

6. Conclusion

The prediction of energy price variability is of immense interest to both practitioners and the academic literature. Nonetheless, most of the studies focus on the usage of daily data and rely on popular GARCH-type models when predicting the volatility of energy prices. In this paper, we couple the recently developed realized measures with popular artificial neural networks to forecast energy price variability.

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 (MCS) and Superior Predictive Ability (SPA). 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) measures of volatility. Motivated by the possible reduction of the model uncertainty, we also experiment with the linear combination of forecasts from the popular HAR model and ANN. This experiment is found to yield the lowest error uniformly through all tested periods. These errors also translate to economic benefits in terms of VaR.

Our main finding is that coupling realized measures with artificial neural networks results in both statistical and economic gains. Importantly, the methodology reduced the tendency to over-predict the volatility confirmed by previous research. Even in the cases when the model is fit on the data coming from the period of high uncertainty and forecasts the period of reduced uncertainty, the results hold. Therefore, the findings hold uniformly through the tested periods, and the methodology yields large advances to previously used methodologies, which tend to over-predict the volatility. In addition, median realized volatility is preferred as a computationally simple measure delivering best forecasting performance.

The results are of great importance for market participants, as they allow a reduction in risk. It will be interesting to see further results in academic literature, coupling realized measures with more sophisticated machine learning frameworks.

References

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

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

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

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

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

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

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

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

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

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

Barndorff-Nielsen, O. E. and N. Shephard (2007). Variation, jumps, market frictions and high frequency data in financial econometrics. In *Advances in Economics and Econometrics. Theory and Applications, Ninth World Congress*, Volume 3, pp. 328–372.

Barunik, J. and L. Vacha (2014). Realized wavelet-based estimation of integrated variance and jumps in the presence of noise. *To Appear in Quantittive Finance*.

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. and R. W. Faff (1996). An evaluation of volatility forecasting techniques. *Journal of Banking & Finance 20*(3), 419–438.

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

Deo, R. S., C. M. Hurvich, and Y. Lu (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. and I. M. Johnstone (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. and Y. Wang (2007). Multi-scale jump and volatility analysis for high-frequency financial data. *Journal of the American Statistical Association 102*(480), 1349–1362.

Fan, Y., Q. Liang, and Y.-M. Wei (2008). A generalized pattern matching approach for multi-step prediction of crude oil price. *Energy Economics 30*(3), 889–904.

Giacomini, R. and I. Komunjer (2005). Evaluation and combination of conditional quantile forecasts. *Journal of Business & Economic Statistics 23*(4), 416–431.

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

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

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics 23*(4).

Hansen, P. R., A. Lunde, and J. M. Nason (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., M. Stinchcombe, and H. White (1989). Multilayer feedforward networks are universal approximators. *Neural Networks 2*(5), 359–366.

Jammazi, R. and C. Aloui (2012). Crude oil price forecasting: Experimental evidence from wavelet decomposition and neural network modeling. *Energy Economics 34*(3), 828–841.

Kang, S. H. and S.-M. Yoon (2013). Modeling and forecasting the volatility of petroleum futures prices. *Energy Economics 36*, 354–362.

Kuper, G. H. and D. P. van Soest (2006). Does oil price uncertainty affect energy use? *Energy*

Journal 27(1).

Linn, S. C. and Z. Zhu (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. and G. Zumbach. (2003). Market heterogeneities and the causal structure of volatility. *Quantitative Finance* (3), 320–331.

McAleer, M. and M. Medeiros (2008). Realized volatility: A review. *Econometric Reviews* (27), 10–45.

McAleer, M. and M. C. Medeiros (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. and V. Zarnowitz (1969). *The evaluation of economic forecasts.* New York: National Bureau of Economic Research.

Mohammadi, H. and L. Su (2010). International evidence on crude oil price dynamics: Applications of arima-garch models. *Energy Economics 32*(5), 1001–1008.

Müller, U., M. Dacorogna, R. Dav, R. Olsen, O. Pictet, and J. von Weizsacker (1997). Volatilities of different time resolutions - analysing the dynamics of market components. *ournal of Empirical Finance* (4), 213–239.

Nomikos, N. K. and P. K. Pouliasis (2011). Forecasting petroleum futures markets volatility: The role of regimes and market conditions. *Energy Economics 33*(2), 321–337.

Panella, M., F. Barcellona, and R. L. D'Ecclesia (2012). Forecasting energy commodity prices using neural networks. *Advances in Decision Sciences 2012.*

Papadimitriou, T., P. Gogas, and E. Stathakis (2014). Forecasting energy markets using support vector machines. *Energy Economics*.

Patton, A. J. and K. Sheppard (2009). Evaluating volatility and correlation forecasts. In *Handbook of financial time series*, pp. 801–838. Springer.

Percival, D. B. and H. Mofjeld (1997). Analysis of subtidal coastal sea level fluctuations using wavelets. *Journal of the American Statistical Association 92*(439), 886–880.

Percival, D. B. and A. T. Walden (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.

Riedmiller, M. and H. Braun (1993). A direct adaptive method for faster backpropagation learning: The rprop algorithm. In *Neural Networks, 1993., IEEE International Conference on*, pp. 586–591. IEEE.

Wang, Y. (1995). Jump and sharp cusp detection via wavelets. *Biometrika 82*(2), 385–397.

Wang, Y. and C. Wu (2012). Forecasting energy market volatility using garch models: Can multivariate models beat univariate models? *Energy Economics 34*(6), 2167–2181.

Wei, Y., Y. Wang, and D. Huang (2010). Forecasting crude oil market volatility: Further evidence using garch-class models. *Energy Economics 32*(6), 1477–1484.

Wilson, B., R. Aggarwal, and C. Inclan (1996). Detecting volatility changes across the oil sector. *Journal of Futures Markets 16*(3), 313–330.

Xiong, T., Y. Bao, and Z. Hu (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., M.-J. Hwang, and B.-N. Huang (2002). An analysis of factors affecting price volatility of the us oil market. *Energy Economics 24*(2), 107–119.

Yu, L., S. Wang, and K. K. Lai (2008). Forecasting crude oil price with an emd-based neural network ensemble learning paradigm. *Energy Economics 30*(5), 2623–2635.

Zhang, L., P. Mykland, and Y. Aït-Sahalia (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.

Asset	Estimator	N	Min.	Max.	Std.	Mean	Ex. Kurt.	Skew.	LB(5)	LB(20)
Crude Oil	TSRV	2231	0.19	39.25	3.82	3.22	23.59	4.28	6184.25	22378.68
	RV	2231	0.23	39.53	3.79	3.26	23.00	4.17	5903.97	21564.16
	RK	2231	0.19	46.68	3.98	3.20	26.31	4.45	5657.72	20320.46
	JWTSRV	2231	0.20	36.79	3.54	3.07	22.80	4.19	6607.72	23738.65
	CBV	2231	0.22	39.53	3.72	3.16	24.34	4.28	5929.74	21723.68
	MedRV	2231	0.15	43.52	3.46	2.96	26.50	4.32	5785.12	20964.45
Heating Oil	TSRV	2222	0.15	27.54	2.58	2.79	12.98	2.84	5078.28	17381.43
	RV	2222	0.18	30.62	2.69	2.85	15.33	2.96	4370.84	15185.33
	RK	2222	0.17	36.14	2.74	2.79	17.92	3.21	4286.88	14519.21
	JWTSRV	2222	0.14	22.61	2.35	2.62	9.69	2.57	5921.63	20005.77
	CBV	2222	0.18	30.62	2.56	2.73	13.19	2.80	4727.61	16486.29
	MedRV	2222	0.14	27.05	2.42	2.55	11.40	2.69	4506.88	15461.91
Natural Gas	TSRV	2219	0.53	98.97	5.72	6.03	42.84	4.56	1636.47	3475.85
	RV	2219	0.57	93.58	6.28	6.29	34.03	4.44	1471.91	2992.58
	RK	2219	0.46	94.89	5.90	5.88	38.51	4.56	1357.20	2828.15
	JWTSRV	2219	0.53	53.31	4.20	5.21	19.69	3.16	3791.07	8473.81
	CBV	2219	0.42	66.29	4.97	5.65	27.02	3.78	2882.75	5923.80
	MedRV	2219	0.46	68.69	4.96	5.28	44.97	4.86	2459.53	5028.12

Table 1: Descriptive statistics for crude oil, heating oil and natural gas for the sample period from July, 6, 2006 until December 31, 2012. Minimum, maximum, standard deviation and mean are multiplied by $\times 10^4$ for convenience. LB(l) is Ljung-Box statistics for l-th lag.

