# Forecasting the Realized Volatility of Agricultural Commodity Prices

## Does Sentiment Matter?

Bonato, Matteo; Cepni, Oguzhan; Gupta, Rangan; Pierdzioch, Christian

*Document Version*
Final published version

*Published in:*
Journal of Forecasting

*DOI:*
10.1002/for.3106

*Publication date:*
2024



*Citation for published version (APA):*
Bonato, M., Cepni, O., Gupta, R., & Pierdzioch, C. (2024). Forecasting the Realized Volatility of Agricultural Commodity Prices: Does Sentiment Matter? *Journal of Forecasting*, *43*(6), 2088-2125.
https://doi.org/10.1002/for.3106

Received: 6 July 2023 | Revised: 1 November 2023 | Accepted: 6 February 2024

DOI: 10.1002/for.3106

RESEARCH ARTICLE

WILEY

# Forecasting the realized volatility of agricultural commodity prices: Does sentiment matter?

**Matteo Bonato**[1,2] | **Oguzhan Cepni**[3,4] | **Rangan Gupta**[5] | **Christian Pierdzioch**[6]

[1]Department of Economics and Econometrics, University of Johannesburg, Johannesburg, South Africa

[2]IPAG Business School, Paris, France

[3]Department of Economics, Copenhagen Business School, Frederiksberg, Denmark

[4]Ostim Technical University, Ankara, Turkiye

[5]Department of Economics, University of Pretoria, Pretoria, South Africa

[6]Department of Economics, Helmut Schmidt University, Hamburg, Germany

**Correspondence**

Oguzhan Cepni, Department of Economics, Copenhagen Business School, Porcelæ nshaven 16A, Frederiksberg DK-2000, Denmark; Ostim Technical University, Ankara, Turkiye.

Email: oce.eco@cbs.dk

**Funding information**

The authors declare that they did not receive any funding for this research.

**Abstract**

We analyze the out-of-sample predictive power of sentiment for the realized volatility of agricultural commodity price returns. We use high-frequency intra-day data covering the period from 2009 to 2020 to estimate realized volatility. Our baseline forecasting model is a heterogeneous autoregressive (HAR) model, which we extend to include sentiment. We further enhance this model by incorporating various key realized moments such as leverage, realized skewness, realized kurtosis, realized upside ("good") volatility, realized downside ("bad") volatility, realized jumps, realized upside tail risk, and realized downside tail risk. In order to setup a forecasting model, we use (i) forward and backward stepwise predictor selection and (ii) a model-based averaging algorithm. The forecasting models constructed through these algorithms outperform both the baseline HAR-RV model and the HAR-RV-sentiment model. We conclude that, for the agricultural commodities studied in our research, realized moments play a more significant role in forecasting realized volatility compared to sentiment.

**KEYWORDS**

agricultural commodities, forecasting, realized moments, realized volatility, sentiment

## 1 | INTRODUCTION

In the wake of the Global Financial Crisis (GFC) of 2007–2009, researchers have argued that agricultural (non-energy) commodities markets (just like energy-based commodities markets) have become increasingly financialized (Aït-Youcef, 2019; Bonato, 2019) and that institutional investors have raised their holdings in agricultural commodities relative to traditional assets. Undeniably, accurately modeling and forecasting the volatility of agricultural commodities price movements holds paramount importance for investors. Volatility plays a crucial role in guiding investment and portfolio allocation decisions, risk management, derivatives pricing, and assessing hedging performance. Additionally, because agricultural commodities are vital components of household consumption, the volatility of their prices has substantial implications for food security, particularly affecting the more economically vulnerable segments of the population (Ordu et al., 2018). Hence, it is also of tremendous policy-related value to develop models and derive accurate predictions of food price volatility, so that policymakers can prepare for periods of large price fluctuations and design preventative policies in response (Greb & Prakash, 2017).

The issue of financialization of agricultural commodities markets has led some researchers to question the sole importance of demand and supply factors in explaining fluctuations in food prices, with focus now also on behavioral factors such as investor sentiment, speculation, financial, and macroeconomic uncertainties (see Ji et al., 2020, and Akyildirim et al., 2022, for a detailed discussion), dealing primarily with the first moment of agricultural commodity prices and/or returns (see, e.g., Bahloul, 2018; Balcilar et al., 2022; Borgards & Czudaj, 2022; Mišecka et al., 2019; Xu & Hsu, 2022). In this regard, Bahloul and Bouri (2016) and Bahloul et al. (2018) have provided some in-sample evidence of the role of such predictors in driving volatility in the prices of agricultural commodities as well, using Generalized Autoregressive Conditional Heteroscedasticity (GARCH) and higher order causality-in-quantiles models. We aim to extend this line of research by concentrating on an out-of-sample analysis, which, as Campbell (2008) points out, is the ultimate test of any predictive model. In addition to shedding light on the statistical significance of forecasting food price volatility based on behavioral factors (specifically sentiment), the value of real-time volatility forecasts for agricultural commodities is evident for investors and policymakers. Such forecasts, as opposed to predictions derived based on full-sample predictive analyses, enable more timely and informed decision-making, considering the large price fluctuations experienced by agricultural commodities since 2008, leading to periods of both high and low volatility regimes (Greb & Prakash, 2015).

Against this backdrop, and given that rich information contained in intraday data can produce more accurate estimates and forecasts of daily (realized) volatility (McAleer & Medeiros, 2008), we augment the Heterogeneous Autoregressive (HAR) model developed by Corsi (2009) to include sentiment to forecast the daily realized volatility ($RV$), as computed from 5-min-interval data, of 14 (cocoa, coffee, corn, cotton, feeder cattle, lean hogs, live cattle, orange juice, rough rice, soybean, soybean meal, soybean oil, sugar, and wheat) important agricultural commodities price returns over the period of September 2009 to May 2020. In order to capture investor sentiment, we utilize the daily agricultural commodity-specific Thomson Reuters MarketPsych Indices (TRMI). This unique dataset employs textual analysis techniques to recover investor sentiment from diverse sources, such as news articles, social media content, press releases, and regulatory filings.[1] Consequently, it offers a comprehensive measure of investor sentiment, surpassing the commonly used Google search-based sentiment measures found in earlier literature.

At this stage, it is crucial to underscore the advantage of employing $RV$ as a measure of volatility for agricultural commodities price returns. This advantage stems from the fact that $RV$ is an observable and unconditional metric of "volatility," unlike the latent process underlying GARCH and stochastic volatility (SV) models that have been utilized by many researchers for modeling and forecasting agricultural commodity price volatility (see Degiannakis et al., 2022, and Luo et al., 2019, for an in-depth discussion of this literature).[2] Furthermore, it is noteworthy that the benchmark HAR-RV framework has the ability to capture the long-memory and multi-scaling properties of agricultural commodities price returns volatility, as documented by Gil-Alana et al. (2012) and Živkov et al. (2019), despite its simplistic structure. This characteristic contributes to the model's popularity in the literature. Additionally, the HAR-RV model, which employs $RV$ at different time resolutions to forecast the $RV$ of agricultural commodities price returns, is theoretically grounded in the heterogeneous market hypothesis (Müller et al., 1997). This hypothesis posits that various groups of market participants, differing in their sensitivity to information flows at different time horizons, populate the markets for agricultural commodities.

While our primary focus is on investigating the role of investor sentiment in forecasting the $RV$ of multiple agricultural commodities price returns, it is also essential to compare the performance of sentiment with that of realized moments. The literature on forecasting of agricultural commodities price returns has emphasized the importance of realized moments, such as leverage, realized skewness, realized kurtosis, realized upside volatility, realized downside volatility, realized jumps, realized upside tail risk, and realized downside tail risk (see, e.g., Bonato et al., 2022; Chatziantoniou et al., 2021; Degiannakis et al., 2022; Luo et al., 2019; Marfatia et al., 2022; Shiba et al., 2022; Tian et al., 2017a, 2017b; Yang et al., 2017). Given that we consider several realized moments as candidates for forecasting $RV$, we construct HAR-RV-Sentiment-Moments forecasting models by two alternative approaches: the forward and backward stepwise predictor selection algorithm (see the textbook by Hastie et al., 2009) and a Model-Based Averaging (MOBA) algorithm (Bonato et al., 2023). These two algorithms allow us to effectively incorporate the relevant predictors and enhance forecast accuracy.

Our results suggest that incorporating investor sentiment, as captured by the TRMI, into the forecasting models for the $RV$ of agricultural commodity price returns has limited impact on improving forecast accuracy. While the addition of sentiment slightly enhances forecasts for certain agricultural commodities, the improvement is marginal, and overall, there is no systematic difference between the models with and without sentiment. By contrast, the inclusion of realized moments, such as leverage, skewness, kurtosis, jumps, and tail risks, significantly improves forecast accuracy in comparison to our baseline models without those realized moments. Hence, realized moments provide more direct and quantifiable indicators of market conditions and capture the inherent characteristics of price returns distributions, enhancing the models' ability to capture volatility dynamics.

Furthermore, the results demonstrate that the HAR-RV-Sentiment-Moments model consistently outperforms the baseline models and the HAR-RV-Sentiment model across various agricultural commodities. The inclusion of realized moments in addition to sentiment leads to notable improvements in forecast accuracy, indicating the added value of these moments. The findings hold true across different our two predictor selection algorithms and various robustness checks. Overall, our research highlights the limited informational value of sentiment in forecasting agricultural commodity price volatility and underscores the importance of incorporating realized moments as significant drivers of forecast accuracy.

We contribute to the literature by investigating the role of investor sentiment and realized moments, such as leverage, skewness, kurtosis, jumps, and tail risks, in forecasting the *RV* of agricultural commodities price returns. By doing so, we delve into the domain of behavioral finance, adding another layer of analysis to the traditional economic and financial models. This is in line with the recent concerns raised by Ordu et al. (2018) about the financialization of agricultural markets, as it helps to illuminate the influence of institutional investors' sentiment on the volatility of agricultural commodity prices. Additionally, the finding that sentiment indices offer limited predictive power may help in better understanding the mechanisms of price formation and the complexities of market speculation.

Additionally, our research builds upon prior research conducted by Yang et al. (2017), Luo et al. (2019), and Degiannakis et al. (2022) by demonstrating the superiority of the HAR-RV-Sentiment-Moments model in forecasting agricultural commodity price volatility. It emphasizes the importance of incorporating realized moments in forecasting models to accurately capture the inherent characteristics of price return distributions. Furthermore, our research contributes to the literature on long-memory dynamics in commodity futures volatility by addressing the scarcity of studies on out-of-sample forecasting, despite the extensive history of modeling approaches noted by Giot and Laurent (2003). By providing a comprehensive framework that effectively integrates both sentiment and realized moments, our research offers valuable insights for accurate forecasting and sheds light on the potential for future improvements in risk management techniques.

The remaining sections of our paper are structured as follows: In Section 2, we provide a description of the data used in our study, while Section 3 outlines our forecasting models. The baseline results are presented in Section 4, followed by the presentation of robustness checks results in Section 5. Finally, we conclude our paper in Section 6.

## 2 | DATA

We source our intraday commodity futures prices from this online resource (https://www.kibot.com/). These futures' data maintain a continuous format where, nearing the expiration of a contract; the position is rolled over to the next available contract provided that activity has increased. The data are collected in 5-min increments throughout the day. The dataset encapsulates three categories of agricultural commodities, namely, grains, softs, and livestock. The Food and Agriculture Organization (FAO) of the United Nations (UN) typically identifies these commodities as highly traded within the agricultural sector.[3]

We consider the classical estimator of realized variance, that is, the sum of squared intraday returns (Andersen & Bollerslev, 1998), expressed as

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

where $r_{t,i}$ denotes the intraday $M \times 1$ return vector and $i = 1, \ldots, M$ is the number of intraday returns. In our forecasting models, we mainly study realized volatility as defined as the square root of realized variance (but also report results for realized variance as a robustness check).

As for investor sentiment, Thomson Reuters MarketPsych Indices (TRMIs) provide daily data on agricultural commodities sentiment. The information is collected using advanced artificial intelligence and machine learning methods by MarketPsych Analytics from several news outlets and social media sources such as The Financial Times, The Wall Street Journal, The New York Times, Twitter, Reddit, and Seeking Alpha. The technology quantifies the tone of expressions, resulting in TRMIs that represent various sentiment dimensions instead of a single measure, as opposed to the singular measure found in previous media-based sentiment indices.

As detailed in the TRMI white paper, this database is made up of 31 sentiment indices specifically designed for agricultural commodities. Table A1 lists the definition of the all specific sentiment indices based on TRMI white paper. Each TRMI sentiment theme is composed of a collection of words or tokens that contribute to a particular TRMI. The MarketPsych's dictionaries contain over 60,000 words, expressions, verb forms, and other terminologies. These words are organized into 4,000 topics, tones, and meanings based on their context and then combined to form a TRMI.

The *Buzz* index, which signifies the total sum of all TRMI-contributing words for a particular agricultural commodity on a specific day, is calculated to generate TRMI sentiment data. If $W$ is the set of all words underlying any TRMI sentiment theme, where $a$ is an agricultural commodity and $C(a)$ is the set of all TRMI-contributing words, the Buzz of $a$ can be determined using the equation:

$$\mathrm{Buzz}(a) = \sum_{c \in C(a), w \in W} |\mathrm{Words}_{c,w}|. \tag{2}$$

The TRMI score for each sentiment theme is then calculated by multiplying the scores of the relevant words by the corresponding values of the *Words* variable, which can be either additive (+1) or subtractive (−1). If $W(t)$ denotes the set of all words related to a specific TRMI sentiment theme $t$:

$$I(t,w) = \begin{cases} 1 & \text{if additive} \\ -1 & \text{if subtractive} \end{cases}$$

for all $w \in W(t)$. The TRMI of sentiment topic $t$ for a particular agricultural commodity ($a$) can then be computed as follows:

$$\mathrm{TRMI}_t(a) = \frac{\sum_{c \in C(a), w \in W(t)} (I(t,w) \times \mathrm{Words}_{c,w})}{\mathrm{Buzz(a)}}. \tag{3}$$

TRMIs are multidimensional, and a single news article or social media post may be associated with multiple sentiment categories. For instance, an article discussing risk and expressing anger towards a regulatory body like the SEC might be linked to both "market risk" and "anger" dimensions. It is important to note that the sentiment for each dimension is determined using distinct words or tokens, depending on the sentiment context. Words such as happy, pleased, jubilant, and elated are some of the most frequent tokens contributing to the "joy" sentiment, while irate, angry, mad, and furious are common terms related to the "anger" sentiment. The TRMI methodology assigns weights to various sentiment categories based on the intensity of the words in the text, giving greater weight to strong words and less to weak ones.

These 31 sentiment indices, while distinct, may contain overlapping or correlated information. For instance, the sentiments of "joy" and "optimism" may often trend together, leading to redundancy in the data. Notwithstanding, it is interesting not only to analyze overall sentiment but also the disaggregated sentiment indices. When analyzing the disaggregated sentiment indices (Subsection 5.3), we apply the algorithms we lay out in Section 3 and, in addition, also transform these subindices into a smaller set of uncorrelated variables by applying principal component analysis (PCA). We then extract the first three principal components, which capture the majority of the information contained within the original dataset in a compact and manageable form. This not only simplifies the analysis but can also improve the interpretability of the results by highlighting the most significant patterns or trends in the data.

We plot the realized volatilities for the 14 agricultural commodities in our sample in Figure 1 and overall sentiment in Figure 2. In addition, we summarize the ticker symbols along with the start date and the end date of the respective sample periods for every agricultural commodity in Table 1.

## 3 | METHODS

### 3.1 | Forecasting models

Our baseline forecasting model is the HAR-RV model proposed by Corsi (2009). This model, estimated by the ordinary-least-squares technique, is given by the following equation[4]:

$$RV_{t+h} = \beta_0 + \beta_1 RV_t + \beta_2 RV_{w,t} + \beta_3 RV_{m,t} + u_{t+h}, \tag{4}$$

where $\beta_j, j = 0,..,3$ are the coefficients to be estimated, $u_{t+h}$ denotes a disturbance term, and $RV_{t+h}$ is the average realized volatility over the forecast horizon, $h$, where we set $h = 1,5,22$. The predictors are the daily realized volatility, $RV_t$; the weekly realized volatility, $RV_{t,w}$; and the monthly realized volatility, $RV_{t,m}$. The weekly realized volatility is defined as the average realized volatility from period $t-5$ to period $t-1$, and the monthly realized volatility defined as the average realized volatility from period $t-22$ to period $t-1$.

