Zhang *et al. Financial Innovation* (2022) 8:56
https://doi.org/10.1186/s40854-022-00360-7

Financial Innovation

RESEARCH

Open Access

To jump or not to jump: momentum of jumps in crude oil price volatility prediction

Yaojie Zhang[1], Yudong Wang[1*], Feng Ma[2] and Yu Wei[3]

*Correspondence: wangyudongnj@126.com
[1] School of Economics and Management, Nanjing University of Science and Technology, Xiaolingwei 200, Xuanwu District, Nanjing 210094, China
Full list of author information is available at the end of the article

Abstract

A well-documented finding is that explicitly using jumps cannot efficiently enhance the predictability of crude oil price volatility. To address this issue, we find a phenomenon, "momentum of jumps" (MoJ), that the predictive ability of the jump component is persistent when forecasting the oil futures market volatility. Specifically, we propose a strategy that allows the predictive model to switch between a benchmark model without jumps and an alternative model with a jump component according to their recent past forecasting performance. The volatility data are based on the intraday prices of West Texas Intermediate. Our results indicate that this simple strategy significantly outperforms the individual models and a series of competing strategies such as forecast combinations and shrinkage methods. A mean–variance investor who targets a constant Sharpe ratio can realize the highest economic gains using the MoJ-based volatility forecasts. Our findings survive a wide variety of robustness tests, including different jump measures, alternative volatility measures, various financial markets, and extensive model specifications.

Keywords: Oil futures market, Volatility forecasting, Momentum of jumps, Model switching, Portfolio exercise

JEL Classifications: C22, C53, Q47, G17

Introduction

The jump component is an essential and useful determinant in the prediction of the volatility dynamics of various asset prices, such as exchange rates, stock returns, and bond yields (see, e.g., Andersen et al. 2007; Corsi et al. 2010; Duong and Swanson 2015; Patton and Sheppard 2015; Clements and Liao 2017). However, an influential paper by Prokopczuk et al. (2016) argues that explicitly using jumps cannot efficiently enhance the out-of-sample forecast accuracy for the volatility of the crude oil futures market. Prokopczuk et al. (2016) provide a plausible explanation that unpredictable events such as political unrest and natural disasters in oil-exporting countries always trigger jumps in oil prices. Theoretically, jumps do not occur at each point in time. When there is no jump in the oil price, incorporating the jump component into the predictive model will probably lead to overfitting, in which the in-sample forecasting performance improves but the out-of-sample performance worsens. Hence, it is critical to investigate whether the jump component should be included in volatility models in real time. The main purpose of this study is to selectively use jumps (or jump model) and thereby discover useful forecasting information that is hidden in jumps. To this end, we propose a simple but successful strategy to improve the forecasting ability of the jump model. This is our main contribution to the literature on forecasting crude oil market volatility.

Our new strategy makes an optimal predictive model switch between the benchmark of the heterogeneous autoregressive realized variance (HAR-RV)[1] model pioneered by Corsi (2009) and the HAR-CJ jump model pioneered by Andersen et al. (2007), which is an extension that adds the jump component to the HAR-RV specification. The switching behavior is conditional on the relative predictive performances of these two models during a recent period of past time. Specifically, if the recent past predictive ability of the HAR-CJ model is superior to that of the HAR-RV model, we continue to select the HAR-CJ model to predict the oil futures market realized variance (RV) in the near future. Otherwise, the HAR-RV benchmark is used to generate volatility forecasts. The motivation is straightforward: We believe that the relatively strong forecasting power of the HAR-CJ model is persistent. That is, a jump model with a good past performance will have a good forecasting performance in the near future. We term this phenomenon "momentum of jumps" (MoJ), in the context of forecasting oil futures market volatility. To be precise, MoJ refers to the momentum (or persistence) of the predictive ability of a jump model.

Consistent with the empirical findings by Prokopczuk et al. (2016), our full-sample estimation results show that the overall impact of the jump component on the oil futures market RV is not pronounced. In other words, the benefits from the use of jumps appear to be negligible. This in-sample evidence also suggests that using the jump model (i.e., the HAR-CJ model) alone is unlikely to be feasible for out-of-sample forecasting exercises. Nevertheless, we observe that the HAR-CJ model outperforms the HAR-RV model during part of the out-of-sample period. More importantly, several relatively good periods are clustered and show a continuous pattern. Using a formal test proposed by Wang et al. (2018), we document the existence of the MoJ between the HAR-RV and HAR-CJ models for the oil futures market. The existence of the MoJ phenomenon suggests that if the jump model (i.e., the HAR-CJ model) could produce more accurate RV forecasts than the HAR-RV benchmark over a recent past period, the jump model would continuously produce more accurate RV forecasts in the near future. This lays the foundation for the success of our MoJ strategy in forecasting oil futures RV.

In our MoJ strategy, we use the mean squared error (MSE) loss function to assess the past predictive performance of the individual HAR-RV and HAR-CJ models during a look-back period. The MoJ strategy uses the model that shows a relatively good past forecasting performance to generate an RV forecast in the next forecast step. To rule out the concern that the success of the MoJ method is due to the forecast combination,[2] we further consider the simple mean combination as a competing model, which takes the equally weighted average of the individual HAR-RV and HAR-CJ forecasts. In terms of the out-of-sample evaluation, we rely on the statistic test of the model confidence set

[1] The reason for using the HAR framework is that the HAR models are very tractable and useful in the prediction of financial market RV. In the robustness test below, we also rely on the MIDAS models to forecast oil RV and obtain similar results.

[2] Particularly, the MoJ model can be regarded as a special combination approach.

(MCS) originated by Hansen et al. (2011). We find convincing evidence that the MoJ strategy consistently exhibits a substantially stronger out-of-sample predictive ability than the competing models of the HAR-RV, HAR-CJ, and mean combination for not only the 1-day, 5-day, 10-day, and 22-day forecast horizons, but also for six widely used loss functions.

We further present a multitude of robustness tests and extensions. The results are summarized in seven streams. First, we consider the robustness regarding the jump detection test and jump model. Specifically, we additionally use two jump models, namely, the HAR-J model proposed by Andersen et al. (2007), which uses a simple measure of the jump component without a jump detection test, and the HAR-TCJ model proposed by Corsi et al. (2010), in which a threshold jump measure is used. In terms of the jump detection test, we consider three confidence levels: 1%, 0.5%, and 0.1%.

The second type of robustness test entails how to evaluate past forecasting performance in our MoJ strategy. We calculate the forecast error of past RV forecasts during 1-day, 5-day, 10-day, and 22-day look-back periods. Furthermore, the average MoJ forecasts are generated by the individual MoJ forecasts based on the various look-back periods. In addition, we use three different evaluation criteria to assess past forecasting performance.

Third, we consider the robustness of the model specification. The MoJ method is mainly based on the linear HAR models. Alternatively, we consider not only the nonlinear HAR models cast in logarithmic and standard deviation forms, but also the MIDAS model, which is regarded as a generalized version of the HAR framework.

Fourth, we consider alternative volatility estimators and forecast evaluation methods. The volatility estimators include the RV and realized kernel (RK). In addition, we use six different loss functions to evaluate forecast accuracy. Moreover, both the MCS test by Hansen et al. (2011) and the Diebold–Mariano (DM) test by Diebold and Mariano (1995) are used to calculate the significance level of forecast accuracy.

Fifth, we consider alternative estimation windows in out-of-sample forecasting exercises. On the one hand, we use both the rolling and expanding estimation windows. On the other hand, we consider different window sizes (i.e., various lengths of out-of-sample evaluation periods).

Sixth, we extend our competing models by considering other similar strategies, including alternative forecast combination approaches that also depend on individual models' past forecasting performance and the three widely used shrinkage approaches of the ridge, lasso, and elastic net.

Finally, the MoJ strategy is extended to the stock market. That is, we use the MoJ strategy as well as the competing models to predict the stock market RV. Fortunately, we observe consistent results for all the above-mentioned robustness tests and extensions, which greatly alleviate the concern of data mining.

In a portfolio exercise, we explore the economic significance of the volatility forecasts of the MoJ strategy and the competing models. Specifically, we follow Bollerslev et al. (2018) and consider a specific case, in which a mean–variance investor who targets a constant Sharpe ratio allocates her wealth between a risky asset (i.e., oil futures) and a risk-free asset (i.e., risk-free bills). The corresponding results suggest that the four forecasting models used in this study deliver sizeable utility gains relative to the ones from a static model that uses the rolling sample average of past RVs. More importantly, the utility gains from our MoJ model relative to the ones from the static model are always highest. The mean–variance investor would be glad to pay at least 56 basis points to employ the MoJ model rather than the simple static model, which is, of course, economically significant.

We organize the paper as follows. 'Related literature and our contribution" section reviews the related literature and highlights the paper's contribution. "RV and HAR models" section details the methodology of RV and HAR models. "Data and in-sample results" section presents the data and in-sample results, while "Out-of-sample analyses" section provides the out-of-sample analyses. "Robustness checks" section provides a wide variety of robustness checks. "Extension and application" section presents extensions and an economic application of the MoJ strategy. "Conclusion" section concludes this paper.

Related literature and our contribution

In this section, we review the related literature on (a) jump behavior in the crude oil market, (b) jumps and crude oil volatility forecasting, and (c) the momentum of predictability (MoP). Moreover, we separately discuss the innovative work of this paper in the three aspects.

Jump behavior in the crude oil market

Crude oil prices are characterized by jump behavior. Gronwald (2012) argues that a large quantity of total oil price volatility is triggered by jumps. Wilmot and Mason (2013) document that jumps help to improve a model's ability to explain crude oil prices. Bouri (2019) finds that the jumps in the sovereign risks of major oil-exporting countries are significantly driven by oil price volatility jumps. Bouri and Gupta (2020) present that crude oil price jumps and macroeconomic news surprises are likely to occur synchronously, indicating the sensitivity of crude oil prices to macroeconomic news. Bouri et al. (2021) provide evidence that the spillover effect of jumps in crude oil and other financial markets is notable. In contrast, this paper provides new insights into the jump behavior of crude oil prices. That is, we find a novel phenomenon, MoJ, in the forecasting of oil futures market volatility. The predictive ability of jumps is confirmed to be persistent.

Jumps and crude oil volatility forecasting

Andersen et al. (2007) is perhaps the first study that uses jump information to forecast the RV of financial assets. Following this seminal work, a growing number of studies rely on jump information to improve the predictability of crude oil volatility (see, e.g., Liu et al. 2018; Ma et al. 2019; Dutta et al. 2021). In contrast, the work by Prokopczuk et al. (2016) most closely relates to this paper. Prokopczuk et al. (2016) explore the role of jumps in forecasting energy market volatility and find that explicitly modeling jumps does not significantly improve the forecast accuracy for the volatility of the oil futures market. However, their study is silent on how to improve the accuracy of oil price volatility forecasts. Our paper contributes to their study by providing a solution to the problem of improving forecast accuracy. Specifically, we propose the MoJ strategy, which selectively uses the jump model and thereby successfully captures the useful forecasting component contained in jumps.

MoP

Our paper also contributes to the literature on MoP (see Wang et al. 2018; Zhang et al. 2019a). Wang et al. (2018) find the MoP that a univariate predictive regression with one macroeconomic variable, which generates more accurate return forecasts than the benchmark of historical average during several past months, can continue to successfully predict stock market returns in the near future. Zhang et al. (2019a) document the existence of the MoP between low- and high-frequency forecasting models in the case of forecasting stock market volatility, thus establishing a new mixed-frequency model. In this sense, the MoJ proposed by our paper is not completely new, as it has been empirically confirmed by the related studies of Wang et al. (2018) and Zhang et al. (2019a). However, the MoP findings of the two related studies, in other words, motivate and support us to investigate the MoJ. In contrast, our paper provides a new study that focuses on the momentum of the forecasting performance of the jump model in predicting oil futures market volatility. This is necessary and meaningful because the role of jumps in forecasting oil futures market volatility is found to be limited. To address this issue, we rely on the fundamental of the MoP and thereby present an efficient model, i.e., the MoJ strategy.