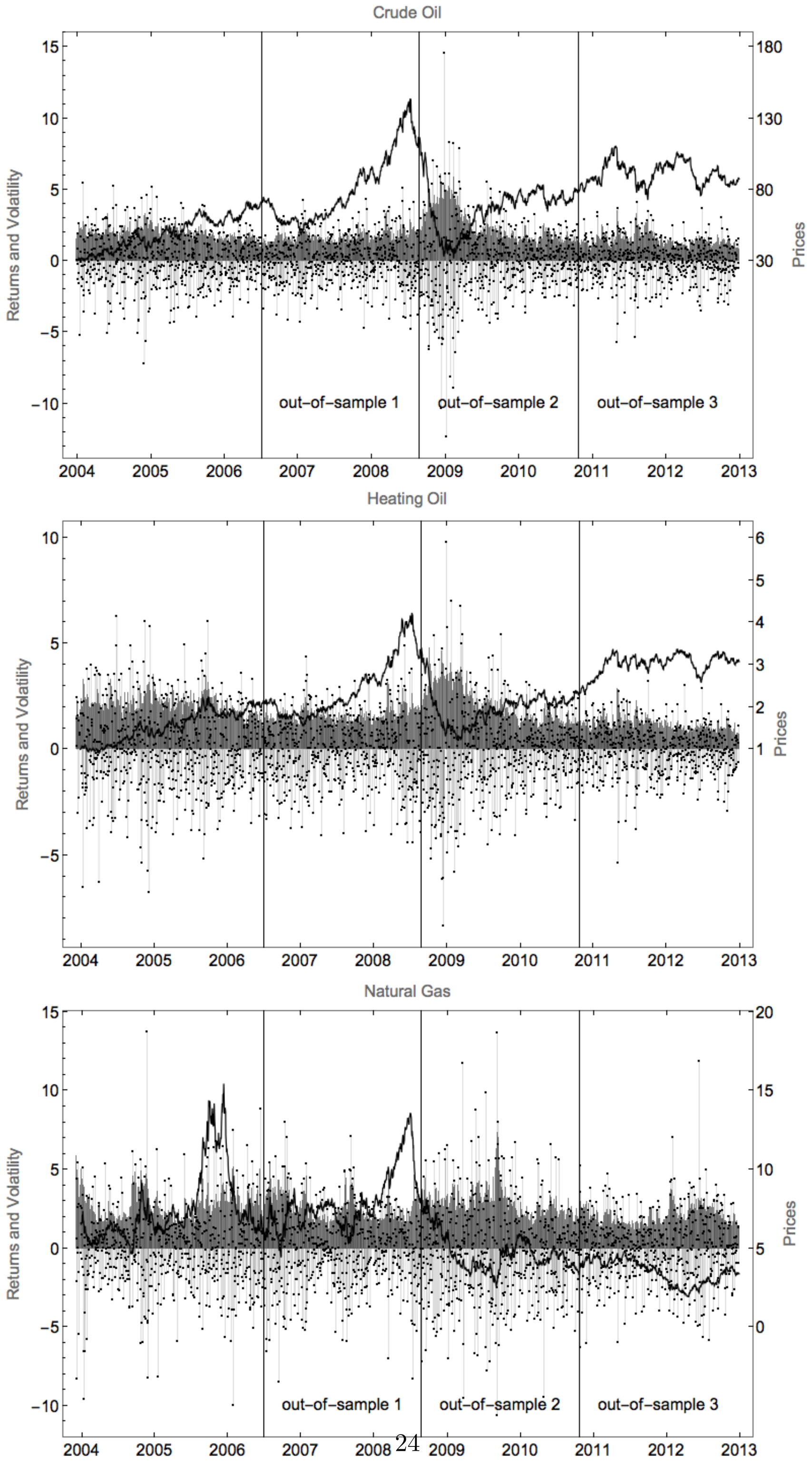

Figure 1: Realized volatility, returns, and prices of crude oil, heating oil and natural gas. The forecast period is divided into three equal sample periods from July, 6, 2006 until August 31, 2008, from September 1, 2008 until October 31, 2010 and from November 1, 2010 until December 31, 2012.

Table 2: **Before September 2008**: RMSE/MAE. 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. Note that numbers are multiplied by $\times 10^2$.

Crude Oil

	TSRV	RV	RK	JWTSRV	CBV	MedRV
RMSE			$h=1$			
ARFIMA	0.365^{b}	**0.357**^{b}	**0.379**^{b}	0.315	0.338	**0.326**^{a,b}
HAR	**0.360**^{b}	**0.351**^{b}	**0.375**^{b}	0.309	**0.332**	**0.323**^{a,b}
ANN	**0.357**^{b}	**0.349**^{b}	**0.373**^{b}	**0.305**^{b}	**0.329**^{b}	**0.321**^{a,b}
HAR-ANN	**0.357**^{b}	**0.349**^{b}	**0.373**^{b}	**0.306**^{b}	**0.329**^{b}	**0.321**^{a,b}
			$h=5$			
ARFIMA	0.606	0.582	0.619	0.519^{a}	0.542^{a}	0.529^{a}
HAR	**0.571**^{b}	**0.546**^{b}	**0.581**^{b}	**0.480**^{b}	**0.504**^{b}	**0.497**^{a,b}
ANN	**0.568**^{b}	**0.545**^{b}	**0.581**^{b}	**0.479**^{b}	**0.502**^{b}	**0.500**^{a,b}
HAR-ANN	**0.567**^{b}	**0.543**^{b}	**0.579**^{b}	**0.478**^{b}	**0.501**^{b}	**0.495**^{a,b}
			$h=10$			
ARFIMA	0.853^{a}	0.830^{a}	0.870^{a}	0.746^{a}	0.774^{a}	0.753^{a}
HAR	**0.769**^{b}	**0.751**^{b}	**0.782**^{b}	**0.658**^{b}	**0.696**^{a,b}	**0.689**^{a,b}
ANN	**0.761**^{b}	**0.745**^{a,b}	**0.778**^{a,b}	**0.652**^{a,b}	**0.692**^{a,b}	**0.694**^{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}
MAE			$h=1$			
ARFIMA	**0.260**^{b}	**0.262**^{b}	**0.274**^{b}	**0.233**^{b}	**0.247**^{b}	**0.241**^{a,b}
HAR	**0.262**^{b}	**0.265**^{b}	**0.276**^{b}	**0.231**^{b}	**0.249**^{b}	**0.243**^{a,b}
ANN	**0.260**^{b}	**0.266**^{b}	**0.275**^{b}	**0.231**^{b}	**0.250**^{b}	**0.243**^{a,b}
HAR-ANN	**0.260**^{b}	**0.265**^{b}	**0.275**^{b}	**0.230**^{b}	**0.248**^{b}	**0.242**^{a,b}
			$h=5$			
ARFIMA	**0.428**^{b}	0.430^{b}	0.443^{b}	**0.381**^{b}	**0.404**^{b}	**0.389**^{a,b}
HAR	**0.406**^{b}	**0.404**^{b}	**0.415**^{b}	**0.359**^{b}	**0.386**^{b}	**0.372**^{a,b}
ANN	**0.406**^{b}	**0.405**^{b}	**0.417**^{b}	**0.361**^{b}	**0.386**^{b}	**0.374**^{a,b}
HAR-ANN	**0.404**^{b}	**0.402**^{b}	**0.414**^{b}	**0.358**^{b}	**0.384**^{b}	**0.370**^{a,b}
			$h=10$			
ARFIMA	**0.612**^{b}	**0.614**^{b}	**0.628**^{a,b}	0.559^{a,b}	**0.580**^{a,b}	**0.556**^{a,b}
HAR	**0.567**^{b}	**0.570**^{b}	**0.583**^{b}	**0.507**^{b}	**0.538**^{b}	**0.526**^{a,b}
ANN	**0.568**^{b}	**0.571**^{b}	**0.583**^{b}	**0.509**^{b}	**0.540**^{b}	**0.539**^{a,b}
HAR-ANN	**0.566**^{b}	**0.568**^{b}	**0.582**^{b}	**0.506**^{b}	**0.537**^{b}	**0.529**^{a,b}