The next step is to extend the HAR-RV model to include sentiment, denoted as $SENT_t$. This variable represents the comprehensive sentiment index linked to each specific agricultural commodity. Hence, we derive the extended model (also estimated by the ordinary-least squares technique), referred to as the HAR-RV-Sentiment model, as detailed below:

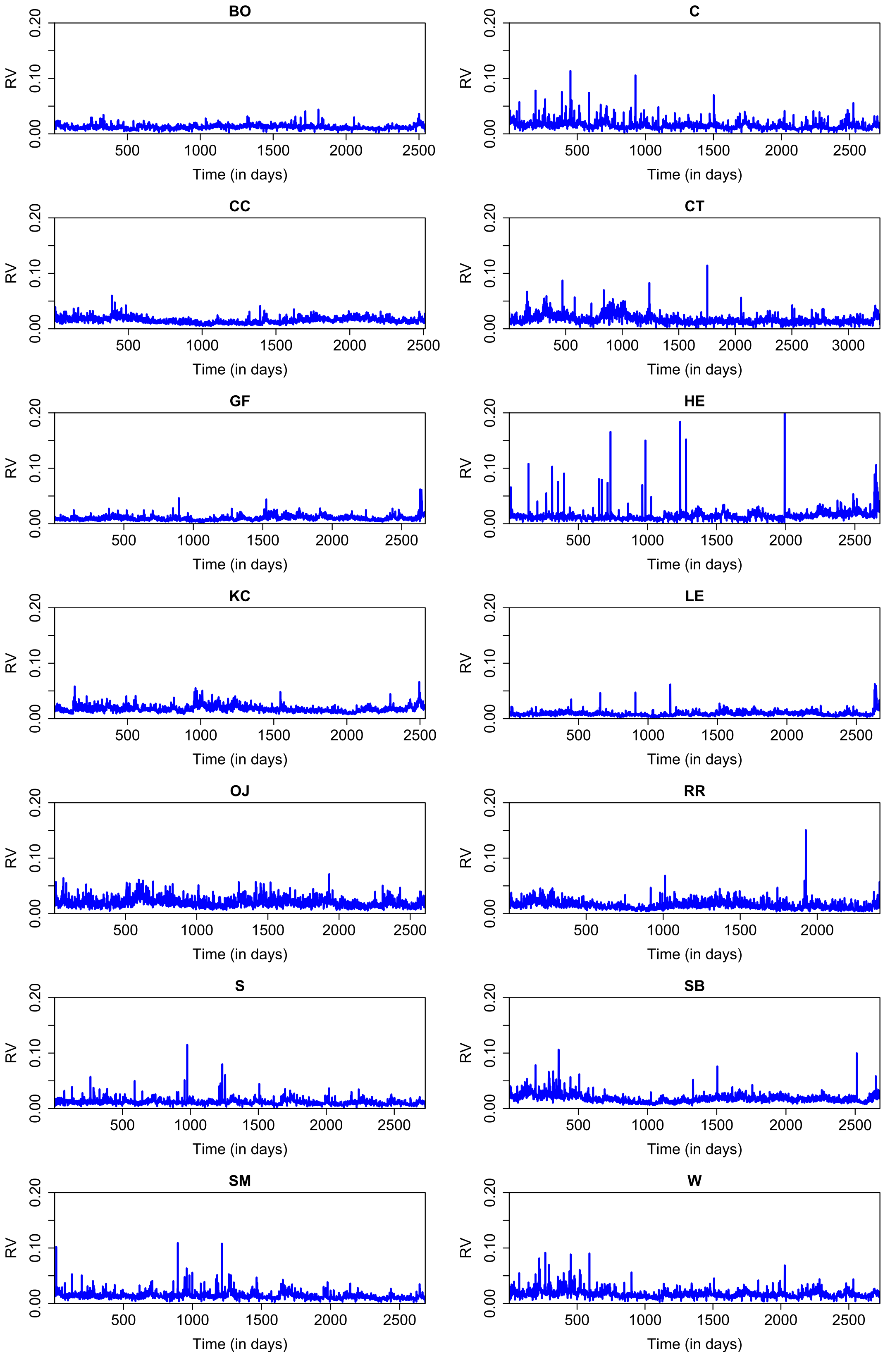


**FIGURE 1** Realized volatility.

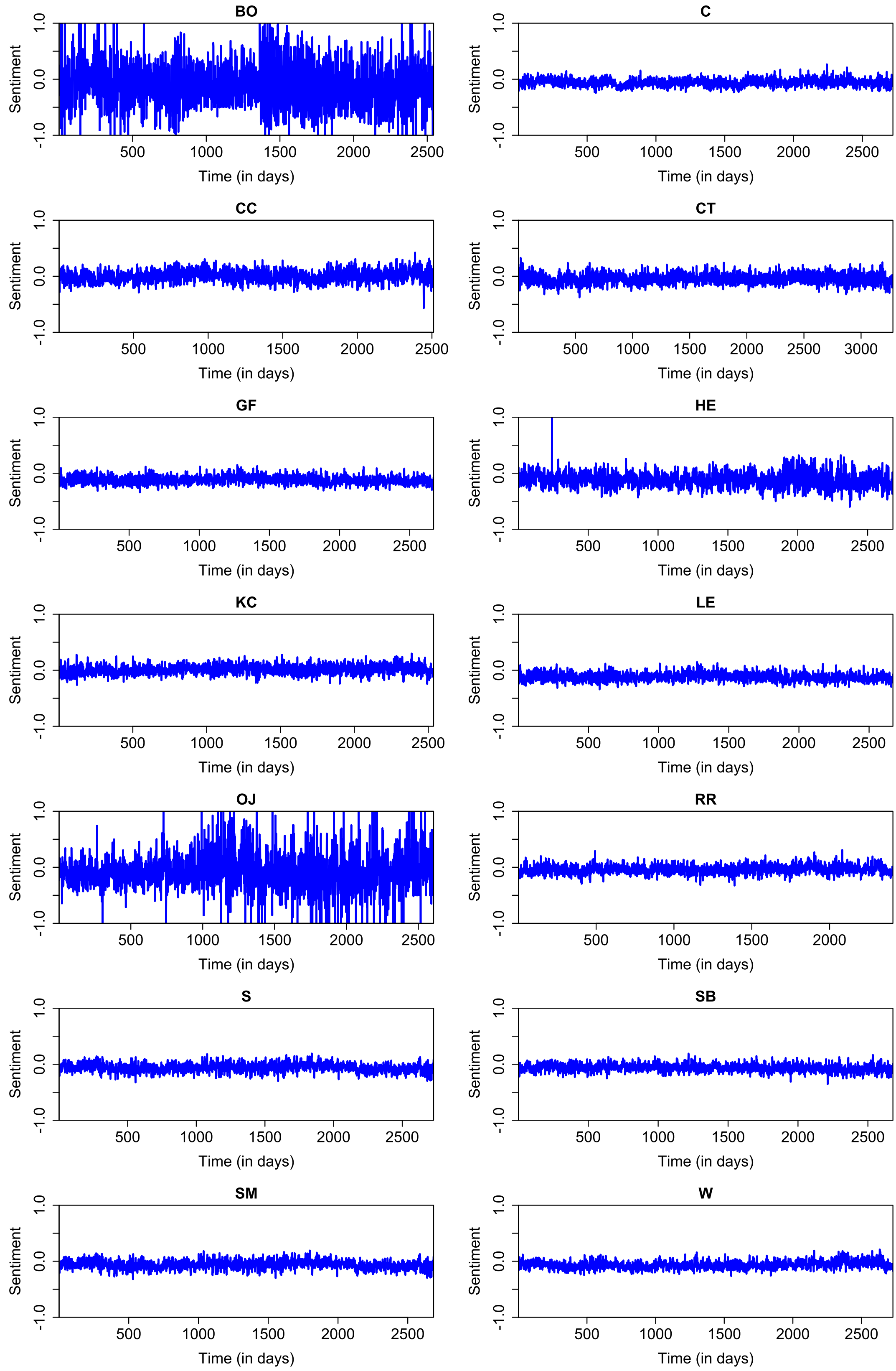


**FIGURE 2** Sentiment.

**TABLE 1** Commodity summary statistics.

| Commodity | Ticker | Sample starts | Sample ends |
|---|---|---|---|
| Soybean oil | BO | 9/28/2009 | 5/18/2020 |
| Corn | C | 9/28/2009 | 5/18/2020 |
| Cocoa | CC | 9/28/2009 | 5/15/2020 |
| Cotton | CT | 5/16/2007 | 5/18/2020 |
| Feeder cattle | GF | 9/28/2009 | 5/15/2020 |
| Lean hogs | HE | 9/28/2009 | 5/15/2020 |
| Coffee | KC | 9/28/2009 | 5/15/2020 |
| Live cattle | LE | 9/28/2009 | 5/15/2020 |
| Orange juice | OJ | 9/28/2009 | 5/15/2020 |
| Rough rice | RR | 9/28/2009 | 5/15/2020 |
| Soybeans | S | 9/28/2009 | 5/18/2020 |
| Sugar | SB | 9/28/2009 | 5/15/2020 |
| Soybean meal | SM | 9/28/2009 | 5/18/2020 |
| Chicago wheat | W | 9/28/2009 | 5/18/2020 |

$$RV_{t+h} = \beta_0 + \beta_1 RV_t + \beta_2 RV_{w,t} + \beta_3 RV_{m,t} + \beta_4 SENT_t + u_{t+h}. \tag{5}$$

Finally, we include various realized moments in our forecasting model. Upon letting the vector, $X_t$, represents the realized moments, we obtain a HAR-RV-Sentiment-Moments model of the following format:

$$RV_{t+h} = \beta_0 + \beta_1 RV_t + \beta_2 RV_{w,t} + \beta_3 RV_{m,t} + \beta_4 SENT_t + \beta_5 X_t + u_{t+h}, \tag{6}$$

where $\beta_5$ denotes an appropriately dimensioned coefficient vector. We consider eight widely studied realized moments: leverage, realized skewness, realized kurtosis, realized upside ("good") volatility, realized downside ("bad") volatility, realized jumps, realized upside tail risk, and realized downside tail risk.

We utilize both a recursive-estimation window and a rolling-estimation window to estimate our various forecasting models. In the case of the recursive-estimation window, we establish a training period to initialize the estimations. Then, we progressively expand the estimation window in a stepwise manner until we reach the end of the sample period. Conversely, for the rolling-estimation window, we begin with a training period and subsequently add (delete) one observation at the end (beginning) of the training period. We continue this process, maintaining a constant-length estimation window, until we reach the end of the sample period. We employ a training period (rolling-estimation window) consisting of 1,000 observations.

Rather than simply including all eight realized moments in the forecasting model, or estimating all $2^8$ possible HAR-RV-Sentiment-Moments, we use two alternative algorithms to setup an "optimal" HAR-RV-Sentiment-Moments forecasting model. Using two alternative algorithms has the advantage that we can compare the results across the two algorithms so as to check the robustness of our results.

## 3.2 | Best stepwise predictor selection algorithm

As our first algorithm, we utilize a forward and backward stepwise predictor selection algorithm (for a detailed explanation, refer to Hastie et al., 2009, chapter 3). To begin, we start with the model provided in Equation (5). With the forward stepwise predictor selection, we estimate forecasting models that incorporate one of the eight realized moments in addition to the predictors mentioned in Equation (5). We store the model that yields the minimum residual sum of squares. Using this model as the new starting point, we estimate all forecasting models that include two realized moments (the one selected in the first step plus one additional realized moment). Again, we store the forecasting model that minimizes the residual sum of squares. We continue this process, gradually adding realized moments, until we reach the complete forecasting model described in Equation (6), which encompasses all realized moments. This stepwise predictor selection procedure provides us with a sequence of forecasting models with increasing complexity, and we must choose the "optimal" forecasting model among them. To determine the best model, we employ three information criteria. Specifically, we select the forecasting model that (i) maximizes the adjusted $R^2$ statistic, (ii) minimizes the Bayesian Information Criterion (BIC), or (iii) minimizes Mallow's CP criterion.

In total, we have a set of five different forecasting models: the baseline HAR-RV model specified in Equation (4), the HAR-RV-Sentiment model presented in Equation (5), and three HAR-RV-Sentiment-Moments models selected using forward stepwise predictor selection based on the adjusted $R^2$ statistic, the BIC, and the CP criterion, respectively. It is worth noting that the first two models are nested versions of the models identified through the forward stepwise predictor selection algorithm.[5]

Moreover, we employ a backward stepwise predictor selection algorithm following the same procedure as the forward stepwise predictor selection algorithm, but now starting from the model featuring all eight realized moments. We iteratively remove realized moments and identify "optimal" HAR-RV-Sentiment-Moments models based on the adjusted $R^2$ statistic, the BIC, and the CP criterion.

## 3.3 | MOBA algorithm

As our second algorithm, we use a variant of the MOBA algorithm recently studied by Bonato et al. (2023). This algorithm is easy to implement and computationally efficient, which is particularly advantageous considering our study involves a total of 14 agricultural commodities.

To explain our variant of the MOBA algorithm, let us denote the length of a rolling estimation window as $R$. We split $R$ into an estimation sample, $E$, and a validation, $V$, sample, both of equal length. The validation sample precedes the estimation sample, $R = V + E$; that is, we use recent data for estimation of the forecasting models and historical data for model validation.[6] To this end, we initiate the following iterative process:

1. We begin with Equation (5) as our "core" model and note that the predictors in the vector $X_t$ can be combined in $2^8$ ways to obtain a HAR-RV-Sentiment-Moments model of the format given in Equation (6).
2. In a first iteration, we sample without replacement from the list of $2^8$ possible combinations of $X_t$. Sampling without replacement ensures that each combination of moment predictors is considered only once to extend the core model and obtain a HAR-RV-Sentiment-Moments model. We estimate Equation (6) using the estimation sample, $E$, and use the validation sample, $V$, to compute predictions of $RV$.
3. We proceed to the next iteration by sampling, again without replacement, a new combination of predictors in $X_t$. We employ the selected predictors to estimate the corresponding new HAR-RV-Sentiment-Moments model on the data in $E$. The validation sample, $V$, provides new predictions of $RV$, which we combine with the predictions from the previous iteration to compute a vector of average predicted realized volatilities.
4. As we proceed from one iteration to the next and average predictions from an increasing number of random HAR-RV-Sentiment-Moments models, we eventually obtain a "stable" vector of average predicted realized volatilities. We terminate the iterative process once the maximum absolute percentage change in the vector of predicted realized volatilities becomes sufficiently small (like Bonato et al., 2022, we use the threshold 0.01 to operationalize "sufficiently small").

In the final step, we utilize the average prediction of realized volatility, computed across all sampled HAR-RV-Sentiment-Moments models, to form an out-of-sample forecast of $RV$.[7] We then advance $R$ by one step in time, create new estimation and validation samples, and restart the MOBA algorithm to generate the next out-of-sample forecast of $RV$. We continue this process until we reach the end of the sample period.

## 3.4 | Forecasting comparison

Equipped with the forecasts of $RV$ generated using both the recursive-estimation window and the rolling-estimation window, along with the two estimation algorithms, we proceed to compare our forecasting models. We employ three evaluation metrics for this purpose: the root-mean-squared forecasting error (RMSFE) statistic, the mean absolute forecast error (MAFE) statistic, and the Clark and West (2007) test to assess the equal predictive performance of nested forecasting models. Specifically, we compare the following nested models:

- The HAR-RV model versus the HAR-RV-Sentiment model,
- the HAR-RV model versus the HAR-RV-Sentiment-Moments model (selected by means of the best stepwise predictor selection algorithm—adjusted $R^2$ statistic, BIC, or CP criterion—or the MOBA algorithm),
- the HAR-RV-Sentiment versus the HAR-RV-Sentiment-Moments model (selected using the best stepwise predictor selection algorithm—adjusted $R^2$ statistic, BIC, or CP criterion—or the MOBA algorithm).

Furthermore, it is interesting to compare the HAR-RV-Sentiment-Moments models selected using the best stepwise predictor selection algorithm with those computed using the MOBA algorithm. It should be noted that the forecasting models selected by the predictor selection and the MOBA algorithms are not necessarily nested versions of each other (but, of course, the HAR-RV model and the HAR-RV-Sentiment model are nested versions of the forecasting models selected by means of the two algorithms). Hence, we compare these models solely based on the RMSFE and MAFE statistics.

To enhance interpretability, we calculate ratios of the RMSFE (MAFE) statistic. A ratio larger than unity indicates that the extended model provides more accurate forecasts than the comparatively more parsimonious nested model. Likewise, the Clark–West test serves as a one-sided test for the null hypothesis of equal predictive accuracy, where the alternative hypothesis is that the larger extended model yields more accurate forecasts than the more parsimonious nested model.

**TABLE 2** RMSFE ratios (forward stepwise predictor selection/recursive-estimation window).

| Commodity/horizon | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| BO/h = 1 | 1.0029 | 1.0126 | 1.0065 | 1.0124 | 1.0097 | 1.0036 | 1.0094 |
| BO/h = 2 | 1.0015 | 1.0163 | 1.0107 | 1.0145 | 1.0148 | 1.0092 | 1.0130 |
| BO/h = 5 | 1.0002 | 1.0309 | 1.0228 | 1.0293 | 1.0307 | 1.0226 | 1.0291 |
| BO/h = 22 | 1.0007 | 1.0217 | 1.0180 | 1.0201 | 1.0210 | 1.0173 | 1.0194 |
| C/h = 1 | 0.9991 | 1.0162 | 1.0126 | 1.0147 | 1.0171 | 1.0134 | 1.0155 |
| C/h = 2 | 0.9993 | 1.0221 | 1.0180 | 1.0208 | 1.0228 | 1.0187 | 1.0215 |
| C/h = 5 | 0.9993 | 1.0271 | 1.0289 | 1.0271 | 1.0279 | 1.0296 | 1.0278 |
| C/h = 22 | 0.9938 | 1.0201 | 1.0194 | 1.0194 | 1.0265 | 1.0258 | 1.0258 |
| CC/h = 1 | 1.0010 | 1.0081 | 1.0069 | 1.0086 | 1.0070 | 1.0059 | 1.0076 |
| CC/h = 2 | 1.0019 | 1.0089 | 1.0051 | 1.0094 | 1.0070 | 1.0032 | 1.0075 |
| CC/h = 5 | 1.0015 | 1.0136 | 1.0078 | 1.0140 | 1.0121 | 1.0063 | 1.0125 |
| CC/h = 22 | 1.0006 | 1.0166 | 1.0109 | 1.0151 | 1.0159 | 1.0103 | 1.0145 |
| CT/h = 1 | 0.9998 | 1.0141 | 1.0143 | 1.0155 | 1.0144 | 1.0145 | 1.0157 |
| CT/h = 2 | 0.9999 | 1.0164 | 1.0140 | 1.0161 | 1.0165 | 1.0140 | 1.0162 |
| CT/h = 5 | 0.9999 | 1.0206 | 1.0176 | 1.0213 | 1.0208 | 1.0178 | 1.0215 |
| CT/h = 22 | 0.9982 | 1.0174 | 1.0159 | 1.0177 | 1.0192 | 1.0178 | 1.0195 |
| GF/h = 1 | 0.9997 | 1.0374 | 1.0407 | 1.0378 | 1.0377 | 1.0410 | 1.0381 |
| GF/h = 2 | 1.0001 | 1.0474 | 1.0487 | 1.0478 | 1.0473 | 1.0486 | 1.0477 |
| GF/h = 5 | 0.9998 | 1.0437 | 1.0359 | 1.0442 | 1.0439 | 1.0361 | 1.0445 |
| GF/h = 22 | 0.9999 | 1.0296 | 1.0257 | 1.0288 | 1.0297 | 1.0259 | 1.0289 |
| HE/h = 1 | 1.0010 | 1.0108 | 1.0091 | 1.0113 | 1.0098 | 1.0080 | 1.0102 |
| HE/h = 2 | 1.0020 | 1.0179 | 1.0130 | 1.0174 | 1.0159 | 1.0110 | 1.0154 |
| HE/h = 5 | 1.0021 | 1.0334 | 1.0299 | 1.0335 | 1.0313 | 1.0278 | 1.0314 |
| HE/h = 22 | 1.0051 | 1.0454 | 1.0438 | 1.0427 | 1.0401 | 1.0385 | 1.0375 |
| KC/h = 1 | 0.9986 | 1.0032 | 1.0021 | 1.0035 | 1.0046 | 1.0036 | 1.0050 |
| KC/h = 2 | 0.9982 | 1.0063 | 1.0006 | 1.0056 | 1.0081 | 1.0024 | 1.0074 |
| KC/h = 5 | 0.9983 | 1.0060 | 1.0071 | 1.0058 | 1.0076 | 1.0088 | 1.0075 |
| KC/h = 22 | 0.9990 | 0.9958 | 1.0011 | 0.9977 | 0.9968 | 1.0020 | 0.9987 |
| LE/h = 1 | 0.9999 | 1.0149 | 1.0118 | 1.0145 | 1.0150 | 1.0119 | 1.0146 |
| LE/h = 2 | 1.0003 | 1.0197 | 1.0158 | 1.0183 | 1.0195 | 1.0155 | 1.0181 |
| LE/h = 5 | 0.9999 | 1.0294 | 1.0272 | 1.0290 | 1.0295 | 1.0273 | 1.0291 |
| LE/h = 22 | 0.9995 | 1.0234 | 1.0228 | 1.0234 | 1.0239 | 1.0233 | 1.0239 |
| OJ/h = 1 | 0.9983 | 0.9996 | 0.9981 | 0.9992 | 1.0013 | 0.9998 | 1.0009 |
| OJ/h = 2 | 0.9987 | 1.0022 | 0.9978 | 1.0023 | 1.0035 | 0.9991 | 1.0036 |
| OJ/h = 5 | 0.9987 | 1.0021 | 1.0017 | 1.0014 | 1.0033 | 1.0030 | 1.0027 |
| OJ/h = 22 | 0.9987 | 0.9944 | 0.9932 | 0.9933 | 0.9957 | 0.9945 | 0.9947 |
| RR/h = 1 | 0.9999 | 1.0092 | 1.0017 | 1.0071 | 1.0093 | 1.0018 | 1.0072 |
| RR/h = 2 | 0.9998 | 1.0164 | 1.0008 | 1.0134 | 1.0165 | 1.0009 | 1.0136 |
| RR/h = 5 | 0.9995 | 1.0182 | 1.0002 | 1.0047 | 1.0187 | 1.0007 | 1.0052 |
| RR/h = 22 | 0.9999 | 1.0183 | 1.0025 | 1.0074 | 1.0184 | 1.0027 | 1.0076 |
| S/h = 1 | 1.0000 | 1.0238 | 1.0196 | 1.0239 | 1.0238 | 1.0196 | 1.0238 |
| S/h = 2 | 0.9995 | 1.0384 | 1.0377 | 1.0390 | 1.0389 | 1.0382 | 1.0394 |
| S/h = 5 | 0.9986 | 1.0532 | 1.0520 | 1.0532 | 1.0546 | 1.0535 | 1.0546 |