More broadly, the MoJ strategy, as well as the MoP, is related to the conditional combination approaches that are based on past predictive performance (see, e.g., Stock and Watson 2004; Yang 2004; Giacomini and White 2006). In contrast, the contribution of this paper is not a technical innovation but a novel idea of clustered jumps (i.e., MoJ). Moreover, we document that our MoJ strategy can outperform a popular forecast combination approach that is conditional on past predictive performance.

RV and HAR models

RV and jump

The quadratic variation (QV) of the asset price return process can be given by

$$QV_t = \int_{t-1}^{t} \sigma^2(z)dz + \sum_{t-1<z\leq t} \nu^2(z), \tag{1}$$

where in the continuous-time jump diffusion process, $\sigma(z)$ represents a stochastic volatility process and $\nu(z)$ measures the size of discrete jumps. On the right hand of Eq. (1), the first term is the so-called integrated variance (IV), which is regarded as the continuous sample path component of QV, while the second one represents the jump (discontinuous) component of QV.

Andersen and Bollerslev (1998), Barndorff-Nielsen and Shephard (2002), and Andersen et al. (2003) emphasize that the RV estimator uniformly converges to QV in probability as the sampling frequency increases. The RV measure can be calculated as the summation of squared intraday oil price returns,

$$RV_t = \sum_{j=1}^{N} r_{t,j}^2, \tag{2}$$

where RV_t refers to the realized variance measure on trading day t, $N = 1/\Delta$, Δ is the sampling interval for intraday returns, and $r_{t,j}$ denotes the jth intraday oil futures market return during day t. Since RV converges to QV, we have

$$RV_t \to \int_{t-1}^{t} \sigma^2(z)dz + \sum_{t-1<z\leq t} \nu^2(z), \tag{3}$$

for $N \to \infty$ or $\Delta \to 0$. Barndorff-Nielsen and Shephard (2004) further propose an estimator dubbed realized bipower variation (BPV), which takes the form of

$$BPV_t = \kappa_1^{-2} \sum\nolimits_{j=2}^{N} |r_{t,j}||r_{t,j-1}|, \tag{4}$$

where $\kappa_1 = \sqrt{2/\pi}$. As the sampling frequency increases ($N \to \infty$ or $\Delta \to 0$), BPV is a consistent estimator of IV. That is, we have

$$BPV_t \to \int_{t-1}^{t} \sigma^2(z)dz, \tag{5}$$

for $N \to \infty$. Combining Eqs. (3) and (5), we can consistently estimate the jump (discontinuous) component of QV as

$$RV_t - BPV_t \to \sum_{t-1<z\leq t} \nu^2(z), \tag{6}$$

for $N \to \infty$. To ensure that each jump estimate is nonnegative, Andersen et al. (2007) truncate jump measures at zero, which is also suggested by Barndorff-Nielsen and Shephard (2004). Consequently, the daily jump measure on day t is given by

$$J_t = \max\{RV_t - BPV_t, 0\}. \tag{7}$$

We follow Andersen et al. (2007) and employ the ratio statistic to detect significant jumps. The jump detection test based on the ratio statistic is given by

$$Z_t = \Delta^{-1/2} \frac{(RV_t - BPV_t)/RV_t}{\sqrt{(\kappa_1^{-4} + 2\kappa_1^{-2} - 5)\max\{1, TQ_t/BPV_t^2\}}}, \tag{8}$$

where TQ_t denotes the realized tripower quarticity measure. Statistically, TQ is expressed as

$$TQ_t = \Delta^{-1}\kappa_{4/3}^{-3} \sum\nolimits_{j=3}^{N} |r_{t,j}|^{4/3}|r_{t,j-1}|^{4/3}|r_{t,j-2}|^{4/3}, \tag{9}$$

where $\kappa_{4/3} = 2^{2/3}\Gamma(7/6)/\Gamma(0.5)$. As the test statistic in Eq. (8) closely follows a standard normal distribution, the significant jump (SJ) is naturally expressed as

$$SJ_t = I(Z_t > \Theta_\alpha)(RV_t - BPV_t), \tag{10}$$

where $I(\cdot)$ refers to an indicator function, which equals one if a significant jump happens and zero otherwise, and Θ_α denotes the threshold value that is calculated by the cumulative standard normal distribution at the confidence level of $1-\alpha$. We follow Andersen et al. (2007) and rely on the significance level of 0.5%. To ensure that the sum of the continuous and discontinuous components equals the whole RV, the continuous component is then identified as

$$C_t = I(Z_t \leq \Theta_\alpha)RV_t + I(Z_t > \Theta_\alpha)BPV_t. \tag{11}$$

HAR models

In terms of our forecasting strategies, the benchmark method is naturally the HAR-RV model, which is pioneered by Corsi (2009). The HAR-RV model is probably the most popular volatility model. This is because the model captures some stylized facts of asset return volatility such as long memory and multi-scaling behavior. Furthermore, the HAR-RV model is tractable, as it merely includes three variables without any hyperparameter tuning. Therefore, its model specification is straightforward, and can be shown as

$$RV_{t+1:t+h} = \varphi_0 + \varphi_d RV_t + \varphi_w RV_{t-4:t} + \varphi_m RV_{t-21:t} + \omega_{t+1:t+h}, \tag{12}$$

where $RV_{t+1:t+h} = (1/h)(RV_{t+1} + \cdots + RV_{t+h})$. In particular, RV_t, $RV_{t-4:t}$, and $RV_{t-21:t}$ denote the daily, weekly, and monthly RVs, respectively, all of which are available up to day t.

More importantly, our MoJ strategy requires a jump model. We choose the HAR-CJ model, which is pioneered by Andersen et al. (2007) and has been widely used by a host of literature on predicting asset return volatility (see, e.g., Sévi 2014; Prokopczuk et al. 2016; Wang et al. 2016; Buncic and Gisler 2017; Zhang et al. 2019a, 2021). Mathematically, the HAR-CJ model takes the following form:

$$RV_{t+1:t+h} = \varphi_0 + \varphi_{cd}C_t + \varphi_{cw}C_{t-4:t} + \varphi_{cm}C_{t-21:t} + \varphi_{sd}SJ_t + \varphi_{sw}SJ_{t-4:t} + \varphi_{sm}SJ_{t-21:t} + \omega_{t+1:t+h}. \tag{13}$$

For robustness, we also consider alternative jump measures and jump models in Sect. 6.1.

Data and in-sample results

Data

Following Sévi (2014), Haugom et al. (2014), and Zhang et al. (2022), we choose a well-known oil price benchmark, West Texas Intermediate (WTI). The intraday price data of the WTI futures are obtained from Tick Data. The whole sample period is between January 3, 2012 and May 11, 2018.

The 5-min RV is commonly used by a substantial body of literature on predicting oil futures market RV (see, e.g., Haugom et al. 2014; Sévi 2014; Ma et al. 2019; Yang et al. 2019; Zhang et al. 2019c; Niu et al. 2021). Overall, Liu et al. (2015) argue that it is extremely difficult to outperform the 5-min RV by using any other volatility measures from a wide range of estimators and financial assets. Thus, we rely on the 5-min interval as the sampling frequency to calculate the oil futures market RV.

Table 1 Full-sample estimation results of the HAR-RV and HAR-CJ models

Variables	*h* = 1	*h* = 5	*h* = 10	*h* = 22
Panel A: HAR-RV				
φ_0	0.199***	0.258***	0.314***	0.410***
	(2.803)	(2.804)	(2.836)	(3.491)
φ_d	0.056	0.067**	0.042*	0.041**
	(1.120)	(2.073)	(1.925)	(2.151)
φ_w	0.356***	0.264**	0.225*	0.210*
	(3.123)	(2.303)	(1.673)	(1.763)
φ_m	0.465***	0.510***	0.540***	0.499***
	(4.383)	(4.227)	(3.683)	(3.327)
R^2	0.314	0.544	0.580	0.581
Panel B: HAR-CJ				
φ_0	0.264***	0.316***	0.360***	0.439***
	(4.639)	(4.358)	(3.933)	(3.735)
φ_{cd}	0.116	0.133**	0.095**	0.102***
	(1.609)	(2.449)	(2.458)	(3.258)
φ_{cw}	0.589***	0.505***	0.481***	0.386***
	(3.773)	(3.869)	(3.595)	(3.213)
φ_{cm}	0.256**	0.282**	0.300**	0.309**
	(2.213)	(2.277)	(2.082)	(1.979)
φ_{sd}	0.007	0.013	− 0.003	− 0.007
	(0.186)	(0.571)	(− 0.227)	(− 0.675)
φ_{sw}	− 0.001	− 0.109	− 0.159**	− 0.073
	(− 0.010)	(− 1.262)	(− 2.098)	(− 0.934)
φ_{sm}	0.227	0.342*	0.467*	0.493*
	(1.447)	(1.763)	(1.814)	(1.755)
R^2	0.342	0.600	0.639	0.622

This table provides the full-sample estimation results of the HAR-RV and HAR-CJ models in Panels A and B, respectively. The HAR-RV model is given by

$$RV_{t+1:t+h} = \varphi_0 + \varphi_d RV_t + \varphi_w RV_{t-4:t} + \varphi_m RV_{t-21:t} + \omega_{t+1:t+h},$$

where RV is the realized variance, $RV_{t+1:t+h} = (1/h)(RV_{t+1} + \cdots + RV_{t+h})$. In particular, RV_t, $RV_{t-4:t}$ and $RV_{t-21:t}$ denotes the daily, weekly, and monthly RVs, respectively, which are all available up to day *t*. The HAR-CJ model is expressed as

$$RV_{t+1:t+h} = \varphi_0 + \varphi_{cd} C_t + \varphi_{cw} C_{t-4:t} + \varphi_{cm} C_{t-21:t} + \varphi_{sd} SJ_t + \varphi_{sw} SJ_{t-4:t} + \varphi_{sm} SJ_{t-21:t} + \omega_{t+1:t+h},$$

where C and SJ denote the continuous component and significant jump measure, respectively. Coefficient estimates are reported and their *t*-statistics shown in parentheses below are computed based on a Newey-West correction which allows for serial correlation up to order 5 (*h* = 1), 10 (*h* = 5), 20 (*h* = 10), and 44 (*h* = 22). The R^2s are also reported. *** (**) (*) indicates significance at the 1% (5%) (10%) two-tailed level. The sample period runs from January 3, 2012 to May 11, 2018.

In-sample results

The in-sample estimation results of the individual HAR-RV and HAR-CJ models are reported in Table 1. One striking observation follows the table immediately. The HAR predictors (namely, RV_t, $RV_{t-4:t}$, and $RV_{t-21:t}$) always yield significant coefficients, while the regression coefficients of the lagged daily and weekly SJs are always insignificant. Although the R^2 of HAR-CJ is greater than that of HAR-RV, the improvement in the R^2 is limited. Overall, the full-sample estimation results suggest that the jump components do not contain a powerful explanatory ability for future oil futures RV. This also implies that a straightforward approach of using the jump model (i.e., the HAR-CJ) alone is unlikely to be feasible. This evidence echoes the findings of Prokopczuk et al. (2016), who document that the jump components are not useful for the in-sample predictability of crude oil price volatility.

Out-of-sample analyses

Forecasting methodology of individual HAR regression models

The in-sample estimation analysis only provides the predictive information of the regression models (namely, the HAR-RV and HAR-CJ), while it is silent on our MoJ strategy and the mean combination approach. In real time, financial investors and practitioners pay more attention to the out-of-sample forecasting test as it is more relevant to examining genuine predictive ability. Moreover, an in-sample analysis is probably influenced by econometric issues such as the Stambaugh bias (see Busetti and Marcucci 2013), small-sample size distortion, and over-fitting, whereas an out-of-sample test is less likely to be influenced. Therefore, it is more crucial to assess the out-of-sample forecasting ability of the volatility models used.

In this study, we generate the out-of-sample RV forecasts for the individual HAR-RV and HAR-CJ models by employing a rolling estimation window. Specifically, we decompose the whole sample period into an in-sample training period and an out-of-sample forecasting period. The former contains the initial 819 observations, while the latter contains the remaining 800 observations. When we obtain each out-of-sample RV forecast, we roll the estimation window forward by not only discarding the first used observation, but by also including one new observation. Finally, the MoJ and mean combination forecasts are produced by the individual HAR-RV- and HAR-CJ forecasts.