Heating Oil

	TSRV	RV	RK	JWTSRV	CBV	MedRV
RMSE			$h=1$			
ARFIMA	**0.352**^{b}	**0.373**^{b}	**0.367**^{b}	**0.283**^{b}	**0.353**^{a,b}	**0.364**^{a,b}
HAR	**0.354**^{b}	**0.373**^{a,b}	**0.369**^{a,b}	**0.286**^{a,b}	**0.353**^{a,b}	**0.365**^{a,b}
ANN	**0.354**^{b}	**0.375**^{a,b}	**0.371**^{a,b}	0.288^{a,b}	**0.355**^{a,b}	**0.367**^{a,b}
HAR-ANN	**0.353**^{b}	**0.372**^{a,b}	**0.369**^{a,b}	**0.287**^{a,b}	**0.352**^{a,b}	**0.365**^{a,b}
			$h=5$			
ARFIMA	**0.538**^{a}	**0.567**^{a}	**0.553**^{a}	**0.444**^{a}	**0.531**^{a}	**0.569**^{a}
HAR	**0.550**^{a}	**0.574**^{a,b}	**0.558**^{a,b}	0.458^{a}	**0.537**^{a,b}	**0.580**^{a,b}
ANN	0.562^{a}	0.591^{a,b}	0.577^{a,b}	0.468^{a}	0.554^{a,b}	0.596^{a,b}
HAR-ANN	0.554^{a,b}	**0.579**^{a,b}	**0.564**^{a,b}	0.461^{a,b}	**0.543**^{a,b}	**0.586**^{a,b}
			$h=10$			
ARFIMA	**0.693**	**0.721**	**0.712**	**0.591**	**0.677**^{a,b}	**0.735**^{a}
HAR	0.721	**0.739**^{b}	**0.725**^{b}	0.625	**0.698**^{a,b}	0.766^{a}
ANN	0.728	0.755^{b}	0.743^{b}	0.636	0.713^{a,b}	0.786^{a}
HAR-ANN	0.721^{b}	**0.742**^{b}	**0.730**^{b}	0.628^{b}	**0.702**^{a,b}	0.773^{a,b}
MAE			$h=1$			
ARFIMA	**0.264**^{b}	**0.271**	**0.278**^{b}	**0.221**	**0.259**	**0.251**^{a}
HAR	**0.262**^{b}	**0.270**^{b}	**0.276**^{b}	**0.221**^{b}	**0.258**^{b}	**0.249**^{a,b}
ANN	**0.264**^{b}	0.275^{b}	**0.278**^{b}	0.225^{b}	0.262^{b}	0.253^{a,b}
HAR-ANN	**0.262**^{b}	**0.272**^{b}	**0.276**^{b}	**0.222**^{b}	**0.259**^{b}	**0.250**^{a,b}
			$h=5$			
ARFIMA	**0.386**	**0.406**	**0.393**	**0.333**	**0.389**	**0.392**^{a}
HAR	**0.396**	**0.412**	**0.395**	**0.339**^{a}	**0.393**^{a}	**0.398**^{a}
ANN	0.409^{b}	0.433^{b}	0.416^{b}	0.354^{a,b}	0.413^{a,b}	0.416^{a,b}
HAR-ANN	0.401^{b}	0.420^{b}	0.404^{b}	0.345^{a,b}	0.402^{a,b}	0.405^{a,b}
			$h=10$			
ARFIMA	**0.509**	**0.537**	**0.517**	**0.448**	**0.516**^{a}	**0.540**^{a}
HAR	**0.528**^{b}	**0.543**^{b}	**0.520**^{b}	0.474^{a}	**0.525**^{a,b}	0.561^{a}
ANN	0.544^{b}	0.565^{b}	0.543^{b}	0.494^{a}	0.547^{a,b}	0.587^{a}
HAR-ANN	**0.533**^{b}	**0.550**^{b}	**0.527**^{b}	0.481^{a,b}	**0.533**^{a,b}	0.571^{a,b}

Natural Gas

	TSRV	RV	RK	JWTSRV	CBV	MedRV
RMSE			$h=1$			
ARFIMA	0.584	0.630	**0.636**^{b}	**0.509**^{b}	**0.545**^{a,b}	**0.556**^{a,b}
HAR	**0.575**^{b}	**0.622**^{b}	**0.629**^{b}	**0.504**^{b}	**0.539**^{a,b}	**0.550**^{a,b}
ANN	**0.575**^{b}	**0.621**^{b}	**0.631**^{b}	**0.501**^{b}	**0.540**^{a,b}	**0.550**^{a,b}
HAR-ANN	**0.573**^{b}	**0.620**^{b}	**0.628**^{b}	**0.501**^{b}	**0.538**^{a,b}	**0.548**^{a,b}
			$h=5$			
ARFIMA	0.974	1.032	1.038	0.881^{a,b}	0.915^{a}	0.893^{a}
HAR	0.920	**0.975**^{b}	**0.979**^{b}	**0.848**^{b}	0.872^{a}	0.859^{a}
ANN	**0.906**^{b}	**0.969**^{b}	**0.975**^{b}	**0.853**^{b}	**0.858**^{a,b}	**0.859**^{a,b}
HAR-ANN	**0.906**^{b}	**0.963**^{b}	**0.970**^{b}	**0.845**^{b}	**0.858**^{a,b}	**0.850**^{a,b}
			$h=10$			
ARFIMA	1.370	1.415	1.432	1.240^{a,b}	1.294^{a}	1.247^{a}
HAR	1.273	**1.305**	**1.328**^{b}	**1.177**^{a,b}	**1.216**^{a}	**1.181**^{a}
ANN	**1.238**^{b}	**1.277**^{b}	**1.315**^{a,b}	**1.163**^{a,b}	**1.188**^{a,b}	**1.162**^{a,b}
HAR-ANN	**1.244**^{b}	**1.276**^{b}	**1.309**^{b}	**1.161**^{a,b}	**1.190**^{a,b}	**1.158**^{a,b}
MAE			$h=1$			
ARFIMA	0.429^{b}	0.449^{b}	**0.457**^{b}	**0.378**^{b}	**0.405**^{a}	**0.409**^{a}
HAR	**0.421**^{b}	**0.442**^{b}	**0.452**^{b}	**0.374**^{b}	**0.400**^{a}	**0.403**^{a}
ANN	**0.419**^{b}	**0.441**^{b}	**0.453**^{b}	**0.372**^{b}	0.403^{a,b}	0.406^{a,b}
HAR-ANN	**0.418**^{b}	**0.439**^{b}	**0.451**^{b}	**0.372**^{b}	**0.400**^{a,b}	**0.403**^{a,b}
			$h=5$			
ARFIMA	0.770	0.794	0.793	0.698	0.726^{a}	0.709^{a}
HAR	0.714	**0.736**^{b}	**0.748**^{b}	**0.666**^{b}	0.688^{a}	0.669^{a}
ANN	**0.703**^{b}	**0.733**^{b}	**0.750**^{b}	**0.669**^{b}	**0.672**^{a,b}	**0.668**^{a,b}
HAR-ANN	**0.701**^{b}	**0.725**^{b}	**0.741**^{b}	**0.663**^{b}	**0.673**^{a,b}	**0.663**^{a,b}
			$h=10$			
ARFIMA	1.135	1.159	1.162^{b}	1.044^{a}	1.080^{a}	1.041^{a}
HAR	1.043	1.056	**1.094**^{b}	**0.982**^{b}	1.019^{a}	0.973^{a}
ANN	**0.996**^{b}	**1.027**^{b}	**1.069**^{b}	**0.956**^{b}	**0.979**^{a,b}	**0.940**^{a,b}
HAR-ANN	**1.011**^{b}	**1.028**^{b}	**1.071**^{b}	**0.963**^{b}	**0.991**^{a,b}	**0.947**^{a,b}

Table 3: **September 2008 – November 2010**: RMSE/MAE. 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 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. Note that numbers are multiplied by $\times 10^2$.

Crude Oil

	TSRV	RV	RK	JWTSRV	CBV	MedRV
RMSE			$h = 1$			
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}
HAR	**0.522**^{a,b}	**0.525**^{a,b}	**0.569**^{a,b}	**0.489**^{a,b}	**0.526**^{a,b}	**0.517**^{a,b}
ANN	**0.524**^{b}	**0.530**^{b}	**0.573**^{b}	**0.533**^{a,b}	**0.530**^{a,b}	0.524^{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}
			$h = 5$			
ARFIMA	**0.877**^{a}	0.879^{a,b}	**0.942**^{a}	**0.852**^{a}	0.882^{a,b}	**0.874**^{a}
HAR	**0.779**^{a}	**0.776**^{a,b}	**0.848**^{a}	**0.764**^{a}	**0.772**^{a,b}	**0.780**^{a}
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}
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}
			$h = 10$			
ARFIMA	**1.345**^{a,b}	**1.336**^{a,b}	**1.420**^{a,b}	**1.312**^{a,b}	**1.344**^{a,b}	**1.313**^{a}
HAR	**1.114**^{b}	**1.108**^{b}	**1.215**^{a,b}	**1.088**^{a,b}	**1.105**^{a,b}	**1.096**^{a}
ANN	1.135^{b}	**1.115**^{b}	**1.234**^{b}	**1.086**^{a,b}	**1.113**^{a,b}	1.134^{a,b}
HAR-ANN	**1.110**^{b}	**1.099**^{b}	**1.204**^{b}	**1.075**^{a,b}	**1.098**^{a,b}	**1.097**^{a,b}
MAE			$h = 1$			
ARFIMA	**0.364**	**0.365**^{b}	**0.400**	**0.343**^{b}	**0.357**^{b}	**0.354**^{a}
HAR	**0.359**	**0.364**^{b}	**0.398**^{b}	**0.339**^{b}	**0.359**^{b}	**0.357**^{a,b}
ANN	0.362^{b}	**0.365**^{b}	0.404^{b}	**0.351**^{b}	**0.361**^{b}	0.364^{a,b}
HAR-ANN	**0.356**^{b}	**0.362**^{b}	**0.396**^{b}	**0.343**^{b}	**0.358**^{a,b}	**0.358**^{a,b}
			$h = 5$			
ARFIMA	**0.593**	**0.600**	**0.636**	**0.584**	**0.600**	**0.593**^{a}
HAR	**0.559**	**0.563**	**0.612**^{b}	**0.547**	**0.560**	**0.558**^{a,b}
ANN	0.580^{b}	0.576^{b}	0.637^{b}	0.568^{b}	0.574^{b}	0.583^{a,b}
HAR-ANN	**0.562**^{b}	**0.563**^{b}	**0.616**^{b}	**0.549**^{b}	**0.561**^{b}	**0.561**^{a,b}
			$h = 10$			
ARFIMA	**0.916**^{b}	**0.920**^{b}	**0.964**^{b}	**0.902**^{b}	**0.928**^{b}	**0.923**^{a,b}
HAR	**0.813**^{b}	**0.807**^{b}	**0.884**^{b}	**0.792**^{b}	**0.804**^{b}	**0.813**^{a,b}
ANN	0.830^{b}	0.812^{b}	**0.896**^{b}	**0.788**^{a,b}	**0.814**^{a,b}	**0.831**^{a,b}
HAR-ANN	**0.810**^{b}	**0.795**^{b}	**0.877**^{b}	**0.779**^{b}	**0.798**^{a,b}	**0.809**^{a,b}