TABLE 2 (Continued)

| Commodity/horizon | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| S/h = 22 | 0.9983 | 1.0277 | 1.0267 | 1.0277 | 1.0295 | 1.0284 | 1.0294 |
| SB/h = 1 | 0.9991 | 1.0107 | 1.0171 | 1.0109 | 1.0116 | 1.0181 | 1.0119 |
| SB/h = 2 | 0.9990 | 1.0222 | 1.0221 | 1.0226 | 1.0232 | 1.0230 | 1.0236 |
| SB/h = 5 | 0.9976 | 1.0251 | 1.0273 | 1.0273 | 1.0276 | 1.0299 | 1.0298 |
| SB/h = 22 | 0.9989 | 1.0217 | 1.0195 | 1.0230 | 1.0228 | 1.0206 | 1.0241 |
| SM/h = 1 | 0.9994 | 1.0232 | 1.0193 | 1.0224 | 1.0237 | 1.0199 | 1.0229 |
| SM/h = 2 | 0.9993 | 1.0454 | 1.0422 | 1.0443 | 1.0462 | 1.0429 | 1.0450 |
| SM/h = 5 | 0.9990 | 1.0613 | 1.0565 | 1.0610 | 1.0623 | 1.0575 | 1.0620 |
| SM/h = 22 | 0.9991 | 1.0341 | 1.0306 | 1.0335 | 1.0350 | 1.0315 | 1.0344 |
| W/h = 1 | 0.9989 | 1.0090 | 1.0049 | 1.0104 | 1.0101 | 1.0060 | 1.0116 |
| W/h = 2 | 0.9997 | 1.0209 | 1.0214 | 1.0210 | 1.0212 | 1.0217 | 1.0212 |
| W/h = 5 | 1.0009 | 1.0278 | 1.0238 | 1.0293 | 1.0268 | 1.0229 | 1.0284 |
| W/h = 22 | 0.9981 | 1.0036 | 1.0028 | 1.0041 | 1.0056 | 1.0047 | 1.0060 |

*Note*: Columns: (1) The HAR-RV model versus the HAR-RV-Sentiment model. (2) The HAR-RV model versus the HAR-RV-Sentiment-Moments model (selected by means of the adjusted $R^2$ statistic). (3) The HAR-RV model versus the HAR-RV-Sentiment-Moments model (selected by means of the BIC). (4) The HAR-RV model versus the HAR-RV-Sentiment-Moments model (selected by means of the CP criterion). (5) The HAR-RV-Sentiment model versus the HAR-RV-Sentiment-Moments model (selected by means of the adjusted $R^2$ statistic). (6) The HAR-RV-Sentiment model versus the HAR-RV-Sentiment-Moments model (selected by means of the BIC). (7) The HAR-RV model versus the HAR-RV-Sentiment-Moments model (selected by means of the CP criterion).

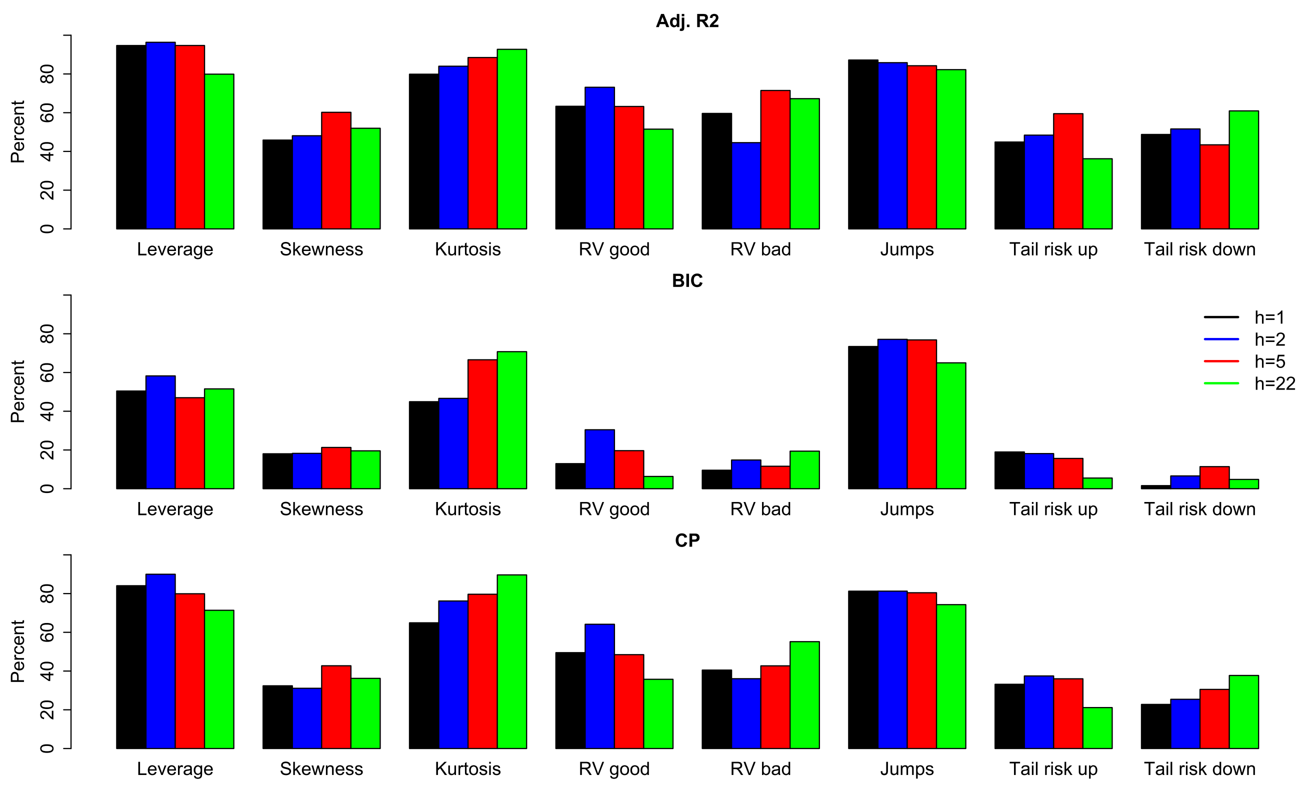


FIGURE 3 Inclusion of realized moments in the forecasting models. Inclusion of realized moments in the forecasting models is averaged across the agricultural commodities. Results are for forward stepwise predictor selection and a recursive-estimation window.

**TABLE 3** MAFE ratios (forward stepwise predictor selection/recursive-estimation window).

| Commodity/horizon | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| BO/h = 1 | 1.0022 | 1.0067 | 0.9988 | 1.0058 | 1.0045 | 0.9966 | 1.0035 |
| BO/h = 2 | 1.0018 | 1.0066 | 1.0022 | 1.0057 | 1.0048 | 1.0004 | 1.0040 |
| BO/h = 5 | 1.0014 | 1.0255 | 1.0159 | 1.0227 | 1.0241 | 1.0145 | 1.0213 |
| BO/h = 22 | 1.0022 | 1.0258 | 1.0236 | 1.0244 | 1.0236 | 1.0213 | 1.0222 |
| C/h = 1 | 1.0001 | 1.0302 | 1.0311 | 1.0288 | 1.0301 | 1.0310 | 1.0287 |
| C/h = 2 | 1.0002 | 1.0315 | 1.0269 | 1.0281 | 1.0313 | 1.0267 | 1.0279 |
| C/h = 5 | 0.9991 | 1.0309 | 1.0301 | 1.0305 | 1.0318 | 1.0310 | 1.0314 |
| C/h = 22 | 0.9971 | 1.0317 | 1.0307 | 1.0310 | 1.0347 | 1.0337 | 1.0340 |
| CC/h = 1 | 1.0010 | 1.0063 | 1.0070 | 1.0073 | 1.0052 | 1.0060 | 1.0063 |
| CC/h = 2 | 1.0016 | 1.0086 | 1.0053 | 1.0085 | 1.0070 | 1.0036 | 1.0069 |
| CC/h = 5 | 0.9992 | 1.0041 | 1.0022 | 1.0039 | 1.0049 | 1.0031 | 1.0047 |
| CC/h = 22 | 0.9994 | 1.0122 | 1.0078 | 1.0108 | 1.0127 | 1.0084 | 1.0114 |
| CT/h = 1 | 0.9997 | 1.0095 | 1.0070 | 1.0107 | 1.0099 | 1.0073 | 1.0110 |
| CT/h = 2 | 0.9998 | 1.0049 | 1.0016 | 1.0046 | 1.0051 | 1.0018 | 1.0047 |
| CT/h = 5 | 0.9997 | 1.0114 | 1.0060 | 1.0113 | 1.0116 | 1.0063 | 1.0116 |
| CT/h = 22 | 0.9997 | 1.0058 | 1.0061 | 1.0066 | 1.0061 | 1.0065 | 1.0069 |
| GF/h = 1 | 0.9995 | 1.0129 | 1.0136 | 1.0119 | 1.0134 | 1.0141 | 1.0124 |
| GF/h = 2 | 0.9986 | 1.0181 | 1.0198 | 1.0178 | 1.0195 | 1.0213 | 1.0193 |
| GF/h = 5 | 0.9981 | 1.0169 | 1.0134 | 1.0173 | 1.0188 | 1.0154 | 1.0192 |
| GF/h = 22 | 0.9989 | 0.9972 | 0.9974 | 0.9963 | 0.9983 | 0.9985 | 0.9974 |
| HE/h = 1 | 1.0020 | 1.0340 | 1.0256 | 1.0326 | 1.0320 | 1.0236 | 1.0306 |
| HE/h = 2 | 1.0059 | 1.0364 | 1.0286 | 1.0361 | 1.0303 | 1.0226 | 1.0300 |
| HE/h = 5 | 1.0047 | 1.0461 | 1.0419 | 1.0464 | 1.0413 | 1.0370 | 1.0416 |
| HE/h = 22 | 1.0082 | 1.0488 | 1.0499 | 1.0471 | 1.0403 | 1.0414 | 1.0386 |
| KC/h = 1 | 0.9948 | 0.9955 | 0.9986 | 0.9961 | 1.0007 | 1.0038 | 1.0013 |
| KC/h = 2 | 0.9960 | 0.9981 | 0.9952 | 0.9979 | 1.0020 | 0.9992 | 1.0019 |
| KC/h = 5 | 0.9944 | 0.9985 | 0.9998 | 0.9986 | 1.0041 | 1.0054 | 1.0042 |
| KC/h = 22 | 0.9959 | 0.9911 | 0.9988 | 0.9923 | 0.9952 | 1.0029 | 0.9964 |
| LE/h = 1 | 0.9993 | 1.0141 | 1.0154 | 1.0138 | 1.0148 | 1.0160 | 1.0144 |
| LE/h = 2 | 0.9993 | 1.0058 | 1.0055 | 1.0064 | 1.0065 | 1.0062 | 1.0071 |
| LE/h = 5 | 0.9995 | 1.0144 | 1.0122 | 1.0143 | 1.0149 | 1.0127 | 1.0148 |
| LE/h = 22 | 0.9986 | 1.0175 | 1.0176 | 1.0170 | 1.0189 | 1.0190 | 1.0184 |
| OJ/h = 1 | 0.9969 | 0.9894 | 0.9935 | 0.9893 | 0.9925 | 0.9967 | 0.9924 |
| OJ/h = 2 | 0.9973 | 0.9949 | 0.9924 | 0.9958 | 0.9975 | 0.9950 | 0.9984 |
| OJ/h = 5 | 0.9984 | 0.9966 | 0.9970 | 0.9965 | 0.9983 | 0.9986 | 0.9982 |
| OJ/h = 22 | 0.9980 | 0.9916 | 0.9902 | 0.9909 | 0.9936 | 0.9921 | 0.9928 |
| RR/h = 1 | 0.9998 | 1.0129 | 1.0094 | 1.0118 | 1.0132 | 1.0097 | 1.0120 |
| RR/h = 2 | 0.9996 | 1.0156 | 1.0052 | 1.0118 | 1.0161 | 1.0056 | 1.0123 |
| RR/h = 5 | 1.0002 | 1.0193 | 1.0068 | 1.0103 | 1.0191 | 1.0065 | 1.0101 |
| RR/h = 22 | 1.0016 | 1.0182 | 1.0107 | 1.0136 | 1.0166 | 1.0091 | 1.0120 |
| S/h = 1 | 0.9990 | 1.0428 | 1.0332 | 1.0414 | 1.0438 | 1.0342 | 1.0424 |
| S/h = 2 | 1.0004 | 1.0468 | 1.0465 | 1.0473 | 1.0464 | 1.0461 | 1.0469 |
| S/h = 5 | 0.9988 | 1.0528 | 1.0522 | 1.0533 | 1.0540 | 1.0534 | 1.0546 |

TABLE 3 (Continued)

| Commodity/horizon | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| S/h = 22 | 0.9979 | 1.0315 | 1.0322 | 1.0320 | 1.0336 | 1.0344 | 1.0342 |
| SB/h = 1 | 1.0019 | 1.0044 | 1.0089 | 1.0033 | 1.0026 | 1.0070 | 1.0014 |
| SB/h = 2 | 1.0034 | 1.0124 | 1.0101 | 1.0120 | 1.0089 | 1.0066 | 1.0086 |
| SB/h = 5 | 1.0002 | 1.0142 | 1.0160 | 1.0150 | 1.0140 | 1.0158 | 1.0148 |
| SB/h = 22 | 0.9986 | 1.0143 | 1.0125 | 1.0164 | 1.0158 | 1.0140 | 1.0179 |
| SM/h = 1 | 0.9990 | 1.0422 | 1.0344 | 1.0413 | 1.0433 | 1.0355 | 1.0424 |
| SM/h = 2 | 1.0004 | 1.0630 | 1.0610 | 1.0618 | 1.0626 | 1.0606 | 1.0614 |
| SM/h = 5 | 0.9992 | 1.0835 | 1.0768 | 1.0838 | 1.0844 | 1.0776 | 1.0846 |
| SM/h = 22 | 0.9997 | 1.0419 | 1.0357 | 1.0415 | 1.0422 | 1.0361 | 1.0419 |
| W/h = 1 | 0.9952 | 1.0043 | 1.0019 | 1.0081 | 1.0091 | 1.0067 | 1.0129 |
| W/h = 2 | 0.9992 | 1.0108 | 1.0112 | 1.0119 | 1.0116 | 1.0120 | 1.0127 |
| W/h = 5 | 1.0026 | 1.0266 | 1.0277 | 1.0278 | 1.0240 | 1.0250 | 1.0252 |
| W/h = 22 | 0.9982 | 1.0101 | 1.0070 | 1.0101 | 1.0119 | 1.0088 | 1.0119 |

*Note*: Columns: (1) The HAR-RV model versus the HAR-RV-Sentiment model. (2) The HAR-RV model versus the HAR-RV-Sentiment-Moments model (selected by means of the adjusted $R^2$ statistic). (3) The HAR-RV model versus the HAR-RV-Sentiment-Moments model (selected by means of the BIC). (4) The HAR-RV model versus the HAR-RV-Sentiment-Moments model (selected by means of the CP criterion). (5) The HAR-RV-Sentiment model versus the HAR-RV-Sentiment-Moments model (selected by means of the adjusted $R^2$ statistic). (6) The HAR-RV-Sentiment model versus the HAR-RV-Sentiment-Moments model (selected by means of the BIC). (7) The HAR-RV model vs. the HAR-RV-Sentiment-Moments model (selected by means of the CP criterion).

## 3.5 | Realized moments

The realized moments of agricultural commodity price returns play a central role in our forecasting experiment. Hence, while the basic steps in the computation of realized moments are well known in the intraday literature, we deem it important to next lay out the technical details of how we compute eight widely used realized moments. Specifically, we study the following realized moments: downward ("bad," *RVB*) and upward ("good," *RVG*) realized variance (that is, the semi-variances), realized skewness (*RSK*), realized kurtosis (*RKU*), realized jumps (*JUMPS*), and realized upside ($TR_u$) and downside tail risks ($TR_d$).[8]

To capture potential sign asymmetries in the realized-variance process, we employ the "good" and "bad" realized variance measures. Following the approach of Barndorff-Nielsen et al. (2010), we estimate good and bad realized variance as follows:

$$RVB_t = \sum_{i=1}^{M} r_{t,i}^2 \, \mathbf{1}_{[(r_{t,i})<0]}, \tag{7}$$

$$RVG_t = \sum_{i=1}^{M} r_{t,i}^2 \, \mathbf{1}_{[(r_{t,i})>0]}, \tag{8}$$

where $\mathbf{1}$ denotes the indicator function.

Next, we introduce *RSK* to capture the asymmetry of the returns distribution and *RKU* to measure the tails of the distribution (see, e.g., Amaya et al., 2015). The calculations for *RSK* and *RKU* are as follows:

$$RSK_t = \frac{\sqrt{M}\sum_{i=1}^{M} r_{(i,t)}^3}{RV_t^{3/2}}, \tag{9}$$

$$RKU_t = \frac{M\sum_{i=1}^{M} r_{(i,t)}^4}{RV_t^2}. \tag{10}$$

Taking into account the fact that realized variance comprises both a discontinuous (jump) component and a permanent component, we make use of the formula developed by Barndorff-Nielsen and Shephard (2004) to identify and quantify the realized jumps. The expression for the realized jumps is as follows:

$$\lim_{M\to\infty} RV_t = \int_{t-1}^{t} \sigma^2(s)ds + \sum_{j=1}^{N_t} k_{t,j}^2, \tag{11}$$

where $N_t$ denotes the number of jumps within day $t$ and $k_{t,j}$ denotes the jump size. It follows from Equation (11) that $RV_t$ is a consistent estimator of the jump contribution plus the integrated variance $\int_{t-1}^{t}\sigma^2(s)ds$.