Forecasting methodology based on the MoJ

Wang et al. (2018) present a similar phenomenon, termed MoP, in which the stock return predictability of univariate regressions is persistent. Specifically, a superior past predicting performance of a univariate regression model that uses a single economic variable is commonly followed by a superior future predicting performance. Furthermore, Zhang et al. (2019a) document that the MoP also exists between GARCH-class and HAR-RV-type models. Along the same lines, we propose the MoJ. To be precise, the MoJ refers to the MoP between the volatility forecasting models with and without jumps. In our case, we have two strands of RV forecasts, separately given by the HAR-RV and HAR-CJ models. We then continue to employ the volatility model whose past predictive performance is relatively good. Following Zhang et al. (2019a) and Wang et al. (2018), we assess whether the past predictive performance of the HAR-CJ model is superior to that of the HAR-RV model as follows.

$$
\begin{aligned}
pp_{t+1:t+h}(k) = & I\left(\sum_{i=t-h-k+1}^{t-h}(RV_{i+1:i+h}-\widehat{RV}^{CJ}_{i+1:i+h})^2\right. \\
& \left. -\sum_{i=t-h-k+1}^{t-h}(RV_{i+1:i+h}-\widehat{RV}^{RV}_{i+1:i+h})^2<0\right),
\end{aligned}
\tag{14}
$$

where k denotes the length of the look-back evaluation period,[3] $I(\cdot)$ denotes an indicator function, $RV_{t+1:t+h}$ is the true RV on days $t+1:t+h$, and $\widehat{RV}^{CJ}_{i+1:i+h}$ and $\widehat{RV}^{RV}_{i+1:i+h}$ are the HAR-CJ and HAR-RV forecasts, respectively, for $RV_{i+1:i+h}$. Based on the relative past performance, as defined by $pp_{t+1:t+h}(k)$, we can readily obtain the corresponding MoJ forecast as follows.

$$\widehat{RV}^{MoJ}_{t+1:t+h}(k) = \begin{cases} \widehat{RV}^{CJ}_{t+1:t+h}, & \text{if } pp_{t+1:t+h}(k) = 1 \\ \widehat{RV}^{RV}_{t+1:t+h}, & \text{if } pp_{t+1:t+h}(k) = 0. \end{cases} \tag{15}$$

Simple mean forecast combination

Our MoJ strategy switches between the benchmark and jump models by observing their relative past forecasting performances. This model selection approach can be treated as a particular combination approach, in which the weight of each model is a binary variable that equals either 0 or 1. For comparison, we also use the equal-weighted combination forecasts, $1/2(\widehat{RV}^{CJ}_{t+1:t+h} + \widehat{RV}^{RV}_{t+1:t+h})$. Here, we do not consider any more complex weighting schemes when combining forecasts. This is because the famous "forecast combination puzzle" suggests that the simple mean cannot be systematically outperformed by many other sophisticated combination methods in out-of-sample prediction exercises (see, e.g., Stock and Watson 2004; Rapach et al. 2010).

Evaluation framework

We employ six commonly used loss functions to provide a quantitative assessment of the out-of-sample predictive performance for different volatility forecasting strategies: the Quasi-Likelihood (QLIKE), mean squared error (MSE), mean absolute error (MAE), mean squared percentage error (MSPE), mean absolute percentage error (MAPE), and mean squared logarithmic error (MSE-LOG) loss functions, which are statistically expressed as

$$QLIKE : L(\widehat{RV}_{t+1:t+h}, RV_{t+1:t+h}) = \log(\widehat{RV}_{t+1:t+h}) + \frac{RV_{t+1:t+h}}{\widehat{RV}_{t+1:t+h}}, \tag{16}$$

$$MSE : L(\widehat{RV}_{t+1:t+h}, RV_{t+1:t+h}) = \left(\widehat{RV}_{t+1:t+h} - RV_{t+1:t+h}\right)^2, \tag{17}$$

$$MAE : L(\widehat{RV}_{t+1:t+h}, RV_{t+1:t+h}) = \left|\widehat{RV}_{t+1:t+h} - RV_{t+1:t+h}\right|, \tag{18}$$

$$MSPE : L(\widehat{RV}_{t+1:t+h}, RV_{t+1:t+h}) = \left(1 - \frac{\widehat{RV}_{t+1:t+h}}{RV_{t+1:t+h}}\right)^2, \tag{19}$$

[3] The forecasting results reported below is based on the look-back period of $k=5$ (i.e., one week) for the MoJ strategy. A robustness check in Sect. 6.2 considers other reasonable values of k.

$$MAPE: L(\widehat{RV}_{t+1:t+h}, RV_{t+1:t+h}) = \left|1 - \frac{\widehat{RV}_{t+1:t+h}}{RV_{t+1:t+h}}\right|, \tag{20}$$

and

$$MSE{-}LOG: L(\widehat{RV}_{t+1:t+h}, RV_{t+1:t+h}) = \left(\log(\widehat{RV}_{t+1:t+h}) - \log(RV_{t+1:t+h})\right)^2, \tag{21}$$

respectively, where $RV_{t+1:t+h}$ is the true RV for days $t+1:t+h$, and $\widehat{RV}_{t+1:t+h}$ denotes the RV forecast given by one of the predictive strategies. Patton (2011) recommends the use of the QLIKE and MSE loss functions because the two are robust to the presence of noise in the volatility proxy. Nonetheless, we employ more loss functions to show a comprehensive test.

To ascertain the confidence level of the different models' out-of-sample forecast accuracies, we follow extensive literature on predicting RV (see, e.g., Patton and Sheppard 2009; Liu et al. 2015; Gong and Lin 2018; Zhang et al. 2019c, 2020; Calzolari et al. 2021; Dai et al. 2022) and employ the MCS econometric method pioneered by Hansen et al. (2011). An MCS refers to a subset of all the used models into which the best model falls with a specific confidence level. Generally, a model that delivers a larger MCS p value is more likely to show the best forecasting performance. Following Hansen et al. (2011) and Zhang et al. (2019c), we choose the confidence (significance) level of 90% (10%). In other words, a model whose MCS p value is greater than 0.1 falls into the MCS. Finally, it should be noted that the MCS p values we report below are all calculated based on the range statistic; however, the results are similar when we rely on the semi-quadratic statistic.

Forecasting performance

Table 2 presents the MCS test results. We summarize the table with one key observation. Our MoJ strategy always produces the highest MCS p value (i.e., 1). In contrast, the HAR-RV, HAR-CJ, and mean combination models generate substantially lower MCS p values, most of which are lower than 0.1, indicating that the corresponding models cannot enter the MCS at the 10% significance level. Overall, the reported MCS p values indicate that our MoJ strategy exhibits significantly better forecasting performance than the competing models of the individual HAR-RV and HAR-CJ models as well as the mean combination. Furthermore, we observe that the relatively powerful predictive ability of the MoJ strategy consistently exists not only across various loss functions but also across various forecast horizons.

Testing the MoJ

Wang et al. (2018) and Zhang et al. (2019a) both highlight that the success of their MoP strategies relies on the presence of the MoP. Therefore, we need to investigate whether the superiority of our MoJ strategy is supported by the existence of the MoJ. More precisely, the MoJ refers to the momentum of the predictive ability of the forecasting strategy with jumps relative to the one without jumps. That is, we should examine whether a better past predictive performance of the HAR-CJ models relative to that of the HAR-RV

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

Models	QLIKE	MSE	MAE	MSPE	MAPE	MSE-LOG
Panel A: 1-day horizon						
HAR-RV	**0.306**	**0.105**	0.005	0.018	0.001	0.001
HAR-CJ	0.075	**0.389**	**0.301**	0.018	0.001	0.001
Mean	**0.306**	**0.276**	0.022	0.018	0.000	0.001
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Panel B: 5-day horizon						
HAR-RV	0.000	0.007	0.001	0.000	0.001	0.000
HAR-CJ	0.000	0.007	0.001	0.000	0.000	0.000
Mean	0.000	0.007	0.001	0.000	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Panel C: 10-day horizon						
HAR-RV	0.000	0.003	0.000	0.000	0.001	0.002
HAR-CJ	0.000	0.003	0.000	0.000	0.000	0.000
Mean	0.000	0.003	0.000	0.000	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Panel D: 22-day horizon						
HAR-RV	0.000	0.000	0.000	0.032	0.001	0.001
HAR-CJ	0.000	0.000	0.000	0.000	0.000	0.000
Mean	0.000	0.000	0.000	0.000	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

This table provides the MCS *p* values of the four used models. Panels A, B, C, and D report the corresponding results for 1-day, 5-day, 10-day, and 22-day horizons, respectively. The mean combination uses the equally weighted average (that is, simple mean) of the individual HAR-RV and HAR-CJ forecasts, while our MoJ strategy switches between the HAR-RV and HAR-CJ forecasts based on their relatively past forecasting performance. The past forecasting performance is evaluated by a 5-day look-back period. The six considered loss functions are QLIKE, MSE, MAE, MSPE, MAPE, and MSE-LOG. Bold numbers highlight important instances in which the corresponding model falls into the MCS with the 10% significance level. The entire sample period consisting of 1619 observations spans January 3, 2012 to May 11, 2018, while the out-of-sample forecasting period contains the last 800 observations

model can generally result in a better future performance. Statistically, the future predictive performance of the HAR-CJ relative to that of the HAR-RV over days $t+1:t+h$ is defined as

$$fp_{t+1:t+h} = I\Big((RV_{t+1:t+h} - \widehat{RV}^{CJ}_{t+1:t+h})^2 - (RV_{t+1:t+h} - \widehat{RV}^{RV}_{t+1:t+h})^2 < 0\Big). \quad (22)$$

In a statistical sense, a cross-sectional dependence of $pp_{t+1:t+h}(k)$ and $fp_{t+1:t+h}$ implies the existence of the MoJ. Following the related studies of Zhang et al. (2019a) and Wang et al. (2018), we rely on the chi-square statistic proposed by Pesaran and Timmermann (2009) to test the null hypothesis that $pp_{t+1:t+h}(k)$ and $fp_{t+1:t+h}$ are not cross-sectional dependent in the presence of serial dependencies for each series itself against the alternative hypothesis that the two time-series variables are cross-sectional dependent. In this sense, if the null hypothesis of no dependence between $pp_{t+1:t+h}(k)$ and $fp_{t+1:t+h}$ is rejected, we statistically prove that the MoJ exists.

We follow Wang et al. (2018) and report the *p*-values for the Pesaran and Timmermann (2009) statistics in Table 3. As expected, all the *p* values for the different lengths of the look-back periods and forecast horizons are less than 0.001. That is, the null hypothesis of independence between $pp_{t+1:t+h}(k)$ and $fp_{t+1:t+h}$ is rejected below the 0.1% significance level. This evidence suggests that the MoJ does exist between the

Table 3 Testing results of the momentum of jumps

Look-back period	$h=1$	$h=5$	$h=10$	$h=22$
$k=1$	< 0.001	< 0.001	< 0.001	< 0.001
$k=5$	< 0.001	< 0.001	< 0.001	< 0.001
$k=10$	< 0.001	< 0.001	< 0.001	< 0.001
$k=22$	< 0.001	< 0.001	< 0.001	< 0.001

This table provides the *p* values of the Pesaran and Timmermann (2009) chi-square statistic that is used to test the existence of the momentum of jumps (MoJ). The MoJ refers to that the forecasting model with jump information (i.e., the HAR-CJ) which outperforms the benchmark model without jump information (i.e., the HAR-RV) over a recent past period is able to show better forecasting performance in the near future. Statistically, the future forecasting performance of the HAR-CJ relative to the HAR-RV for time $t+1:t+h$ is defined as

$$fp_{t+1:t+h} = I\left((RV_{t+1:t+h} - \widehat{RV}^{CJ}_{t+1:t+h})^2 - (RV_{t+1:t+h} - \widehat{RV}^{RV}_{t+1:t+h})^2 < 0 \right),$$

where $I(\cdot)$ refers to an indicator function, $\widehat{RV}^{CJ}_{i+1:i+h}$ and $\widehat{RV}^{RV}_{i+1:i+h}$ are the HAR-CJ and HAR-RV forecasts, respectively, for $RV_{i+1:i+h}$. Similarly, the past forecasting performance of the HAR-CJ relative to the HAR-RV for time $t+1:t+h$ is defined as

$$pp_{t+1:t+h}(k) = I\left(\sum_{i=t-h-k+1}^{t-h} (RV_{i+1:i+h} - \widehat{RV}^{CJ}_{i+1:i+h})^2 - \sum_{i=t-h-k+1}^{t-h} (RV_{i+1:i+h} - \widehat{RV}^{RV}_{i+1:i+h})^2 < 0 \right),$$

where *k* refers to the length of the look-back period. In a statistical sense, the cross-sectional dependence between $pp_{t+1:t+h}(k)$ and $fp_{t+1:t+h}$ equates with the existence of MoJ. The chi-square statistic of Pesaran and Timmermann (2009) is used to test the null hypothesis that $pp_{t+1:t+h}(k)$ and $fp_{t+1:t+h}$ are not cross-sectional dependent in the presence of serial dependencies for each series itself against the alternative hypothesis that the two time series are cross-sectional dependent. The corresponding *p* values are reported.

