Energy Economics 80 (2019) 423–433

Contents lists available at ScienceDirect

Energy Economics

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

Forecasting oil price volatility: Forecast combination versus shrinkage method

Yaojie Zhang [a], Yu Wei [b,*], Yi Zhang [c,d], Daxiang Jin [d]

[a] School of Economics and Management, Nanjing University of Science and Technology, Nanjing, China
[b] School of Finance, Yunnan University of Finance and Economics, Kunming, China
[c] China Audit Intelligence Center, Nanjing Audit University, Nanjing, China
[d] School of Economics and Management, Southwest Jiaotong University, Chengdu, China

ARTICLE INFO

Article history:
Received 3 August 2018
Received in revised form 12 January 2019
Accepted 16 January 2019
Available online 28 January 2019

JEL classification:
C53
C58
Q47
G11
G17

Keywords:
Oil price volatility
HAR model
Forecast combination
Elastic net
Lasso

ABSTRACT

In this paper, we compare the predictive ability between forecast combination and shrinkage method in the prediction of oil price volatility. Our investigation is based on the heterogeneous autoregressive (HAR) framework. Five combination approaches combine the individual forecasts generated by the HAR model and its various extensions, while two prevailing shrinkage methods, the elastic net and lasso, employ all the predictors in our HAR framework to generate the forecast of oil price volatility. The model confidence set (MCS) test shows that the elastic net and lasso have significantly better out-of-sample forecasting performance than not only the individual extended HAR models but also the combination approaches. This result is robust across a wide range of checks. In addition, we document that the elastic net and lasso also exhibit substantially higher directional accuracy. Furthermore, a mean-variance investor can realize sizeable economic gains by using the volatility forecasts based on the shrinkage methods to allocate her portfolio.

1. Introduction

Crude oil is one of the important commodities and has a crucial impact on the global economy. Also, oil price volatility is central to asset pricing, asset allocation, and risk management. Therefore, an increasing number of works pay attention to the prediction of oil price volatility. For example, Haugom et al. (2014), Sévi (2014), Liu et al. (2018), and Ma et al. (2018c), among others, employ the heterogeneous autoregressive model for realized variance (HAR-RV) pioneered by Corsi (2009) as well as its various extensions to predict the volatility of oil futures market. Along the same lines, this paper relies on the prevailing HAR-RV framework to make a further investigation into the prediction of oil price volatility.

It is well known that forecast combination has impressive predictive ability in financial forecasting (see, e.g., Stock and Watson, 2004; Becker and Clements, 2008; Rapach et al., 2010). Meanwhile, shrinkage methods such as least absolute shrinkage and selection operator (lasso) have found to perform well in many applications (see, e.g., Li et al., 2015; Audrino and Knaus, 2016; Li and Tsiakas, 2017; Zhang et al., 2018b). However, the two widely used forecasting strategies have not yet received much attention in the extant literature on oil price volatility forecasting. For this straightforward motivation, this paper attempt to use forecast combination and shrinkage method to obtain reliable oil price volatility forecasts. Further, we make a comparison between the two forecasting strategies with respect to their out-of-sample forecasting performance in both statistical and economic senses.

We follow a related study by Wang et al. (2016) and use eight popular HAR-RV-type models. On one hand, based on the individual forecasts generated by the HAR-RV-type models, we employ five prevailing combination approaches suggested by Rapach et al. (2010) to produce combination forecasts. On the other hand, based on all the predictors from the HAR-RV-type models, we employ two simple but efficient shrinkage methods, the elastic net pioneered by Zou and

* Corresponding author.
E-mail address: weiyusy@126.com (Y. Wei).

https://doi.org/10.1016/j.eneco.2019.01.010
0140-9883/

Hastie (2005) and the lasso pioneered by Tibshirani (1996), to produce accurate forecasts.[1]

Our empirical results provide several notable findings. First, the model confidence set (MCS) test proposed by Hansen et al. (2011) shows that the shrinkage methods of the elastic net and lasso generate significantly more accurate forecasts of oil price volatility than not only the individual HAR-RV-type models but also the combination approaches. Second, the results for the superiority of the elastic net and lasso are robust across a wide range of checks, including various sample sizes used to estimate shrinkage factors, alternative estimation windows (expanding or rolling window), different out-of-sample periods, and other combination approaches. Third, the Direction-of-Change (DoC) test suggested by Degiannakis and Filis (2017) provides evidence that the elastic net and lasso also have substantially higher directional accuracy. Fourth, a mean-variance investor prefers to use the elastic net and lasso instead of the combination approaches to guide her asset allocation because the elastic net and lasso forecasts of oil price volatility can yield sizeable economic gains related to the competing forecasts. In summary, we conclude that the shrinkage methods consistently outperform the combination approaches in the prediction of oil price volatility.

Our paper is related to the work of Audrino and Knaus (2016) as the two studies both rely on the framework of lasso and HAR-RV. In contrast, we can provide at least three new contributions or major differences. First, Audrino and Knaus (2016) make an empirical analysis on several individual stocks, while our paper focuses on the crude oil futures market. Second, Audrino and Knaus (2016) explore whether the lag structure imposed by the simple HAR-RV model can be recovered by the lasso. However, we collect all the predictors from not only the original HAR-RV model but also its various extensions and investigate whether all these predictors can be efficiently used by the lasso as well as the elastic net. Third and more importantly, according to a comprehensive comparison between the shrinkage methods and combination approaches, we answer an important question which one is more reliable for predicting oil price volatility under the HAR-RV framework. In addition, this paper further does some additional work relative to Audrino and Knaus (2016), such as a wide range of robustness checks, the DoC test, and the portfolio exercise. Therefore, our paper contributes new insights into the existing literature on oil price volatility forecasting.

The remainder of the paper is organized as follows. Section 2 provides the econometric specifications. Section 3 describes our data. Section 4 presents the empirical results. Section 5 details a series of robustness checks and extensions. Section 6 provides a portfolio exercise. Finally, Section 7 concludes.

2. Econometric specifications

2.1. HAR-RV-type models

With the availability of high-frequency data, we can use the intraday oil futures prices to measure the daily realized volatility or variance (RV) proposed by Andersen and Bollerslev (1998).[2] Specifically, RV can be defined as the summation of the intraday squared returns,

$$RV_t = \sum_{i=1}^{M} r_{t,i}^2, \tag{1}$$

where $r_{t,\,i}$ represents the i-th intraday oil futures return on day t, $M = 1/\Delta$, and Δ is the sampling frequency.

[1] The two shrinkage methods of elastic net and lasso are widely used in financial forecasting. See, for example, Elliott et al. (2013), Li et al. (2015), Li and Tsiakas (2017), and Zhang et al. (2018b).

[2] Following Andersen et al. (2007), we use the terms of realized volatility and realized variation interchangeably in this paper.

Following the related study of Wang et al. (2016), we consider eight widely used HAR-RV-type models in this paper.[3] The first is the HAR-RV model originally proposed by Corsi (2009). The HAR-RV model has become one of the most popular models to predict the dynamics of RV. This model describes some of the stylized facts found in the asset return volatility such as long memory and multi-scaling behavior. Furthermore, the HAR-RV model is tractable as it only contains three predictors including daily, weekly, and monthly RVs. The HAR-RV model specification can be expressed as

$$RV_t = \beta_0 + \beta_d RV_{t-1} + \beta_w RV_{t-5:t-1} + \beta_m RV_{t-22:t-1} + \varepsilon_t, \tag{2}$$

where $RV_{t-h:t-1} = (1/h)(RV_{t-h} + \cdots + RV_{t-1})$; particularly, $RV_{t-5:t-1}$ and $RV_{t-22:t-1}$ represent weekly and monthly RVs, respectively.

Given the importance of jumps, Andersen et al. (2007) extend the HAR-RV model by incorporating a jump component. The extended model is termed HAR-RV-J, which is given by

$$RV_t = \beta_0 + \beta_d RV_{t-1} + \beta_w RV_{t-5:t-1} + \beta_m RV_{t-22:t-1} + \beta_J J_{t-1} + \varepsilon_t, \tag{3}$$

where the jump component $J_{t-1} = \max\{RV_{t-1} - BPV_{t-1}, 0\}$, $BPV_t = u_1^{-2}\sum_{i=2}^{M}|r_{t,i}||r_{t,i-1}|$ is the realized bi-power variation (BPV), and $u_1 = \sqrt{2/\pi}$. Moreover, Andersen et al. (2007) additionally introduce the realized measure of significant jump (SJ) and propose another new HAR-RV-type model termed HAR-RV-CJ, which is expressed as

$$\begin{aligned} RV_t = {} & \beta_0 + \beta_{cd} C_{t-1} + \beta_{cw} C_{t-5:t-1} + \beta_{cm} C_{t-22:t-1} + \beta_{sd} SJ_{t-1} \\ & + \beta_{sw} SJ_{t-5:t-1} + \beta_{sm} SJ_{t-22:t-1} + \varepsilon_t, \end{aligned} \tag{4}$$

where $SJ_{t-1} = I(Z_{t-1} > \Phi_\alpha)(RV_{t-1} - BPV_{t-1})$ and Z_{t-1} is the ratio-statistic of Huang and Tauchen (2005); to ensure that the continuous sample path component and jump component sum to the total realized variance, the continuous sample path component is set to $C_{t-1} = I(Z_{t-1} \le \Phi_\alpha)RV_{t-1} + I(Z_{t-1} > \Phi_\alpha)BPV_{t-1}$.