TABLE 4 Tests (forward stepwise predictor selection/recursive-estimation window).

| Commodity/horizon | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| BO/h = 1 | 0.0026 | 0.0000 | 0.0002 | 0.0000 | 0.0001 | 0.0023 | 0.0000 |
| BO/h = 2 | 0.0279 | 0.0000 | 0.0001 | 0.0000 | 0.0000 | 0.0002 | 0.0001 |
| BO/h = 5 | 0.2399 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| BO/h = 22 | 0.1097 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| C/h = 1 | 0.7707 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| C/h = 2 | 0.4526 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| C/h = 5 | 0.4465 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| C/h = 22 | 0.2698 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| CC/h = 1 | 0.0322 | 0.0000 | 0.0001 | 0.0000 | 0.0001 | 0.0004 | 0.0000 |
| CC/h = 2 | 0.0123 | 0.0000 | 0.0023 | 0.0000 | 0.0001 | 0.0141 | 0.0001 |
| CC/h = 5 | 0.0285 | 0.0000 | 0.0009 | 0.0000 | 0.0000 | 0.0105 | 0.0000 |
| CC/h = 22 | 0.0103 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| CT/h = 1 | 0.9164 | 0.0050 | 0.0044 | 0.0036 | 0.0050 | 0.0043 | 0.0035 |
| CT/h = 2 | 0.5606 | 0.0029 | 0.0045 | 0.0032 | 0.0029 | 0.0045 | 0.0032 |
| CT/h = 5 | 0.5585 | 0.0069 | 0.0116 | 0.0066 | 0.0068 | 0.0113 | 0.0064 |
| CT/h = 22 | 0.7750 | 0.0086 | 0.0155 | 0.0077 | 0.0085 | 0.0156 | 0.0076 |
| GF/h = 1 | 0.4548 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| GF/h = 2 | 0.1562 | 0.0003 | 0.0002 | 0.0004 | 0.0003 | 0.0002 | 0.0004 |
| GF/h = 5 | 0.2682 | 0.0002 | 0.0005 | 0.0002 | 0.0002 | 0.0005 | 0.0002 |
| GF/h = 22 | 0.5157 | 0.0009 | 0.0007 | 0.0010 | 0.0009 | 0.0007 | 0.0010 |
| HE/h = 1 | 0.0708 | 0.0000 | 0.0001 | 0.0000 | 0.0000 | 0.0002 | 0.0000 |
| HE/h = 2 | 0.0181 | 0.0000 | 0.0002 | 0.0000 | 0.0000 | 0.0001 | 0.0000 |
| HE/h = 5 | 0.0447 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| HE/h = 22 | 0.0030 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| KC/h = 1 | 0.4590 | 0.0007 | 0.0030 | 0.0005 | 0.0004 | 0.0018 | 0.0003 |
| KC/h = 2 | 0.3527 | 0.0003 | 0.0015 | 0.0003 | 0.0002 | 0.0013 | 0.0002 |
| KC/h = 5 | 0.2306 | 0.0002 | 0.0001 | 0.0001 | 0.0001 | 0.0000 | 0.0000 |
| KC/h = 22 | 0.4603 | 0.2351 | 0.0513 | 0.1524 | 0.2414 | 0.0337 | 0.1441 |
| LE/h = 1 | 0.4310 | 0.0055 | 0.0187 | 0.0075 | 0.0066 | 0.0201 | 0.0083 |
| LE/h = 2 | 0.1561 | 0.0047 | 0.0101 | 0.0065 | 0.0048 | 0.0108 | 0.0067 |
| LE/h = 5 | 0.5284 | 0.0001 | 0.0003 | 0.0002 | 0.0001 | 0.0003 | 0.0001 |
| LE/h = 22 | 0.9928 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 |
| OJ/h = 1 | 0.6858 | 0.0230 | 0.0715 | 0.0314 | 0.0167 | 0.0484 | 0.0229 |
| OJ/h = 2 | 0.5726 | 0.0061 | 0.0465 | 0.0070 | 0.0046 | 0.0329 | 0.0055 |
| OJ/h = 5 | 0.4437 | 0.0052 | 0.0072 | 0.0056 | 0.0040 | 0.0057 | 0.0042 |
| OJ/h = 22 | 0.9811 | 0.1279 | 0.3430 | 0.2097 | 0.0935 | 0.2679 | 0.1580 |
| RR/h = 1 | 0.6137 | 0.0472 | 0.1229 | 0.0665 | 0.0469 | 0.1193 | 0.0661 |
| RR/h = 2 | 0.7090 | 0.0486 | 0.2324 | 0.0668 | 0.0478 | 0.2265 | 0.0658 |
| RR/h = 5 | 0.7685 | 0.0496 | 0.2995 | 0.0092 | 0.0471 | 0.2722 | 0.0065 |
| RR/h = 22 | 0.4604 | 0.0236 | 0.0814 | 0.0002 | 0.0253 | 0.0907 | 0.0004 |
| S/h = 1 | 0.2681 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| S/h = 2 | 0.4485 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| S/h = 5 | 0.5744 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

**TABLE 4** (Continued)

| Commodity/horizon | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| S/h = 22 | 0.9977 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| SB/h = 1 | 0.6358 | 0.0239 | 0.0391 | 0.0270 | 0.0258 | 0.0412 | 0.0290 |
| SB/h = 2 | 0.4289 | 0.0161 | 0.0247 | 0.0189 | 0.0195 | 0.0287 | 0.0226 |
| SB/h = 5 | 0.5608 | 0.0141 | 0.0249 | 0.0124 | 0.0169 | 0.0281 | 0.0149 |
| SB/h = 22 | 0.9855 | 0.0048 | 0.0156 | 0.0043 | 0.0037 | 0.0129 | 0.0034 |
| SM/h = 1 | 0.8516 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| SM/h = 2 | 0.6625 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| SM/h = 5 | 0.9756 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| SM/h = 22 | 0.9217 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| W/h = 1 | 0.6913 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| W/h = 2 | 0.7305 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| W/h = 5 | 0.1005 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| W/h = 22 | 0.9074 | 0.0031 | 0.0186 | 0.0045 | 0.0012 | 0.0081 | 0.0019 |

*Note*: The table summarizes $p$ values (based on robust standard errors) of the Clark–West test. Columns: (1) The HAR-RV model versus the HAR-RV-Sentiment model. (2) The HAR-RV model versus the HAR-RV-Sentiment-Moments model (selected by means of the adjusted $R^2$ statistic). (3) The HAR-RV model versus the HAR-RV-Sentiment-Moments model (selected by means of the BIC). (4) The HAR-RV model versus the HAR-RV-Sentiment-Moments model (selected by means of the CP criterion). (5) The HAR-RV-Sentiment model versus the HAR-RV-Sentiment-Moments model (selected by means of the adjusted $R^2$ statistic). (6) The HAR-RV-Sentiment model versus the HAR-RV-Sentiment-Moments model (selected by means of the BIC). (7) The HAR-RV model versus the HAR-RV-Sentiment-Moments model (selected by means of the CP criterion).

Building on asymptotics, Barndorff-Nielsen and Shephard (2004, 2006) show that

$$\lim_{M\to\infty} BV_t = \int_{t-1}^{t} \sigma^2(s)ds, \tag{12}$$

where $BV_t$ denotes the daily realized bipolar variation, which is defined as

$$BV_t = \mu_1^{-2}\left(\frac{M}{M-1}\right)\sum_{i=2}^{M} |r_{t,i-1}||r_{i,t}| = \frac{\pi}{2}\sum_{i=2}^{M} |r_{t,i-1}||r_{i,t}|, \tag{13}$$

where we define

$$\mu_a = E(|Z|^a), Z \sim N(0,1), a > 0. \tag{14}$$

Upon using the continuous component of realized variance, a consistent estimator of the pure daily jump contribution is defined as

$$J_t = RV_t - BV_t. \tag{15}$$

To evaluate the statistical significance of the jump component, we employ the formal test estimator proposed by Barndorff-Nielsen and Shephard (2006). The test statistic used for this purpose is as follows:

$$JT_t = \frac{RV_t - BV_t}{(v_{bb}-v_{qq})\frac{1}{N}QP_t}, \tag{16}$$

where $v_{bb} = \left(\frac{\pi}{2}\right) + \pi - 3$ and $v_{qq} = 2$ and $QP_t$ is defined as the daily Tri-Power Quarticity:

$$TP_t = M\frac{M}{M-2}\left(\frac{\Gamma(0.5)}{2^{2/3}\Gamma(7/6)}\right)\sum_{i=3}^{M} |r_{t,i}|^{4/3}|r_{t,i-1}|^{4/3}|r_{t,i-2}|^{4/3}, \tag{17}$$

which converges to

$$TP_t \to \int_{t-1}^{t} \sigma^4(s)ds, \tag{18}$$

even in the presence of jumps. For each $t$, $JT_t \sim N(0,1)$ as $M \to \infty$.

Equation (15) makes clear that the jump contribution to $RV_t$ is non-negative. Hence, in order to rule out negative empirical contributions, we redefine the jump measure as (see Zhou & Zhu, 2012):

$$RJ_t = \max(RV_t - BV_t; 0). \tag{19}$$

Last, we compute two measures of tail risk using the Hill estimator (Hill, 1975). We construct $X_{t,i}$, the set of reordered intraday returns $r_{t,i}$, in such a way that

TABLE 5 Tests results (backward stepwise predictor selection/recursive-estimation window).

| Commodity/horizon | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| BO/h = 1 | 0.0026 | 0.0000 | 0.0002 | 0.0000 | 0.0001 | 0.0017 | 0.0000 |
| BO/h = 2 | 0.0279 | 0.0000 | 0.0001 | 0.0000 | 0.0000 | 0.0002 | 0.0001 |
| BO/h = 5 | 0.2399 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| BO/h = 22 | 0.1097 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| C/h = 1 | 0.7707 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| C/h = 2 | 0.4526 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| C/h = 5 | 0.4465 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| C/h = 22 | 0.2698 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| CC/h = 1 | 0.0322 | 0.0000 | 0.0001 | 0.0000 | 0.0001 | 0.0004 | 0.0000 |
| CC/h = 2 | 0.0123 | 0.0000 | 0.0023 | 0.0000 | 0.0001 | 0.0141 | 0.0001 |
| CC/h = 5 | 0.0285 | 0.0000 | 0.0001 | 0.0000 | 0.0000 | 0.0008 | 0.0000 |
| CC/h = 22 | 0.0103 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| CT/h = 1 | 0.9164 | 0.0049 | 0.0041 | 0.0036 | 0.0049 | 0.0040 | 0.0036 |
| CT/h = 2 | 0.5606 | 0.0029 | 0.0038 | 0.0032 | 0.0029 | 0.0037 | 0.0032 |
| CT/h = 5 | 0.5585 | 0.0069 | 0.0120 | 0.0068 | 0.0068 | 0.0117 | 0.0066 |
| CT/h = 22 | 0.7750 | 0.0086 | 0.0143 | 0.0078 | 0.0085 | 0.0143 | 0.0076 |
| GF/h = 1 | 0.4548 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| GF/h = 2 | 0.1562 | 0.0003 | 0.0002 | 0.0004 | 0.0003 | 0.0002 | 0.0004 |
| GF/h = 5 | 0.2682 | 0.0002 | 0.0005 | 0.0002 | 0.0002 | 0.0004 | 0.0002 |
| GF/h = 22 | 0.5157 | 0.0009 | 0.0006 | 0.0010 | 0.0009 | 0.0006 | 0.0010 |
| HE/h = 1 | 0.0708 | 0.0000 | 0.0001 | 0.0000 | 0.0000 | 0.0002 | 0.0000 |
| HE/h = 2 | 0.0181 | 0.0000 | 0.0002 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| HE/h = 5 | 0.0447 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| HE/h = 22 | 0.0030 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| KC/h = 1 | 0.4590 | 0.0007 | 0.0019 | 0.0005 | 0.0004 | 0.0011 | 0.0003 |
| KC/h = 2 | 0.3527 | 0.0003 | 0.0001 | 0.0003 | 0.0002 | 0.0001 | 0.0002 |
| KC/h = 5 | 0.2306 | 0.0002 | 0.0001 | 0.0001 | 0.0001 | 0.0000 | 0.0000 |
| KC/h = 22 | 0.4603 | 0.2346 | 0.0551 | 0.1754 | 0.2410 | 0.0366 | 0.1648 |
| LE/h = 1 | 0.4310 | 0.0055 | 0.0167 | 0.0077 | 0.0066 | 0.0181 | 0.0085 |
| LE/h = 2 | 0.1561 | 0.0047 | 0.0103 | 0.0065 | 0.0048 | 0.0108 | 0.0068 |
| LE/h = 5 | 0.5284 | 0.0001 | 0.0003 | 0.0002 | 0.0001 | 0.0003 | 0.0001 |
| LE/h = 22 | 0.9928 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 |
| OJ/h = 1 | 0.6858 | 0.0234 | 0.0715 | 0.0321 | 0.0171 | 0.0484 | 0.0234 |
| OJ/h = 2 | 0.5726 | 0.0061 | 0.0465 | 0.0070 | 0.0046 | 0.0329 | 0.0055 |
| OJ/h = 5 | 0.4437 | 0.0053 | 0.0072 | 0.0056 | 0.0041 | 0.0057 | 0.0042 |
| OJ/h = 22 | 0.9811 | 0.1279 | 0.3430 | 0.2097 | 0.0935 | 0.2679 | 0.1580 |
| RR/h = 1 | 0.6137 | 0.0391 | 0.1229 | 0.0797 | 0.0376 | 0.1193 | 0.0771 |
| RR/h = 2 | 0.7090 | 0.0175 | 0.2324 | 0.1891 | 0.0164 | 0.2265 | 0.1837 |
| RR/h = 5 | 0.7685 | 0.0497 | 0.2693 | 0.0088 | 0.0472 | 0.2432 | 0.0064 |
| RR/h = 22 | 0.4604 | 0.0232 | 0.0814 | 0.0002 | 0.0248 | 0.0907 | 0.0003 |
| S/h = 1 | 0.2681 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| S/h = 2 | 0.4485 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| S/h = 5 | 0.5744 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

TABLE 5 (Continued)

| Commodity/horizon | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| S/h = 22 | 0.9977 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| SB/h = 1 | 0.6358 | 0.0239 | 0.0391 | 0.0270 | 0.0258 | 0.0412 | 0.0290 |
| SB/h = 2 | 0.4289 | 0.0161 | 0.0247 | 0.0189 | 0.0195 | 0.0287 | 0.0226 |
| SB/h = 5 | 0.5608 | 0.0141 | 0.0249 | 0.0124 | 0.0169 | 0.0281 | 0.0149 |
| SB/h = 22 | 0.9855 | 0.0044 | 0.0156 | 0.0044 | 0.0034 | 0.0129 | 0.0034 |
| SM/h = 1 | 0.8516 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| SM/h = 2 | 0.6625 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| SM/h = 5 | 0.9756 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| SM/h = 22 | 0.9217 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| W/h = 1 | 0.6913 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| W/h = 2 | 0.7305 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| W/h = 5 | 0.1005 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| W/h = 22 | 0.9074 | 0.0032 | 0.0186 | 0.0045 | 0.0013 | 0.0081 | 0.0019 |

*Note*: The table summarizes $p$ values (based on robust standard errors) of the Clark–West test. Columns: (1) The HAR-RV model versus the HAR-RV-Sentiment model. (2) The HAR-RV model versus the HAR-RV-Sentiment-Moments model (selected by means of the adjusted $R^2$ statistic). (3) The HAR-RV model versus the HAR-RV-Sentiment-Moments model (selected by means of the BIC). (4) The HAR-RV model versus the HAR-RV-Sentiment-Moments model (selected by means of the CP criterion). (5) The HAR-RV-Sentiment model versus the HAR-RV-Sentiment-Moments model (selected by means of the adjusted $R^2$ statistic). (6) The HAR-RV-Sentiment model versus the HAR-RV-Sentiment-Moments model (selected by means of the BIC). (7) The HAR-RV model versus the HAR-RV-Sentiment-Moments model (selected by means of the CP criterion).

$$X_{t,i} \geq X_{t,j} \text{ for } i<j. \tag{20}$$

The Hill (1975) positive tail risk estimator (our predictor $TR_u$) is computed as

$$H_t^{up} = \frac{1}{k}\sum_{j=1}^{k} \ln(X_{t,i}) - \ln(X_{t,k}) \tag{21}$$

and the negative tail risk estimator (our predictor $TR_d$) as

$$H_t^{down} = \frac{1}{k}\sum_{j=n-k}^{n} \ln(X_{t,i}) - \ln(X_{t,n-k}), \tag{22}$$

where $k$ is the observation denoting the chosen $\alpha$ tail interval.

## 4 | BASELINE EMPIRICAL RESULTS

We begin our presentation of the results by summarizing the outcomes for the best stepwise predictor selection algorithm as applied to the overall sentiment index. In Table 2, we present the RMSFE ratios obtained through a recursive-estimation window and forward stepwise selection of the realized moments. Column (1) of the table showcases a comparison between the HAR-RV model and the HAR-RV-Sentiment model. The overall observation indicates that the RMSFE ratio remains close to unity. While the addition of sentiment to the HAR-RV model slightly enhances forecast accuracy for certain agricultural commodities such as BO and CC, the improvement is marginal. For several other agricultural commodities, however, the HAR-RV-Sentiment model produces forecasts that are somewhat less accurate than those of the baseline HAR-RV model, again by a rather small margin. Consequently, there appears to be no systematic difference between the HAR-RV model and the HAR-RV-Sentiment model, suggesting limited informational value of sentiment in forecasting subsequent realized volatility. The nature of agricultural commodities themselves can potentially explain the limited role of sentiment in predicting future volatility. Agricultural commodities are primarily driven by fundamental factors such as weather conditions, production levels, and governmental policies. These elements, which can cause significant price fluctuations, are largely immune to investor sentiment. This distinct characteristic of agricultural commodities sets them apart from other investment vehicles like stocks and bonds, where sentiment can play a more substantial role.