HAR-CJ model with and the HAR-RV model without jump information. In other words, we statistically document that a better past performance of the HAR-CJ model is always associated with a better future forecasting performance. Of course, the existence of the MoJ phenomenon is the fundamental driving force of our MoJ method.

To further provide a visual impression of the model switching between the HAR-CJ and HAR-RV models, we plot the dynamics of the model selection between the two models in Fig. 1 for various forecast horizons, which is based on the case of $k=5$. We summarize this graphical device with two major observations. First, we observe that our MoJ strategy sometimes selects the HAR-RV model and sometimes selects the HAR-CJ model. This implies that the jump component cannot always provide useful information for forecasting the oil futures market RV during the entire out-of-sample period; however, it contains useful forecasting information during part of the out-of-sample period. That is, the HAR-RV and HAR-CJ models cannot outperform each other completely. This evidences the potential success of our MoJ strategy (which uses both the HAR-RV and HAR-CJ models) in selecting the relatively good model. Second and more importantly, the model selection between the HAR-RV and HAR-CJ is highly persistent. That is, we observe the momentum of model selection. To be precise, the MoJ strategy persistently selects one model between the HAR-RV and HAR-CJ for a relatively long period. Therefore, the model that shows a relatively good past forecasting performance tends to yield a relatively good future performance. This appealing selection pattern contributes to the success of the MoJ strategy.

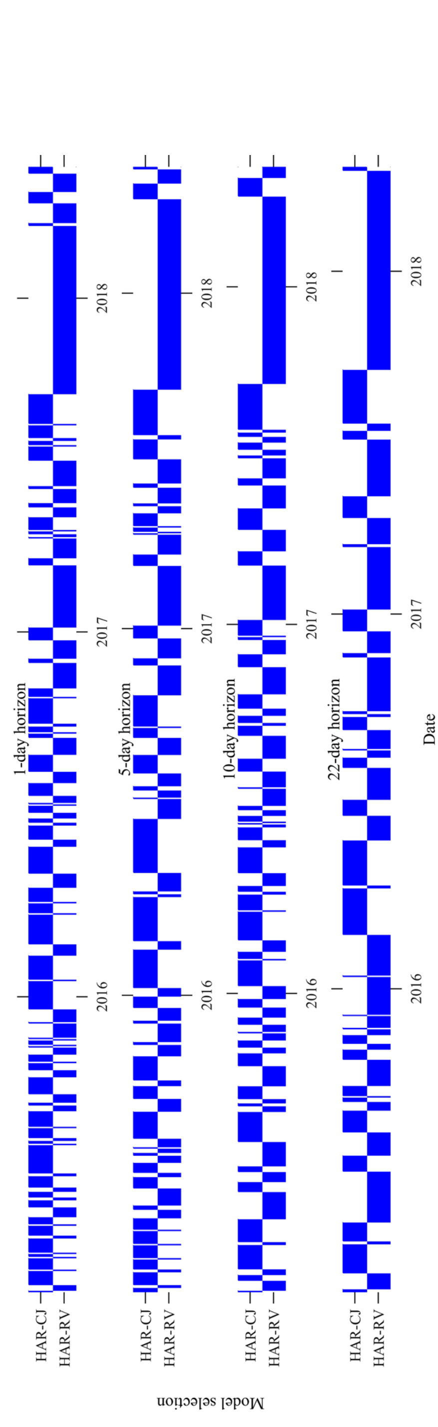

Fig. 1 Model selection between the HAR-RV and HAR-CJ during the out-of-sample forecasting period. The HAR-RV or HAR-CJ model is selected by the MoJ strategy when its past forecasting performance is relatively good. A selected (discarded) model at each forecast step is shown as a blue (white) bar

Table 4 MCS out-of-sample forecasting test using the HAR-J jump model

Models	QLIKE	MSE	MAE	MSPE	MAPE	MSE-LOG
Panel A: 1-day horizon						
HAR-RV	0.079	0.026	0.000	0.032	0.000	0.000
HAR-J	**0.566**	**0.434**	**1.000**	**0.736**	**0.309**	**0.171**
Mean	**0.875**	0.063	0.004	0.075	0.000	0.080
MoJ	**1.000**	**1.000**	**0.852**	**1.000**	**1.000**	**1.000**
Panel B: 5-day horizon						
HAR-RV	0.014	**0.125**	0.011	0.038	0.010	0.013
HAR-J	0.032	**1.000**	**0.855**	0.064	0.044	0.022
Mean	0.014	**0.235**	**0.211**	0.038	0.010	0.013
MoJ	**1.000**	**0.754**	**1.000**	**1.000**	**1.000**	**1.000**
Panel C: 10-day horizon						
HAR-RV	0.003	0.021	0.003	0.005	0.001	0.002
HAR-J	0.005	**0.250**	0.013	0.005	0.001	0.002
Mean	0.001	0.021	0.001	0.005	0.000	0.001
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Panel D: 22-day horizon						
HAR-RV	0.001	0.071	0.000	0.000	0.002	0.004
HAR-J	0.000	0.071	0.000	0.000	0.000	0.000
Mean	0.000	0.031	0.000	0.000	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

This table provides the MCS *p* values of the four used models. Panels A, B, C, and D report the corresponding results for 1-day, 5-day, 10-day, and 22-day horizons, respectively. The mean combination uses the equally weighted average (that is, simple mean) of the individual HAR-RV and HAR-J forecasts, while our MoJ strategy switches between the HAR-RV and HAR-J forecasts based on their relatively past forecasting performance. The past forecasting performance is evaluated by a 5-day look-back period. The six considered loss functions are QLIKE, MSE, MAE, MSPE, MAPE, and MSE-LOG. Bold numbers highlight important instances in which the corresponding model falls into the MCS with the 10% significance level

Robustness checks

The primary forecasting performance reported in Table 2 shows that our out-of-sample results are robust to a multitude of loss functions and forecast horizons. Furthermore, we provide ten robustness tests in this section. These robustness tests alleviate the concern about data mining, thus validating our results.

Alternative jump models

We use the prevailing HAR-CJ model to incorporate the jump component. However, there are many other jump models used to predict financial market RV. To alleviate the concern about the arbitrary use of the jump model, we additionally consider two popular forecasting models that also use the jump component. The first new jump model is termed HAR-J, which is also originated by Andersen et al. (2007). The HAR-J model specification is given by

$$RV_{t+1:t+h} = \varphi_0 + \varphi_d RV_t + \varphi_w RV_{t-4:t} + \varphi_m RV_{t-21:t} + \varphi_J J_t + \omega_{t+1:t+h}, \tag{23}$$

The second new jump model is termed HAR-TCJ, which is pioneered by Corsi et al. (2010). To detect jumps, Corsi et al. (2010) depend not only on a new test statistic, termed *C-Tz*, but also on the threshold bipower variation (TBPV) to calculate the threshold jump measure as $TJ_t = I(C{-}Tz_t > \Theta_\alpha)(RV_t - TBPV_t)$. The continuous

Table 5 MCS out-of-sample forecasting test using the HAR-TCJ jump model

Models	QLIKE	MSE	MAE	MSPE	MAPE	MSE-LOG
Panel A: 1-day horizon						
HAR-RV	**0.489**	0.046	0.001	0.010	0.000	0.001
HAR-TCJ	**0.291**	**1.000**	**1.000**	0.010	0.040	0.019
Mean	**0.568**	**0.640**	0.014	0.010	0.000	0.002
MoJ	**1.000**	**0.640**	**0.247**	**1.000**	**1.000**	**1.000**
Panel B: 5-day horizon						
HAR-RV	0.000	0.003	0.000	0.004	0.000	0.000
HAR-TCJ	0.000	**1.000**	0.039	0.027	0.000	0.000
Mean	0.000	**0.188**	0.000	0.004	0.000	0.000
MoJ	**1.000**	**0.998**	**1.000**	**1.000**	**1.000**	**1.000**
Panel C: 10-day horizon						
HAR-RV	0.000	0.000	0.000	0.005	0.000	0.000
HAR-TCJ	0.000	**0.209**	0.000	0.009	0.000	0.000
Mean	0.000	0.004	0.000	0.005	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Panel D: 22-day horizon						
HAR-RV	0.000	0.001	0.000	0.006	0.000	0.000
HAR-TCJ	0.000	0.055	0.000	0.006	0.000	0.000
Mean	0.000	0.001	0.000	0.001	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

This table provides the MCS *p* values of the four used models. Panels A, B, C, and D report the corresponding results for 1-day, 5-day, 10-day, and 22-day horizons, respectively. The mean combination uses the equally weighted average (that is, simple mean) of the individual HAR-RV and HAR-TCJ forecasts, while our MoJ strategy switches between the HAR-RV and HAR-TCJ forecasts based on their relatively past forecasting performance. The past forecasting performance is evaluated by a 5-day look-back period. The six considered loss functions are QLIKE, MSE, MAE, MSPE, MAPE, and MSE-LOG. Bold numbers highlight important instances in which the corresponding model falls into the MCS with the 10% significance level

counterpart is calculated as $TC_t = RV_t - TJ_t$. Consequently, the HAR-TCJ model takes the regression form of

$$RV_{t+1:t+h} = \varphi_0 + \varphi_{tcd} TC_t + \varphi_{tcw} TC_{t-4:t} + \varphi_{tcm} TC_{t-21:t} + \varphi_{TJ} TJ_t + \omega_{t+1:t+h}. \quad (24)$$

Thus far, we have the HAR-RV benchmark model and two new jump models of the HAR-J and HAR-TCJ. We can generate two new MoJ and mean combination forecasts by separately using the two new jump models. Tables 4 and 5 provide the forecasting results when we use the HAR-J and HAR-TCJ jump models, respectively. In the HAR-J case, we find that the MoJ model falls into the MCS at the 10% significance level for all the 24 cases (6 different loss functions and 4 different forecast horizons). Furthermore, our MoJ model generates the greatest MCS p value (i.e., 1) for 22 out of the 24 cases. The HAR-J model generates the greatest MCS p value for only 2 cases and survives in the MCS test for several cases. However, the HAR-J model as well as the other two competing models fails to remain in the MCS for most of the 24 cases. The results suggest that the MoJ method has significantly stronger predictive power than the competing methods for most of the cases (i.e., most loss functions and forecast horizons). We observe similar results when using the HAR-TCJ as the jump model. Thus, our forecasting results are robust to different jump models.