The fourth HAR-RV-type model used in this study is the HAR-RV-TCJ model originated by Corsi et al. (2010). In this model, Corsi et al. (2010) propose a novel test statistic $C\text{-}Tz$ for jump detection and further employ the threshold bi-power variation (TBPV) to measure the threshold jump as $TJ_{t-1} = I(C\text{-}Tz_{t-1} > \Phi_\alpha)(RV_{t-1} - TBPV_{t-1})$. The corresponding continuous part is defined as $TC_{t-1} = RV_{t-1} - TJ_{t-1}$. Consequently, the HAR-RV-TCJ model is given by

$$RV_t = \beta_0 + \beta_{tcd} TC_{t-1} + \beta_{tcw} TC_{t-5:t-1} + \beta_{tcm} TC_{t-22:t-1} + \beta_{TJ} TJ_{t-1} + \varepsilon_t. \tag{5}$$

In contrast to the three above-mentioned models that pay attention to the role of jumps, the following HAR-RV-type models focus on the asymmetric effect (i.e., leverage effect). The last four models are proposed by Patton and Sheppard (2015). Patton and Sheppard (2015) introduce some signed realized measures to describe the leverage effect. The first model termed HAR-RV-RS-I decomposes the daily RV into two realized semi-variances (RS) and extends the HAR-RV model as

$$RV_t = \beta_0 + \beta_d^+ RS_{t-1}^+ + \beta_d^- RS_{t-1}^- + \beta_w RV_{t-5:t-1} + \beta_m RV_{t-22:t-1} + \varepsilon_t, \tag{6}$$

where $RS_{t-1}^+ = \sum_{i=1}^{M} r_{t-1,i}^2 I(r_{t-1,i} \ge 0)$ and $RS_{t-1}^- = \sum_{i=1}^{M} r_{t-1,i}^2 I(r_{t-1,i} < 0)$. The second model (HAR-RV-RS-II) augments the first model with a term that interacts the lagged daily RV with an indicator for negative

[3] Wang et al. (2016) is an influential paper that also uses the eight popular HAR-RV-type models to forecast the realized volatility of the S&P 500 index. Furthermore, the eight used HAR-RV-type models are proposed by four seminal papers, namely, Corsi (2009), Andersen et al. (2007), Corsi et al. (2010), and Patton and Sheppard (2015). As of 2018, the four papers have got 2793 (nearly 700 on average) Google citations. In particular, there are over 200 Google citations for the most recent paper of Patton and Sheppard (2015). Therefore, the eight HAR-RV-type models are literally influential and widely used.

lagged daily returns, $RV_{t-1}I(r_{t-1}<0)$. Thus, the HAR-RV-RS-II model specification can be expressed as

$$RV_t = \beta_0 + \beta_d^+ RS_{t-1}^+ + \beta_d^- RS_{t-1}^- + \gamma RV_{t-1}I(r_{t-1}<0) + \beta_w RV_{t-5:t-1} + \beta_m RV_{t-22:t-1} + \varepsilon_t. \tag{7}$$

Patton and Sheppard (2015) further use signed jump variation, $\Delta J_{t-1} = RS_{t-1}^+ - RS_{t-1}^-$, and realized bi-power variation (BPV) to develop a new model, which is termed HAR-RV-SJ-I,

$$RV_t = \beta_0 + \beta_{\Delta J}\Delta J_{t-1} + \beta_{bpv} BPV_{t-1} + \beta_w RV_{t-5:t-1} + \beta_m RV_{t-22:t-1} + \varepsilon_t. \tag{8}$$

The last model for capturing the leverage effect is termed HAR-RV-SJ-II,

$$RV_t = \beta_0 + \beta_{\Delta J}^+ \Delta J_{t-1}^+ + \beta_{\Delta J}^- \Delta J_{t-1}^- + \beta_{bpv} BPV_{t-1} + \beta_w RV_{t-5:t-1} + \beta_m RV_{t-22:t-1} + \varepsilon_t, \tag{9}$$

where $\Delta J_{t-1}^+ = \Delta J_{t-1} I(\Delta J_{t-1} \geq 0)$ and $\Delta J_{t-1}^- = \Delta J_{t-1} I(\Delta J_{t-1} < 0)$.

2.2. Forecast combination

In the above subsection, we introduce eight prevailing HAR-RV-type models. Hence, we can obtain eight individual volatility forecasts. However, it has been well known that the out-of-sample forecasting performance of an individual model is unstable because of model uncertainty (see, e.g., Avramov, 2002; Stock and Watson, 2004; Becker and Clements, 2008; Rapach et al., 2010). It is thus too risky to rely on the forecasts produced by an individual model. With this in mind, we use five popular combination approaches recommended by Rapach et al. (2010) to address this issue. These combination approaches are widely used by a host of related literature (see, e.g., Paye, 2012; Zhu and Zhu, 2013; Wang et al., 2016; Zhang et al., 2019). Statistically, the combination volatility forecasts can be calculated as

$$\widehat{RV}_{c,t+1} = \sum_{k=1}^{N} \omega_{k,t} \widehat{RV}_{k,t+1}, \tag{10}$$

where $\widehat{RV}_{c,t+1}$ denotes the combination forecast for oil price volatility on day $t+1$, $\widehat{RV}_{k,t+1}$ denotes the individual forecast generated by the k-th HAR-RV-type model, $\omega_{k,t}$ denotes the combining weight for the k-th individual forecast formed at t, and N is the number of all the used individual models, namely, eight. The detailed description of the five combination weighting schemes is provided as follows.

- Mean combination. The mean combination forecast is the equal-weighted average of the N individual forecasts.
- Median combination. This combination approach uses the median of the N individual forecasts.
- Trimmed mean combination. The trimmed mean combination forecast sets $\omega_{k,t} = 0$ for the largest and smallest individual forecasts and $\omega_{k,t} = 1/(N-2)$ for the remaining individual forecasts.
- Discount mean square prediction error (DMSPE) combining method. In the DMSPE method, the combining weight of the k-th individual forecast on day t is calculated as $\omega_{k,t} = \phi_{k,t}^{-1} / \sum_{\ell=1}^{N} \phi_{\ell,t}^{-1}$, where $\phi_{\ell,t} = \sum_{s=m+1}^{t} \theta^{t-s} \left(RV_s - \widehat{RV}_{\ell,s}\right)^2$, RV_s is the actual RV on day s, m is the number of observations in the initial training sample, and θ denotes a discount factor. Following Rapach et al. (2010), Zhu and Zhu (2013), Zhang et al. (2019) and Zhang et al. (2018a), among others, we consider two values of θ, namely, 1 and 0.9. Hence, two DMSPE methods, DMSPE(1) and DMSPE(0.9), are employed in this study.

2.3. Shrinkage methods

In this paper, we follow some related works (e.g., Li et al., 2015; Li and Tsiakas, 2017; Zhang et al., 2018b) and employ two most popular shrinkage methods, namely, the elastic net pioneered by Zou and Hastie (2005) and the lasso pioneered by Tibshirani (1996). Statistically, the lasso forecasts of oil price volatility are computed as

$$\widehat{RV}_{t+1} = \hat{\beta}_0 + \sum_{i=1}^{K} \hat{\beta}_i x_{i,t}, \tag{11}$$

where

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left(\frac{1}{2(t-1)} \sum_{l=1}^{t-1} \left(RV_{l+1} - \beta_0 - \sum_{i=1}^{K} \beta_i x_{i,l} \right)^2 + \lambda \sum_{i=1}^{K} |\beta_i| \right), \tag{12}$$

RV_{t+1} is the RV of oil prices on day $t+1$, $x_{i,t}$ denotes the i-th predictor available on day t, K is the number of all the used predictors, $\hat{\beta}$ is the shrinkage estimator of regression coefficients in the lasso, which are estimated by the data available up to day t, and λ is the nonnegative regularization parameter serving as the penalty function of β. It is noteworthy that we use the same predictors to impartially compare the predictive ability between combination approaches and shrinkage methods. That is, the shrinkage methods include all the predictors from the eight individual HAR-RV-type models.

Equivalently, Eq. (12) can be also written as ordinary least squares (OLS) estimators

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \frac{1}{2} \sum_{l=1}^{t-1} \left(RV_{l+1} - \beta_0 - \sum_{i=1}^{N} \beta_i x_{i,l} \right)^2 \tag{13}$$

with a L_1 penalty function of

$$\sum_{i=1}^{K} |\beta_i| < \psi, \tag{14}$$

where the parameter ψ controls for the amount of shrinkage, playing the same role as λ. When the constraint is not binding for a sufficiently large (small) value of ψ (λ), the lasso estimator reduces to the OLS estimator. The constraint is not linear due to the absolute value operator $|\cdot|$, and therefore, a closed form solution of the coefficient estimator is not available.