TABLE 6 Tests results (forward stepwise predictor selection/rolling-estimation window).

| Commodity/horizon | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| BO/h = 1 | 0.0054 | 0.0001 | 0.0002 | 0.0002 | 0.0012 | 0.0048 | 0.0033 |
| BO/h = 2 | 0.0789 | 0.0000 | 0.0002 | 0.0001 | 0.0001 | 0.0003 | 0.0002 |
| BO/h = 5 | 0.2093 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 |
| BO/h = 22 | 0.0687 | 0.0003 | 0.0011 | 0.0000 | 0.0008 | 0.0045 | 0.0002 |
| C/h = 1 | 0.5635 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| C/h = 2 | 0.3000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| C/h = 5 | 0.1419 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| C/h = 22 | 0.0148 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| CC/h = 1 | 0.0328 | 0.0001 | 0.0002 | 0.0000 | 0.0002 | 0.0007 | 0.0001 |
| CC/h = 2 | 0.0128 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0002 | 0.0000 |
| CC/h = 5 | 0.0494 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0000 |
| CC/h = 22 | 0.0053 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| CT/h = 1 | 0.8179 | 0.0001 | 0.0013 | 0.0002 | 0.0001 | 0.0011 | 0.0001 |
| CT/h = 2 | 0.5049 | 0.0001 | 0.0001 | 0.0000 | 0.0001 | 0.0001 | 0.0000 |
| CT/h = 5 | 0.6089 | 0.0001 | 0.0008 | 0.0002 | 0.0001 | 0.0009 | 0.0002 |
| CT/h = 22 | 0.6317 | 0.0081 | 0.0056 | 0.0073 | 0.0087 | 0.0059 | 0.0078 |
| GF/h = 1 | 0.6129 | 0.0000 | 0.0001 | 0.0000 | 0.0000 | 0.0001 | 0.0000 |
| GF/h = 2 | 0.1893 | 0.0003 | 0.0001 | 0.0002 | 0.0003 | 0.0001 | 0.0002 |
| GF/h = 5 | 0.4280 | 0.0002 | 0.0013 | 0.0003 | 0.0002 | 0.0011 | 0.0003 |
| GF/h = 22 | 0.9402 | 0.0006 | 0.0024 | 0.0011 | 0.0002 | 0.0009 | 0.0005 |
| HE/h = 1 | 0.0483 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0000 |
| HE/h = 2 | 0.0018 | 0.0000 | 0.0001 | 0.0000 | 0.0000 | 0.0004 | 0.0000 |
| HE/h = 5 | 0.0197 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| HE/h = 22 | 0.0147 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| KC/h = 1 | 0.3362 | 0.0024 | 0.0112 | 0.0025 | 0.0021 | 0.0110 | 0.0023 |
| KC/h = 2 | 0.0942 | 0.0001 | 0.0005 | 0.0004 | 0.0002 | 0.0009 | 0.0007 |
| KC/h = 5 | 0.1466 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| KC/h = 22 | 0.2558 | 0.0351 | 0.0019 | 0.0140 | 0.0418 | 0.0021 | 0.0174 |
| LE/h = 1 | 0.7386 | 0.0001 | 0.0014 | 0.0001 | 0.0001 | 0.0013 | 0.0001 |
| LE/h = 2 | 0.2130 | 0.0002 | 0.0008 | 0.0002 | 0.0003 | 0.0010 | 0.0003 |
| LE/h = 5 | 0.4574 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| LE/h = 22 | 0.9278 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| OJ/h = 1 | 0.7536 | 0.1010 | 0.4000 | 0.1386 | 0.0767 | 0.3154 | 0.1069 |
| OJ/h = 2 | 0.7450 | 0.0248 | 0.2479 | 0.0383 | 0.0183 | 0.1841 | 0.0281 |
| OJ/h = 5 | 0.7021 | 0.0119 | 0.0510 | 0.0118 | 0.0082 | 0.0298 | 0.0081 |
| OJ/h = 22 | 0.7325 | 0.0502 | 0.1444 | 0.0481 | 0.0415 | 0.1193 | 0.0395 |
| RR/h = 1 | 0.2773 | 0.0362 | 0.0376 | 0.0347 | 0.0374 | 0.0392 | 0.0361 |
| RR/h = 2 | 0.1806 | 0.0000 | 0.2118 | 0.0003 | 0.0000 | 0.2421 | 0.0004 |
| RR/h = 5 | 0.2526 | 0.0038 | 0.0133 | 0.0205 | 0.0044 | 0.0164 | 0.0236 |
| RR/h = 22 | 0.0882 | 0.0005 | 0.0079 | 0.0006 | 0.0034 | 0.0367 | 0.0045 |
| S/h = 1 | 0.2918 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| S/h = 2 | 0.6164 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| S/h = 5 | 0.7781 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

**TABLE 6** (Continued)

| Commodity/horizon | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| S/h = 22 | 0.9958 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| SB/h = 1 | 0.6125 | 0.0088 | 0.0051 | 0.0113 | 0.0072 | 0.0034 | 0.0095 |
| SB/h = 2 | 0.5554 | 0.0000 | 0.0019 | 0.0000 | 0.0000 | 0.0008 | 0.0000 |
| SB/h = 5 | 0.5247 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| SB/h = 22 | 0.0359 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| SM/h = 1 | 0.8021 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| SM/h = 2 | 0.7232 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| SM/h = 5 | 0.9581 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| SM/h = 22 | 0.9097 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| W/h = 1 | 0.7423 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| W/h = 2 | 0.5235 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| W/h = 5 | 0.3104 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| W/h = 22 | 0.8484 | 0.0213 | 0.0209 | 0.0180 | 0.0060 | 0.0047 | 0.0040 |

*Note*: The table summarizes *p* values (based on robust standard errors) of the Clark–West test. Columns: (1) The HAR-RV model versus the HAR-RV-Sentiment model. (2) The HAR-RV model versus the HAR-RV-Sentiment-Moments model (selected by means of the adjusted $R^2$ statistic). (3) The HAR-RV model versus the HAR-RV-Sentiment-Moments model (selected by means of the BIC). (4) The HAR-RV model versus the HAR-RV-Sentiment-Moments model (selected by means of the CP criterion). (5) The HAR-RV-Sentiment model versus the HAR-RV-Sentiment-Moments model (selected by means of the adjusted $R^2$ statistic). (6) The HAR-RV-Sentiment model versus the HAR-RV-Sentiment-Moments model (selected by means of the BIC). (7) The HAR-RV model versus the HAR-RV-Sentiment-Moments model (selected by means of the CP criterion).

In Columns (2) to (4) of the table, we conduct a comparison between the HAR-RV model and the HAR-RV-Sentiment-Moments model. To obtain three versions of the latter, we employ three information criteria, namely, the adjusted $R^2$ statistic, the BIC, and the CP statistic. This comparison produces RMSFE ratios that exceed unity in almost all cases. While the RMSFE ratios for certain agricultural commodities like KC and OJ are only slightly larger than unity, overall, the RMSFE ratios exceed unity by a significant margin across all forecast horizons. Consequently, the inclusion of realized moments enhances forecast accuracy when compared to the baseline HAR-RV model. Since the HAR-RV-Sentiment model demonstrates only marginal, if any, improvement over the baseline HAR-RV model, it is unsurprising that the HAR-RV-Sentiment-Moments model outperforms the HAR-RV-Sentiment model in the majority of cases, as evidenced by the RMSFE ratios reported in Columns (5) to (7) of the table. One plausible explanation could be that realized moments, such as leverage, realized skewness, realized kurtosis, realized jumps, and realized tail risks, among others, offer more direct and quantifiable indicators of market conditions. These moments capture the inherent characteristics of price return distributions, including volatility clustering, skewness, fat tails, and extreme events. By incorporating measures such as realized skewness and kurtosis, the model becomes more adept at capturing any asymmetry and fat-tailed nature of the returns distribution. Similarly, incorporating realized jumps and tail risks enhances the model's ability to anticipate abrupt and extreme fluctuations in volatility. In contrast, sentiment, being a softer and less quantifiable measure, may provide a less direct and potentially weaker signal of future price movements.

In Figure 3, we depict the percentage frequency of inclusion of various realized moments in the HAR-RV-Sentiment-Moments model using the forward stepwise predictor selection algorithm with a recursive-estimation window. We report results for the three information criteria, averaged across the 14 agricultural commodities in our sample. As expected, the utilization of the adjusted $R^2$ statistic (and the CP criterion) leads to a larger forecasting model compared to model selection based on the BIC. An important finding is that, irrespective of the information criterion employed, leverage, realized kurtosis, and realized jumps emerge as the "top three" realized moments.[9]

Considering that infrequent large peaks are a distinctive characteristic of realized volatility, as documented in Figure 1, it is interesting to study the MAFE ratios summarized in Table 3. Corroborating the results for the RMSFE ratios, we find that the differences in forecast accuracy between the baseline HAR-RV model and the HAR-RV-Sentiment model are relatively small. The HAR-RV-Sentiment-Moments model, in turn,

TABLE 7 Tests results (realized variance/forward stepwise predictor selection/recursive-estimation window).

| Commodity/horizon | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| BO/h = 1 | 0.0069 | 0.0002 | 0.0136 | 0.0003 | 0.0016 | 0.0540 | 0.0022 |
| BO/h = 2 | 0.0312 | 0.0003 | 0.0019 | 0.0004 | 0.0011 | 0.0051 | 0.0013 |
| BO/h = 5 | 0.3030 | 0.0014 | 0.0021 | 0.0018 | 0.0025 | 0.0036 | 0.0031 |
| BO/h = 22 | 0.1262 | 0.0001 | 0.0008 | 0.0000 | 0.0004 | 0.0038 | 0.0003 |
| C/h = 1 | 0.6441 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| C/h = 2 | 0.2985 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| C/h = 5 | 0.5054 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| C/h = 22 | 0.2943 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| CC/h = 1 | 0.0310 | 0.0001 | 0.0006 | 0.0000 | 0.0002 | 0.0025 | 0.0001 |
| CC/h = 2 | 0.0099 | 0.0000 | 0.0060 | 0.0001 | 0.0001 | 0.0576 | 0.0012 |
| CC/h = 5 | 0.0237 | 0.0000 | 0.0008 | 0.0000 | 0.0000 | 0.0074 | 0.0001 |
| CC/h = 22 | 0.0175 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| CT/h = 1 | 0.8779 | 0.0152 | 0.0236 | 0.0141 | 0.0162 | 0.0249 | 0.0151 |
| CT/h = 2 | 0.2082 | 0.0094 | 0.0132 | 0.0105 | 0.0101 | 0.0141 | 0.0113 |
| CT/h = 5 | 0.0343 | 0.0159 | 0.0159 | 0.0185 | 0.0163 | 0.0163 | 0.0190 |
| CT/h = 22 | 0.8764 | 0.0300 | 0.0314 | 0.0296 | 0.0293 | 0.0307 | 0.0289 |
| GF/h = 1 | 0.2603 | 0.0557 | 0.0689 | 0.0597 | 0.0539 | 0.0675 | 0.0578 |
| GF/h = 2 | 0.1482 | 0.0441 | 0.0576 | 0.0406 | 0.0445 | 0.0591 | 0.0411 |
| GF/h = 5 | 0.1895 | 0.0132 | 0.0207 | 0.0149 | 0.0133 | 0.0210 | 0.0153 |
| GF/h = 22 | 0.0557 | 0.0911 | 0.1040 | 0.0908 | 0.0955 | 0.1090 | 0.0953 |
| HE/h = 1 | 0.9773 | 0.3395 | 0.4416 | 0.3684 | 0.1983 | 0.2550 | 0.2087 |
| HE/h = 2 | 0.8095 | 0.0516 | 0.0542 | 0.0542 | 0.0307 | 0.0327 | 0.0327 |
| HE/h = 5 | 0.9293 | 0.0123 | 0.0170 | 0.0125 | 0.0058 | 0.0083 | 0.0057 |
| HE/h = 22 | 0.1196 | 0.0017 | 0.0031 | 0.0026 | 0.0017 | 0.0030 | 0.0025 |
| KC/h = 1 | 0.5644 | 0.0034 | 0.0142 | 0.0027 | 0.0016 | 0.0060 | 0.0012 |
| KC/h = 2 | 0.4873 | 0.0012 | 0.0008 | 0.0012 | 0.0011 | 0.0007 | 0.0011 |
| KC/h = 5 | 0.3240 | 0.0007 | 0.0001 | 0.0007 | 0.0002 | 0.0000 | 0.0003 |
| KC/h = 22 | 0.3401 | 0.2752 | 0.0218 | 0.1775 | 0.3359 | 0.0131 | 0.2057 |
| LE/h = 1 | 0.2886 | 0.3162 | 0.3147 | 0.2968 | 0.3202 | 0.3204 | 0.3009 |
| LE/h = 2 | 0.0972 | 0.2854 | 0.3301 | 0.2963 | 0.2884 | 0.3335 | 0.2995 |
| LE/h = 5 | 0.0441 | 0.0727 | 0.0831 | 0.0725 | 0.0740 | 0.0846 | 0.0738 |
| LE/h = 22 | 0.0148 | 0.0145 | 0.0159 | 0.0143 | 0.0148 | 0.0162 | 0.0145 |
| OJ/h = 1 | 0.5957 | 0.0260 | 0.1320 | 0.0281 | 0.0220 | 0.1121 | 0.0243 |
| OJ/h = 2 | 0.5546 | 0.0287 | 0.0351 | 0.0204 | 0.0189 | 0.0198 | 0.0123 |
| OJ/h = 5 | 0.4714 | 0.0344 | 0.1328 | 0.0445 | 0.0222 | 0.0925 | 0.0288 |
| OJ/h = 22 | 0.9816 | 0.1746 | 0.3549 | 0.1850 | 0.1431 | 0.3011 | 0.1521 |
| RR/h = 1 | 0.0073 | 0.0017 | 0.0026 | 0.0015 | 0.0018 | 0.0030 | 0.0016 |
| RR/h = 2 | 0.0202 | 0.0232 | 0.0106 | 0.0133 | 0.0230 | 0.0123 | 0.0152 |
| RR/h = 5 | 0.8307 | 0.0663 | 0.0016 | 0.0486 | 0.0608 | 0.0026 | 0.0432 |
| RR/h = 22 | 0.9533 | 0.0840 | 0.0004 | 0.0748 | 0.0768 | 0.0000 | 0.0670 |
| S/h = 1 | 0.1250 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| S/h = 2 | 0.6081 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| S/h = 5 | 0.6190 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

TABLE 7 (Continued)

| Commodity/horizon | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| S/h = 22 | 0.9943 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| SB/h = 1 | 0.6005 | 0.0377 | 0.0452 | 0.0561 | 0.0413 | 0.0472 | 0.0604 |
| SB/h = 2 | 0.4498 | 0.0253 | 0.0499 | 0.0241 | 0.0292 | 0.0535 | 0.0278 |
| SB/h = 5 | 0.3363 | 0.0150 | 0.0374 | 0.0144 | 0.0204 | 0.0443 | 0.0195 |
| SB/h = 22 | 0.6707 | 0.0050 | 0.0210 | 0.0055 | 0.0043 | 0.0193 | 0.0047 |
| SM/h = 1 | 0.8181 | 0.0004 | 0.0001 | 0.0002 | 0.0003 | 0.0001 | 0.0002 |
| SM/h = 2 | 0.6802 | 0.0017 | 0.0035 | 0.0015 | 0.0017 | 0.0034 | 0.0015 |
| SM/h = 5 | 0.9367 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| SM/h = 22 | 0.8387 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| W/h = 1 | 0.7163 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| W/h = 2 | 0.1994 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| W/h = 5 | 0.0127 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| W/h = 22 | 0.2766 | 0.0000 | 0.0001 | 0.0000 | 0.0000 | 0.0001 | 0.0000 |

*Note*: The table summarizes *p* values (based on robust standard errors) of the Clark–West test. Columns: (1) The HAR-RV model versus the HAR-RV-Sentiment model. (2) The HAR-RV model versus the HAR-RV-Sentiment-Moments model (selected by means of the adjusted $R^2$ statistic). (3) The HAR-RV model versus the HAR-RV-Sentiment-Moments model (selected by means of the BIC). (4) The HAR-RV model versus the HAR-RV-Sentiment-Moments model (selected by means of the CP criterion). (5) The HAR-RV-Sentiment model versus the HAR-RV-Sentiment-Moments model (selected by means of the adjusted $R^2$ statistic). (6) The HAR-RV-Sentiment model versus the HAR-RV-Sentiment-Moments model (selected by means of the BIC). (7) The HAR-RV model versus the HAR-RV-Sentiment-Moments model (selected by means of the CP criterion).

produces slightly less accurate forecasts than the HAR-RV model for few agricultural commodities, like KC and OJ, but outperforms the latter for several other agricultural commodities. The margin by which the forecasts produced by the HAR-RV-Sentiment-Moments model are more accurate than the forecasts produced by the HAR-RV model is small for some agricultural commodities, like CC, but is seizable for others, like C, HE, S, and SM. Similar to the RMSFE ratios, the strong performance of the HAR-RV-Sentiment-Moments model in relation to the baseline HAR-RV model translates into notable outperformance compared to the HAR-RV-Sentiment model.

The results we report in Table 4 shed light on the statistical significance (in terms of *p* values of the Clark–West test) of the differences in forecast accuracy across forecasting models. In line with the results for the RMSFE and MAFE ratios, we observe only a few significant test results when we compare the baseline HAR-RV model with the HAR-RV-Sentiment model. Hence, across all 14 agricultural commodities in our sample, adding sentiment to the forecasting model does not improve systematically forecast accuracy. In contrast, when comparing the HAR-RV or the HAR-RV-Moments model with the HAR-RV-Sentiment-Moments, the majority of test results are statistically significant, with only a few exceptions. The test results thus help to build confidence in our main result that realized moments rather than sentiment are a major driver of forecast accuracy.