In this subsection, we use two alternative jump measures to explore the robustness of the MoJ strategy. Additionally, sampling frequency is an important factor in detecting jumps

Table 6 MCS out-of-sample forecasting test based on various look-back periods

Models	QLIKE	MSE	MAE	MSPE	MAPE	MSE-LOG
Panel A: 1-day horizon						
HAR-RV	0.071	0.086	0.003	0.019	0.001	0.000
HAR-CJ	0.016	**0.192**	0.077	0.019	0.001	0.000
Mean	0.071	**0.114**	0.004	0.001	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Panel B: 5-day horizon						
HAR-RV	0.001	0.004	0.000	0.000	0.002	0.000
HAR-CJ	0.001	0.004	0.000	0.000	0.000	0.000
Mean	0.000	0.004	0.000	0.000	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Panel C: 10-day horizon						
HAR-RV	0.000	0.001	0.000	0.000	0.002	0.002
HAR-CJ	0.000	0.001	0.000	0.000	0.000	0.000
Mean	0.000	0.001	0.000	0.000	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Panel D: 22-day horizon						
HAR-RV	0.001	0.000	0.001	0.039	0.003	0.003
HAR-CJ	0.000	0.000	0.000	0.000	0.000	0.000
Mean	0.000	0.000	0.000	0.000	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

This table provides the MCS *p* values of the four used models. Panels A, B, C, and D report the corresponding results for 1-day, 5-day, 10-day, and 22-day horizons, respectively. The mean combination uses the equally weighted average (that is, simple mean) of the individual HAR-RV and HAR-CJ forecasts, while our MoJ strategy switches between the HAR-RV and HAR-CJ forecasts based on their relatively past forecasting performance. The past forecasting performance is separately evaluated by 1-day, 5-day, 10-day, and 22-day look-back periods. Then, the final MoJ forecast reported in this table is equal to the simple mean of the four individual MoJ forecasts based on the four different look-back periods. The six considered loss functions are QLIKE, MSE, MAE, MSPE, MAPE, and MSE-LOG. Bold numbers highlight important instances in which the corresponding model falls into the MCS with the 10% significance level

(see, e.g., Lyócsa et al. 2020; Maneesoonthorn et al. 2020). However, we leave it for future research due to data constraints.

Alternative look-back periods

Our MoJ strategy relies on recent past forecasting performance that is assessed based on a look-back period whose length is defined as k. The previously reported forecasting results of the MoJ model are based on a weekly ($k=5$) look-back period. In this subsection, we follow Wang et al. (2018) and use a few reasonable look-back periods to generate an average MoJ forecast. More specifically, we consider daily (1-day), weekly (5-day), biweekly (10-day), and monthly (22-day) look-back periods, and thereby generate $\widehat{RV}^{MoJ}_{t+1:t+h}(k)$ for $k=1$, 5, 10, 22. The average MoJ forecast is then given by

$$\widehat{RV}^{MoJ-AVG}_{t+1:t+h} = \frac{1}{4} \sum_{k\in\{1,5,10,22\}} \widehat{RV}^{MoJ}_{t+1:t+h}(k). \tag{25}$$

Table 6 presents the corresponding MCS results when the average MoJ strategy is used. Expectedly, the MoJ strategy continues to deliver the highest MCS p values for all the 24 cases and, of course, consistently falls into the MCS at the 10% significance level. In contrast, the competing models of the HAR-CJ and mean combination survive in the MCS for

Table 7 MCS out-of-sample test using the logarithmic HAR-RV and HAR-CJ models

Models	QLIKE	MSE	MAE	MSPE	MAPE	MSE-LOG
Panel A: 1-day horizon						
HAR-RV	**0.127**	0.038	0.000	0.073	0.000	0.002
HAR-CJ	**0.788**	**0.827**	**0.851**	**1.000**	**0.349**	**0.387**
Mean	**1.000**	**0.739**	0.046	**0.208**	0.001	0.056
MoJ	**0.788**	**1.000**	**1.000**	**0.856**	**1.000**	**1.000**
Panel B: 5-day horizon						
HAR-RV	0.000	0.002	0.000	0.000	0.000	0.000
HAR-CJ	0.000	0.033	0.002	0.001	0.000	0.000
Mean	0.000	0.003	0.000	0.000	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Panel C: 10-day horizon						
HAR-RV	0.000	0.001	0.000	0.000	0.000	0.000
HAR-CJ	0.000	0.003	0.000	0.000	0.000	0.000
Mean	0.000	0.001	0.000	0.000	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Panel D: 22-day horizon						
HAR-RV	0.000	0.000	0.000	0.000	0.000	0.000
HAR-CJ	0.000	0.000	0.000	0.000	0.000	0.000
Mean	0.000	0.000	0.000	0.000	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

This table provides the MCS *p* values of the four used models. Particularly, the HAR-RV and HAR-CJ models used in this table are cast in logarithmic form. Panels A, B, C, and D report the corresponding results for 1-day, 5-day, 10-day, and 22-day horizons, respectively. The mean combination uses the equally weighted average (that is, simple mean) of the individual HAR-RV and HAR-CJ forecasts, while our MoJ strategy switches between the HAR-RV and HAR-CJ forecasts based on their relatively past forecasting performance. The past forecasting performance is evaluated by a 5-day look-back period. The six considered loss functions are QLIKE, MSE, MAE, MSPE, MAPE, and MSE-LOG. Bold numbers highlight important instances in which the corresponding model falls into the MCS with the 10% significance level

only 1 out of 24 cases (i.e., the 1-day forecast horizon and MSE loss function). The robust results suggest that our MoJ strategy consistently outperforms the competing models. Finally, it should be noted that the MoJ forecasting results based on the individual look-back periods (i.e., $k=1$, 10, 22) are tabulated in our Additional file 1. The forecasting results are robust to alternative look-back periods.

Nonlinear HAR models

A commonly considered issue for model specification is whether to use linear or nonlinear HAR models. With this in mind, we further employ nonlinear HAR models that are cast in logarithmic and standard deviation forms (see also Andersen et al. 2007; Corsi et al. 2010; Prokopczuk et al. 2016). Mathematically, the logarithmic HAR-RV and HAR-CJ models are given by

$$\ln(RV_{t+1:t+h}) = \varphi_0 + \varphi_d \ln(RV_t) + \varphi_w \ln(RV_{t-4:t}) + \varphi_m \ln(RV_{t-21:t}) + \omega_{t+1:t+h} \tag{26}$$

and

$$\begin{aligned} \ln(RV_{t+1:t+h}) = {} & \varphi_0 + \varphi_{cd} \ln(C_t) + \varphi_{cw} \ln(C_{t-4:t}) + \varphi_{cm} \ln(C_{t-21:t}) \\ & + \varphi_{sd} \ln(SJ_t + 1) + \varphi_{sw} \ln(SJ_{t-4:t} + 1) + \varphi_{sm} \ln(SJ_{t-21:t} + 1) + \omega_{t+1:t+h}, \end{aligned} \tag{27}$$

Table 8 MCS out-of-sample test using the square-root HAR-RV and HAR-CJ models

Models	QLIKE	MSE	MAE	MSPE	MAPE	MSE-LOG
Panel A: 1-day horizon						
HAR-RV	**0.118**	0.040	0.000	0.055	0.000	0.001
HAR-CJ	**0.436**	**0.317**	**0.604**	**0.669**	0.098	0.061
Mean	**0.436**	**0.139**	0.004	0.091	0.000	0.001
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Panel B: 5-day horizon						
HAR-RV	0.000	0.004	0.000	0.003	0.000	0.000
HAR-CJ	0.005	0.032	0.011	0.026	0.000	0.001
Mean	0.000	0.004	0.000	0.003	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Panel C: 10-day horizon						
HAR-RV	0.000	0.001	0.000	0.004	0.000	0.000
HAR-CJ	0.005	0.041	0.002	0.013	0.000	0.001
Mean	0.000	0.001	0.000	0.004	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Panel D: 22-day horizon						
HAR-RV	0.000	0.004	0.000	0.002	0.000	0.000
HAR-CJ	0.000	0.004	0.000	0.002	0.000	0.000
Mean	0.000	0.001	0.000	0.001	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

This table provides the MCS *p* values of the four used models. Particularly, the HAR-RV and HAR-CJ models used in this table are cast in standard deviation form. Panels A, B, C, and D report the corresponding results for 1-day, 5-day, 10-day, and 22-day horizons, respectively. The mean combination uses the equally weighted average (that is, simple mean) of the individual HAR-RV and HAR-CJ forecasts, while our MoJ strategy switches between the HAR-RV and HAR-CJ forecasts based on their relatively past forecasting performance. The past forecasting performance is evaluated by a 5-day look-back period. The six considered loss functions are QLIKE, MSE, MAE, MSPE, MAPE, and MSE-LOG. Bold numbers highlight important instances in which the corresponding model falls into the MCS with the 10% significance level

respectively. The square-root counterparts are expressed as

$$RV_{t+1:t+h}^{1/2} = \varphi_0 + \varphi_d RV_t^{1/2} + \varphi_w RV_{t-4:t}^{1/2} + \varphi_m RV_{t-21:t}^{1/2} + \omega_{t+1:t+h} \tag{28}$$

and

$$\begin{aligned} RV_{t+1:t+h}^{1/2} =& \varphi_0 + \varphi_{cd} C_t^{1/2} + \varphi_{cw} C_{t-4:t}^{1/2} + \varphi_{cm} C_{t-21:t}^{1/2} \\ &+ \varphi_{sd} SJ_t^{1/2} + \varphi_{sw} SJ_{t-4:t}^{1/2} + \varphi_{sm} SJ_{t-21:t}^{1/2} + \omega_{t+1:t+h}, \end{aligned} \tag{29}$$

respectively.

Tables 7 and 8 present the corresponding MCS results when we employ the logarithmic and square-root HAR models, respectively. In the logarithmic version, our MoJ strategy generates the highest *p* values for 22 out of 24 cases and falls into the MCS at the 10% significance level across all the 24 cases. Moreover, we observe a better forecasting performance of the MoJ strategy than the competing models in the square-root version. Overall, the MoJ strategy still shows a significantly stronger predictive ability than the competing models when nonlinear HAR models are employed. Our forecasting results are thus robust to the use of linear and nonlinear HAR models.

MIDAS model

A growing number of studies rely on the MIDAS regression model to forecast financial market volatility (see, e.g., Ghysels et al. 2006, 2007; Forsberg and Ghysels 2007; Santos and Ziegelmann 2014; Ma et al. 2019). The HAR model imposes constant weights on lagged RVs, while the MIDAS model allows a more flexible weighting scheme. In this sense, the HAR model appears to be a special case of the MIDAS model. Therefore, we further examine the forecasting ability of the MoJ strategy based on the MIDAS model. Specifically, we use our MoJ strategy to switch between the MIDAS-RV and MIDAS-CJ models.

The MIDAS-RV model can be shown as

$$RV_{t+1:t+h} = \mu + \beta \sum_{k=0}^{k^{\max}-1} b(k, \theta_1, \theta_2) RV_{t-k} + \omega_{t+1:t+h}, \tag{30}$$

where $k^{\max}$ refers to the maximal lag length of the included RVs and the weighting function, $b(k, \theta_1, \theta_2)$, provides the lag coefficients of lagged RVs. Consistent with the previously used HAR models, we use $k^{\max} = 22$.[4] The weighting function, $b(k, \theta_1, \theta_2)$, is given by

$$b(k, \theta_1, \theta_2) = \frac{g\left(\frac{k}{k^{\max}}, \theta_1, \theta_2\right)}{\sum_{j=1}^{k^{\max}} g\left(\frac{j}{k^{\max}}, \theta_1, \theta_2\right)}, \tag{31}$$

where $g(x, y, z) = x^{y-1}(1-x)^{z-1} \big/ f(y, z)$ and $f(y, z) = \Gamma(y)\Gamma(z) \big/ \Gamma(y+z)$. It should be noted that the weighting scheme always delivers positive weights, which ensures that the RV forecasts are positive. We refer the reader to Ghysels et al. (2007) for more details regarding the weighting scheme.