In contrast, the elastic net forecasts are also computed by Eq. (11), while the corresponding regression coefficients are estimated as

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left(\frac{1}{2(t-1)} \sum_{l=1}^{t-1} \left(RV_{l+1} - \beta_0 - \sum_{i=1}^{N} \beta_i x_{i,l} \right)^2 + \lambda \sum_{i=1}^{N} \left((1-\alpha)\beta_i^2 + \alpha|\beta_i| \right) \right), \tag{15}$$

where α is a positive constant strictly between 0 and 1. Particularly, the elastic net reduces to the lasso when $\alpha = 1$, and the elastic net approaches the ridge regression when α shrinks towards 0. Hence, it is evident that the elastic net introduces both L_1 and L_2 penalties.

When employing the lasso and elastic net to forecast oil price RV on day $t+1$, we should ex ante obtain the optimal estimates of α and λ using the data up to t. In this study, we use an efficient estimation algorithm proposed by Zhang et al. (2018b) to address this issue.[4] Specifically, when we use the elastic net and lasso to generate the RV

[4] Zhang et al. (2018b) document that, in contrast to other popular estimation algorithms such as cross validation, their algorithm can not only yield comparable or better forecasting performance but also reduce computational burden. See their online Internet Appendix for more related details.

forecast on day $t+1$, the estimation algorithm involves the following steps:

1. Given an initial set of values of α and λ,[5] say $(\alpha^{(1)},\lambda^{(1)})$, we further divide the estimation period (i.e., the period from day 1 to t) into an "in-sample" estimation period and an "out-of-sample" evaluation period.[6]
2. We estimate the regression coefficients $\hat{\beta}$ of the lasso and elastic net by solving the optimization problem of Eqs. (12) and (15), respectively, using the "in-sample" data from day 1 to j ($j<t$) with $\alpha=\alpha^{(1)}$ and $\lambda=\lambda^{(1)}$. The coefficient estimation used in this step is based on the alternating direction method of multipliers (ADMM) algorithm (Boyd et al., 2011).
3. We use the estimated coefficients $\hat{\beta}$ to produce "out-of-sample" forecasts from day $j+1$ to t and then compute the mean squared forecast error (MSFE) based on these forecasts, which is conditional on $(\alpha^{(1)},\lambda^{(1)})$.
4. We repeat steps 1 to 3 with different values of α and λ, say $(\alpha^{(2)},\lambda^{(2)})$, $(\alpha^{(3)},\lambda^{(3)})$,…, and select the optimal combination $(\hat{\alpha},\hat{\lambda})$ that provides the lowest value of MSFE.
5. Given $(\hat{\alpha},\hat{\lambda})$, we implement the ADMM algorithm to estimate the regression coefficients $\hat{\beta}$ using all the data over the entire estimation period from day 1 to t, instead of only its "in-sample" part from day 1 to j. By now the RV forecast on day $t+1$ is obtained based on Eq. (11).

2.4. Forecast evaluation

To quantitatively compare the out-of-sample performance among the forecasting models used by this paper, we employ three popular loss functions of QLIKE, MSE, and MAE. In particular, Patton (2011) demonstrates that, in terms of the ranking of competing volatility forecasts, QLIKE and MSE are robust to the presence of noise in the volatility proxy. Statistically, QLIKE, MSE, and MAE can be expressed as

$$QLIKE=\frac{1}{q}\sum_{t=m+1}^{m+q}\left(\log\left(\widehat{RV}_t\right)+\frac{RV_t}{\widehat{RV}_t}\right), \tag{16}$$

$$MSE=\frac{1}{q}\sum_{t=m+1}^{m+q}\left(RV_t-\widehat{RV}_t\right)^2, \tag{17}$$

and

$$MAE=\frac{1}{q}\sum_{t=m+1}^{m+q}\left|RV_t-\widehat{RV}_t\right|, \tag{18}$$

respectively, where RV_t is the actual RV on day t, $\widehat{RV}_t$ is the RV forecast based on one of the forecasting models, and m and q are the length of in-sample estimation period and out-of-sample evaluation period, respectively.

Following the convention in the literature on volatility prediction (see, e.g., Patton and Sheppard, 2009; Liu et al., 2015; Wang et al., 2016; Liu et al., 2018; Ma et al., 2018a; Ma et al., 2018c), we use the methodology of model confidence set (MCS) proposed by Hansen et al. (2011) to ascertain whether the used models have a statistically significant difference in out-of-sample performance. A MCS is a subset of models that contains the best model with a given level of confidence. The interpretation of a MCS p-value is analogous to that of a classical p-value (Hansen et al., 2011). It is evident that a model with a larger MCS p-value shows stronger predictive ability. Following Hansen et al. (2011), Wang et al. (2016), and Liu et al. (2018), among others, we consider the significance (confidence) level of 10% (90%). That is, a model with a MCS p-value that is larger than 0.1 will be included in the MCS.

3. Data

In this study, following Haugom et al. (2014), Sévi (2014), and Liu et al. (2018), among others, we use a prevailing benchmark in crude oil pricing, i.e., West Texas Intermediate (WTI), which is also known as Texas light sweet. To be precise, we use the high-frequency data for the front-month contract of WTI oil futures traded on the New York Mercantile Exchange. The trading of the front-month contract for WTI futures ends on the third business day prior to the 25th of the month prior to delivery. Our data of high-frequency crude oil futures prices are available from the Thomson Reuters Tick History Database. The whole sample period spans from January 2, 2007, to July 15, 2016.[7] After removing days with a shortened trading session or too few transactions, we obtain 2357 observations. To generate out-of-sample forecasts, we divide the whole sample period into an in-sample estimation period consisting of the first 1457 observations and an out-of-sample evaluation period consisting of the remaining 900 observations.

Liu et al. (2015) find little evidence that the 5-min RV is outperformed by any other measures from 400 volatility estimators for 31 different financial assets spanning five asset classes. In addition, the 5-min RV is also widely used and recommended in a large number of studies on forecasting oil market volatility (see, e.g., Haugom et al., 2014; Sévi, 2014; Liu et al., 2018; Ma et al., 2018c). Along the same lines, we choose the 5-min interval as our sampling frequency to measure the RV of crude oil prices.

The descriptive statistics of all the variables are reported in Table 1. All of the variables are right skewed and leptokurtic except for ΔJ^-, which is left skewed. The Jarque-Bera statistics suggest that RV and all the predictors have significantly non-normal distributions. With respect to serial correlation, the Ljung-Box statistics for most of these variables always support the rejection of the null hypothesis of no autocorrelation up to the fifth order. The long memory property of RV motivates us to use the HAR-RV-type models.

4. Empirical results

4.1. Out-of-sample forecasting performance

In this study, we focus on the out-of-sample test and ignore the in-sample analysis. We have two major reasons. First, the in-sample estimation is only suitable for the individual HAR-RV-type models instead of the combination approaches and shrinkage methods, while the paper's main purpose is to compare the predictive ability between the combination approaches and shrinkage methods. Second, compared to the in-sample predictability, the out-of-sample predictability is a more stringent test and is thus of greater interest to financial practitioners and researchers. We use the first 1457 observations as the initial training sample and generate the out-of-sample volatility forecasts based on a recursive (expanding) estimation window.[8]

Panel A of Table 2 reports the out-of-sample performance for all the forecasting models, including the mean value of loss functions and the MCS p-values based on both the range and semi-quadratic statistics.[9]

[5] With respect to the lasso, α is always equal to 1.

[6] The "out-of-sample" evaluation period accounts for approximately 40% of the entire estimation period. In the robustness check below, we consider other reasonable out-of-sample sizes. In addition, it is important to note that the so-called "in-sample" period and "out-of-sample" period are based on the estimation period instead of the whole sample period. In other words, we only use the data from day 1 to t to validate the shrinkage factors of α and λ when generating forecast on day $t+1$. This guarantees that our elastic net and lasso forecasts do not have a look-ahead bias.

[7] The actual sample period used by this paper begins from February 1, 2007, due to the construction of monthly measures.

[8] In the robustness check below, we also rely on a rolling window to generate volatility forecasts. The results are qualitative similar for alternative estimation windows.

[9] All of the MCS p-values reported in this paper are based on the stationary bootstrap. We obtain qualitatively similar results of MCS p-values when using the block bootstrap. To save space, we do not report these results, but they are available upon request.

Table 1
Descriptive statistics.

Variables	Mean	Std. Dev.	Skewness	Kurtosis	Jarque-Bera	Q(5)
RV	5.794	10.432	17.406	526.209	27,196,867.020***	2437.632***
Weekly RV	5.794	7.848	5.044	40.407	169,630.363***	9049.718***
Monthly RV	5.795	7.073	3.419	13.795	23,189.938***	11,603.870***
J	0.779	7.148	37.583	1581.946	245,283,884.689***	8.065
C	5.314	6.665	3.267	12.934	20,539.845***	7682.030***
Weekly C	5.314	6.178	2.867	9.205	11,506.921***	10,688.825***
Monthly C	5.316	5.887	2.809	8.582	10,293.962***	11,666.754***
SJ	0.480	7.141	37.822	1594.803	249,285,285.165***	2.298
Weekly SJ	0.480	3.201	16.645	309.494	9,475,756.796***	3027.723***
Monthly SJ	0.479	1.640	7.924	69.234	493,349.272***	10,064.527***
TC	4.930	6.210	3.245	12.840	20,247.695***	7420.023***
Weekly TC	4.930	5.714	2.745	8.092	9356.132***	10,713.934***
Monthly TC	4.931	5.454	2.701	7.686	8635.221***	11,667.769***
TJ	0.864	7.510	35.054	1430.421	200,574,834.984***	11.817**
RS^+	3.030	8.342	28.819	1097.510	118,119,018.921***	534.325***
RS^-	2.764	3.510	3.341	14.150	23,951.984***	7337.552***
Daily return	−0.010	2.401	−0.038	3.082	928.159***	21.113***
ΔJ	0.265	7.416	35.883	1496.663	219,559,383.564***	5.205
BPV	5.050	6.325	3.224	12.488	19,322.352***	7638.925***
ΔJ^+	0.731	7.279	37.786	1602.871	251,807,023.336***	3.607*
ΔJ^-	−0.466	1.153	−6.283	61.406	384,203.925***	451.240***

This table reports the descriptive statistics of all the used variables. See Section 2.1 for more details about their definitions. The Jarque-Bera statistic is used to test the null hypothesis of normal distribution. Q(5) is the Ljung-Box statistic for up to the fifth order serial correlation. All the measures are based on the intraday oil returns in percentage. The entire sample period is from February 1, 2007, to July 15, 2016.
*** Indicates significance at the 1% level.
** Indicates significance at the 5% level.
* Indicates significance at the 10% level.