Heating Oil

	TSRV	RV	RK	JWTSRV	CBV	MedRV
RMSE			$h = 1$			
ARFIMA	**0.420**	**0.421**	**0.466**	**0.378**	**0.411**	**0.395**^{a}
HAR	**0.416**^{b}	**0.412**	**0.464**^{b}	**0.376**^{b}	**0.403**^{b}	**0.388**^{a,b}
ANN	0.420^{b}	0.419^{b}	0.471^{b}	0.381^{b}	0.410^{b}	0.395^{a,b}
HAR-ANN	**0.416**^{b}	**0.413**^{b}	**0.464**^{b}	**0.377**^{b}	**0.404**^{b}	**0.389**^{a,b}
			$h = 5$			
ARFIMA	**0.684**	**0.691**^{a}	**0.751**^{a}	**0.650**^{a}	**0.677**^{a}	**0.653**^{a}
HAR	**0.643**^{b}	**0.639**^{b}	**0.715**^{b}	**0.630**^{b}	**0.630**^{a,b}	**0.614**^{a,b}
ANN	0.680^{a,b}	0.667^{a,b}	0.762^{a,b}	0.658^{a,b}	0.664^{a,b}	0.651^{a,b}
HAR-ANN	**0.650**^{a,b}	**0.643**^{a,b}	**0.726**^{a,b}	**0.636**^{a,b}	**0.638**^{a,b}	**0.621**^{a,b}
			$h = 10$			
ARFIMA	**0.997**	**1.025**	**1.075**^{a}	**0.957**^{a}	**1.003**^{a}	**0.964**^{a}
HAR	**0.904**^{b}	**0.918**	**0.988**^{b}	**0.903**^{a,b}	**0.905**^{a}	**0.880**^{a,b}
ANN	0.953^{b}	0.963^{b}	1.064^{b}	0.945^{a,b}	0.942^{a,b}	0.962^{a,b}
HAR-ANN	**0.911**^{b}	**0.926**^{b}	**1.005**^{b}	**0.911**^{a,b}	**0.910**^{a,b}	**0.899**^{a,b}
MAE			$h = 1$			
ARFIMA	**0.304**^{b}	0.302	**0.341**	**0.278**^{b}	**0.295**	**0.291**^{a}
HAR	**0.304**^{b}	**0.294**	**0.340**^{b}	**0.278**^{b}	**0.288**	**0.288**^{a,b}
ANN	**0.305**^{b}	0.299^{b}	0.345^{b}	**0.279**^{b}	0.293^{b}	0.294^{a,b}
HAR-ANN	**0.302**^{b}	**0.294**^{b}	**0.340**^{b}	**0.277**^{b}	**0.288**^{b}	**0.289**^{a,b}
			$h = 5$			
ARFIMA	**0.495**	**0.501**	**0.536**	**0.487**	**0.498**	**0.486**^{a}
HAR	**0.469**^{b}	**0.466**^{b}	**0.523**^{b}	**0.465**^{b}	**0.461**^{b}	**0.450**^{a,b}
ANN	0.498^{b}	0.490^{b}	0.555^{b}	0.491^{b}	0.491^{b}	0.486^{a,b}
HAR-ANN	**0.475**^{b}	**0.470**^{b}	**0.530**^{b}	**0.473**^{b}	**0.467**^{b}	**0.457**^{a,b}
			$h = 10$			
ARFIMA	**0.736**	**0.759**	**0.782**	**0.727**	**0.752**	**0.734**^{a}
HAR	**0.681**^{b}	**0.681**	**0.740**^{b}	**0.680**^{b}	**0.679**	**0.669**^{a,b}
ANN	0.712^{b}	0.722^{b}	0.801^{b}	0.719^{b}	0.717^{b}	0.731^{a,b}
HAR-ANN	**0.682**^{b}	**0.688**^{b}	**0.751**^{b}	**0.685**^{b}	**0.686**^{b}	**0.682**^{a,b}

Natural Gas

	TSRV	RV	RK	JWTSRV	CBV	MedRV
RMSE			$h = 1$			
ARFIMA	0.897	0.971	0.931	**0.545**^{b}	**0.724**^{b}	**0.766**^{b}
HAR	0.872	0.936	0.907	**0.548**^{b}	**0.716**^{b}	**0.759**^{b}
ANN	**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	**0.859**^{a,b}	**0.918**^{a,b}	**0.894**^{a,b}	**0.548**^{a,b}	**0.716**^{a,b}	**0.757**^{a,b}
			$h = 5$			
ARFIMA	**1.339**^{b}	**1.403**^{b}	**1.342**^{b}	**1.030**^{a,b}	**1.225**^{a,b}	**1.277**^{a,b}
HAR	**1.327**^{b}	**1.387**^{b}	**1.325**^{b}	**1.061**^{b}	**1.250**^{a,b}	**1.312**^{a,b}
ANN	**1.355**^{b}	**1.409**^{b}	**1.349**^{b}	**1.082**^{b}	**1.308**^{a,b}	**1.369**^{a,b}
HAR-ANN	**1.335**^{b}	**1.390**^{b}	**1.329**^{b}	**1.065**^{b}	**1.271**^{a,b}	**1.329**^{a,b}
			$h = 10$			
ARFIMA	**1.864**^{b}	**1.979**^{b}	**1.862**^{b}	**1.501**^{b}	**1.761**^{a,b}	**1.793**^{a,b}
HAR	**1.861**^{b}	**1.975**^{b}	**1.845**^{b}	**1.538**^{b}	**1.790**^{b}	**1.826**^{b}
ANN	**1.882**^{b}	**2.010**^{b}	**1.831**^{b}	**1.549**^{b}	**1.821**^{b}	**1.805**^{b}
HAR-ANN	**1.851**^{b}	**1.968**^{b}	**1.810**^{b}	**1.531**^{b}	**1.775**^{b}	**1.798**^{b}
MAE			$h = 1$			
ARFIMA	0.607	0.653	0.642^{b}	**0.409**^{b}	**0.509**^{b}	**0.523**^{b}
HAR	0.597	0.641	**0.632**^{b}	**0.410**^{b}	**0.509**^{b}	**0.525**^{b}
ANN	**0.582**^{b}	**0.625**^{b}	**0.623**^{b}	**0.414**^{b}	**0.512**^{b}	**0.530**^{b}
HAR-ANN	**0.585**^{b}	**0.629**^{b}	**0.623**^{b}	**0.411**^{b}	**0.509**^{b}	**0.526**^{b}
			$h = 5$			
ARFIMA	**0.940**^{b}	**0.982**^{b}	**0.947**^{b}	**0.718**	**0.841**^{a}	**0.852**^{a}
HAR	**0.942**^{b}	**0.996**^{b}	**0.956**^{b}	**0.744**^{b}	**0.877**^{a}	**0.894**^{a}
ANN	**0.957**^{b}	**1.001**^{b}	**0.966**^{b}	0.757^{b}	0.893^{a}	0.915^{a}
HAR-ANN	**0.944**^{b}	**0.994**^{b}	**0.955**^{b}	**0.745**^{b}	**0.880**^{a,b}	**0.897**^{a,b}
			$h = 10$			
ARFIMA	**1.319**^{b}	**1.404**^{b}	**1.309**^{b}	**1.013**	**1.195**^{a,b}	**1.180**^{a,b}
HAR	**1.378**^{b}	**1.445**^{b}	**1.331**^{b}	**1.078**	**1.265**^{b}	**1.253**^{b}
ANN	**1.358**^{b}	**1.442**^{b}	**1.289**^{b}	1.079	**1.254**^{b}	**1.237**^{b}
HAR-ANN	**1.352**^{b}	**1.424**^{b}	**1.291**^{b}	**1.071**^{b}	**1.244**^{b}	**1.231**^{b}

Table 4: **After November 2010**: RMSE/MAE. 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 $\widehat{\mathcal{M}}^*_{10\%}$. 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. Note that the numbers are multiplied by $\times 10^2$.

Crude Oil

	TSRV	RV	RK	JWTSRV	CBV	MedRV
RMSE			$h=1$			
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}
			$h=5$			
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}
			$h=10$			
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			$h=1$			
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}
			$h=5$			
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}
			$h=10$			
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			$h=1$			
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}
			$h=5$			
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}
			$h=10$			
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	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			$h=1$			
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**	**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}
			$h=5$			
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}
			$h=10$			
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			$h=1$			
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}
			$h=5$			
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}
			$h=10$			
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}	**1.030**^{a}
HAR-ANN	**1.124**^{b}	**1.183**^{b}	**1.144**^{b}	0.953^{b}	1.062^{a,b}	1.035^{a,b}
MAE			$h=1$			
ARFIMA	0.468	0.487	0.480	**0.286**	0.366^{a}	0.345^{a}
HAR	0.447	0.462	0.465	**0.290**	**0.360**^{a,b}	**0.344**^{a,b}
ANN	**0.433**^{b}	**0.443**^{b}	**0.460**^{b}	0.290	**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}
			$h=5$			
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}
			$h=10$			
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: **Whole period**: RMSE/MAE. 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 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. Note that numbers are multiplied by $\times 10^2$.