Consistent with Santos and Ziegelmann (2014) and Ma et al. (2019), the MIDAS-CJ model can be shown as

$$RV_{t+1:t+h} = \mu + \beta_1 \sum_{k=0}^{k^{\max}-1} b(k, \theta_1^C, \theta_2^C) C_{t-k} + \beta_2 \sum_{k=0}^{k^{\max}-1} b(k, \theta_1^{SJ}, \theta_2^{SJ}) SJ_{t-k} + \omega_{t+1:t+h}. \tag{32}$$

We provide the forecasting results in Table 9 when using the MIDAS-RV and MIDAS-CJ models to replace the HAR-RV and HAR-CJ models, respectively. The MoJ model consistently falls into the MCS at the 10% significance level for all the cases. Conversely, the competing models of the MIDAS-RV, MIDAS-CJ, and mean combination approach hardly fall into the MCS. This indicates that the MoJ model exhibits substantially better predictive ability than the competing models. Thus, the forecasting results remain robust to the alternative use of the MIDAS or HAR frameworks.

[4] Forsberg and Ghysels (2007), Santos and Ziegelmann (2014), and Ma et al. (2019) use lag length in the range between 40 and 60. The forecasting results are qualitatively similar for alternative lag lengths, which is provided in the Additional file 1.

Table 9 MCS out-of-sample test based on the MIDAS-RV and MIDAS-CJ models

Models	QLIKE	MSE	MAE	MSPE	MAPE	MSE-LOG
Panel A: 1-day horizon						
MIDAS-RV	**0.259**	0.067	0.002	0.002	0.000	0.000
MIDAS-CJ	0.027	**0.152**	0.024	0.002	0.000	0.000
Mean	**0.259**	**0.137**	0.002	0.002	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Panel B: 5-day horizon						
MIDAS-RV	0.000	0.002	0.000	0.003	0.000	0.000
MIDAS-CJ	0.000	0.002	0.000	0.000	0.000	0.000
Mean	0.000	0.002	0.000	0.000	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Panel C: 10-day horizon						
MIDAS-RV	0.000	0.000	0.000	0.011	0.000	0.000
MIDAS-CJ	0.000	0.000	0.000	0.000	0.000	0.000
Mean	0.000	0.000	0.000	0.000	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Panel D: 22-day horizon						
MIDAS-RV	0.000	0.012	0.000	0.028	0.000	0.001
MIDAS-CJ	0.000	0.000	0.000	0.000	0.000	0.000
Mean	0.000	0.000	0.000	0.000	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

This table provides the MCS *p* values of the four used models. Particularly, we use the MIDAS-RV and MIDAS-CJ models to replace the HAR-RV and HAR-CJ models, respectively. Panels A, B, C, and D report the corresponding results for 1-day, 5-day, 10-day, and 22-day horizons, respectively. The mean combination uses the equally weighted average (that is, simple mean) of the individual MIDAS-RV and MIDAS-CJ forecasts, while our MoJ strategy switches between the MIDAS-RV and MIDAS-CJ forecasts based on their relatively past forecasting performance. The past forecasting performance is evaluated by a 5-day look-back period. The six considered loss functions are QLIKE, MSE, MAE, MSPE, MAPE, and MSE-LOG. Bold numbers highlight important instances in which the corresponding model falls into the MCS with the 10% significance level

Alternative volatility estimators

It is commonly known that the actual measure of asset price volatility is unobservable. With this in mind, we additionally employ another widely used volatility measure, termed realized kernel (RK), which is originally proposed by Barndorff-Nielsen et al. (2008). RK has the appealing property that it is not affected by market microstructure noise. Statistically, the RK is calculated as

$$RK_t = \sum_{p=-P}^{P} k(\frac{p}{P+1})\gamma_p, \tag{33}$$

where

$$\gamma_p = \sum_{j=|p|+1}^{N} r_{t,j} r_{t,j-|p|} \tag{34}$$

and $k(x)$ refers to the Parzen kernel function. For more details, please refer to Barndorff-Nielsen et al. (2009).

Table 10 presents the corresponding out-of-sample results when we use the RK to predict oil futures market volatility in all the four models used. The MoJ strategy continues to exhibit substantially stronger forecasting power than the competing models.

Table 10 MCS out-of-sample forecasting test based the volatility measure of realized kernel

Models	QLIKE	MSE	MAE	MSPE	MAPE	MSE-LOG
Panel A: 1-day horizon						
HAR-RK	0.016	0.030	0.003	0.001	0.000	0.000
HAR-CJ	**0.253**	**1.000**	**0.258**	0.001	0.000	0.000
Mean	**0.253**	**0.170**	0.042	0.001	0.000	0.000
MoJ	**1.000**	**0.489**	**1.000**	**1.000**	**1.000**	**1.000**
Panel B: 5-day horizon						
HAR-RK	0.000	0.012	0.000	0.000	0.000	0.000
HAR-CJ	0.000	0.012	0.000	0.000	0.000	0.000
Mean	0.000	0.012	0.000	0.000	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Panel C: 10-day horizon						
HAR-RK	0.000	0.003	0.000	0.000	0.001	0.001
HAR-CJ	0.000	0.003	0.000	0.000	0.000	0.000
Mean	0.000	0.003	0.000	0.000	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Panel D: 22-day horizon						
HAR-RK	0.000	0.003	0.000	0.015	0.000	0.000
HAR-CJ	0.000	0.032	0.000	0.006	0.000	0.000
Mean	0.000	0.004	0.000	0.006	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

This table provides the MCS *p* values of the four used models. The volatility estimator to forecast in this table is the realized kernel (RK) instead of the realized variance (RV). Panels A, B, C, and D report the corresponding results for 1-day, 5-day, 10-day, and 22-day horizons, respectively. The mean combination uses the equally weighted average (that is, simple mean) of the individual HAR-RK and HAR-CJ forecasts, while our MoJ strategy switches between the HAR-RK and HAR-CJ forecasts based on their relatively past forecasting performance. The past forecasting performance is evaluated by a 5-day look-back period. The six considered loss functions are QLIKE, MSE, MAE, MSPE, MAPE, and MSE-LOG. Bold numbers highlight important instances in which the corresponding model falls into the MCS with the 10% significance level

Specifically, the MoJ model generates the highest MCS *p* values (i.e., 1) for 23 out of 24 cases and falls into the MCS for all the 24 cases. In contrast, the three competing models enter the MCS for no more than 3 cases. The RK evidence suggests that our out-of-sample results are robust to the use of alternative volatility estimators.

Other robustness tests

In this subsection, we provide additional robustness tests from five different aspects. First, we consider various significance levels for the jump detection test. The previously reported forecasting results are based on the 0.5% significance level, which is suggested by Andersen et al. (2007). For the consideration of robustness, we follow Corsi et al. (2010) and Prokopczuk et al. (2016) and additionally use the 1% and 0.1% significance levels for the jump detection test.

Second, Rossi and Inoue (2012) and Inoue et al. (2017) both present that out-of-sample forecasting performance is often influenced by the choice of forecasting window size. Therefore, we further employ two window sizes. Specifically, the first 1019 and 619 observations are used as the initial training samples, while the rest of the observations are in the out-of-sample period.

Third, while the rolling estimation window can mitigate the impact of structural breaks (see, e.g., Clark and McCracken 2009), the rolling scheme also discards initial

Table 11 MCS out-of-sample forecasting test for stock market

Models	QLIKE	MSE	MAE	MSPE	MAPE	MSE-LOG
Panel A: 1-day horizon						
HAR-RV	**1.000**	**0.690**	0.096	0.002	0.000	0.012
HAR-CJ	0.047	**0.751**	0.096	0.014	0.012	0.012
Mean	**0.906**	**1.000**	0.096	0.001	0.000	0.012
MoJ	**0.945**	**0.690**	**1.000**	**1.000**	**1.000**	**1.000**
Panel B: 5-day horizon						
HAR-RV	0.000	0.010	0.000	0.000	0.000	0.000
HAR-CJ	0.000	0.010	0.001	0.000	0.002	0.001
Mean	0.000	0.068	0.001	0.000	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Panel C: 10-day horizon						
HAR-RV	0.000	0.004	0.000	0.000	0.000	0.000
HAR-CJ	0.000	0.004	0.000	0.014	0.000	0.000
Mean	0.000	0.005	0.000	0.000	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Panel D: 22-day horizon						
HAR-RV	0.000	0.005	0.000	0.000	0.000	0.000
HAR-CJ	0.000	0.000	0.000	0.000	0.000	0.000
Mean	0.000	0.000	0.000	0.000	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

This table provides the MCS *p* values of the four used models. In particular, we forecast the stock market (i.e., the S&P 500 Index) volatility instead of the oil futures market volatility in this table. Panels A, B, C, and D report the corresponding results for 1-day, 5-day, 10-day, and 22-day horizons, respectively. The mean combination uses the equally weighted average (that is, simple mean) of the individual HAR-RV and HAR-CJ forecasts, while our MoJ strategy switches between the HAR-RV and HAR-CJ forecasts based on their relatively past forecasting performance. The past forecasting performance is evaluated by a 5-day look-back period. The six considered loss functions are QLIKE, MSE, MAE, MSPE, MAPE, and MSE-LOG. Bold numbers highlight important instances in which the corresponding model falls into the MCS with the 10% significance level

observations when the window rolls forward. The discarded observations perhaps contain useful information for forecasting future RV. For this consideration, we alternatively employ the recursive (expanding) estimation window to obtain out-of-sample RV forecasts.

Fourth, as shown in Eq. (14), the past predictive performance is evaluated based on the MSE form. For the consideration of robustness, we separately use the QLIKE and MAE forms to evaluate past forecasting performance.

Fifth, in addition to the MCS test, we consider another popular test, the Diebold–Mariano (DM) test (Diebold and Mariano 1995).[5] Based on the DM test, we compare each forecasting model with the HAR-RV benchmark to investigate whether the MoJ model shows the highest forecasting gains.

For the sake of brevity, all the results for the five different types of robustness tests are tabulated in the Additional file 1. In short, we find robust results that the MoJ strategy consistently surpasses the competing models.

[5] Specifically, we use the modified Diebold–Mariano test proposed by Harvey et al. (1997), which considers potential contemporaneous correlation between forecast errors, as well as autocorrelation and heavy-tailed distributions for forecast errors.

Extension and application

Stock market evidence

Stock market volatility forecasting is equally important and popular in the academic literature (see, e.g., Wang et al. 2016; Clements and Liao 2017; Zhang et al. 2020). Therefore, the question arises as to whether our MoJ strategy is useful for forecasting stock market volatility. Thus, we extend the MoJ strategy to the stock market. Specifically, we use the MoJ strategy as well as the three competing models to produce the RV forecasts of the S&P 500 Index. The entire sample period spans January 2, 2009 to December 30, 2016, which includes 2001 observations. The first 1201 observations are used as the initial training sample, while the rest 800 observations are in the out-of-sample period. A rolling estimation window is employed to produce the stock market RV forecasts.

Table 11 presents the corresponding MCS results for forecasting stock market volatility. We find similar results for the crude oil futures and stock markets. The MoJ model is in the MCS at the 10% significance level for all the 24 cases. Furthermore, the MoJ model yields the highest MCS value and significantly beats the competing model in 22 out of 24 cases. This evidence suggests that the MoJ strategy is also feasible and useful for forecasting stock market volatility.

Portfolio performance

We examine the economic value of the RV forecasts of the MoJ strategy and the competing models in an asset allocation exercise. Following Bollerslev et al. (2018), we assume that a mean–variance investor will allocate her wealth between a risky asset (i.e., WTI futures) and a risk-free asset (i.e., risk-free bills) with a constant Sharpe ratio. Compared to the related approaches (see, e.g., Fleming et al. 2001, 2003; Campbell and Thompson 2008; Rapach et al. 2010; Zhang et al. 2019c), which rely on both the return and volatility forecasts, the portfolio exercise proposed by Bollerslev et al. (2018) depends exclusively on the volatility forecast. This is appealing since forecasting returns is notoriously difficult (see, e.g., Campbell and Thompson 2008; Welch and Goyal 2008; Rapach et al. 2010).