An impressive finding is that only the elastic net consistently appears in the MCS with the confidence level of 90% for all the cases. Furthermore, the elastic net always delivers the largest MCS *p*-value of 1. This evidence suggests that the shrinkage method of the elastic net has significantly better out-of-sample forecasting ability than the other forecasting models. In addition, the remaining forecasting models only fall into the MCS based on the MSE loss function, while all of them cannot enter the MCS with the 90% confidence level for the loss functions of QLIKE and MAE. In contrast, the lasso is a good choice except for the elastic net because the lasso usually yields the second largest MCS *p*-value. With respect to the mean value of loss functions, we observe a similar pattern that the elastic net consistently produces the lowest forecast error and the lasso produces the second lowest value for MSE and MAE.

Numerous studies conclude that combination forecasts typically outperform individual forecasts (see, e.g., Stock and Watson, 2004; Becker and Clements, 2008; Rapach et al., 2010; Ma et al., 2018a). However, we cannot draw the conclusion from Panel A of Table 2. Given this, we exclude the two shrinkage methods and just compare the out-of-sample performance between the individual HAR-RV-type models and combination approaches. The corresponding results of the MCS test are provided in Panel B of Table 2. Overall, the combination approaches are more likely to survive in the MCS and generally yield larger MCS *p*-values than the individual HAR-RV-type models. Moreover, the largest MCS *p*-value of 1 always belongs to one of the combination approaches. In particular, none of the individual HAR-RV-type models survive in the MCS with the confidence level of 90% for the QLIKE and MAE loss functions based on the range statistic. Furthermore, we can see that the mean values of the three loss functions of the combination approaches are typically smaller than those of the individual HAR-RV-type models. In summary, these results indicate that the combination forecasts are more accurate than the individual volatility forecasts based on the HAR-RV-type models, which is consistent with the existing literature.

Finally, to directly compare the out-of-sample forecasting performance between forecast combination and shrinkage method, we exclude the individual HAR-RV-type models and just compare the combination approaches with the elastic net and lasso. The corresponding results of the MCS test are provided in Panel C of Table 2. We observe a similar result as reported in Panel A. Only the elastic net consistently enters the MCS with the confidence level of 90% for all the cases and always delivers the largest MCS *p*-value of 1. Another shrinkage method of lasso also generates larger or equal MCS *p*-values relative to those of the combination approaches.[10] This evidence further confirms the superiority of the shrinkage methods in forecasting oil price volatility.

4.2. Potential explanations and further discussions

Why can the shrinkage methods beat the combination approaches? One possible explanation for this is that, as continuous shrinkage methods, the elastic net and lasso usually enhance the prediction accuracy due to the bias-variance tradeoff (see, e.g., Tibshirani, 1996; Zou and Hastie, 2005; Zou, 2006; Li and Tsiakas, 2017). In particular, the ordinary least squares (OLS) estimates are unbiased but have a large variance, whereas the regression slope estimates of the elastic net and lasso sacrifice some bias but reduce the forecast error variance. It should be noted that forecast combination is also very efficient to reduce the forecast variance (see Rapach et al., 2010). We find empirical evidence that the forecast error variance of the elastic net and lasso is smallest, while the combination approaches only generate smaller variance than most of the individual HAR-RV-type models (unreported). This is probably because the individual forecasts do not move towards a common target. Some of the HAR-RV-type models capture the role of jumps, while some others pay attention to the leverage effect. Since the individual HAR-RV-type models have different forecast targets, the combination forecasts are not convergent.

Since the two shrinkage methods of the elastic net and lasso consistently surpass the combination approaches, we may wonder whether a further forecast combination that utilizes the shrinkage forecasts can beat the elastic net and lasso.[11] To answer this interesting question,

[10] In the robustness checks below, the lasso is more likely to enter the MCS and thus exhibits stronger predictive power.
[11] We thank an anonymous referee for suggesting this interesting idea.

Table 2
Out-of-sample performance based on the MCS test.

Forecasting models	QLIKE			MSE			MAE		
	Loss function	Range	Semi-quadratic	Loss function	Range	Semi-quadratic	Loss function	Range	Semi-quadratic
Panel A: all the forecasting models									
HAR-RV	2.096	0.002	0.031	13.625	**0.196**	0.085	1.815	0.001	0.004
HAR-RV-J	2.109	0.002	0.031	9.816	**0.680**	**0.720**	1.504	0.030	0.014
HAR-RV-CJ	9.498	0.002	0.031	9.704	**0.975**	**0.972**	1.605	0.001	0.005
HAR-RV-TCJ	6.989	0.002	0.031	10.078	**0.666**	**0.591**	1.589	0.014	0.010
HAR-RV-RS-I	2.093	0.002	0.031	10.864	**0.196**	**0.281**	1.548	0.030	0.013
HAR-RV-RS-II	2.092	0.002	0.031	11.083	**0.196**	**0.151**	1.563	0.026	0.013
HAR-RV-SJ-II	2.112	0.002	0.031	9.866	**0.680**	**0.672**	1.508	0.030	0.014
HAR-RV-SJ-I	2.120	0.002	0.031	9.829	**0.806**	**0.802**	1.498	0.030	0.014
Mean	2.097	0.002	0.031	9.869	**0.666**	**0.594**	1.495	0.030	0.014
Median	2.109	0.002	0.031	9.703	**0.934**	**0.947**	1.480	0.031	0.015
Trimmed mean	2.109	0.002	0.031	9.803	**0.680**	**0.720**	1.491	0.030	0.014
DMSPE(1)	2.093	0.006	0.031	9.836	**0.680**	**0.672**	1.490	0.030	0.014
DMSPE(0.9)	2.087	0.005	0.031	9.816	**0.680**	**0.716**	1.485	0.031	0.015
Elastic net	2.070	**1.000**	**1.000**	9.608	**1.000**	**1.000**	1.431	**1.000**	**1.000**
Lasso	2.099	0.002	0.031	9.609	**0.975**	**0.974**	1.445	0.031	0.015
Panel B: eight individual HAR-RV-type models and five combination approaches									
HAR-RV	2.096	0.078	**0.181**	13.625	**0.207**	0.094	1.815	0.001	0.011
HAR-RV-J	2.109	0.051	0.063	9.816	**0.584**	**0.773**	1.504	0.076	**0.168**
HAR-RV-CJ	9.498	0.029	0.043	9.704	**0.999**	**0.999**	1.605	0.001	0.018
HAR-RV-TCJ	6.989	0.029	0.043	10.078	**0.584**	**0.642**	1.589	0.055	0.046
HAR-RV-RS-I	2.093	0.078	**0.181**	10.864	**0.240**	**0.302**	1.548	0.076	**0.109**
HAR-RV-RS-II	2.092	0.078	**0.181**	11.083	**0.237**	**0.163**	1.563	0.076	0.073
HAR-RV-SJ-II	2.112	0.051	0.044	9.866	**0.584**	**0.642**	1.508	0.076	**0.157**
HAR-RV-SJ-I	2.120	0.029	0.043	9.829	**0.584**	**0.773**	1.498	0.076	**0.179**
Mean	2.097	0.078	**0.153**	9.869	**0.584**	**0.642**	1.495	0.076	**0.168**
Median	2.109	0.051	0.055	9.703	**1.000**	**1.000**	1.480	**1.000**	**1.000**
Trimmed mean	2.109	0.051	0.079	9.803	**0.584**	**0.773**	1.491	0.076	**0.179**
DMSPE(1)	2.093	0.078	**0.181**	9.836	**0.584**	**0.701**	1.490	0.076	**0.179**
DMSPE(0.9)	2.087	**1.000**	**1.000**	9.816	**0.584**	**0.773**	1.485	**0.695**	**0.695**
Panel C: five combination approaches and two shrinkage methods									
Mean	2.097	0.014	0.015	9.869	**0.464**	**0.252**	1.495	0.028	0.011
Median	2.109	0.014	0.015	9.703	**0.832**	**0.820**	1.480	0.032	0.016
Trimmed mean	2.109	0.014	0.015	9.803	**0.467**	**0.397**	1.491	0.018	0.011
DMSPE(1)	2.093	0.014	0.015	9.836	**0.467**	**0.328**	1.490	0.030	0.011
DMSPE(0.9)	2.087	0.014	0.015	9.816	**0.513**	**0.468**	1.485	0.032	0.016
Elastic net	2.070	**1.000**	**1.000**	9.608	**1.000**	**1.000**	1.431	**1.000**	**1.000**
Lasso	2.099	0.014	0.015	9.609	**0.976**	**0.976**	1.445	0.032	0.016