	Crude Oil					
	TSRV	RV	RK	JWTSRV	CBV	MedRV
RMSE			$h=1$			
ARFIMA	**0.417**^{b}	**0.419**^{b}	**0.443**	**0.378**^{b}	**0.411**^{b}	**0.401**^{a,b}
HAR	**0.413**^{b}	**0.414**^{b}	**0.440**	**0.375**^{b}	**0.406**^{b}	**0.397**^{a,b}
ANN	**0.414**^{b}	0.415^b	0.441^b	**0.394**^{b}	**0.407**^{a,b}	0.399^{a,b}
HAR-ANN	**0.411**^{b}	**0.412**^{b}	**0.438**^{b}	**0.378**^{b}	**0.404**^{b}	**0.396**^{a,b}
			$h=5$			
ARFIMA	**0.709**^{a}	0.705^a	**0.735**^{a}	**0.665**^{a}	0.691^a	**0.680**^{a}
HAR	**0.666**	**0.658**	**0.690**^{a}	**0.625**^{a}	**0.643**^{a}	**0.640**^{a}
ANN	0.673^b	0.664^b	0.702^{a,b}	0.632^{a,b}	0.649^{a,b}	0.657^{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}
			$h=10$			
ARFIMA	**1.045**^{b}	**1.035**^{a,b}	**1.073**^{a}	**0.994**^{a,b}	**1.021**^{a,b}	**0.998**^{a}
HAR	**0.933**^{b}	**0.924**^{b}	**0.966**	**0.887**^{b}	**0.909**^{b}	**0.900**^{a}
ANN	0.941^b	**0.927**^{b}	0.976^b	**0.885**^{a,b}	**0.915**^{a,b}	0.918^{a,b}
HAR-ANN	**0.929**^{b}	**0.917**^{b}	**0.959**^{b}	**0.878**^{b}	**0.905**^{b}	**0.898**^{a,b}
MAE			$h=1$			
ARFIMA	**0.285**	**0.288**^{b}	**0.307**	**0.262**^{b}	**0.278**^{b}	**0.272**^{a}
HAR	**0.287**^{b}	**0.292**^{b}	**0.311**^{b}	**0.263**^{b}	**0.282**^{b}	**0.276**^{a,b}
ANN	0.288^b	0.293^b	0.313^b	**0.266**^{b}	0.284^b	0.278^{a,b}
HAR-ANN	**0.286**^{b}	**0.291**^{b}	**0.310**^{b}	**0.263**^{b}	**0.282**^{b}	**0.276**^{a,b}
			$h=5$			
ARFIMA	**0.475**	**0.481**	**0.494**	**0.450**	**0.469**	**0.455**^{a}
HAR	**0.467**^{b}	**0.474**^{b}	**0.488**^{b}	**0.444**^{b}	**0.464**^{b}	**0.452**^{a,b}
ANN	0.478^b	0.482^b	0.502^b	0.454^b	0.472^b	0.465^{a,b}
HAR-ANN	**0.468**^{b}	**0.475**^{b}	**0.490**^{b}	**0.445**^{b}	**0.465**^{b}	**0.454**^{a,b}
			$h=10$			
ARFIMA	**0.695**	**0.701**	**0.716**	**0.668**^{b}	**0.689**	**0.671**^{a}
HAR	**0.665**	**0.669**	**0.689**	**0.637**^{b}	**0.657**	**0.652**^{a,b}
ANN	0.675^b	0.676^b	0.701^b	**0.639**^{b}	0.665^b	0.666^{a,b}
HAR-ANN	**0.663**^{b}	**0.665**^{b}	**0.688**^{b}	**0.632**^{b}	**0.655**^{b}	**0.652**^{a,b}

	Heating Oil					
	TSRV	RV	RK	JWTSRV	CBV	MedRV
RMSE			$h=1$			
ARFIMA	**0.350**	**0.361**	**0.376**	**0.303**	**0.346**	**0.344**^{a}
HAR	**0.351**^{b}	**0.359**^{b}	**0.378**^{b}	**0.304**^{b}	**0.344**^{b}	**0.342**^{a,b}
ANN	0.353^b	0.363^b	0.382^b	0.308^b	0.348^b	0.347^{a,b}
HAR-ANN	**0.350**^{b}	**0.359**^{b}	**0.378**^{b}	**0.305**^{b}	**0.344**^{b}	**0.343**^{a,b}
			$h=5$			
ARFIMA	**0.564**	**0.581**	**0.596**	**0.514**	**0.558**^{a}	**0.561**^{a}
HAR	**0.562**^{b}	**0.572**^{b}	**0.591**^{b}	**0.519**^{b}	**0.551**^{a,b}	**0.559**^{a,b}
ANN	0.585^b	0.593^b	0.621^b	0.537^b	0.573^{a,b}	0.582^{a,b}
HAR-ANN	**0.567**^{b}	**0.576**^{b}	**0.599**^{b}	**0.523**^{b}	**0.557**^{a,b}	**0.565**^{a,b}
			$h=10$			
ARFIMA	**0.783**	**0.805**	**0.818**	**0.729**^{a}	**0.780**^{a}	**0.780**^{a}
HAR	**0.772**^{b}	**0.783**^{b}	**0.800**	**0.735**^{b}	**0.764**^{b}	**0.775**^{a,b}
ANN	0.806^b	0.816^b	0.850	0.763^{a,b}	0.792^{a,b}	0.821^{a,b}
HAR-ANN	**0.779**^{b}	**0.790**^{b}	0.813^b	**0.741**^{a,b}	**0.770**^{a,b}	**0.786**^{a,b}
MAE			$h=1$			
ARFIMA	**0.253**	**0.256**	**0.275**	**0.223**	**0.247**	**0.242**^{a}
HAR	**0.253**^{b}	**0.253**^{b}	**0.274**^{b}	**0.223**^{b}	**0.244**^{b}	**0.241**^{a,b}
ANN	0.255^b	0.258^b	0.279^b	0.226^b	0.249^b	0.245^{a,b}
HAR-ANN	**0.253**^{b}	**0.254**^{b}	**0.275**^{b}	**0.224**^{b}	**0.245**^{b}	**0.242**^{a,b}
			$h=5$			
ARFIMA	**0.402**	**0.415**	**0.421**	**0.375**	**0.405**	**0.401**^{a}
HAR	**0.409**	**0.417**	**0.426**	**0.381**	**0.405**	**0.401**^{a}
ANN	0.429^b	0.437^b	0.451^b	0.399^b	0.426^b	0.423^{a,b}
HAR-ANN	0.415^b	0.423^b	0.434^b	0.387^b	0.412^b	0.407^{a,b}
			$h=10$			
ARFIMA	**0.566**	**0.589**	**0.582**	**0.534**	**0.576**	**0.576**^{a}
HAR	**0.577**^{b}	**0.585**^{b}	**0.588**	**0.552**^{b}	**0.576**^{b}	**0.582**^{a,b}
ANN	0.604^b	0.616^b	0.631^b	0.581^b	0.603^b	0.620^{a,b}
HAR-ANN	**0.583**^{b}	**0.593**^{b}	0.600^b	**0.559**^{b}	**0.583**^{b}	**0.592**^{a,b}

	Natural Gas					
	TSRV	RV	RK	JWTSRV	CBV	MedRV
RMSE			$h=1$			
ARFIMA	0.714	0.765	0.747	**0.479**^{b}	0.594^a	0.606^a
HAR	0.697	0.744	0.732	**0.480**^{b}	**0.587**^{a,b}	**0.602**^{a,b}
ANN	**0.684**^{b}	**0.725**^{b}	**0.723**^{b}	**0.480**^{b}	**0.588**^{a,b}	**0.601**^{a,b}
HAR-ANN	**0.686**^{b}	**0.729**^{b}	**0.723**^{b}	**0.479**^{b}	**0.586**^{a,b}	**0.600**^{a,b}
			$h=5$			
ARFIMA	**1.060**^{b}	**1.114**^{b}	**1.088**^{b}	**0.861**^{b}	**0.970**^{a,b}	**0.984**^{a,b}
HAR	**1.047**^{b}	**1.099**^{b}	**1.072**^{b}	**0.872**^{b}	**0.978**^{a,b}	**1.000**^{a,b}
ANN	**1.054**^{b}	**1.106**^{b}	**1.081**^{b}	**0.879**^{b}	**0.995**^{a,b}	**1.022**^{a,b}
HAR-ANN	**1.045**^{b}	**1.095**^{b}	**1.070**^{b}	**0.870**^{b}	**0.980**^{a,b}	**1.002**^{a,b}
			$h=10$			
ARFIMA	**1.474**^{b}	**1.547**^{b}	**1.496**^{b}	**1.240**^{b}	**1.390**^{a,b}	**1.383**^{a,b}
HAR	**1.457**^{b}	**1.530**^{b}	**1.471**^{b}	**1.250**^{b}	**1.398**^{b}	**1.397**^{b}
ANN	**1.457**^{b}	**1.539**^{b}	**1.464**^{b}	**1.246**^{b}	**1.397**^{a,b}	**1.375**^{a,b}
HAR-ANN	**1.443**^{b}	**1.517**^{b}	**1.450**^{b}	**1.239**^{b}	**1.378**^{a,b}	**1.372**^{a,b}
MAE			$h=1$			
ARFIMA	0.501	0.530	0.526	**0.358**^{b}	**0.427**^{a}	**0.426**^{a}
HAR	0.488	0.515	0.517	**0.358**^{b}	**0.423**^{a}	**0.424**^{a}
ANN	**0.478**^{b}	**0.503**^{b}	**0.512**^{b}	**0.359**^{b}	0.425^{a,b}	0.427^{a,b}
HAR-ANN	**0.480**^{b}	**0.504**^{b}	**0.511**^{b}	**0.357**^{b}	**0.423**^{a,b}	**0.424**^{a,b}
			$h=5$			
ARFIMA	**0.769**^{b}	**0.800**^{b}	**0.787**	**0.630**	**0.697**^{a,b}	**0.696**^{a,b}
HAR	**0.760**^{b}	**0.796**^{b}	**0.786**^{b}	**0.637**^{b}	**0.706**^{a,b}	**0.707**^{a,b}
ANN	**0.762**^{b}	**0.799**^{b}	0.794^b	0.642^b	**0.705**^{a,b}	**0.713**^{a,b}
HAR-ANN	**0.755**^{b}	**0.791**^{b}	**0.784**^{b}	**0.635**^{b}	**0.700**^{a,b}	**0.705**^{a,b}
			$h=10$			
ARFIMA	**1.091**	**1.137**^{b}	**1.103**	**0.913**^{b}	**1.007**^{a}	**0.982**^{a}
HAR	1.096^b	**1.139**^{b}	1.107^b	**0.929**^{b}	1.034^b	1.007^b
ANN	**1.078**^{b}	**1.132**^{b}	**1.089**^{b}	**0.922**^{b}	**1.010**^{a,b}	**0.985**^{a,b}
HAR-ANN	**1.076**^{b}	**1.123**^{b}	**1.086**^{b}	**0.919**^{b}	**1.012**^{a,b}	**0.987**^{a,b}