In the portfolio exercise of Bollerslev et al. (2018), the investor invests a fraction, w_t, of her current (i.e., time t) portfolio in WTI futures with a return of r_{t+1} and the rest in risk-free bills with a return of r^f_t. Correspondingly, her future portfolio return becomes $r^p_{t+1} = w_t r_{t+1} + (1 - w_t) r^f_t = w_t r^e_{t+1} + r^f_t$, where $r^e_{t+1} = r_{t+1} - r^f_t$. Excluding the constant terms, which depend only on time-t variables, we can approximate the expected utility as

$$U(w_t) = w_t E_t(r^e_{t+1}) - \frac{\gamma}{2} w_t^2 Var(r^e_{t+1}), \tag{35}$$

where γ denotes the investor's risk aversion coefficient and $Var(r^e_{t+1}) = E_t(RV_{t+1})$. To focus exclusively on volatility forecasting, Bollerslev et al. (2018) assume that the conditional Sharpe ratio, which is written as $SR \equiv E_t(r^e_{t+1}) \big/ \sqrt{E_t(RV_{t+1})}$, is constant. Consequently, the expected utility can be rewritten as

$$U(w_t) = w_t SR \sqrt{E_t(RV_{t+1})} - \frac{\gamma}{2} w_t^2 E_t(RV_{t+1}), \tag{36}$$

Table 12 Portfolio performance

Models	$h=1$	$h=5$	$h=10$	$h=22$
Static	2.839	3.100	3.176	3.234
HAR-RV	3.502	3.719	3.756	3.783
HAR-CJ	3.487	3.702	3.741	3.764
Mean	3.504	3.719	3.755	3.776
MoJ	3.505	3.749	3.783	3.799

This table provides the portfolio performance evaluated by the average realized utility (in percentage). In this portfolio exercise, we assume that a mean–variance investor will allocate her portfolio between WTI futures and risk-free bills by using different RV forecasts, which is based on a constant Sharpe ratio of 0.4 and a risk aversion coefficient of 2. The static model simply takes the rolling sample average of in-sample RVs as the RV forecast. The mean combination uses the equally weighted average (that is, simple mean) of the individual HAR-RV and HAR-CJ forecasts, while our MoJ strategy switches between the HAR-RV and HAR-CJ forecasts based on their relatively past forecasting performance. The past forecasting performance is evaluated by a 5-day look-back period

which simply relies on the portfolio weight, w_t, and the expected RV, $E_t(RV_{t+1})$. Maximizing the expected utility given by Eq. (36), we can obtain the optimal portfolio weight for oil futures as follows.

$$w_t^* = \frac{SR/\gamma}{\sqrt{E_t(RV_{t+1})}}. \tag{37}$$

Given Eq. (37), we can derive that the conditional standard deviation of the portfolio's risky part is $\sqrt{Var(w_t^* r_{t+1}^e)} = SR/\gamma$. This indicates that the investor targets an optimal volatility of SR/γ. When the forecast of $\sqrt{E_t(RV_{t+1})}$ is greater than the "risk target" of SR/γ (that is, $w_t^* < 1$), the investor only allocates part of her wealth to the risky asset of oil futures. On the contrary, when the predicted volatility risk of $\sqrt{E_t(RV_{t+1})}$ is smaller than this risk target (that is, $w_t^* > 1$), the investor must rely on leverage to achieve her target.

Substituting Eq. (37) into (36), we can realize an expected utility from the optimally targeted portfolio as follows.

$$U(w_t^*) = \frac{SR^2}{2\gamma}. \tag{38}$$

However, in practice, $E_t(RV_{t+1})$ is not available. Using the RV forecast of $\widehat{RV}_{t+1}$ for day $t+1$, we can realize an expected utility of

$$U(\widehat{RV}_{t+1}) = \frac{SR^2}{\gamma}\left(\frac{\sqrt{RV_{t+1}}}{\sqrt{\widehat{RV}_{t+1}}} - \frac{1}{2}\frac{RV_{t+1}}{\widehat{RV}_{t+1}}\right). \tag{39}$$

We empirically report the average utility during the out-of-sample forecasting period. Accordingly, the reported average utility is calculated as

$$\overline{U}(\widehat{RV}) = \frac{1}{q}\sum_{t=R}^{R+P-1}\frac{SR^2}{\gamma}\left(\frac{\sqrt{RV_{t+1}}}{\sqrt{\widehat{RV}_{t+1}}} - \frac{1}{2}\frac{RV_{t+1}}{\widehat{RV}_{t+1}}\right), \tag{40}$$

where R and P denote the lengths of in- and out-of-sample periods, respectively. Following Bollerslev et al. (2018), we set the risk aversion coefficient and annualized Sharpe ratio to be $\gamma = 2$ and $SR = 0.4$, respectively.[6] Consequently, $U(w_t^*) = 4\%$, implying that the investor is happy to pay 4% of her wealth to obtain the w_t^* portfolio of the risky asset rather than to exclusively invest in risk-free bills.

Table 12 reports the portfolio performance evaluated based on the average realized utility. Particularly, we follow Bollerslev et al. (2018) and additionally use a static model as the benchmark in portfolio performance. Under the assumption that volatility risk is constant, the static model simply takes the rolling sample average of in-sample RVs as the RV forecast.[7] Two important findings emerge. First, all the four forecasting models deliver substantially higher realized utilities than the static model. The utility gains, which are computed as the difference between the realized utilities of our previously used forecasting models and that of the static model, are mostly above 50 basis points. The realized utility can be regarded as the portfolio's profit (or return) adjusted by volatility risk. Therefore, this evidence means that the investor is happy to forego 50 basis points to have access to the four econometric models rather than to simply use the static model. Second and more importantly, the utility gain from the MoJ model is the largest of the gains from all the four models used. This means that the investor is happy to pay more fees to use the MoJ model than to use the other three competing models. In other words, our MoJ model can deliver the largest economic gains for the assumed investor in a real portfolio exercise.

A comparison with alternative strategies

In this subsection, we extend our competing models by considering two strands of similar forecasting strategies. First, the MoJ approach also works like the discount mean squared prediction error (DMSPE) combination method. Both the MoJ and DMSPE strategies depend on the past predictive performances of individual models. The difference is that the MoJ model imposes a binary weight of 0 or 1 on the individual HAR-RV and HAR-CJ models, while the DMSPE method produces a continuous weight between 0 and 1 for the two models based on their past forecasting performances. Several recent studies explicitly show that the DMSPE method can improve out-of-sample forecast accuracy (see, e.g., Rapach et al. 2010; Wang et al. 2019; Dai et al. 2021). Statistically, the DMSPE weight for individual forecast i on days $t+1:t+h$ is given by

$$\omega_{i,t+1:t+h} = \phi_{i,t+1:t+h}^{-1} \Big/ \sum_{\ell \in \{\mathrm{RV},\,\mathrm{CJ}\}} \phi_{\ell,t+1:t+h}^{-1}, \quad \text{for} \quad i = \mathrm{RV}, \mathrm{CJ}, \tag{41}$$

where

[6] That is, the annualized volatility target equals 20%. Other reasonable values of SR and γ will not influence the comparison results of the average realized utility among different RV forecasting models.

[7] Bollerslev et al. (2018) use the expanding sample average of RVs, while we use the rolling sample average, which is consistent with the forecasting scheme of our previously used models. Moreover, the portfolio results are similar when we employ the expanding sample average.

Table 13 A comparison of alternative forecasting models

Models	QLIKE	MSE	MAE	MSPE	MAPE	MSE-LOG
Panel A: 1-day horizon						
DMSPE(1)	0.002	**0.212**	0.000	0.000	0.000	0.000
DMSPE(0.9)	**0.267**	**0.487**	0.000	0.026	0.000	0.001
Ridge	0.002	**0.533**	**0.684**	0.026	0.003	0.002
Lasso	0.002	**0.533**	0.000	0.000	0.000	0.000
Elastic net	0.002	**0.487**	0.000	0.000	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Panel B: 5-day horizon						
DMSPE(1)	0.000	0.011	0.000	0.000	0.000	0.000
DMSPE(0.9)	0.000	0.016	0.000	0.000	0.000	0.000
Ridge	0.000	0.016	0.000	0.000	0.000	0.000
Lasso	0.000	0.016	0.000	0.000	0.000	0.000
Elastic net	0.000	0.016	0.000	0.000	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Panel C: 10-day horizon						
DMSPE(1)	0.000	0.003	0.000	0.000	0.000	0.000
DMSPE(0.9)	0.000	0.003	0.000	0.000	0.000	0.000
Ridge	0.000	0.003	0.000	0.000	0.000	0.000
Lasso	0.000	0.004	0.001	0.000	0.000	0.000
Elastic net	0.000	0.003	0.000	0.000	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Panel D: 22-day horizon						
DMSPE(1)	0.000	0.000	0.000	0.000	0.000	0.000
DMSPE(0.9)	0.000	0.000	0.000	0.000	0.000	0.000
Ridge	0.000	0.000	0.000	0.000	0.000	0.000
Lasso	0.000	0.000	0.000	0.000	0.000	0.000
Elastic net	0.000	0.000	0.000	0.000	0.000	0.000
MoJ	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

This table provides the MCS *p* values for the MoJ and alternative models. Panels A, B, C, and D report the corresponding results for 1-day, 5-day, 10-day, and 22-day horizons, respectively. The combination models of DMSPE(1) and DMSPE(0.9) calculate the weights of the individual HAR-RV and HAR-CJ forecasts based on their past forecasting performance, while our MoJ strategy switches between the HAR-RV and HAR-CJ forecasts based on their relatively past forecasting performance. The shrinkage methods of the ridge, lasso, and elastic net are performed based on the HAR-CJ model. In the MoJ model, the past forecasting performance is evaluated by a 5-day look-back period. The six considered loss functions are QLIKE, MSE, MAE, MSPE, MAPE, and MSE-LOG. Bold numbers highlight important instances in which the corresponding model falls into the MCS with the 10% significance level

$$\phi_{i,t+1:t+h} = \sum_{s=R}^{t-h} \delta^{t-h-s} \left(RV_{s+1:s+h} - \widehat{RV}^{i}_{s+1:s+h} \right)^2, \tag{42}$$

R denotes the length of the initial training period, and δ refers to the discount factor. We follow Rapach et al. (2010) and rely on two values of δ, that is, 1 and 0.9. We then obtain two corresponding approaches, dubbed DMSPE(1) and DMSPE(0.9).

Second, the MoJ strategy is similar to shrinkage approaches, which push the coefficients of jump components toward 0 when the recent past performance of the HAR-CJ is worse than that of the HAR-RV model. Therefore, we compare the forecast accuracy of the MoJ with those of alternative shrinkage methods, including the ridge, lasso, and elastic net.[8] These shrinkage methods have been demonstrated to perform well in forecasting asset price returns and volatilities (see, e.g., Li et al. 2015; Li and Tsiakas 2017; Zhang et al. 2019c).

The corresponding comparison results are shown in Table 13. Our MoJ strategy consistently produces the highest MCS *p*-values (i.e., 1) for all loss functions and forecast horizons. In contrast, the alternative competing models of DMSPE(1), DMSPE(0.9), ridge, lasso, and elastic net produce substantially lower MCS *p* values, most of which are smaller than 0.1, implying that the corresponding models cannot enter the MCS at the 10% significance level. We thus conclude that the MoJ strategy also outperforms the two DMSPE and three shrinkage methods in the out-of-sample forecasting test.

Conclusion

The jump component is not informative for forecasting oil futures market volatility (Prokopczuk et al. 2016). To improve the efficiency of using jump information, we propose the MoJ strategy, which switches between the HAR-RV model without jumps and the HAR-CJ model incorporating jump information based on their relative past forecasting performances. The MoJ approach depends on the momentum of the jump model's predictive ability. More precisely, the MoJ implies that a good past predictive performance of the jump model (i.e., the HAR-CJ model) typically delivers a good future predictive performance.

Empirically, the in-sample estimation results suggest that the jump component does not contain a powerful explanatory ability for future oil futures RV. This evidence implies that a straightforward approach of using the jump model alone is unlikely to be feasible. In addition, based on six prevailing loss functions, the MCS out-of-sample forecasting test provides convincing evidence that the MoJ strategy outperforms the HAR-RV, HAR-CJ, and the mean combination of the two HAR models. Furthermore, we document the existence of the MoJ in forecasting the oil futures market RV; that is, the stronger predictive power of the jump model is persistent. This lays the foundation for the success of the MoJ strategy.