This table reports the out-of-sample performance including the mean value of loss functions and the MCS *p*-values that are calculated based on the range and semi-quadratic statistics. Three loss functions we consider are QLIKE, MSE, and MAE. In Panel A, we compare the out-of-sample predictive performance among all the forecasting models used by this paper, while Panel B includes eight individual HAR-RV-type models and five combination approaches and Panel C includes five combination approaches and two shrinkage methods. Bold figures highlight instances in which the MCS *p*-value is larger than 0.1. The entire sample period containing 2357 observations spans from February 1, 2007, to July 15, 2016, while the length of out-of-sample period is 900.

we provide two kinds of further combination approaches that both incorporate the forecasts of the elastic net and lasso. First, we combine the individual forecasts generated by not only the eight HAR-RV-type models but also the elastic net and lasso. Second, we just combine the forecasts of the five combination approaches and the two shrinkage methods. We then compare all the forecasts of the traditional combination approaches, the two kinds of new combination approaches, the elastic net and lasso. We find that the new combination approaches outperform the traditional counterparts, while the elastic net and lasso still show significantly better out-of-sample forecasting performance than the new combination approaches.[12] This evidence further documents the superiority of the two shrinkage methods.

5. Robustness checks and extensions

In this section, we provide a wide range of robustness checks and extensions. For the sake of brevity, we exclude the eight individual HAR-RV-type models and directly compare the out-of-sample performance between the shrinkage methods and combination approaches when using the MCS test. The results are robust when we further include the individual HAR-RV-type models, as evidenced in Table 2.

5.1. The estimation of shrinkage factors

The shrinkage factors, α and λ, are two critically important parameters for generating the elastic net and lasso forecasts. In this study, we employ the estimation algorithm suggested by Zhang et al. (2018b) to determine α and λ, in which the last 40% of the observations in the estimation period are used to determine the optimal values of α and λ. Our partition of the estimation period can result in a desirable trade-off between the in-sample period that has enough observations to precisely estimate initial parameters and the out-of-sample period that has a relatively long length for the evaluation of α and λ. However, to further mitigate the concern of data snooping, we additionally consider three reasonable sample sizes to assess the performance for determining the values of α and λ. The three window sizes are 20%, 30%, and 50% of the estimation sample length.

Table 3 presents the out-of-sample performance including the mean value of loss functions and the MCS *p*-values when we separately use 20%, 30%, and 50% estimation sample to estimate α and λ. Overall, the results are qualitatively similar for different estimation sample sizes. In contrast to the 40% sample size, a slight difference for other window

[12] For the sake of brevity, we do not report the results. The results for the new combination approaches are available upon request.

Table 3
Out-of-sample performance for different window sizes used to estimate shrinkage factors.

Forecasting models	QLIKE			MSE			MAE		
	Loss function	Range	Semi-quadratic	Loss function	Range	Semi-quadratic	Loss function	Range	Semi-quadratic
Panel A: the window size used to estimate shrinkage factors is 20%									
Mean	2.097	0.013	0.010	9.869	**0.443**	**0.393**	1.495	0.043	0.087
Median	2.109	0.011	0.008	9.703	**0.847**	**0.847**	1.480	**0.360**	**0.209**
Trimmed mean	2.109	0.011	0.008	9.803	**0.450**	**0.529**	1.491	0.043	0.087
DMSPE(1)	2.093	0.011	0.008	9.836	**0.450**	**0.478**	1.490	0.043	**0.102**
DMSPE(0.9)	2.087	0.011	0.008	9.816	**0.450**	**0.529**	1.485	**0.360**	**0.209**
Elastic net	2.068	**1.000**	**1.000**	9.655	**1.000**	**1.000**	1.443	**1.000**	**1.000**
Lasso	2.069	**0.726**	**0.726**	9.900	**0.450**	**0.529**	1.453	**0.360**	**0.209**
Panel B: the window size used to estimate shrinkage factors is 30%									
Mean	2.097	0.039	0.028	9.869	**0.445**	**0.283**	1.495	0.043	0.060
Median	2.109	0.039	0.017	9.703	**0.466**	**0.505**	1.480	**0.199**	**0.118**
Trimmed mean	2.109	0.039	0.019	9.803	**0.466**	**0.395**	1.491	0.043	0.060
DMSPE(1)	2.093	0.039	0.026	9.836	**0.466**	**0.350**	1.490	0.043	0.065
DMSPE(0.9)	2.087	0.039	0.041	9.816	**0.466**	**0.395**	1.485	**0.199**	**0.118**
Elastic net	2.074	**1.000**	**1.000**	9.575	**1.000**	**1.000**	1.436	**1.000**	**1.000**
Lasso	2.079	0.087	0.087	9.664	**0.466**	**0.505**	1.446	**0.199**	**0.152**
Panel C: the window size used to estimate shrinkage factors is 50%									
Mean	2.097	0.003	0.006	9.869	**0.140**	0.040	1.495	0.011	0.009
Median	2.109	0.003	0.006	9.703	**0.140**	0.081	1.480	0.011	0.009
Trimmed mean	2.109	0.009	0.006	9.803	**0.140**	0.045	1.491	0.010	0.009
DMSPE(1)	2.093	0.003	0.006	9.836	**0.140**	0.040	1.490	0.011	0.009
DMSPE(0.9)	2.087	0.003	0.006	9.816	**0.140**	0.042	1.485	0.019	0.020
Elastic net	2.066	**1.000**	**1.000**	9.397	**0.334**	**0.334**	1.424	**0.692**	**0.692**
Lasso	2.067	**0.604**	**0.604**	9.348	**1.000**	**1.000**	1.423	**1.000**	**1.000**

This table reports the out-of-sample performance including the mean value of loss functions and the MCS p-values that are calculated based on the range and semi-quadratic statistics. In this table, we consider different window sizes used to estimate shrinkage factors, including 20% (Panel A), 30% (Panel B), 50% (Panel C). Three used loss functions are QLIKE, MSE, and MAE. We compare the out-of-sample predictive performance among five combination approaches and two shrinkage methods. Bold figures highlight instances in which the MCS p-value is larger than 0.1. The entire sample period containing 2357 observations spans from February 1, 2007, to July 15, 2016, while the length of out-of-sample period is 900.

sizes is that the lasso model generates relatively large MCS p-values. That is, the lasso model exhibits stronger predictive power. In more cases, the MCS p-values of the lasso are greater than 0.1. Particularly, the lasso yields the largest MCS p-value of 1 for the loss functions of MSE and MAE when the sample size is 50%. In addition, the elastic net and lasso consistently generate the two lowest values for all the three loss functions, suggesting their lowest forecast error. Anyhow, the out-of-sample results of the elastic net and lasso are robust to different window sizes used to estimate α and λ.

5.2. Alternative estimation windows

In the out-of-sample test above, we rely on the recursive estimation window to produce the forecasts of oil price volatility. Alternatively, we employ the rolling window to produce out-of-sample forecasts. The corresponding out-of-sample results are reported in Table 4. The elastic net and lasso models consistently enter the MCS, while the combination approaches cannot always survive in the MCS. That is, our results are robust to alternative estimation windows.

5.3. Different out-of-sample evaluation periods

Rossi and Inoue (2012) and Inoue et al. (2017) emphasize that the arbitrary choices of different window sizes may result in quite different out-of-sample results in practical applications. Therefore, the forecasting window size plays a crucial role in out-of-sample evaluation. For this consideration, we additionally consider another two window sizes, where the initial in-sample estimation windows contain 1757 and 1157 observations, so that the corresponding out-of-sample length is 600 and 1200, respectively.

Table 5 reports the out-of-sample forecasting performance for different out-of-sample evaluation periods. The elastic net and lasso exhibit lower forecast error than the combination approaches as the two shrinkage methods consistently generate the two lowest values of all the three loss functions. For the out-of-sample length of 1200, the elastic net model surpasses the remaining models and the lasso model is the second best choice. With respect to the out-of-sample length of 600, both the elastic net and lasso are consistently in the MCS with the confidence level of 90%, implying their significantly better out-of-sample

Table 4
Out-of-sample performance for rolling window.

Forecasting models	QLIKE			MSE			MAE		
	Loss function	Range	Semi-quadratic	Loss function	Range	Semi-quadratic	Loss function	Range	Semi-quadratic
Mean	2.107	0.021	0.024	9.823	**0.464**	**0.281**	1.413	0.047	**0.100**
Median	2.124	0.021	0.019	9.923	**0.464**	**0.281**	1.424	0.030	0.040
Trimmed mean	2.125	0.021	0.020	9.815	**0.464**	**0.309**	1.422	0.007	0.016
DMSPE(1)	2.100	0.021	0.024	9.803	**0.464**	**0.309**	1.409	0.077	**0.235**
DMSPE(0.9)	2.092	0.021	0.027	9.796	**0.464**	**0.309**	1.405	**0.706**	**0.706**
Elastic net	2.072	**1.000**	**1.000**	9.604	**1.000**	**1.000**	1.397	**1.000**	**1.000**
Lasso	2.085	**0.103**	**0.103**	9.611	**0.464**	**0.309**	1.405	**0.224**	**0.405**

This table reports the out-of-sample performance including the mean value of loss functions and the MCS p-values that are calculated based on the range and semi-quadratic statistics. In this table, we employ an alternative estimation window, i.e., rolling window. Three used loss functions are QLIKE, MSE, and MAE. We compare the out-of-sample predictive performance among five combination approaches and two shrinkage methods. Bold figures highlight instances in which the MCS p-value is larger than 0.1. The entire sample period containing 2357 observations spans from February 1, 2007, to July 15, 2016, while the length of out-of-sample period is 900.