## 5 | ROBUSTNESS CHECKS

### 5.1 | Ensuring consistency in predictor selection approaches

As part of several robustness checks, we present additional test results in Tables 5–7. Table 5 provides the test results obtained using a backward stepwise selection algorithm, while Table 6 displays the results obtained from a rolling-estimation window with a forward stepwise predictor selection algorithm.[10] In Table 7, we present the results obtained using a forward stepwise predictor selection algorithm and a recursive-estimation window, focusing on the realized variance of agricultural commodity returns. While the specifics may vary across the tables, the overarching finding remains consistent: realized moments, rather than sentiment, play a crucial role in enhancing forecast accuracy.[11]

In Table A2, we demonstrate that our primary conclusion—namely, that realized moments significantly influence forecasting performance more than sentiment—remains consistent when our models are

**TABLE 8** RMSFE ratios (MOBA algorithm/$V = E = 1{,}000$).

| Commodity/horizon | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| BO/h = 1 | 0.9988 | 1.0107 | 1.0071 | 1.0054 | 1.0099 |
| BO/h = 2 | 1.0098 | 1.0184 | 0.9916 | 1.0032 | 0.9982 |
| BO/h = 5 | 1.0304 | 1.0352 | 0.9895 | 1.0088 | 0.9902 |
| BO/h = 22 | 1.0139 | 1.0156 | 0.9828 | 0.9979 | 0.9918 |
| C/h = 1 | 1.0227 | 1.0218 | 1.0004 | 0.9954 | 0.9986 |
| C/h = 2 | 1.0212 | 1.0224 | 0.9972 | 0.9967 | 0.9976 |
| C/h = 5 | 1.0302 | 1.0320 | 0.9944 | 1.0039 | 0.9949 |
| C/h = 22 | 1.0205 | 1.0293 | 0.9975 | 1.0087 | 0.9996 |
| CC/h = 1 | 1.0119 | 1.0123 | 0.9979 | 1.0006 | 0.9972 |
| CC/h = 2 | 1.0196 | 1.0177 | 0.9962 | 1.0037 | 0.9968 |
| CC/h = 5 | 1.0165 | 1.0160 | 0.9943 | 0.9995 | 0.9971 |
| CC/h = 22 | 1.0139 | 1.0106 | 1.0006 | 1.0072 | 1.0002 |
| CT/h = 1 | 1.0102 | 1.0108 | 0.9998 | 1.0008 | 1.0003 |
| CT/h = 2 | 1.0128 | 1.0126 | 0.9995 | 1.0024 | 0.9995 |
| CT/h = 5 | 1.0155 | 1.0153 | 0.9978 | 1.0046 | 0.9992 |
| CT/h = 22 | 1.0062 | 1.0065 | 0.9997 | 0.9988 | 1.0003 |
| GF/h = 1 | 1.0256 | 1.0260 | 0.9999 | 1.0091 | 1.0000 |
| GF/h = 2 | 1.0278 | 1.0295 | 0.9941 | 0.9926 | 0.9902 |
| GF/h = 5 | 1.0151 | 1.0193 | 0.9966 | 0.9994 | 1.0002 |
| GF/h = 22 | 0.9931 | 1.0000 | 1.0084 | 1.0100 | 1.0082 |
| HE/h = 1 | 1.0341 | 1.0278 | 1.0007 | 1.0057 | 1.0007 |
| HE/h = 2 | 1.0354 | 1.0230 | 1.0012 | 1.0006 | 0.9946 |
| HE/h = 5 | 1.0442 | 1.0359 | 0.9839 | 0.9909 | 0.9874 |
| HE/h = 22 | 1.0279 | 1.0215 | 0.9906 | 0.9900 | 0.9905 |
| KC/h = 1 | 1.0025 | 1.0047 | 1.0086 | 1.0128 | 1.0089 |
| KC/h = 2 | 1.0025 | 1.0056 | 1.0018 | 1.0065 | 1.0053 |
| KC/h = 5 | 1.0026 | 1.0075 | 1.0008 | 0.9968 | 0.9971 |
| KC/h = 22 | 1.0058 | 1.0053 | 0.9988 | 0.9941 | 0.9960 |
| LE/h = 1 | 1.0161 | 1.0161 | 0.9904 | 0.9923 | 0.9891 |
| LE/h = 2 | 1.0120 | 1.0121 | 0.9932 | 0.9945 | 0.9921 |
| LE/h = 5 | 1.0152 | 1.0168 | 0.9949 | 1.0021 | 0.9921 |
| LE/h = 22 | 1.0028 | 1.0061 | 0.9969 | 0.9969 | 0.9980 |
| OJ/h = 1 | 0.9959 | 0.9966 | 1.0142 | 1.0095 | 1.0125 |
| OJ/h = 2 | 0.9949 | 0.9959 | 1.0112 | 1.0111 | 1.0104 |
| OJ/h = 5 | 1.0026 | 1.0036 | 1.0094 | 1.0061 | 1.0075 |
| OJ/h = 22 | 0.9961 | 0.9965 | 1.0026 | 0.9969 | 1.0017 |
| RR/h = 1 | 1.0123 | 1.0121 | 0.9994 | 1.0001 | 1.0026 |
| RR/h = 2 | 1.0137 | 1.0143 | 1.0052 | 1.0064 | 1.0067 |
| RR/h = 5 | 1.0128 | 1.0139 | 0.9915 | 0.9965 | 0.9954 |
| RR/h = 22 | 1.0073 | 1.0033 | 1.0038 | 0.9927 | 1.0049 |
| S/h = 1 | 1.0217 | 1.0211 | 1.0006 | 1.0108 | 0.9999 |
| S/h = 2 | 1.0278 | 1.0296 | 1.0053 | 1.0148 | 1.0069 |
| S/h = 5 | 1.0336 | 1.0389 | 1.0078 | 1.0201 | 1.0086 |

**TABLE 8** (Continued)

| Commodity/horizon | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| S/h = 22 | 1.0158 | 1.0253 | 1.0045 | 1.0017 | 1.0016 |
| SB/h = 1 | 1.0318 | 1.0323 | 1.0104 | 1.0135 | 1.0131 |
| SB/h = 2 | 1.0137 | 1.0154 | 0.9983 | 0.9950 | 1.0029 |
| SB/h = 5 | 1.0138 | 1.0158 | 0.9799 | 0.9867 | 0.9803 |
| SB/h = 22 | 1.0331 | 1.0338 | 0.9872 | 1.0083 | 0.9884 |
| SM/h = 1 | 1.0240 | 1.0253 | 1.0028 | 1.0061 | 1.0038 |
| SM/h = 2 | 1.0374 | 1.0401 | 1.0056 | 1.0070 | 1.0064 |
| SM/h = 5 | 1.0242 | 1.0280 | 1.0162 | 1.0109 | 1.0148 |
| SM/h = 22 | 0.9925 | 0.9978 | 1.0157 | 1.0060 | 1.0097 |
| W/h = 1 | 1.0196 | 1.0208 | 0.9974 | 1.0023 | 1.0020 |
| W/h = 2 | 1.0290 | 1.0305 | 0.9988 | 0.9968 | 0.9985 |
| W/h = 5 | 1.0395 | 1.0400 | 0.9968 | 0.9977 | 0.9939 |
| W/h = 22 | 1.0052 | 1.0120 | 0.9943 | 0.9993 | 0.9961 |

*Note*: Columns: (1) The HAR-RV model versus the HAR-RV-Sentiment-MOBA-Moments model. (2) The HAR-RV-Sentiment model versus the HAR-RV-Sentiment-MOBA-Moments model. (3) The HAR-RV-Sentiment-Moments /Adj. $R^2$ model versus the HAR-RV-Sentiment-MOBA-Moments model. (4) The HAR-RV-Sentiment-Moments/BIC model versus HAR-RV-Sentiment-MOBA-Moments model. (5) The HAR-RV-Sentiment-Moments/BIC model versus the HAR-RV-Sentiment-MOBA-Moments model. (6) The HAR-RV-Sentiment-Moments/CP model versus the HAR-RV-Sentiment-MOBA-Moments model.

estimated using another popular method: the elastic-net shrinkage estimator.[12] We employ a rolling-estimation window consisting of 1,000 observations. In terms of the RMSFE ratio, the forecasting accuracy of the HAR-RV-Sentiment model hardly differs from that of the HAR-RV model, with both being estimated using the ordinary least squares technique. Conversely, the HAR-RV-Sentiment-Moments model (estimated with the elastic net) outperforms both the HAR-RV and HAR-RV-Sentiment models in the vast majority of combinations of commodities and forecast horizons. This provides additional validation to our primary conclusion.[13]

## 5.2 | MOBA algorithm: Assessing forecast accuracy for different sample sizes

In the next step, we present a summary of the RMSFE ratios obtained for the MOBA algorithm in Table 8 for $V = E = 1{,}000$ and in Table 9 for $V = E = 500$. Two key findings emerge from these results. First, similar to the best predictor selection algorithm, the HAR-RV-Sentiment-MOBA-Moments model consistently outperforms the HAR-RV model across the majority of agricultural commodities, with the exception of OJ. Second, the HAR-RV-Sentiment-MOBA-Moments model demonstrates better forecast accuracy compared to the HAR-RV-Sentiment model, with OJ being the primary exception once again. When we set $V = E = 1{,}000$, the HAR-RV-Sentiment-MOBA-Moments model performs better than the HAR-RV-Sentiment-Moments selected by the best stepwise predictor selection algorithm in 42.86% (60.71%, 46.43%) of cases when considering the adjusted $R^2$ statistic (BIC, CP criterion). However, when $V = E = 500$, the balance shifts in favor of the MOBA algorithm, with the HAR-RV-MOBA-Sentiment-Moments model outperforming the model selected by the best stepwise predictor selection algorithm in 62.50% (66.07%, 60.71%) of cases when considering the adjusted $R^2$ statistic (BIC, CP criterion), where, as one would have expected, the difference in forecasting performance between the MOBA and the predictor selection algorithms is smaller on average than the difference between the HAR-RV/HAR-RV-Sentiment models and the HAR-RV-Sentiment-MOBA-Moments model.[14]

In the appendix, Table A3 for $V = E = 1{,}000$ and Table A4 for $V = E = 500$ summarize similar results for the MAFE ratios. For $V = E = 1{,}000$, the HAR-RV-Sentiment-MOBA-Moments model yields more accurate forecasts than the model selected by the best stepwise predictor selection algorithm in 46.43% (55.36%, 48.21%) of cases when we consider the Adj. $R^2$ statistic (BIC, CP criterion), and for $V = E = 500$ these numbers increase to 78.57%, 75%, and 78.57%. Importantly, the results for the MAFE ratios lend further support to our finding that the realized moments play a more significant role in forecast accuracy than sentiment.[15]

**TABLE 9** RMSFE ratios (MOBA algorithm/$V = E = 500$).

| Commodity/horizon | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| BO/h = 1 | 1.0102 | 1.0063 | 1.0026 | 1.0107 | 1.0010 |
| BO/h = 2 | 1.0122 | 1.0097 | 0.9973 | 1.0011 | 0.9975 |
| BO/h = 5 | 1.0314 | 1.0303 | 0.9936 | 1.0019 | 0.9975 |
| BO/h = 22 | 1.0135 | 1.0132 | 0.9935 | 1.0000 | 0.9957 |
| C/h = 1 | 1.0154 | 1.0168 | 1.0192 | 1.0137 | 1.0140 |
| C/h = 2 | 1.0149 | 1.0167 | 1.0196 | 1.0157 | 1.0153 |
| C/h = 5 | 1.0210 | 1.0221 | 1.0161 | 1.0119 | 1.0147 |
| C/h = 22 | 1.0211 | 1.0223 | 1.0028 | 1.0029 | 1.0035 |
| CC/h = 1 | 1.0098 | 1.0090 | 1.0039 | 1.0094 | 1.0027 |
| CC/h = 2 | 1.0123 | 1.0104 | 1.0012 | 1.0131 | 1.0043 |
| CC/h = 5 | 1.0103 | 1.0091 | 0.9951 | 1.0024 | 0.9976 |
| CC/h = 22 | 1.0142 | 1.0122 | 0.9987 | 1.0095 | 1.0015 |
| CT/h = 1 | 0.9779 | 0.9791 | 1.0423 | 0.9726 | 1.1058 |
| CT/h = 2 | 1.0089 | 1.0105 | 0.9990 | 0.9967 | 0.9986 |
| CT/h = 5 | 1.0071 | 1.0088 | 0.9983 | 0.9992 | 0.9997 |
| CT/h = 22 | 1.0035 | 1.0066 | 1.0046 | 1.0033 | 1.0001 |
| GF/h = 1 | 1.0226 | 1.0243 | 1.0231 | 1.0061 | 1.0244 |
| GF/h = 2 | 1.0341 | 1.0350 | 1.0095 | 0.9950 | 0.9949 |
| GF/h = 5 | 1.0299 | 1.0313 | 0.9984 | 0.9948 | 0.9998 |
| GF/h = 22 | 1.0064 | 1.0073 | 1.0122 | 1.0081 | 1.0092 |
| HE/h = 1 | 1.0142 | 1.0142 | 0.9992 | 0.9995 | 0.9999 |
| HE/h = 2 | 0.9678 | 0.9633 | 0.9520 | 0.9543 | 0.9541 |
| HE/h = 5 | 1.0240 | 1.0219 | 1.0014 | 1.0031 | 1.0000 |
| HE/h = 22 | 0.9960 | 0.9929 | 0.9763 | 0.9725 | 0.9739 |
| KC/h = 1 | 1.0022 | 1.0025 | 1.0109 | 1.0151 | 1.0131 |
| KC/h = 2 | 1.0116 | 1.0101 | 1.0105 | 1.0132 | 1.0077 |
| KC/h = 5 | 1.0120 | 1.0126 | 1.0049 | 1.0021 | 1.0056 |
| KC/h = 22 | 0.9990 | 1.0019 | 0.9997 | 1.0000 | 0.9997 |
| LE/h = 1 | 1.0094 | 1.0105 | 0.9996 | 0.9977 | 0.9989 |
| LE/h = 2 | 1.0076 | 1.0080 | 0.9968 | 1.0058 | 0.9993 |
| LE/h = 5 | 1.0170 | 1.0172 | 1.0053 | 0.9960 | 1.0048 |
| LE/h = 22 | 1.0073 | 1.0090 | 1.0045 | 1.0056 | 1.0029 |
| OJ/h = 1 | 0.9977 | 1.0000 | 1.0120 | 1.0084 | 1.0159 |
| OJ/h = 2 | 0.9985 | 1.0005 | 1.0080 | 1.0042 | 1.0085 |
| OJ/h = 5 | 0.9988 | 1.0013 | 1.0073 | 1.0036 | 1.0067 |
| OJ/h = 22 | 0.9983 | 0.9990 | 1.0039 | 1.0028 | 1.0023 |
| RR/h = 1 | 1.0049 | 1.0055 | 1.0475 | 1.0490 | 1.0441 |
| RR/h = 2 | 1.0045 | 1.0056 | 0.9955 | 0.9919 | 0.9910 |
| RR/h = 5 | 1.0161 | 1.0167 | 1.0073 | 1.0072 | 1.0042 |
| RR/h = 22 | 0.9703 | 0.9717 | 1.0848 | 0.9703 | 0.9708 |
| S/h = 1 | 1.0163 | 1.0169 | 1.0037 | 1.0053 | 1.0050 |
| S/h = 2 | 1.0296 | 1.0306 | 1.0032 | 1.0061 | 1.0016 |
| S/h = 5 | 1.0451 | 1.0450 | 0.9983 | 1.0060 | 0.9988 |

**TABLE 9** (Continued)

| Commodity/horizon | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| S/h = 22 | 1.0193 | 1.0220 | 0.9993 | 1.0005 | 1.0021 |
| SB/h = 1 | 0.9941 | 0.9953 | 1.0569 | 0.9892 | 1.0444 |
| SB/h = 2 | 0.9731 | 0.9747 | 1.0946 | 0.9627 | 1.0923 |
| SB/h = 5 | 1.0188 | 1.0209 | 0.9941 | 0.9905 | 0.9910 |
| SB/h = 22 | 1.0110 | 1.0158 | 0.9956 | 0.9970 | 0.9956 |
| SM/h = 1 | 1.0112 | 1.0130 | 1.0072 | 1.0101 | 1.0055 |
| SM/h = 2 | 1.0364 | 1.0381 | 1.0063 | 1.0097 | 1.0042 |
| SM/h = 5 | 1.0488 | 1.0504 | 0.9963 | 1.0070 | 0.9991 |
| SM/h = 22 | 1.0264 | 1.0299 | 1.0018 | 1.0011 | 1.0020 |
| W/h = 1 | 1.0107 | 1.0120 | 1.0106 | 1.0039 | 1.0113 |
| W/h = 2 | 1.0147 | 1.0165 | 1.0058 | 1.0080 | 1.0067 |
| W/h = 5 | 1.0223 | 1.0254 | 1.0042 | 1.0066 | 1.0026 |
| W/h = 22 | 0.9988 | 1.0023 | 0.9965 | 0.9925 | 0.9924 |

*Note*: Columns: (1) The HAR-RV model versus the HAR-RV-Sentiment-MOBA-Moments model. (2) The HAR-RV-Sentiment model versus the HAR-RV-Sentiment-MOBA-Moments model. (3) The HAR-RV-Sentiment-Moments /Adj. $R^2$ model versus the HAR-RV-Sentiment-MOBA-Moments model. (4) The HAR-RV-Sentiment-Moments/BIC model versus HAR-RV-Sentiment-MOBA-Moments model. (5) The HAR-RV-Sentiment-Moments/BIC model versus the HAR-RV-Sentiment-MOBA-Moments model. (6) The HAR-RV-Sentiment-Moments/CP model versus the HAR-RV-Sentiment-MOBA-Moments model.

### 5.3 | Alternative measures of sentiment

In addition, we incorporate robustness checks by using different measures of sentiment. Instead of relying solely on the TRMI overall sentiment index, we concentrate on the 30 sub-sentiment indices described in Table A1 (excluding the overall sentiment). Our aim is to gain a more detailed understanding of sentiment patterns. To achieve this, we employ the forward stepwise predictor selection algorithm. Additionally, to address the complexity arising from the wide range of subindices, we employ principal component analysis (PCA) to transform these variables into a smaller set of uncorrelated variables. We reestimate the PCA as we shift the estimation window across the sample period. By extracting the first three principal components, which capture a significant portion of the original dataset's information, we obtain a more condensed and manageable dataset for our analysis.

The second alternative measure that we consider is the "Buzz" index. Unlike the sentiment variable, the Buzz index quantifies the total sum of all TRMI-contributing words for a particular agricultural commodity on a specific day. The Buzz index acts as an indicator of the total volume of sentiment-related words, thereby reflecting the extent of market participants' attention towards a particular commodity.

On applying these alternative measures, the results demonstrate notable consistency. The results for the "Buzz" index and the PCA are reported in Table A6. When we consider the adjusted $R^2$ statistic, BIC, and CP criterion for model selection, the $p$ values of the Clark–West tests (for the models with realized moments; that is, both HAR-RV-BUZZ-Moments and HAR-RV-PCA-Moments models) remain consistently low, indicating the statistical significance of our results.

The results for the forward stepwise sentiment selection algorithm are summarized in Table A7. Here, we compare HAR-RV models extended to include the sentiment subindices as selected by the algorithm with HAR-RV model constructed by applying the algorithm to the realized moments. The resulting models are not nested and so we compare them in terms of the MAFE and RMSFE ratios. Again, the results confirm the robustness of our findings and suggest that adding realized moments to the model improves the results in all cases, regardless of the alternative measures used.

## 6 | CONCLUDING REMARKS

Based on high-frequency data for the period from 2009 to 2020 for 14 important agricultural commodities, we have shown, using two different algorithms to compute forecasts of realized volatility (and realized variance), that realized moments like leverage, realized kurtosis, and realized jumps are more important drivers of the accuracy of forecasts of realized volatility in markets for agricultural commodities than sentiment, which contributes only little, if anything, beyond what the baseline

HAR-RV model produces in terms of forecast accuracy. Hence, investors, corporate decision makers, and policymakers should closely track realized moments rather than sentiment when they need to produce forecasts of the future realized volatility in markets for agricultural commodities. From the perspective of an econometrician, we confirm that in-sample results depicting predictability of behavioral factors on agricultural price volatility might not necessarily translate into forecasting gains, thus confirming the stringency of out-of-sample tests.

At this stage, it is important to discuss a possible underlying reasoning for our finding that realized moments matter more than sentiment in forecasting variability in agricultural commodities prices. Roache (2010), Johnson (2011), and Sujithan et al. (2014) suggest that the increased volatility in food prices observed post the GFC is primarily a result of rare disaster risks associated with climate change (with other important factors being the production of biofuels, market speculation, and also rising demand coupled with declines in food stocks). In this regard, we believe that the information contained in the realized moments actually captures these rare disaster events (Bouri et al., 2021; Demirer et al., 2022)[16] and translates into better forecast performance compared to sentiment. At the same time, the possibility that the dynamics of realized moments internalizes the information content of behavioral factors of agricultural commodity markets cannot be ruled out either, in light of disaster risks driving investor sentiments of in the asset markets (Manela & Moreira, 2017; Sakariyahu et al., 2023; Shan & Gong, 2012).

Our empirical analysis should help stimulating interesting future research. One avenue for future research is to go beyond the class of linear forecasting models that we have studied in our research, and look at machine learning approaches, such as random forests (see, e.g., Bonato et al., 2022). Random forests are tailored to recover nonlinearities in the data, and this may help to shed light on potential nonlinear links between sentiment and realized volatility. Staying within the context of machine learning, another interesting avenue for future research would be to analyze in more detail the predictive value of the disaggregated measures of various aspects of sentiment for the realized volatility of returns of agricultural commodity prices, as well as cross-market spillovers of sentiments and moments. Random forests could be utilized for such a forecasting experiment, given its ability to handle "big data" setups.

## AKCNOWLEDGMENT

We would like to thank an anonymous referee for many helpful comments. The usual disclaimer applies.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from Refinitiv Eikon. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from the author(s) with the permission of Refinitiv Eikon.