The results of the superiority of our MoJ model are found to be robust to various forecast horizons (ranging from 1-day to 22-day horizons), alternative jump models, various look-back periods, alternative volatility estimators, the use of the HAR or MIDAS framework, the use of linear and nonlinear HAR models, different forecasting windows, and many other robustness perspectives. In addition, we extend the MoJ model to the prediction of stock market RV and obtain consistent forecasting results. Finally, in a portfolio exercise, we explore the economic significance of the RV forecasts of the MoJ strategy and the competing models. A mean–variance investor who targets a constant Sharpe ratio can realize sizeable utility gains relying on the MoJ-based RV forecasts to allocate her portfolio.

Our empirical findings have some useful implications for the participants in the crude oil market. For example, volatility forecasting is commonly used in the applications of

[8] For ridge, we use the Hoerl et al. (1975) algorithm to ascertain the reasonable value of the biasing parameter. In terms of lasso and elastic net, we follow Zhang et al. (2019b) and Zhang et al. (2019c) to estimate the shrinkage parameters. We refer to these references for further details about the estimations of ridge, lasso, and elastic net.

asset allocation and risk management. While directly using jumps is likely to be useless when forecasting crude oil price volatility, the participants still need to indirectly consider jump information. Our MoJ approach is a successful example that can enhance the predictability of oil market volatility and thereby improve the performance of asset allocation and risk management. There are, of course, many other ways to address jump information. This is an interesting field for future research. Machine learning is probably a better choice for discovering more useful information in jumps.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40854-022-00360-7.

Additional file 1. Internet Appendix.

Acknowledgements
We are grateful to Gang Kou (the editor) and six anonymous referees for insightful comments that significantly improved the paper. We also thank seminar participants at Nanjing University of Science and Technology for helpful comments and suggestions.

Author contributions
YZ: conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing—original draft, visualization, funding acquisition. YW: conceptualization, methodology, software, investigation, writing—review and editing, supervision, funding acquisition. FM: methodology, formal analysis, investigation, data curation, visualization. YW: investigation, writing—review and editing, supervision, funding acquisition. All authors read and approved the final manuscript.

Funding
Yaojie Zhang acknowledges the financial support from the National Natural Science Foundation of China (72001110), the Fundamental Research Funds for the Central Universities (30919013232), the Research Fund for Young Teachers of School of Economics and Management, NJUST (JGQN2009). Yudong Wang acknowledges the financial support from the National Natural Science Foundation of China (72071114). Feng Ma acknowledges the support from the National Natural Science Foundation of China (71701170, 72071162). Yu Wei acknowledges the support from the National Natural Science Foundation of China (71671145, 71971191), Science and technology innovation team of Yunnan provincial.

Availability of data and materials
The data used by this paper are available from the corresponding author on reasonable request.

Declaration

Competing interests
The authors declare they have no conflict of interest.

Author details
[1]School of Economics and Management, Nanjing University of Science and Technology, Xiaolingwei 200, Xuanwu District, Nanjing 210094, China. [2]School of Economics and Management, Southwest Jiaotong University, Chengdu, China. [3]School of Finance, Yunnan University of Finance and Economics, Kunming, China.

Received: 6 September 2021 Accepted: 24 April 2022
Published online: 06 June 2022

References

Andersen TG, Bollerslev T (1998) Answering the skeptics: yes, standard volatility models do provide accurate forecasts. Int Econ Rev 39:885–905

Andersen TG, Bollerslev T, Diebold FX, Labys P (2003) Modeling and forecasting realized volatility. Econometrica 71:579–625

Andersen TG, Bollerslev T, Diebold FX (2007) Roughing it up: including jump components in the measurement, modeling, and forecasting of return volatility. Rev Econ Stat 89:701–720

Barndorff-Nielsen OE, Shephard N (2002) Estimating quadratic variation using realized variance. J Appl Econom 17:457–477

Barndorff-Nielsen OE, Shephard N (2004) Power and bipower variation with stochastic volatility and jumps. J Financ Econom 2:1–37

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

Barndorff-Nielsen OE, Hansen PR, Lunde A, Shephard N (2009) Realized kernels in practice: trades and quotes. Econom J 12:1–32
Bollerslev T, Hood B, Huss J, Pedersen LH (2018) Risk everywhere: modeling and managing volatility. Rev Financ Stud 31:2729–2773
Bouri E (2019) The effect of jumps in the crude oil market on the sovereign risks of major oil exporters. Risks 7:118
Bouri E, Gupta R (2020) Jumps in energy and non-energy commodities. OPEC Energy Rev 44:91–111
Bouri E, Lei X, Jalkh N, Xu Y, Zhang H (2021) Spillovers in higher moments and jumps across US stock and strategic commodity markets. Resour Policy 72:102060
Buncic D, Gisler KI (2017) The role of jumps and leverage in forecasting volatility in international equity markets. J Int Money Financ 79:1–19
Busetti F, Marcucci J (2013) Comparing forecast accuracy: a Monte Carlo investigation. Int J Forecast 29:13–27
Calzolari G, Halbleib R, Zagidullina A (2021) A latent factor model for forecasting realized variances. J Financ Econom 19:860–909
Campbell JY, Thompson SB (2008) Predicting excess stock returns out of sample: can anything beat the historical average? Rev Financ Stud 21:1509–1531
Clark TE, McCracken MW (2009) Improving forecast accuracy by combining recursive and rolling forecasts. Int Econ Rev 50:363–395
Clements A, Liao Y (2017) Forecasting the variance of stock index returns using jumps and cojumps. Int J Forecast 33:729–742
Corsi F (2009) A simple approximate long-memory model of realized volatility. J Financ Econom 7:174–196
Corsi F, Pirino D, Reno R (2010) Threshold bipower variation and the impact of jumps on volatility forecasting. J Econom 159:276–288
Dai Z, Zhou H, Kang J, Wen F (2021) The skewness of oil price returns and equity premium predictability. Energy Econ 94:105069
Dai Z, Li T, Yang M (2022) Forecasting stock return volatility: the role of shrinkage approaches in a data-rich environment. J Forecast. https://doi.org/10.1002/for.2841
Diebold FX, Mariano RS (1995) Comparing predictive accuracy. J Bus Econ Stat 13:253–263
Duong D, Swanson NR (2015) Empirical evidence on the importance of aggregation, asymmetry, and jumps for volatility prediction. J Econom 187:606–621
Dutta A, Bouri E, Roubaud D (2021) Modelling the volatility of crude oil returns: jumps and volatility forecasts. Int J Financ Econ 26:889–897
Fleming J, Kirby C, Ostdiek B (2001) The economic value of volatility timing. J Finance 56:329–352
Fleming J, Kirby C, Ostdiek B (2003) The economic value of volatility timing using "realized" volatility. J Financ Econ 67:473–509
Forsberg L, Ghysels E (2007) Why do absolute returns predict volatility so well? J Financ Econom 5:31–67
Ghysels E, Santa-Clara P, Valkanov R (2006) Predicting volatility: getting the most out of return data sampled at different frequencies. J Econom 131:59–95
Ghysels E, Sinko A, Valkanov R (2007) MIDAS regressions: further results and new directions. Econom Rev 26:53–90
Giacomini R, White H (2006) Tests of conditional predictive ability. Econometrica 74:1545–1578
Gong X, Lin B (2018) Structural breaks and volatility forecasting in the copper futures market. J Futures Mark 38:290–339
Gronwald M (2012) A characterization of oil price behavior—evidence from jump models. Energy Econ 34:1310–1317
Hansen PR, Lunde A, Nason JM (2011) The model confidence set. Econometrica 79:453–497
Harvey D, Leybourne S, Newbold P (1997) Testing the equality of prediction mean squared errors. Int J Forecast 13:281–291
Haugom E, Langeland H, Molnár P, Westgaard S (2014) Forecasting volatility of the US oil market. J Bank Finance 47:1–14
Hoerl AE, Kannard RW, Baldwin KF (1975) Ridge regression: some simulations. Commun Stat Theory Methods 4:105–123
Inoue A, Jin L, Rossi B (2017) Rolling window selection for out-of-sample forecasting with time-varying parameters. J Econom 196:55–67
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 Econom 13:293–341
Liu LY, Patton AJ, Sheppard K (2015) Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. J Econom 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
Lyócsa Š, Molnár P, Plíhal T, Širaňová M (2020) Impact of macroeconomic news, regulation and hacking exchange markets on the volatility of bitcoin. J Econ Dyn Control 119:103980
Ma F, Liao Y, Zhang Y, Cao Y (2019) Harnessing jump component for crude oil volatility forecasting in the presence of extreme shocks. J Empir Finance 52:40–55
Maneesoonthorn W, Martin GM, Forbes CS (2020) High-frequency jump tests: which test should we use? J Econom 219:478–487
Niu Z, Liu Y, Gao W, Zhang H (2021) The role of coronavirus news in the volatility forecasting of crude oil futures markets: evidence from China. Resour Policy 73:102173
Patton AJ (2011) Volatility forecast comparison using imperfect volatility proxies. J Econom 160:246–256
Patton AJ, Sheppard K (2009) Optimal combinations of realised volatility estimators. Int J Forecast 25:218–238
Patton AJ, Sheppard K (2015) Good volatility, bad volatility: signed jumps and the persistence of volatility. Rev Econ Stat 97:683–697
Pesaran MH, Timmermann A (2009) Testing dependence among serially correlated multicategory variables. J Am Stat Assoc 104:325–337
Prokopczuk M, Symeonidis L, Wese Simen C (2016) Do jumps matter for volatility forecasting? Evidence from energy markets. J Futures Mark 36:758–792

Rapach D, Strauss JK, Zhou G (2010) Out-of-sample equity premium prediction: combination forecasts and links to the real economy. Rev Financ Stud 23:821–862
Rossi B, Inoue A (2012) Out-of-sample forecast tests robust to the choice of window size. J Bus Econ Stat 30:432–453
Santos DG, Ziegelmann FA (2014) Volatility forecasting via MIDAS, HAR and their combination: an empirical comparative study for IBOVESPA. J Forecast 33:284–299
Sévi B (2014) Forecasting the volatility of crude oil futures using intraday data. Eur J Oper Res 235:643–659
Stock JH, Watson MW (2004) Combination forecasts of output growth in a seven-country data set. J Forecast 23:405–430
Wang Y, Ma F, Wei Y, Wu C (2016) Forecasting realized volatility in a changing world: a dynamic model averaging approach. J Bank Finance 64:136–149
Wang Y, Liu L, Ma F, Diao X (2018) Momentum of return predictability. J Empir Finance 45:141–156
Wang Y, Pan Z, Liu L, Wu C (2019) Oil price increases and the predictability of equity premium. J Bank Finance 102:43–58
Welch I, Goyal A (2008) A comprehensive look at the empirical performance of equity premium prediction. Rev Financ Stud 21:1455–1508
Wilmot NA, Mason CF (2013) Jump processes in the market for crude oil. Energy J 34:33–48
Yang Y (2004) Combining forecasting procedures: some theoretical results. Econom Theor 20:176–222
Yang C, Gong X, Zhang H (2019) Volatility forecasting of crude oil futures: the role of investor sentiment and leverage effect. Resour Policy 61:548–563
Zhang Y, Ma F, Wang T, Liu L (2019a) Out-of-sample volatility prediction: a new mixed-frequency approach. J Forecast 38:669–680
Zhang Y, Ma F, Wang Y (2019b) Forecasting crude oil prices with a large set of predictors: can LASSO select powerful predictors? J Empir Finance 54:97–117
Zhang Y, Wei Y, Zhang Y, Jin D (2019c) Forecasting oil price volatility: forecast combination versus shrinkage method. Energy Econ 80:423–433
Zhang Y, Ma F, Liao Y (2020) Forecasting global equity market volatilities. Int J Forecast 36:1454–1475
Zhang W, Yan K, Shen D (2021) Can the Baidu Index predict realized volatility in the Chinese stock market? Financ Innov 7:7
Zhang Y, Wahab MIM, Wang Y (2022) Forecasting crude oil market volatility using variable selection and common factor. Int J Forecast. https://doi.org/10.1016/j.ijforecast.2021.12.013

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.