Table 5
Out-of-sample performance for different out-of-sample evaluation periods.

Forecasting models	QLIKE			MSE			MAE		
	Loss function	Range	Semi-quadratic	Loss function	Range	Semi-quadratic	Loss function	Range	Semi-quadratic
Panel A: the length of out-of-sample evaluation period is 600									
Mean	2.458	0.099	0.081	14.629	**0.380**	**0.226**	2.045	0.016	0.004
Median	2.468	0.099	0.070	14.367	**0.639**	**0.662**	2.014	0.039	0.030
Trimmed mean	2.468	0.099	0.068	14.520	**0.412**	**0.363**	2.034	0.016	0.004
DMSPE(1)	2.447	0.099	**0.107**	14.577	**0.412**	**0.300**	2.035	0.021	0.007
DMSPE(0.9)	2.446	0.099	**0.109**	14.553	**0.412**	**0.350**	2.031	0.021	0.010
Elastic net	2.428	**1.000**	**1.000**	14.251	**0.639**	**0.662**	1.950	**1.000**	**1.000**
Lasso	2.442	**0.202**	**0.202**	14.229	**1.000**	**1.000**	1.951	**0.774**	**0.774**
Panel B: the length of out-of-sample evaluation period is 1200									
Mean	2.078	0.016	0.015	7.810	**0.560**	**0.297**	1.329	0.044	0.033
Median	2.088	0.016	0.015	7.692	**0.819**	**0.775**	1.319	0.044	0.035
Trimmed mean	2.087	0.016	0.015	7.763	**0.570**	**0.449**	1.328	0.044	0.033
DMSPE(1)	2.078	0.016	0.015	7.787	**0.570**	**0.367**	1.328	0.044	0.033
DMSPE(0.9)	2.070	0.016	0.015	7.769	**0.627**	**0.512**	1.322	0.044	0.035
Elastic net	2.056	**1.000**	**1.000**	7.619	**1.000**	**1.000**	1.283	**1.000**	**1.000**
Lasso	2.082	0.016	0.015	7.626	**0.819**	**0.775**	1.298	0.044	0.035

This table reports the out-of-sample performance including the mean value of loss functions and the MCS *p*-values that are calculated based on the range and semi-quadratic statistics. In this table, the length of out-of-sample evaluation period includes 600 (Panel A) and 1200 (Panel B). Three used loss functions are QLIKE, MSE, and MAE. We compare the out-of-sample predictive performance among five combination approaches and two shrinkage methods. Bold figures highlight instances in which the MCS *p*-value is larger than 0.1. The entire sample period containing 2357 observations spans from February 1, 2007, to July 15, 2016.

forecasting performance relative to the combination approaches. Our MCS results are thus robust to various length of out-of-sample evaluation period.

5.4. Alternative combination approaches

In this subsection, we wonder whether the results are robust to the use of different combination approaches. For this motivation, we further consider four additional combination approaches. More specifically, the four popular combination approaches are from an influential study of Stock and Watson (2004), which are shown as follows.

- Shrinkage combination (SC hereafter). This combining method computes the weights as a weighted average of the recursive OLS estimator (the Granger and Ramanathan (1984) estimator, imposing an intercept of zero) and equal weighting. The shrinkage combination forecasts are evaluated for $\kappa = 0.25, 0.5, 1$, with larger values corresponding to more shrinkage towards equal weighting.[13] As a result, we obtain three different shrinkage combination approaches, namely, SC(0.25), SC(0.5), and SC(1).
- The most recently best (MRB hereafter) forecasts. As in Stock and Watson (2004), this strategy places all weight on the individual forecast that has the best historical performance (i.e., lowest mean squared forecast error) during the previous four periods.

Table 6 reports the forecasting performance when we compare the out-of-sample performance between the new combination approaches and shrinkage methods. We observe a quite similar result as shown in Panel C of Table 2. The four new combination approaches are also outperformed by the elastic net and lasso. This evidence further supports the superiority of the shrinkage methods in the out-of-sample prediction of oil price volatility.

5.5. Direction-of-change

Following Degiannakis and Filis (2017), we employ the Direction-of-Change (DoC) as an additional out-of-sample evaluation criterion. Degiannakis and Filis (2017) state that the DoC is central to the trading strategies of market timing and asset allocation. Specifically, the DoC measures the proportion of forecasts that correctly predict the direction of the volatility movement. We let p_t be a dummy variable that takes the value of one if a model correctly predicts the direction of volatility movement on trading day t, and zero otherwise. Consequently, this dummy variable is given by

$$p_t = \begin{cases} 1 \ \text{ if } RV_t > RV_{t-1} \ \text{ and } \ \widehat{RV}_t > RV_{t-1} \\ 1 \ \text{ if } RV_t < RV_{t-1} \ \text{ and } \ \widehat{RV}_t < RV_{t-1} \\ 0 \ \text{ otherwise.} \end{cases} \tag{19}$$

Further, we define the DoC rate as the proportion of forecasts that correctly predict the direction of the volatility movement. Statistically, the DoC rate is equal to $1/q\sum_{t=m+1}^{m+q} p_t$. To explore the statistical significance of the directional accuracy, we use a nonparametric test proposed by Pesaran and Timmermann (1992) to test the null hypothesis that the DoC rate of a forecasting model of interest is less than or equal to the DoC rate of random walk against the alternative hypothesis that the DoC rate of a forecasting model of interest is larger than the DoC rate of random walk.

Table 7 reports the DoC results for all the forecasting models. Two observations follow the table immediately. First, we reject the null hypothesis of no directional accuracy at the 1% significance level for all the forecasting models, suggesting the success of the HAR-RV model as well as its extended models in the directional prediction. Second and more importantly, the elastic net and lasso yield substantially larger DoC rates than the competing models including the HAR-RV-type models and combination approaches. Also, the confidence levels for the higher DoC rates of the elastic net and lasso are greater than those of the HAR-RV-type models and combination approaches. In summary, the DoC results are consistent with the MCS results. The elastic net and lasso models exhibit substantially higher directional accuracy.

6. Economic significance

Since volatility is one of the key input variables for asset allocation, we further investigate the economic value of oil price volatility forecasting from an asset allocation perspective. Following the extant literature on forecasting oil price and volatility (see, e.g., Yin and Yang, 2016; Ma et al., 2018b; Zhang et al., 2018a), we assume that a mean-variance investor who optimally allocates between crude oil futures and risk-free

[13] The three values of κ are suggested by Stock and Watson (2004) and also used by Ma et al. (2018a) and Zhang et al. (2018a). See Stock and Watson (2004) for the details about computing shrinkage weights.

Table 6
Out-of-sample performance for alternative combination approaches.

Forecasting models	QLIKE			MSE			MAE		
	Loss function	Range	Semi-quadratic	Loss function	Range	Semi-quadratic	Loss function	Range	Semi-quadratic
SC(0.25)	2.087	0.001	0.014	9.835	**0.510**	**0.271**	1.480	0.057	0.032
SC(0.5)	2.087	0.005	0.014	9.835	**0.510**	**0.296**	1.480	0.057	0.032
SC(1)	2.088	0.001	0.014	9.834	**0.510**	**0.352**	1.479	0.057	0.032
MRB	2.087	0.001	0.014	10.277	**0.194**	**0.153**	1.518	0.013	0.012
Elastic net	2.070	**1.000**	**1.000**	9.608	**1.000**	**1.000**	1.431	**1.000**	**1.000**
Lasso	2.099	0.001	0.014	9.609	**0.974**	**0.974**	1.445	0.057	0.032

This table reports the out-of-sample performance including the mean value of loss functions and the MCS p-values that are calculated based on the range and semi-quadratic statistics. In this table, we compare the out-of-sample predictive performance among four different combination approaches and two shrinkage methods. The alternative combination approaches include three shrinkage combination (SC) approaches and a most recently best (MRB) method. Three used loss functions are QLIKE, MSE, and MAE. Bold figures highlight instances in which the MCS p-value is larger than 0.1. The entire sample period containing 2357 observations spans from February 1, 2007, to July 15, 2016, while the length of out-of-sample period is 900.

bills based on the various volatility forecasts. Zhang et al. (2018a) argue that a futures investor usually pays a small proportion of the entire value of crude oil futures to meet margin requirements; therefore, her gains and losses would be magnified by a leverage ratio, θ, which is inversely related to the margin level. Given this, we follow Zhang et al. (2018a) and compute the portfolio return as

$$R_p = w\theta\left(r + r^f\right) + (1-w)r^f, \tag{20}$$

where w is the portfolio weight of crude oil futures, and r and r^f are the excess return of crude oil futures and the risk-free interest rate, respectively. The variance of the portfolio return can thus be expressed as

$$\mathrm{Var}(R_p) = w^2\theta^2\sigma^2, \tag{21}$$

where σ is the volatility of excess returns. The objective function of expected utility is the certainty equivalent return (CER),

$$\begin{aligned} U(R_p) &= E(R_p) - 0.5\gamma \mathrm{Var}(R_p) \\ &= w\theta\left(r + r^f\right) + (1-w)r^f - 0.5\gamma w^2\theta^2\sigma^2, \end{aligned} \tag{22}$$

Table 7
Direction-of-change.