Table 6: **Whole period**: VaR (long). The table reports unconditional coverage. 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				$h=1$						$h=1$						$h=1$		
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}
				$h=5$						$h=5$						$h=5$		
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}
				$h=10$						$h=10$						$h=10$		
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				$h=1$						$h=1$						$h=1$		
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}
				$h=5$						$h=5$						$h=5$		
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}
				$h=10$						$h=10$						$h=10$		
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: **Whole period**: VaR (short). The table reports unconditional coverage. 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					
	TSRV	RV	RK	JWTSRV	CBV	MedRV
99% VaR			$h = 1$			
ARFIMA	**98.651**	**98.590**^{b}	**98.406**^{b}	**98.345**^{a,b}	**98.406**^{a}	**98.099**^{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}
ANN	**98.712**^{b}	**98.774**^{a,b}	**98.651**^{a,b}	**98.651**^{a,b}	**98.651**^{a,b}	**98.099**^{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}
			$h = 5$			
ARFIMA	99.262	99.262	99.262	**99.017**^{b}	**99.017**^{a}	**98.648**^{a,b}
HAR	99.385	99.508	99.447	**99.262**^{b}	**99.447**^{b}	**99.078**^{a,b}
ANN	99.570	99.508	99.508	**99.447**^{b}	**99.447**^{b}	**99.201**^{a,b}
HAR-ANN	**99.570**^{b}	**99.508**^{b}	**99.508**^{b}	**99.385**^{b}	**99.447**^{b}	**99.201**^{a,b}
			$h = 10$			
ARFIMA	99.260	99.260	99.260	99.075	99.260	98.890^{a}
HAR	99.753	99.753	99.692	99.630	99.630	99.383^{a}
ANN	99.630	99.630	99.568	99.445	99.630	99.322^{a}
HAR-ANN	**99.630**^{b}	**99.692**^{b}	**99.630**^{b}	**99.507**^{b}	**99.630**^{b}	**99.322**^{a,b}
95% VaR			$h = 1$			
ARFIMA	**94.482**^{b}	**94.298**^{b}	**94.237**^{a,b}	**93.746**^{a}	**93.930**^{a}	**93.256**^{a,b}
HAR	94.543^{b}	94.359^{b}	94.421^{b}	94.237	94.237	**93.624**^{a,b}
ANN	**94.850**^{b}	94.359^{b}	94.605^{b}	94.543	94.482	**93.930**^{a,b}
HAR-ANN	**94.788**^{b}	**94.421**^{b}	**94.727**^{b}	**94.298**^{b}	**94.359**^{b}	**93.746**^{a,b}
			$h = 5$			
ARFIMA	94.776	94.530	94.345	94.161	94.038^{a}	93.239^{a}
HAR	95.698	95.882	95.698	95.022	95.452	94.407^{a}
ANN	95.636	95.698	95.636	95.022	95.267	94.345^{a}
HAR-ANN	**95.698**^{b}	**95.882**^{b}	**95.698**^{b}	**95.083**^{b}	**95.452**^{b}	**94.407**^{a,b}
			$h = 10$			
ARFIMA	95.191	95.376	95.006	94.698	94.575	93.711^{a}
HAR	95.993	95.993	96.054	95.746	95.869	95.314^{a}
ANN	95.993	95.931	96.239	95.869	95.869	95.253^{a}
HAR-ANN	**95.993**^{b}	**96.054**^{b}	**96.054**^{b}	**95.808**^{b}	**95.993**^{b}	**95.191**^{a,b}

	Heating Oil					
	TSRV	RV	RK	JWTSRV	CBV	MedRV
99% VaR			$h = 1$			
ARFIMA	**98.767**^{b}	**98.644**^{b}	**98.767**^{b}	**98.705**^{b}	**98.582**^{b}	**98.274**^{a,b}
HAR	**98.767**^{b}	**98.520**^{b}	**98.890**^{b}	**98.582**^{b}	**98.582**^{b}	**98.335**^{a,b}
ANN	**98.829**^{b}	**98.582**^{b}	**98.829**^{b}	**98.644**^{b}	**98.459**^{b}	**98.335**^{a,b}
HAR-ANN	**98.705**^{b}	**98.582**^{b}	**98.890**^{b}	**98.582**^{b}	**98.582**^{b}	**98.274**^{a,b}
			$h = 5$			
ARFIMA	99.320	99.320	99.320	99.258	99.258	**99.073**^{a}
HAR	99.444	99.506	99.567	99.197	99.382	**99.073**^{a}
ANN	99.506	99.444	99.506	99.258	99.382	99.197^{a}
HAR-ANN	**99.567**^{b}	**99.444**^{b}	**99.506**^{b}	**99.197**^{b}	**99.320**^{b}	**99.135**^{a,b}
			$h = 10$			
ARFIMA	99.814	99.876	99.814	99.628	99.876	99.442^{a}
HAR	99.814	99.752	99.814	99.628	99.752	99.442^{a}
ANN	99.690	99.752	99.752	99.566	99.690	99.504^{a}
HAR-ANN	**99.752**^{b}	**99.752**^{b}	**99.814**^{b}	**99.628**^{b}	**99.752**^{b}	**99.442**^{a,b}
95% VaR			$h = 1$			
ARFIMA	**95.253**^{b}	**95.191**^{b}	**95.253**^{b}	**94.636**^{b}	**94.698**^{b}	**93.896**^{a,b}
HAR	**95.068**^{b}	**95.068**^{b}	**95.006**^{b}	**94.451**^{b}	**94.451**^{b}	**93.527**^{a,b}
ANN	**95.438**^{b}	**95.314**^{b}	**95.314**^{b}	**94.760**^{b}	**94.575**^{b}	**93.773**^{a,b}
HAR-ANN	**95.191**^{b}	**95.129**^{b}	**95.191**^{b}	**94.513**^{b}	**94.513**^{b}	**93.835**^{a,b}
			$h = 5$			
ARFIMA	95.797	95.859	95.921	95.426	95.612	95.179^{a}
HAR	95.921	95.921	95.983	95.488	95.612	95.056^{a}
ANN	95.859	95.921	95.921	95.488	95.797	95.117^{a}
HAR-ANN	**95.797**^{b}	**95.921**^{b}	**95.921**^{b}	**95.426**^{b}	**95.797**^{b}	**95.117**^{a,b}
			$h = 10$			
ARFIMA	95.970	96.156	96.032	95.598	95.970	95.040^{a}
HAR	95.846	95.908	95.970	95.350	95.474	95.102^{a}
ANN	95.908	96.156	96.156	95.598	96.032	95.226^{a}
HAR-ANN	**95.846**^{b}	**96.094**^{b}	**95.970**^{b}	**95.350**^{b}	**95.784**^{b}	**95.102**^{a,b}