## ORCID

*Oguzhan Cepni* https://orcid.org/0000-0003-0711-8880
*Rangan Gupta* https://orcid.org/0000-0001-5002-3428

## ENDNOTES

[1] The TRMI incorporates textual analysis from the top 2,000 global news outlets and 800 global financial/social media sites.

[2] Framing our empirical analysis in terms of *RV* rather than a GARCH model has several other advantages. First, a preliminary full-sample analysis of our data (detailed results are available from the authors upon request) showed that *RV* indeed tracks squared returns better than a GARCH(1, 1) model. Second, the HAR-RV model can be estimated by the ordinary-least-squares technique, implying that we can efficiently estimate a large number of models in a reasonable amount of time. Third, the models can be easily analyzed by means of the predictor selection and MOBA algorithms, which makes the selection of the predictors straightforward. Fourth, the GARCH zoo is populated by a large variety of GARCH models, implying that it is necessary to select either a preferred model in advance or to estimate a large number of competing GARCH models, further inflating the number of models to be estimated. Fifth, integrating the type of weekly and monthly *RV* predictors characteristic of the HAR-RV model in a GARCH-style model would be difficult.

[3] See https://www.fao.org/faostat/en/#home for details.

[4] We emphasize that we mainly study realized volatility (the square root of realized variance) to mitigate the influence of the usual large peaks in *RV* on our results. We report results for the realized variance in Table 7 as a robustness check.

[5] For our empirical research, we utilize the R language and environment for statistical computing (R Core Team, 2023). We employ the "leaps" add-on package by Thomas Lumley and based on Fortran code by Miller (2020) to implement the best stepwise predictor selection algorithm.

[6] Using a validation sample rather than a training sample to assess how model predictions evolve from iteration to iteration increases the number of models to be estimated before convergence to a "stable" vector of average predicted realized volatilities is achieved. A larger number of models leads to a fuller extraction of the predictive value for realized volatility embedded in the realized moments and, on average across the agricultural commodities, to slightly more accurate forecasts. Importantly, irrespective of whether we use a validation or a training sample to assess convergence of the algorithm, the realized moments dominate sentiment as a predictor of subsequent realized volatility.

[7] The predictions from the MOBA algorithm, thus, can be interpreted as resulting from a "thick modeling" approach (Granger & Jeon, 2004). It also should be noted that combining the estimated models into a single model, averaging the coefficients of the various predictors, and then setting the resulting coefficients of the predictors to zero gives the nested model (e.g., HAR-RV model), so that the Clark–West test provides an appropriate framework for comparing the predictions from the MOBA algorithm and the nested model. In this regard, it also is important to note that our main finding (realized moments matter while sentiment does not) does not hinge on a single specific test. Rather, the RMSFE/MAFE ratios, the predictor selection algorithm (and also the elastic-net shrinkage estimator; see Section 5.1) all support our main finding (realized moments matter much more for forecasting performance than sentiment).

[8] When we forecast realized volatility, we take the square root of *RVB* and *RVG* to obtain the respective bad and good realized volatilities.

[9] This finding is consistent with results obtained using a rolling-estimation window (not included here due to space constraints but available from the authors upon request) and the forward stepwise predictor selection algorithm.

[10] The results in Column (1) of Table 5 are identical to those in Column (1) of Table 4 since no stepwise predictor selection is involved when estimating the baseline HAR-RV model and the HAR-RV-Sentiment model. Additionally, when examining the realized variance, we employ the realized "good" and realized "bad" variances as predictors.

[11] As an alternative to forward and backward stepwise predictor selectiond (and, hence, as another robustness check), we also studied a sequential replacement algorithm, which gave results (not reported for reasons of space but available upon request from the authors) that are qualitatively similar to those that we report in Table 4.

[12] The elastic-net model is estimated using the R add-on package "glmnet" (Friedman et al., 2010), with a mixing parameter set at 0.5 and employing tenfold cross-validation.

[13] Interestingly, the realized good and bad volatilities play a more prominent role with the elastic-net estimator than with the predictor selection algorithm. (Detailed results are not reported but are available upon request from the authors.)

[14] At this juncture in our analysis, it is vital to emphasize that we employ the predictor selection and MOBA algorithms primarily to demonstrate the robustness of our results, given the choice and details of the algorithm used to analyze our data. Consequently, our forecasting experiment is not a horserace between the two algorithms but rather a comparison between sentiment and realized moments. In this context, our results robustly indicate that realized moments play a more significant role in forecasting realized volatility than sentiment does.

[15] Also, in the appendix, we summarize in Table A5 MAFE ratios that we obtain when we consider a scenario in which $V \neq E$. Specifically, we set $E = 1{,}000$, ensuring the validation sample always starts at the beginning of the sample period. This means the validation sample expands recursively as we move the rolling-estimation window through the data. These results further support our primary conclusion: realized moments generally hold greater significance for forecasting performance than sentiment.

[16] The key assumption underlying rare-disaster models is that the entire universe of assets in an economy is exposed to an aggregate jump-risk factor (Barro, 2006, 2009), which, in turn, is known to be an important component of volatility (Giot et al., 2010).

## REFERENCES


Aït-Youcef, C. (2019). How index investment impacts commodities: A story about the financialization of agricultural commodities. *Economic Modelling*, *80*, 23–33.

Akyildirim, E., Çepni, O., Pham, L., & Uddin, G. S. (2022). How connected is the agricultural commodity market to the news-based investor sentiment? *Energy Economics*, *113*(C), 106174. https://doi.org/10.1016/j.eneco.2022.106174

Amaya, D., Christoffersen, P., Jacobs, K., & Vasquez, A. (2015). Does realized skewness predict the cross-section of equity returns? *Journal of Financial Economics*, *118*, 135–167. https://doi.org/10.1016/j.jfineco.2015.02.009

Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, *39*(4), 885–905. https://doi.org/10.2307/2527343

Bahloul, W. (2018). Short-term contrarian and sentiment by traders' types on futures markets: Evidence from the DCOT traders' positions. *Review of Behavioral Finance*, *10*(4), 298–319. https://doi.org/10.1108/RBF-07-2017-0063

Bahloul, W., Balcilar, M., Cunado, J., & Gupta, R. (2018). The role of economic and financial uncertainties in predicting commodity futures returns and volatility: Evidence from a nonparametric causality-in-quantiles test. *Journal of Multinational Financial Management*, *45*(C), 52–71. https://doi.org/10.1016/j.mulfin.2018.04.002

Bahloul, W., & Bouri, A. (2016). The impact of investor sentiment on returns and conditional volatility in US futures markets. *Journal of Multinational Financial Management*, *36*(C), 89–102. https://doi.org/10.1016/j.mulfin.2016.07.003

Balcilar, M., Sertoglu, K., & Agan, B. (2022). The COVID-19 effects on agricultural commodity markets. *Agrekon*, *61*(3), 239–265. https://doi.org/10.1080/03031853.2022.2078381

Barndorff-Nielsen, O. E., Kinnebrouk, S., & Shephard, N. (2010). Measuring downside risk: realised semivariance. In T. Bollerslev, J. Russell, & M. Watson (Eds.), *Volatility and time series econometrics: Essays in honor of Robert F. Engle* (pp. 117–136). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199549498.003.0007

Barndorff-Nielsen, O. E., & Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, *2*, 1–37. https://doi.org/10.1093/jjfinec/nbh001

Barndorff-Nielsen, O. E., & Shephard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics*, *4*, 1–30. https://doi.org/10.1093/jjfinec/nbi022

Barro, R. J. (2006). Rare disasters and asset markets in the twentieth century. *Quarterly Journal of Economics*, *121*(3), 823–866. https://doi.org/10.1162/qjec.121.3.823

Barro, R. J. (2009). Rare disasters, asset prices, and welfare costs. *American Economic Review*, *99*(1), 243–264. https://doi.org/10.1257/aer.99.1.243

Bonato, M. (2019). Realized correlations, betas and volatility spillover in the agricultural commodity market: What has changed? *Journal of International Financial Markets Institutions and Money*, *62*(C), 184–202. https://doi.org/10.1016/j.intfin.2019.07.005

Bonato, M., Çepni, O., Gupta, R., & Pierdzioch, C. (2022). El Niño, La Niña, and forecastability of the realized variance of agricultural commodity prices: Evidence from a machine learning approach. *Journal of Forecasting*, *42*, 785–801. https://doi.org/10.1002/for.2914

Bonato, M., Cepni, O., Gupta, R., & Pierdzioch, C. (2023). Climate risks and state-level stock-market realized volatility. *Journal of Financial Markets*, *66*, 100854.

Borgards, O., & Czudaj, R. L. (2022). Long-short speculator sentiment in agricultural commodity markets. *International Journal of Finance and Economics*, *28*, 3511–3528. https://doi.org/10.1002/ijfe.2605

Bouri, E., Gupta, R., Pierdzioch, C., & Salisu, A. A. (2021). El Niño and forecastability of oil-price realized volatility. *Theoretical and Applied Climatology*, *144*, 1173–1180. https://doi.org/10.1007/s00704-021-03569-1

Campbell, J. Y. (2008). Viewpoint: Estimating the equity premium. *Canadian Journal of Economics*, *41*(1), 1–21. https://doi.org/10.1111/j.1365-2966.2008.00453.x

Chatziantoniou, I., Degiannakis, S., Filis, G., & Lloyd, T. (2021). Oil price volatility is effective in predicting food price volatility. Or is it? *Energy Journal*, *42*, 2. https://doi.org/10.5547/01956574.42.6.icha

Clark, T. D., & West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, *138*, 291–311. https://doi.org/10.1016/j.jeconom.2006.05.023

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*, 174–196. https://doi.org/10.1093/jjfinec/nbp001

Degiannakis, S., Filis, G., Klein, T., & Walther, T. (2022). Forecasting realized volatility of agricultural commodities. *International Journal of Forecasting*, *38*(1), 74–96. https://doi.org/10.1016/j.ijforecast.2019.08.011

Demirer, R., Gupta, R., Nel, J., & Pierdzioch, C. (2022). Effect of rare disaster risks on crude oil: Evidence from El Niño from over 140 years of data. *Theoretical and Applied Climatology*, *147*, 691–699. https://doi.org/10.1007/s00704-021-03856-x

Friedman, J., Tibshirani, R., & Hastie, T. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22. https://doi.org/10.18637/jss.v033.i01

Gil-Alana, L. A., Cunado, J., & de Gracia, F. P. (2012). Persistence, long memory, and unit roots in commodity prices. *Canadian Journal of Agricultural Economics*, *60*(4), 451–468. https://doi.org/10.1111/j.1744-7976.2012.01253.x

Giot, P., & Laurent, S. (2003). Market risk in commodity markets: A VaR approach. *Energy Economics*, *25*(5), 435–457. https://doi.org/10.1016/S0140-9883(03)00052-5

Giot, P., Laurent, S., & Petitjean, M. (2010). Trading activity, realized volatility and jumps. *Journal of Empirical Finance*, *17*(1), 168–175. https://doi.org/10.1016/j.jempfin.2009.07.001

Granger, C. W., & Jeon, Y. (2004). Thick modeling. *Economic Modelling*, *21*(2), 323–343. https://doi.org/10.1016/S0264-9993(03)00017-8

Greb, F., & Prakash, A. (2015). *Has price volatility changed? Food outlook*. Food and Agriculture Organization of the United Nations.

Greb, F., & Prakash, A. (2017). Assessing volatility patterns in food crops. In FAO commodity and trade policy research working paper series. Available for download from: http://www.fao.org/3/a-i7066e.pdf.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning. Data mining, inference, and prediction* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, *3*, 1163–1173. https://doi.org/10.1214/aos/1176343247

Ji, Q., Bahloul, W., Geng, J.-B., & Gupta, R. (2020). Trading behaviour connectedness across commodity markets: Evidence from the hedgers' sentiment perspective. *Research in International Business and Finance*, *52*(C), 101114. https://doi.org/10.1016/j.ribaf.2019.101114

Johnson, T. (2011). Food price volatility and insecurity. Council on Foreign Relations (CFR). Available for download from: https://www.cfr.org/backgrounder/foodprice-volatility-and-insecurity.

Lumley, T. based on Fortran code by A. Miller (2020). leaps: Regression subset selection. R package version 3.1. Available for download from: https://CRAN.R-project.org/package=leaps.

Luo, J., Klein, T., Ji, Q., & Hou, C. (2019). Forecasting realized volatility of agricultural commodity futures with infinite hidden Markov HAR models. *International Journal of Forecasting*, *38*(1), 51–73. https://doi.org/10.1016/j.ijforecast.2019.08.007

Manela, A., & Moreira, A. (2017). News implied volatility and disaster concerns. *Journal of Financial Economics*, *123*(1), 137–162. https://doi.org/10.1016/j.jfineco.2016.01.032

Marfatia, H. A., Luo, J., & Ji, Q. (2022). Forecasting the volatility of agricultural commodity futures: The role of co-volatility and oil volatility. *Journal of Forecasting*, *41*(2), 383–404. https://doi.org/10.1002/for.2811

McAleer, M., & Medeiros, M. C. (2008). Realized volatility: A review. *Econometric Reviews*, *27*, 10–45. https://doi.org/10.1080/07474930701853509

Mišecka, T., Ciaian, P., Rajcaniova, M., & Pokrivcak, J. (2019). In search of attention in agricultural commodity markets. *Economics Letters*, *184*(3), 108668. https://doi.org/10.1016/j.econlet.2019.108668

Müller, U. A., Dacorogna, M. M., Davé, R. D., Olsen, R. B., & Pictet, O. V. (1997). Volatilities of different time resolutions—Analyzing the dynamics of market components. *Journal of Empirical Finance*, *4*, 213–239. https://doi.org/10.1016/S0927-5398(97)00007-8

Ordu, B. M., Oran, A., & Soytas, U. (2018). Is food financialized? Yes, but only when liquidity is abundant. *Journal of Banking* *and Finance*, *95*, 82–96. https://doi.org/10.1016/j.jbankfin.2017.06.001

R Core Team. (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Roache, S. K. (2010). What explains the rise in food price volatility? International Monetary Fund Working Paper No. 2010/129.

Sakariyahu, R., Lawal, R., Oyekola, O., Dosumu, O. E., & Adigun, R. (2023). Natural disasters, investor sentiments and stock market reactions: Evidence from Turkey–Syria earthquakes. *Economics Letters*, *228*, 111153. https://doi.org/10.1016/j.econlet.2023.111153

Shan, L., & Gong, S. X. (2012). Investor sentiment and stock returns: Wenchuan earthquake. *Finance Research Letters*, *9*(1), 36–47. https://doi.org/10.1016/j.frl.2011.07.002

Shiba, S., Aye, G. C., Gupta, R., & Goswami, S. (2022). Forecastability of agricultural commodity futures realised volatility with daily infectious disease-related uncertainty. *Journal of Risk and Financial Management*, *15*(11), 525. https://doi.org/10.3390/jrfm15110525

Sujithan, K. O., Avouyi-Dovi, S., & Koliai, L. (2014). On the determinants of food price volatility. International Conference on Food Price Volatility: Causes & Challenges, 25–26 February, 2014, Rabat. Available for download from: https://www.imf.org/external/np/seminars/eng/2014/food/pdf/Avouyi-Dovi.pdf.

Tian, F., Yang, K., & Chen, L. (2017a). Realized volatility forecasting of agricultural commodity futures using long memory and regime switching. *Journal of Forecasting*, *36*(4), 421–430. https://doi.org/10.1002/for.2443

Tian, F., Yang, K., & Chen, L. (2017b). Realized volatility forecasting of agricultural commodity futures using the HAR model with time-varying sparsity. *International Journal of Forecasting*, *33*(1), 132–152. https://doi.org/10.1016/j.ijforecast.2016.08.002

Xu, J.-L., & Hsu, Y.-L. (2022). The impact of news sentiment indicators on agricultural product prices. *Computational Economics*, *59*(2), 1645–1657. https://doi.org/10.1007/s10614-021-10189-4

Yang, K., Tian, F., Chen, L., & Li, S. (2017). Realized volatility forecast of agricultural futures using the HAR models with bagging and combination approaches. *International Review of Economics and Finance*, *49*, 276–291. https://doi.org/10.1016/j.iref.2017.01.030

Zhou, H., & Zhu, J. Q. (2012). An empirical examination of jump risk in asset pricing and volatility forecasting in China's equity and bond markets. *Pacific-Basin Finance Journal*, *20*(5), 857–880. https://doi.org/10.1016/j.pacfin.2009.05.005

Živkov, D., Njegić, J., & Pećanac, M. (2019). Multiscale interdependence between the major agricultural commodities. *Agricultural Economics – Czech*, *65*(2), 82–92. https://doi.org/10.17221/147/2018-AGRICECON

## AUTHOR BIOGRAPHIES

**Matteo Bonato** holds an MSc in Statistics from the University of Padua, Italy, and a PhD in Finance from the Swiss Finance Institute and the University of Zürich. He is a senior research associate at the Department of Economics and Econometrics at the University of Johannesburg, South Africa, and IPAG Busines School in Paris. Dr. Bonato has 11 years of experience in the financial industry. He began with an internship in the Research Department of the Swiss National Bank in Zürich. He then took positions in the Independent Model Validation units of Credit Suisse and UBS. He is a co-founder of Valdon Group, a Fintech start-up based in Zurich, Switzerland.

**Oguzhan Cepni** is a visiting researcher at Copenhagen Business School (CBS). His primary research interests include applied macroeconomics, empirical asset pricing, and forecasting.

**Rangan Gupta** is currently a professor at the Department of Economics, University of Pretoria, South Africa. He completed his PhD in May 2005 from the Department of Economics, University of Connecticut, United States. He then joined the Department of Economics, University of Pretoria, as a senior lecturer in August 2005. His academic interests are mainly monetary theory and policy and time series econometrics. He has published widely in international journals in these areas of research.

**Christian Pierdzioch** is a professor of Economics at Helmut-Schmidt-University Hamburg, Germany, Department of Economics. His research interests include macroeconomics, monetary economics, forecasting, financial econometrics, and sports economics.