	DoC rate	PT statistic
HAR-RV	0.620***	8.222
HAR-RV-J	0.591***	5.341
HAR-RV-CJ	0.575***	4.378
HAR-RV-TCJ	0.583***	4.860
HAR-RV-RS-I	0.594***	5.593
HAR-RV-RS-II	0.600***	6.037
HAR-RV-SJ-II	0.594***	5.552
HAR-RV-SJ-I	0.597***	5.786
Mean	0.615***	6.829
Median	0.602***	6.036
Trimmed mean	0.608***	6.438
DMSPE(1)	0.620***	7.098
DMSPE(0.9)	0.621***	7.164
Elastic net	**<u>0.665***</u>**	**<u>10.105</u>**
Lasso	**<u>0.651***</u>**	**<u>9.099</u>**

This table reports the Direction-of-Change (DoC) rates and the PT statistics of Pesaran and Timmermann (1992). We consider all the forecasting models used by this paper, including eight individual HAR-RV-type models, five combination approaches, and two shrinkage methods. Bold and underlined figures highlight instances in which the DoC rates as well as the PT statistics are the largest two. Statistical significance for DoC rate is based on the p-values of the PT statistic. The entire sample period containing 2357 observations spans from February 1, 2007, to July 15, 2016, while the length of out-of-sample period is 900.

*** Indicates significance at the 1% level.

where γ is the investor's coefficient of relative risk aversion. Maximizing the objective function, we can obtain the optimal portfolio weight given by

$$w_t = \frac{1}{\gamma}\frac{\theta\hat{r}_{t+1} + (\theta-1)r^f_{t+1}}{\theta^2\hat{\sigma}^2_{t+1}}, \tag{23}$$

where $\hat{r}_{t+1}$ is the forecast of oil futures excess return on day $t+1$, and $\hat{\sigma}^2_{t+1}$ is the RV forecast of crude oil futures on day $t+1$. We use the prevailing historical average to estimate $\hat{r}_{t+1}$ as it is an asset return benchmark that is very difficult to be outperformed (see, e.g., Campbell and Thompson, 2008; Welch and Goyal, 2008). Accordingly, the portfolio weights in Eq. (23) differ only due to the volatility forecasts from our forecasting models. It is noteworthy that when the leverage ratio $\theta = 1$, the optimal portfolio weights computed by Eq. (23) will reduce to the ones with no margin such as stocks (see, e.g., Campbell and Thompson, 2008; Rapach et al., 2010; Neely et al., 2014; Rapach et al., 2016). Furthermore, we assume that three different investors will maintain their margin accounts at the 10%, 12.5%, and 16.7% levels; that is, the corresponding leverage ratios $\theta = 10$, 8, and 6, respectively.[14] In addition, we restrict w_t to lie between -1.5 and 1.5.[15]

The mean-variance investor who allocates assets using Eq. (23) can realize a certainty equivalent return (CER) of

$$\mathrm{CER} = \overline{R}_p - 0.5\gamma\sigma^2_p, \tag{24}$$

where $\overline{R}_p$ and σ^2_p are the sample mean and variance, respectively, of the portfolio return over the out-of-sample evaluation period. Also, the Sharpe ratio is given by.

$$SR = \frac{\overline{R}^e_p}{\sigma^e_p}, \tag{25}$$

where $\overline{R}^e_p$ and σ^e_p are the sample mean and standard deviation, respectively, of the excess portfolio return over the out-of-sample evaluation period.

Table 8 reports the portfolio performance for all the forecasting models. Surprisingly, the simple HAR-RV model yields larger CER and Sharpe ratio than its extensions. In particular, the HAR-RV-CJ and HAR-RV-TCJ models show extremely poor portfolio performance. As expected, the elastic net and lasso models always yield higher economic values for both the CER and Sharpe ratio. Moreover, this result is robust

[14] The results are qualitatively similar for other reasonable values of leverage ratio. In particular, Zhang et al. (2018a) assume that the leverage ratio is 10. For the consideration of robustness, this paper considers more possible leverage ratios.

[15] In the application of stock market, Campbell and Thompson (2008), Huang et al. (2015), Neely et al. (2014), and Jiang et al. (2017) restrict the portfolio weights to lie between 0 and 1.5 because of short-sale constraint. However, it is easily possible to take a short position for futures, so that we relax the weight constraint to lie between -1.5 and 1.5.

Table 8
Portfolio performance.

Forecasting models	Leverage ratio = 10		Leverage ratio = 8		Leverage ratio = 6	
	CER	Sharpe ratio	CER	Sharpe ratio	CER	Sharpe ratio
Panel A: the risk aversion coefficient is one						
HAR-RV	10.836	0.464	10.837	0.464	10.839	0.464
HAR-RV-J	8.079	0.477	9.641	0.490	9.995	0.485
HAR-RV-CJ	−35.723	0.171	−20.135	0.249	−9.918	0.284
HAR-RV-TCJ	−61.686	−0.109	−48.522	−0.064	−34.221	−0.018
HAR-RV-RS-I	3.253	0.382	2.454	0.365	2.645	0.359
HAR-RV-RS-II	3.290	0.377	2.590	0.362	1.716	0.341
HAR-RV-SJ-II	8.057	0.480	9.614	0.492	10.110	0.488
HAR-RV-SJ-I	3.672	0.429	5.789	0.445	7.301	0.450
Mean	4.381	0.417	4.580	0.411	5.364	0.411
Median	5.422	0.443	7.137	0.456	7.483	0.448
Trimmed mean	2.161	0.400	3.638	0.408	5.252	0.415
DMSPE(1)	6.438	0.442	5.983	0.429	7.029	0.434
DMSPE(0.9)	7.425	0.449	6.798	0.436	7.595	0.438
Elastic net	**13.647**	**0.525**	**13.666**	**0.525**	**13.533**	**0.523**
Lasso	**17.501**	**0.594**	**15.346**	**0.560**	**12.377**	**0.509**
Panel B: the risk aversion coefficient is three						
HAR-RV	3.666	0.464	3.667	0.464	3.667	0.464
HAR-RV-J	2.942	0.499	2.943	0.499	2.944	0.499
HAR-RV-CJ	−37.931	0.012	−34.920	−0.028	−28.884	−0.024
HAR-RV-TCJ	−55.182	−0.302	−48.286	−0.305	−37.511	−0.252
HAR-RV-RS-I	1.257	0.389	1.258	0.389	1.260	0.389
HAR-RV-RS-II	1.231	0.382	1.232	0.382	1.233	0.382
HAR-RV-SJ-II	2.575	0.489	2.576	0.489	2.577	0.489
HAR-RV-SJ-I	0.586	0.433	0.587	0.433	0.590	0.433
Mean	2.143	0.452	2.144	0.452	2.146	0.452
Median	1.792	0.456	1.793	0.456	1.795	0.456
Trimmed mean	1.400	0.446	1.401	0.446	1.403	0.446
DMSPE(1)	2.631	0.465	2.632	0.465	2.633	0.465
DMSPE(0.9)	2.743	0.461	2.744	0.461	2.745	0.461
Elastic net	**4.603**	**0.525**	**4.604**	**0.525**	**4.604**	**0.525**
Lasso	**6.762**	**0.637**	**6.762**	**0.637**	**6.763**	**0.637**
Panel C: the risk aversion coefficient is five						
HAR-RV	2.232	0.464	2.233	0.464	2.233	0.464
HAR-RV-J	1.798	0.499	1.798	0.499	1.799	0.499
HAR-RV-CJ	−22.969	0.179	−22.898	0.103	−22.712	0.012
HAR-RV-TCJ	−47.075	−0.312	−39.845	−0.311	−33.064	−0.302
HAR-RV-RS-I	0.787	0.389	0.788	0.389	0.788	0.389
HAR-RV-RS-II	0.771	0.382	0.771	0.382	0.772	0.382
HAR-RV-SJ-II	1.577	0.489	1.578	0.489	1.579	0.489
HAR-RV-SJ-I	0.384	0.433	0.385	0.433	0.386	0.433
Mean	1.319	0.452	1.319	0.452	1.320	0.452
Median	1.108	0.456	1.108	0.456	1.110	0.456
Trimmed mean	0.873	0.446	0.873	0.446	0.874	0.446
DMSPE(1)	1.611	0.465	1.611	0.465	1.612	0.465
DMSPE(0.9)	1.678	0.461	1.679	0.461	1.679	0.461
Elastic net	**2.795**	**0.525**	**2.795**	**0.525**	**2.795**	**0.525**
Lasso	**4.090**	**0.637**	**4.090**	**0.637**	**4.090**	**0.637**

This table reports the annualized certainty equivalent return (CER) in percentage and Sharpe ratio for a mean-variance investor who allocates between oil futures and risk-free bills using various oil price volatility forecasts. The mean-variance investor may have various risk aversion coefficients, e.g., one (Panel A), three (Panel B), and five (Panel C). Three leverage ratios of 10, 8, and 6 are considered. The futures weight is constrained to lie between −1.5 and 1.5. We consider all the forecasting models used by this paper, including eight individual HAR-RV-type models, five combination approaches, and two shrinkage methods. Bold and underlined figures highlight instances in which the economic values are the largest two. The entire sample period containing 2357 observations spans from February 1, 2007, to July 15, 2016, while the length of out-of-sample period is 900.

not only to alternative risk aversion coefficients but also to alternative leverage ratios. The evidence suggests that the statistically accurate volatility forecasts of the elastic net and lasso can generate sizeable economic gains for a mean-variance investor in the practical application.