	Natural Gas					
	TSRV	RV	RK	JWTSRV	CBV	MedRV
99% VaR			$h = 1$			
ARFIMA	**98.703**^{b}	**98.826**^{b}	**98.888**^{b}	**98.023**^{b}	**98.394**^{a,b}	**98.147**^{a,b}
HAR	**98.765**^{b}	**98.950**^{b}	**98.888**^{b}	**97.962**^{a,b}	**98.456**^{a,b}	**98.085**^{a,b}
ANN	**98.641**^{b}	**98.888**^{b}	**98.888**^{a,b}	**97.900**^{a,b}	**98.394**^{a,b}	**98.147**^{a,b}
HAR-ANN	**98.765**^{b}	**98.950**^{b}	**98.888**^{a,b}	**97.962**^{a,b}	**98.394**^{a,b}	**97.962**^{a,b}
			$h = 5$			
ARFIMA	99.257	99.319	99.257	98.576	98.824^{a}	98.638^{a}
HAR	99.257	99.319	99.257	98.638	98.700^{a}	98.700^{a}
ANN	99.195	99.381	99.195	98.638	98.824^{a}	98.700^{a}
HAR-ANN	**99.257**^{b}	**99.443**^{b}	**99.257**^{b}	**98.638**^{b}	**98.824**^{a,b}	**98.700**^{a,b}
			$h = 10$			
ARFIMA	99.689	99.627	99.627	99.317	99.565^{a}	99.255^{a}
HAR	99.627	99.627	99.565	99.379	99.441^{a}	99.379^{a}
ANN	99.565	99.379	99.503	99.317	99.441^{a}	99.193^{a}
HAR-ANN	**99.565**^{b}	**99.441**^{b}	**99.503**^{b}	**99.379**^{b}	**99.503**^{a,b}	**99.317**^{a,b}
95% VaR			$h = 1$			
ARFIMA	**94.441**^{b}	**94.565**^{b}	**94.194**^{b}	**93.206**^{b}	**93.638**^{a,b}	**92.959**^{a,b}
HAR	**94.194**^{b}	**94.194**^{b}	**93.885**^{b}	**93.144**^{b}	**93.391**^{a,b}	**92.588**^{a,b}
ANN	**94.132**^{b}	**94.194**^{b}	**94.256**^{b}	**92.835**^{b}	**93.391**^{a,b}	**92.712**^{a,b}
HAR-ANN	**94.009**^{b}	**94.132**^{b}	**94.132**^{b}	**92.959**^{b}	**93.515**^{a,b}	**92.773**^{a,b}
			$h = 5$			
ARFIMA	95.604	96.037	95.789	94.737	95.480^{a}	94.923^{a}
HAR	96.223	96.285	96.223	94.923	95.604^{a}	95.046^{a}
ANN	96.037	96.285	95.975	94.737	95.356^{a}	95.046^{a}
HAR-ANN	**96.223**^{b}	**96.285**^{b}	**96.161**^{b}	**94.799**^{b}	**95.418**^{a,b}	**95.046**^{a,b}
			$h = 10$			
ARFIMA	97.267	97.267	f96.957	96.025	96.584^{a}	95.963^{a}
HAR	97.516	97.516	97.143	96.211	96.646^{a}	96.087^{a}
ANN	97.205	97.267	96.957	96.087	96.584^{a}	96.149^{a}
HAR-ANN	**97.267**^{b}	**97.453**^{b}	**97.081**^{b}	**96.087**^{b}	**96.646**^{a,b}	**96.211**^{a,b}