7. Conclusion

In this paper, we predict the crude oil price volatility based on the HAR-RV framework, which includes eight popular HAR-RV-type models. On one hand, we employ the combination approaches recommended by Rapach et al. (2010) to combine individual forecasts generated by the HAR-RV-type models. On the other hand, we employ two prevailing shrinkage methods, the elastic net and lasso, in which all the predictors from the eight HAR-RV-type models are used to produce oil price volatility.

The MCS test suggests that the elastic net and lasso exhibit significantly better out-of-sample forecasting performance than not only the individual HAR-RV-type models but also the combination approaches. Our results are robust to a wide range of settings, including various sample sizes used to estimate shrinkage factors, alternative estimation windows (expanding or rolling window), different out-of-sample periods, and other combination approaches. Furthermore, we document that the elastic net and lasso also generate substantially higher DoC rates, suggesting the higher directional accuracy. Finally, a mean-variance investor can realize sizeable economic gains when using the volatility forecasts generated by the elastic net and lasso to allocate her portfolio. In conclusion, the shrinkage methods can consistently beat the combination approaches in the out-of-sample prediction of oil price volatility from both the statistical and economic perspectives.

Acknowledgements

The authors are grateful to the editor and anonymous referees for insightful comments that significantly improved the paper. This work is supported by the National Natural Science Foundation of China [71371157, 71671145], the Humanities and Social Science Fund of Ministry of Education of China [16YJA790062, 17YJA790015, 17XJA790002, 18YJC790132, 18XJA790002], key laboratory of service computing and safety management of Yunnan provincial universities Doctoral Innovation Fund Program of Southwest Jiaotong University [D-CX201724], and Service Science and Innovation Key Laboratory of Sichuan Province [KL1704].

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eneco.2019.01.010.

References

Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. Int. Econ. Rev. 39, 885–905.
Andersen, T.G., Bollerslev, T., Diebold, F.X., 2007. Roughing it up: including jump components in the measurement, modeling, and forecasting of return volatility. Rev. Econ. Stat. 89, 701–720.
Audrino, F., Knaus, S.D., 2016. Lassoing the HAR model: a model selection perspective on realized volatility dynamics. Econ. Rev. 35, 1485–1521.
Avramov, D., 2002. Stock return predictability and model uncertainty. J. Financ. Econ. 64, 423–458.
Becker, R., Clements, A.E., 2008. Are combination forecasts of S&P 500 volatility statistically superior? Int. J. Forecast. 24, 122–133.
Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Financ. Mach. Learn. 3, 1–122.
Campbell, J.Y., Thompson, S.B., 2008. Predicting excess stock returns out of sample: can anything beat the historical average? Rev. Financ. Stud. 21, 1509–1531.
Corsi, F., 2009. A simple approximate long-memory model of realized volatility. J. Financ. Economet. 7, 174–196.
Corsi, F., Pirino, D., Reno, R., 2010. Threshold bipower variation and the impact of jumps on volatility forecasting. J. Econ. 159, 276–288.
Degiannakis, S., Filis, G., 2017. Forecasting oil price realized volatility using information channels from other asset classes. J. Int. Money Financ. 76, 28–49.
Elliott, G., Gargano, A., Timmermann, A., 2013. Complete subset regressions. J. Econ. 177, 357–373.
Granger, C.W., Ramanathan, R., 1984. Improved methods of combining forecasts. J. Forecast. 3, 197–204.
Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79, 453–497.
Haugom, E., Langeland, H., Molnár, P., Westgaard, S., 2014. Forecasting volatility of the US oil market. J. Bank. Financ. 47, 1–14.
Huang, X., Tauchen, G., 2005. The relative contribution of jumps to total price variance. J. Financ. Economet. 3, 456–499.
Huang, D., Jiang, F., Tu, J., Zhou, G., 2015. Investor sentiment aligned: a powerful predictor of stock returns. Rev. Financ. Stud. 28, 791–837.

Inoue, A., Jin, L., Rossi, B., 2017. Rolling window selection for out-of-sample forecasting with time-varying parameters. J. Econ. 196, 55–67.
Jiang, F., Lee, J.A., Martin, X., Zhou, G., 2017. Manager sentiment and stock returns. J. Financ. Econ. https://doi.org/10.1016/j.jfineco.2018.10.001.
Li, J., Tsiakas, I., 2017. Equity premium prediction: the role of economic and statistical constraints. J. Financ. Mark. 36, 56–75.
Li, J., Tsiakas, I., Wang, W., 2015. Predicting exchange rates out of sample: can economic fundamentals beat the random walk? J. Financ. Economet. 13, 293–341.
Liu, L.Y., Patton, A.J., Sheppard, K., 2015. Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. J. Econ. 187, 293–311.
Liu, J., Ma, F., Yang, K., Zhang, Y., 2018. Forecasting the oil futures price volatility: large jumps and small jumps. Energy Econ. 72, 321–330.
Ma, F., Li, Y., Liu, L., Zhang, Y., 2018a. Are low-frequency data really uninformative? A forecasting combination perspective. N. Am. J. Econ. Financ. 44, 92–108.
Ma, F., Liu, J., Wahab, M.I.M., Zhang, Y., 2018b. Forecasting the aggregate oil price volatility in a data-rich environment. Econ. Model. 72, 320–332.
Ma, F., Wei, Y., Liu, L., Huang, D., 2018c. Forecasting realized volatility of oil futures market: a new insight. J. Forecast. 37, 419–436.
Neely, C.J., Rapach, D.E., Tu, J., Zhou, G., 2014. Forecasting the equity risk premium: the role of technical indicators. Manag. Sci. 60, 1772–1791.
Patton, A.J., 2011. Volatility forecast comparison using imperfect volatility proxies. J. Econ. 160, 246–256.
Patton, A.J., Sheppard, K., 2009. Optimal combinations of realised volatility estimators. Int. J. Forecast. 25, 218–238.
Patton, A.J., Sheppard, K., 2015. Good volatility, bad volatility: signed jumps and the persistence of volatility. Rev. Econ. Stat. 97, 683–697.
Paye, B.S., 2012. 'Déjà vol': predictive regressions for aggregate stock market volatility using macroeconomic variables. J. Financ. Econ. 106, 527–546.
Pesaran, M.H., Timmermann, A., 1992. A simple nonparametric test of predictive performance. J. Bus. Econ. Stat. 10, 461–465.
Rapach, D.E., Strauss, J.K., Zhou, G., 2010. Out-of-sample equity premium prediction: combination forecasts and links to the real economy. Rev. Financ. Stud. 23, 821–862.
Rapach, D.E., Ringgenberg, M.C., Zhou, G., 2016. Short interest and aggregate stock returns. J. Financ. Econ. 121, 46–65.
Rossi, B., Inoue, A., 2012. Out-of-sample forecast tests robust to the choice of window size. J. Bus. Econ. Stat. 30, 432–453.
Sévi, B., 2014. Forecasting the volatility of crude oil futures using intraday data. Eur. J. Oper. Res. 235, 643–659.
Stock, J.H., Watson, M.W., 2004. Combination forecasts of output growth in a seven-country data set. J. Forecast. 23, 405–430.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B Methodol. 58, 267–288.
Wang, Y., Ma, F., Wei, Y., Wu, C., 2016. Forecasting realized volatility in a changing world: a dynamic model averaging approach. J. Bank. Financ. 64, 136–149.
Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. Rev. Financ. Stud. 21, 1455–1508.
Yin, L., Yang, Q., 2016. Predicting the oil prices: do technical indicators help? Energy Econ. 56, 338–350.
Zhang, Y., Ma, F., Shi, B., Huang, D., 2018a. Forecasting the prices of crude oil: an iterated combination approach. Energy Econ. 70, 472–483.
Zhang, Y., Ma, F., Wang, Y., 2018b. Forecasting Crude Oil Prices with a Large Set of Predictors: Can LASSO Select Powerful Predictors? (Working paper)
Zhang, Y., Zeng, Q., Ma, F., Shi, B., 2019. Forecasting stock returns: do less powerful predictors help? Econ. Model. https://doi.org/10.1016/j.econmod.2018.09.014.
Zhu, X., Zhu, J., 2013. Predicting stock returns: a regime-switching combination approach and economic links. J. Bank. Financ. 37, 4120–4133.
Zou, H., 2006. The adaptive lasso and its oracle properties. J. Am. Stat. Assoc. 101, 1418–1429.
Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. R. Stat. Soc. Ser. B (Stat Methodol.) 67, 301–320.