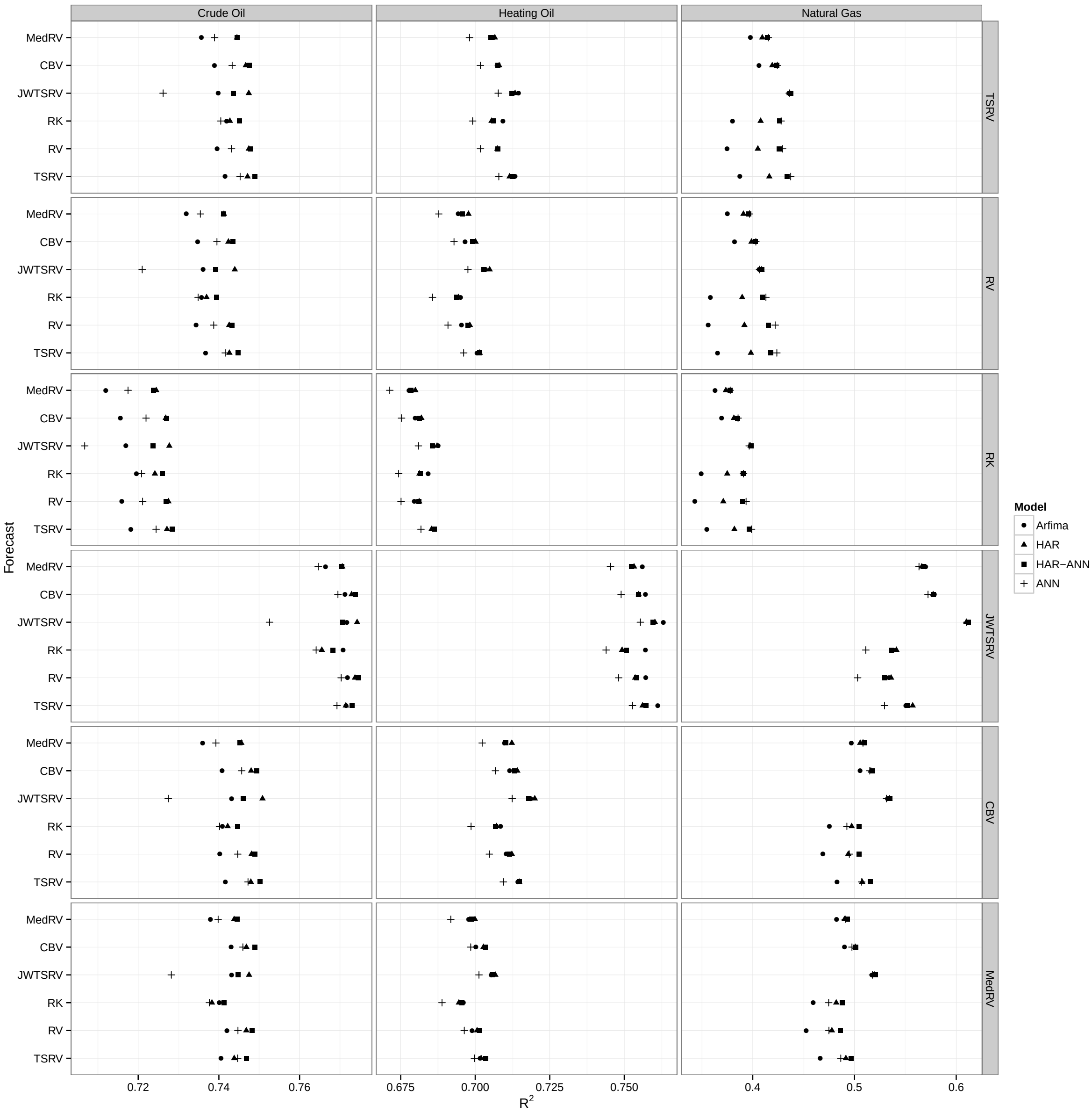

Figure 2: Whole period: R^2 from the Mincer Zarnowitz regressions $h = 1$

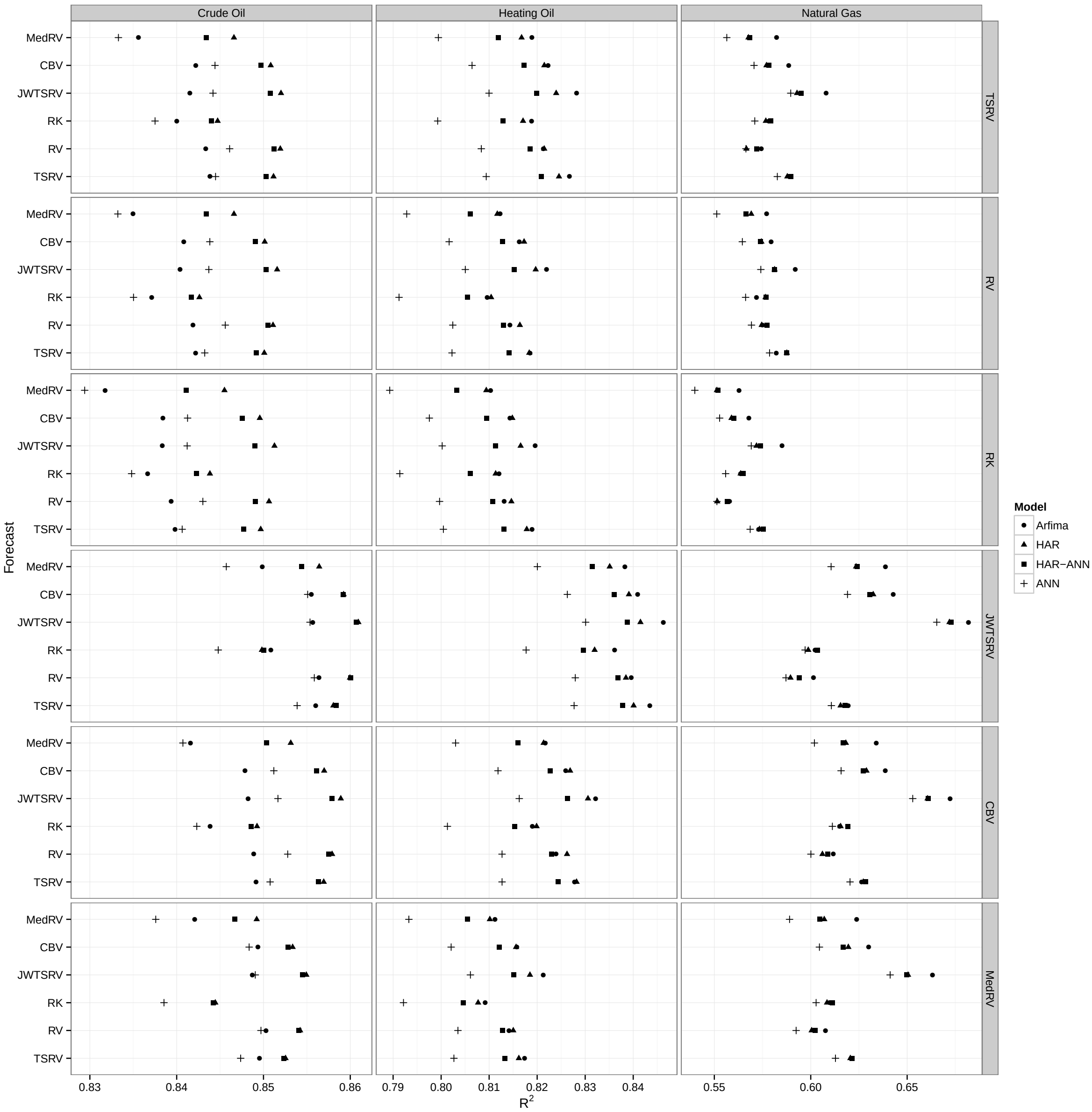

Figure 3: Whole period: R^2 from the Mincer Zarnowitz regressions $h = 5$

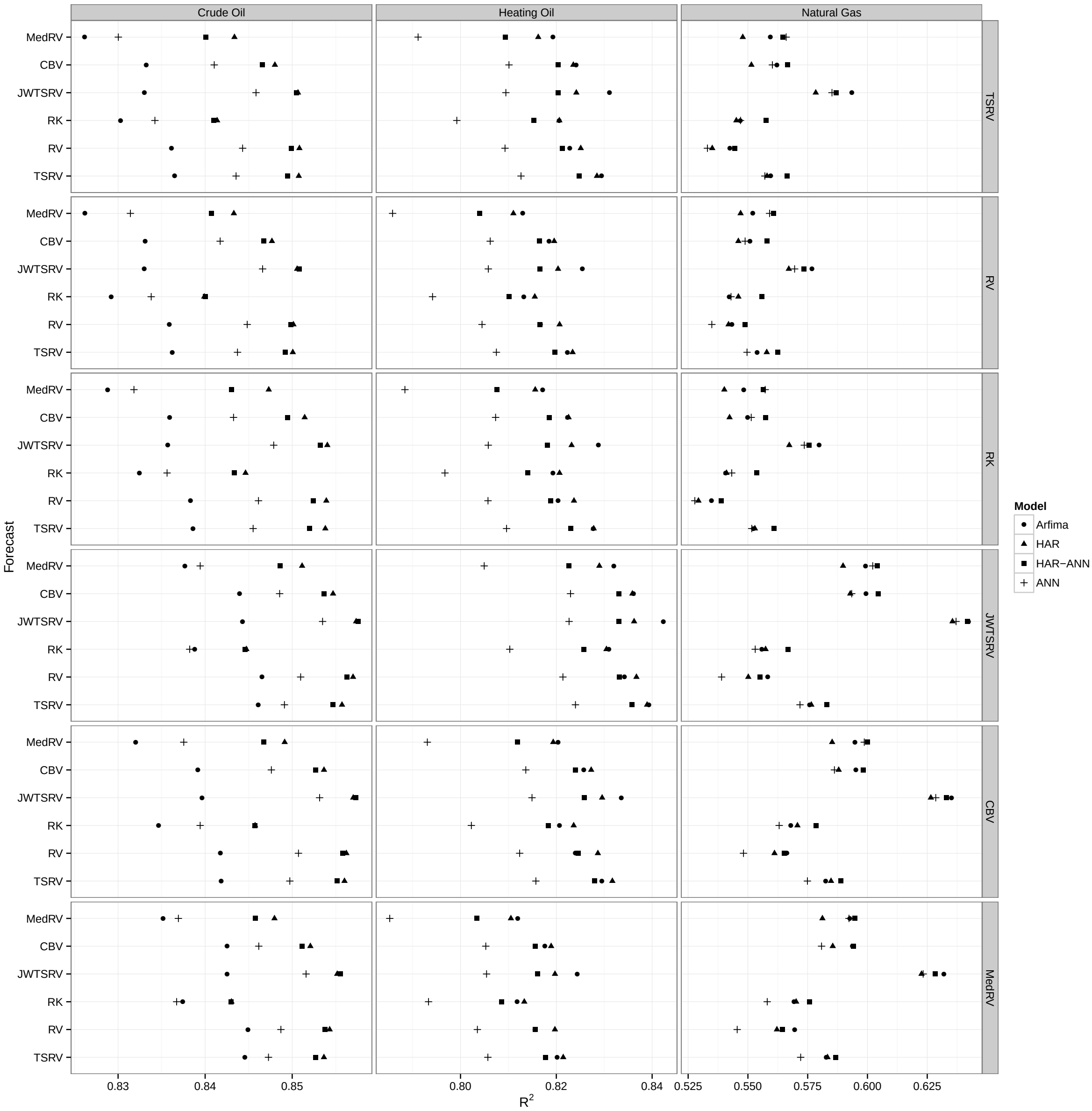

Figure 4: Whole period: R^2 from the Mincer Zarnowitz regressions $h = 10$

IES Working Paper Series

2014

1. Diana Zigraiova, Petr Jakubik: *Systemic Event Prediction by Early Warning System*
2. Pavla Blahova, Karel Janda, Ladislav Kristoufek: *The Perspectives for Genetically Modified Cellulosic Ethanol in the Czech Republic*
3. Oxana Babecká Kucharčuková: *The impact of exchange rate volatility on trade: Evidence for the Czech Republic*
4. Oliver Polyák: *The Impact of Euro Adoption on Export Performance: Comparison of the Czech Republic and Slovakia*
5. Lucie Davidová, Vladimír Benáček: *Determinants of Austrian International Trade: Analysis Based on the Gravity Model*
6. Julie Chytilová, Václav Korbel: *Individual and Group Cheating Behavior: A Field Experiment with Adolescents*
7. Petr Janský: *Consumer Demand System Estimation and Value Added Tax Reforms in the Czech Republic*
8. Pavel Hait, Petr Janský: *Inflation Differentials among Czech Households*
9. Hana Džmuráňová, Petr Teplý: *Risk management of savings accounts*
10. Tomáš Fiala, Tomáš Havránek: *Ailing Mothers, Healthy Daughters? Contagion in the Central European Banking Sector*
11. Petra Andrlíková: *Is Barrier version of Merton model more realistic? Evidence from Europe*
12. Martin Pospíšil, Jiří Schwarz: *Bankruptcy, Investment, and Financial Constraints: Evidence from a Post-Transition Economy*
13. Matt Hunt, Goran Serdarević, Tom Ovington, Clive Kenny: *Evidence for a ladder of investment in Central and Eastern European countries*
14. Petra Andrlíková: *Bayesian default probability models*
15. Petra Buzková: *Has the Relationship Between Market and Model CDS Price Changed during the EMU Debt Crisis?*
16. Jakub Černý, Jiří Witzany: *Interest Rate Swap Credit Valuation Adjustment*
17. Martin Gregor: *Receiver's access fee for a single sender*
18. Pavel Ryska: *Deflation and Economic Growth in Long-Term Perspective*
19. Michal Paulus, Eva Michalíková, Vladimír Benáček: *German International Trade: Interpreting Export Flows According to the Gravity Model*
20. Michal Bauer, Nathan Fiala, Ian Levely: *Trusting Former Rebels: An Experimental Approach to Understanding Reintegration after Civil War*
21. Michal Skořepa: *Family Matters: Concurrent Capital Buffers in a Banking Group*
22. Martin Gregor: *Access Fees for Competing Lobbies*
23. Jozef Baruník, František Čech: *On the Modelling and Forecasting Multivariate Realized Volatility: Generalized Heterogeneous Autoregressivee (GHAR) Model*
24. Roman Horváth, Ayaz Zeynalov: *The Natural Resource Curse and Institutions in Post-Soviet Countries*
25. Petr Janský: *Impact of Excise Duties on Households in the Czech Republic*
26. Petr Jakubík, Thomas Reininger: *What Are the Key Determinants of Nonperforming Loans in CESEE?*

27. Jaromír Baxa, Tomáš Olešňaník: *Current Accounts in the Eurozone Countries: The Role of Euro, Fiscal Policies and Financial Developments*
28. Petr Polák: *The Productivity Paradox: A Meta-analysis*
29. Tomáš Havránek, Zuzana Irsová, Karel Janda, David Zilberman: *Selective Reporting and the Social Cost of Carbon*
30. Jozef Baruník, Tomáš Křehlík: *Coupling High-Frequency Data with Nonlinear Models in Multiple-Step-ahead Forecasting of Energy Markets' Volatility*

All papers can be downloaded at: http://ies.fsv.cuni.cz

Univerzita Karlova v Praze, Fakulta sociálních věd
Institut ekonomických studií [UK FSV – IES] Praha 1, Opletalova 26
E-mail : ies@fsv.cuni.cz http://ies.fsv.cuni.cz