**How to cite this article:** Bonato, M., Cepni, O., Gupta, R., & Pierdzioch, C. (2024). Forecasting the realized volatility of agricultural commodity prices: Does sentiment matter? *Journal of Forecasting*, *43*(6), 2088–2125. https://doi.org/10.1002/for.3106

# APPENDIX A

**TABLE A1** Description of TRMI indices.

| Index | Description: references in news and social media to… | Range |
|---|---|---|
| sentiment | Overall positive references, net of negative references | −1 to 1 |
| optimism | Optimism, net of references to pessimism | −1 to 1 |
| joy | Happiness and affection | 0 to 1 |
| loveHate | Love, net of references to hate | −1 to 1 |
| trust | Trustworthiness, net of references connoting corruption | −1 to 1 |
| anger | Anger and disgust | 0 to 1 |
| conflict | Disagreement and swearing net of agreement and conciliation | −1 to 1 |
| fear | Fear and anxiety | 0 to 1 |
| gloom | Gloom and negative future outlook | 0 to 1 |
| stress | Arousal and intensity, weighted towards distress | 0 to 1 |
| surprise | Unexpected events and surprise | 0 to 1 |
| timeUrgency | Urgency and timeliness, net of references to tardiness and delays | −1 to 1 |
| uncertainty | Uncertainty and confusion | 0 to 1 |
| violence | Violent crime, terrorism, and war | 0 to 1 |
| emotionVsFact | All emotional sentiments, net of all factual and topical references | −1 to 1 |
| longShort | Buying, net of references to shorting or selling | −1 to 1 |
| longShortForecast | Forecasts of buying, net of references to forecasts of shorting or selling | −1 to 1 |
| priceDirection | Price increases, net of references to price decreases | −1 to 1 |
| priceForecast | Forecasts of asset price rises, net of references to forecasts of asset price drops | −1 to 1 |
| volatility | Volatility in market prices or business conditions | 0 to 1 |
| consumptionVolume | Increasing, net of decreasing, commodity consumption | −1 to 1 |
| productionVolume | Increasing, net of decreasing, commodity production | −1 to 1 |
| regulatoryIssues | Regulatory changes affecting the commodity | 0 to 1 |
| supplyVsDemand | Surplus supply and lack of demand, net of references to supply | −1 to 1 |
| | Shortage and high demand | |
| supplyVsDemandForecast | Expectations of supply outstripping demand, net of references to | −1 to 1 |
| | Expectations of demand outstripping supply | |
| acreageCultivated | Increases in acreage and crop cultivation, net or references to decreases | −1 to 1 |
| | In acreage and crop cultivation | |
| agDisease | Commodity disease | 0 to 1 |
| agStress | Production stress related to disease, water, or weather | 0 to 1 |
| subsidies | Subsidies affecting commodity prices | 0 to 1 |
| subsidiesSentiment | Increases in subsidies, net of references to decreases in subsidies | -1 to 1 |
| weatherDamage | Commodity weather risk and damage | 0 to 1 |

*Note*: The TRMI values are normalized to fit the interval [0, 1] or [−1, 1]. For instance, while "optimism" may be expressed by numbers ranging from −1 to 1, depending on whether the predominate sentiments are optimistic or pessimistic, "Fear" conveys information regarding fear and anxiety, and as a result, it is represented by numbers ranging from 0 to 1.

TABLE A2 RMSFE ratios for an elastic net estimator (rolling-estimation window).

| Commodity/horizon | (1) | (2) | (3) |
|---|---|---|---|
| BO/h = 1 | 1.0021 | 1.0043 | 1.0022 |
| BO/h = 2 | 0.9999 | 1.0122 | 1.0123 |
| BO/h = 5 | 0.9996 | 1.0274 | 1.0278 |
| BO/h = 22 | 1.0007 | 1.0180 | 1.0173 |
| C/h = 1 | 0.9994 | 1.0214 | 1.0220 |
| C/h = 2 | 0.9994 | 1.0226 | 1.0232 |
| C/h = 5 | 0.9995 | 1.0275 | 1.0280 |
| C/h = 22 | 1.0002 | 1.0248 | 1.0246 |
| CC/h = 1 | 1.0008 | 1.0054 | 1.0046 |
| CC/h = 2 | 1.0015 | 1.0078 | 1.0062 |
| CC/h = 5 | 1.0007 | 1.0120 | 1.0114 |
| CC/h = 22 | 1.0018 | 1.0156 | 1.0138 |
| CT/h = 1 | 0.9996 | 1.0113 | 1.0118 |
| CT/h = 2 | 0.9997 | 1.0157 | 1.0160 |
| CT/h = 5 | 0.9996 | 1.0187 | 1.0191 |
| CT/h = 22 | 0.9973 | 1.0087 | 1.0114 |
| GF/h = 1 | 0.9994 | 1.0197 | 1.0203 |
| GF/h = 2 | 0.9997 | 1.0287 | 1.0290 |
| GF/h = 5 | 0.9986 | 1.0263 | 1.0278 |
| GF/h = 22 | 0.9967 | 1.0053 | 1.0086 |
| HE/h = 1 | 1.0013 | 1.0085 | 1.0071 |
| HE/h = 2 | 1.0029 | 1.0049 | 1.0020 |
| HE/h = 5 | 1.0023 | 1.0079 | 1.0056 |
| HE/h = 22 | 1.0039 | 1.0194 | 1.0155 |
| KC/h = 1 | 0.9989 | 1.0022 | 1.0034 |
| KC/h = 2 | 0.9995 | 1.0051 | 1.0056 |
| KC/h = 5 | 0.9996 | 1.0098 | 1.0102 |
| KC/h = 22 | 1.0003 | 0.9993 | 0.9990 |
| LE/h = 1 | 0.9995 | 1.0220 | 1.0225 |
| LE/h = 2 | 1.0001 | 1.0141 | 1.0140 |
| LE/h = 5 | 0.9996 | 1.0223 | 1.0228 |
| LE/h = 22 | 0.9981 | 1.0107 | 1.0127 |
| OJ/h = 1 | 0.9982 | 0.9955 | 0.9973 |
| OJ/h = 2 | 0.9983 | 1.0012 | 1.0029 |
| OJ/h = 5 | 0.9977 | 1.0016 | 1.0040 |
| OJ/h = 22 | 0.9992 | 0.9953 | 0.9961 |
| RR/h = 1 | 1.0001 | 1.0081 | 1.0080 |
| RR/h = 2 | 1.0002 | 1.0114 | 1.0112 |
| RR/h = 5 | 1.0001 | 1.0111 | 1.0110 |
| RR/h = 22 | 1.0018 | 1.0067 | 1.0049 |
| S/h = 1 | 0.9997 | 1.0215 | 1.0218 |
| S/h = 2 | 0.9984 | 1.0329 | 1.0346 |
| S/h = 5 | 0.9962 | 1.0481 | 1.0521 |

(Continues)

**TABLE A2** (Continued)

| Commodity/horizon | (1) | (2) | (3) |
|---|---|---|---|
| S/h = 22 | 0.9939 | 1.0207 | 1.0269 |
| SB/h = 1 | 0.9993 | 1.0037 | 1.0044 |
| SB/h = 2 | 0.9987 | 1.0115 | 1.0128 |
| SB/h = 5 | 0.9989 | 1.0126 | 1.0137 |
| SB/h = 22 | 1.0009 | 1.0228 | 1.0219 |
| SM/h = 1 | 0.9990 | 1.0173 | 1.0182 |
| SM/h = 2 | 0.9985 | 1.0403 | 1.0418 |
| SM/h = 5 | 0.9980 | 1.0499 | 1.0521 |
| SM/h = 22 | 0.9972 | 1.0184 | 1.0212 |
| W/h = 1 | 0.9989 | 1.0149 | 1.0160 |
| W/h = 2 | 0.9994 | 1.0222 | 1.0229 |
| W/h = 5 | 0.9996 | 1.0305 | 1.0308 |
| W/h = 22 | 0.9971 | 1.0006 | 1.0036 |

*Note*: Columns: (1) The HAR-RV model versus the HAR-RV-Sentiment model. (2) The HAR-RV model versus the HAR-RV-Sentiment-Moments model. (3) The HAR-RV-Sentiment model versus the HAR-RV-Sentiment-Moments model. The HAR-RV model and the HAR-RV-Sentiment model are estimated by means of the ordinary-least-squares technique, while the HAR-RV-Sentiment-Moments model is estimated by the elastic-net estimator.

**TABLE A3** MAFE ratios (MOBA algorithm/$V = E = 1{,}000$).

| Commodity/horizon | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| BO/h = 1 | 0.9858 | 1.0056 | 1.0099 | 1.0122 | 1.0121 |
| BO/h = 2 | 0.9924 | 1.0125 | 0.9996 | 1.0090 | 1.0045 |
| BO/h = 5 | 1.0068 | 1.0162 | 0.9989 | 1.0054 | 0.9999 |
| BO/h = 22 | 1.0101 | 1.0106 | 0.9902 | 1.0010 | 0.9953 |
| C/h = 1 | 1.0418 | 1.0411 | 0.9969 | 0.9853 | 0.9857 |
| C/h = 2 | 1.0229 | 1.0255 | 0.9953 | 0.9917 | 0.9956 |
| C/h = 5 | 1.0277 | 1.0320 | 0.9957 | 0.9984 | 0.9930 |
| C/h = 22 | 1.0224 | 1.0338 | 0.9959 | 1.0042 | 0.9955 |
| CC/h = 1 | 1.0135 | 1.0123 | 0.9949 | 0.9999 | 0.9956 |
| CC/h = 2 | 1.0247 | 1.0197 | 0.9946 | 1.0063 | 0.9957 |
| CC/h = 5 | 1.0084 | 1.0135 | 0.9980 | 1.0032 | 0.9993 |
| CC/h = 22 | 1.0081 | 1.0076 | 1.0003 | 1.0056 | 0.9999 |
| CT/h = 1 | 1.0059 | 1.0057 | 1.0012 | 1.0038 | 1.0034 |
| CT/h = 2 | 1.0052 | 1.0041 | 1.0021 | 1.0031 | 1.0008 |
| CT/h = 5 | 1.0099 | 1.0110 | 0.9997 | 1.0066 | 1.0012 |
| CT/h = 22 | 1.0026 | 1.0050 | 1.0031 | 0.9973 | 1.0025 |
| GF/h = 1 | 1.0109 | 1.0132 | 1.0076 | 1.0081 | 1.0084 |
| GF/h = 2 | 1.0045 | 1.0116 | 0.9978 | 0.9968 | 0.9967 |
| GF/h = 5 | 1.0007 | 1.0098 | 1.0015 | 0.9971 | 1.0013 |
| GF/h = 22 | 0.9800 | 1.0001 | 1.0056 | 1.0086 | 1.0080 |
| HE/h = 1 | 1.0283 | 1.0260 | 1.0110 | 1.0060 | 1.0074 |
| HE/h = 2 | 1.0305 | 1.0219 | 1.0000 | 0.9980 | 0.9951 |
| HE/h = 5 | 1.0424 | 1.0319 | 0.9954 | 0.9979 | 0.9929 |
| HE/h = 22 | 1.0337 | 1.0211 | 0.9978 | 0.9962 | 0.9968 |

TABLE A3 (Continued)

| Commodity/horizon | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| KC/h = 1 | 0.9996 | 1.0056 | 1.0056 | 1.0081 | 1.0048 |
| KC/h = 2 | 1.0004 | 1.0058 | 1.0038 | 1.0051 | 1.0055 |
| KC/h = 5 | 0.9958 | 1.0011 | 1.0027 | 0.9970 | 1.0007 |
| KC/h = 22 | 1.0084 | 1.0073 | 0.9996 | 0.9968 | 1.0000 |
| LE/h = 1 | 1.0155 | 1.0173 | 0.9964 | 0.9945 | 0.9960 |
| LE/h = 2 | 1.0045 | 1.0069 | 0.9990 | 0.9978 | 0.9988 |
| LE/h = 5 | 1.0099 | 1.0137 | 0.9963 | 1.0037 | 0.9932 |
| LE/h = 22 | 1.0060 | 1.0137 | 0.9998 | 0.9975 | 0.9991 |
| OJ/h = 1 | 0.9910 | 0.9916 | 1.0075 | 1.0035 | 1.0070 |
| OJ/h = 2 | 0.9933 | 0.9942 | 1.0124 | 1.0060 | 1.0090 |
| OJ/h = 5 | 1.0045 | 1.0051 | 1.0079 | 1.0065 | 1.0055 |
| OJ/h = 22 | 0.9972 | 0.9964 | 1.0011 | 0.9978 | 1.0000 |
| RR/h = 1 | 1.0114 | 1.0113 | 0.9985 | 0.9916 | 0.9981 |
| RR/h = 2 | 1.0143 | 1.0140 | 1.0028 | 1.0017 | 1.0028 |
| RR/h = 5 | 1.0204 | 1.0155 | 0.9857 | 0.9944 | 0.9899 |
| RR/h = 22 | 1.0119 | 1.0113 | 0.9985 | 0.9928 | 1.0007 |
| S/h = 1 | 1.0259 | 1.0277 | 1.0019 | 1.0072 | 1.0011 |
| S/h = 2 | 1.0366 | 1.0394 | 1.0099 | 1.0226 | 1.0104 |
| S/h = 5 | 1.0324 | 1.0360 | 1.0082 | 1.0176 | 1.0071 |
| S/h = 22 | 1.0269 | 1.0305 | 1.0003 | 0.9968 | 0.9949 |
| SB/h = 1 | 1.0306 | 1.0323 | 1.0028 | 1.0071 | 1.0034 |
| SB/h = 2 | 1.0144 | 1.0161 | 0.9921 | 1.0003 | 0.9961 |
| SB/h = 5 | 1.0191 | 1.0209 | 0.9924 | 0.9953 | 0.9917 |
| SB/h = 22 | 1.0231 | 1.0223 | 0.9914 | 1.0045 | 0.9937 |
| SM/h = 1 | 1.0291 | 1.0287 | 1.0043 | 1.0020 | 1.0050 |
| SM/h = 2 | 1.0429 | 1.0436 | 1.0006 | 1.0060 | 1.0008 |
| SM/h = 5 | 1.0371 | 1.0375 | 1.0109 | 1.0091 | 1.0109 |
| SM/h = 22 | 1.0020 | 1.0003 | 1.0100 | 1.0050 | 1.0028 |
| W/h = 1 | 1.0157 | 1.0241 | 0.9949 | 0.9947 | 0.9936 |
| W/h = 2 | 1.0174 | 1.0235 | 0.9999 | 0.9995 | 1.0011 |
| W/h = 5 | 1.0299 | 1.0349 | 0.9911 | 0.9943 | 0.9912 |
| W/h = 22 | 1.0001 | 1.0121 | 0.9926 | 0.9997 | 0.9946 |

*Note*: Columns: (1) The HAR-RV model versus the HAR-RV-Sentiment-MOBA-Moments model. (2) The HAR-RV-Sentiment model versus the HAR-RV-Sentiment-MOBA-Moments model. (3) The HAR-RV-Sentiment-Moments/Adj. $R^2$ model versus the HAR-RV-Sentiment-MOBA-Moments model. (4) The HAR-RV-Sentiment-Moments/BIC model versus HAR-RV-Sentiment-MOBA-Moments model. (5) The HAR-RV-Sentiment-Moments/BIC model versus the HAR-RV-Sentiment-MOBA-Moments model. (6) The HAR-RV-Sentiment-Moments/CP model versus the HAR-RV-Sentiment-MOBA-Moments model.

**TABLE A4** MAFE ratios (MOBA algorithm/$V = E = 500$).

| Commodity/horizon | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| BO/h = 1 | 1.0022 | 1.0059 | 1.0073 | 1.0124 | 1.0035 |
| BO/h = 2 | 1.0035 | 1.0078 | 1.0047 | 1.0035 | 1.0028 |
| BO/h = 5 | 1.0203 | 1.0178 | 0.9998 | 1.0063 | 1.0016 |
| BO/h = 22 | 1.0140 | 1.0104 | 0.9970 | 1.0011 | 0.9980 |
| C/h = 1 | 1.0303 | 1.0313 | 1.0166 | 1.0106 | 1.0149 |
| C/h = 2 | 1.0211 | 1.0222 | 1.0088 | 1.0046 | 1.0087 |
| C/h = 5 | 1.0222 | 1.0234 | 1.0052 | 1.0068 | 1.0080 |
| C/h = 22 | 1.0207 | 1.0224 | 0.9997 | 1.0006 | 1.0017 |
| CC/h = 1 | 1.0066 | 1.0044 | 1.0091 | 1.0094 | 1.0060 |
| CC/h = 2 | 1.0099 | 1.0085 | 1.0042 | 1.0129 | 1.0051 |
| CC/h = 5 | 1.0021 | 1.0020 | 0.9974 | 1.0022 | 1.0006 |
| CC/h = 22 | 1.0077 | 1.0075 | 1.0024 | 1.0126 | 1.0043 |
| CT/h = 1 | 0.9964 | 0.9989 | 1.0097 | 0.9937 | 1.0146 |
| CT/h = 2 | 1.0028 | 1.0070 | 1.0057 | 0.9990 | 1.0047 |
| CT/h = 5 | 1.0060 | 1.0098 | 1.0020 | 1.0056 | 1.0032 |
| CT/h = 22 | 0.9997 | 1.0035 | 1.0049 | 1.0014 | 1.0018 |
| GF/h = 1 | 1.0133 | 1.0157 | 1.0137 | 1.0107 | 1.0164 |
| GF/h = 2 | 1.0168 | 1.0185 | 1.0063 | 1.0039 | 1.0034 |
| GF/h = 5 | 1.0132 | 1.0147 | 1.0052 | 1.0038 | 1.0040 |
| GF/h = 22 | 1.0013 | 1.0035 | 1.0130 | 1.0087 | 1.0123 |
| HE/h = 1 | 1.0142 | 1.0179 | 0.9972 | 0.9959 | 0.9958 |
| HE/h = 2 | 1.0111 | 1.0076 | 0.9881 | 0.9907 | 0.9890 |
| HE/h = 5 | 1.0228 | 1.0203 | 0.9936 | 0.9962 | 0.9931 |
| HE/h = 22 | 1.0094 | 1.0062 | 0.9973 | 0.9929 | 0.9950 |
| KC/h = 1 | 1.0028 | 1.0044 | 1.0094 | 1.0141 | 1.0126 |
| KC/h = 2 | 1.0114 | 1.0098 | 1.0126 | 1.0092 | 1.0104 |
| KC/h = 5 | 1.0121 | 1.0144 | 1.0086 | 1.0034 | 1.0074 |
| KC/h = 22 | 1.0019 | 1.0051 | 0.9987 | 0.9988 | 0.9990 |
| LE/h = 1 | 1.0056 | 1.0096 | 1.0100 | 1.0035 | 1.0057 |
| LE/h = 2 | 1.0083 | 1.0099 | 1.0054 | 1.0085 | 1.0101 |
| LE/h = 5 | 1.0159 | 1.0166 | 1.0040 | 1.0004 | 1.0032 |
| LE/h = 22 | 1.0154 | 1.0166 | 1.0019 | 1.0070 | 1.0008 |
| OJ/h = 1 | 0.9950 | 0.9973 | 1.0142 | 1.0070 | 1.0139 |
| OJ/h = 2 | 0.9958 | 0.9980 | 1.0096 | 1.0009 | 1.0080 |
| OJ/h = 5 | 0.9979 | 1.0010 | 1.0047 | 1.0014 | 1.0052 |
| OJ/h = 22 | 1.0020 | 1.0007 | 1.0035 | 1.0034 | 1.0012 |
| RR/h = 1 | 1.0035 | 1.0055 | 1.0204 | 1.0129 | 1.0172 |
| RR/h = 2 | 1.0057 | 1.0071 | 1.0025 | 1.0016 | 0.9992 |
| RR/h = 5 | 1.0169 | 1.0165 | 1.0043 | 1.0085 | 0.9997 |
| RR/h = 22 | 1.0015 | 0.9993 | 1.0089 | 0.9934 | 0.9959 |
| S/h = 1 | 1.0287 | 1.0323 | 1.0061 | 1.0121 | 1.0084 |
| S/h = 2 | 1.0344 | 1.0370 | 1.0077 | 1.0074 | 1.0067 |
| S/h = 5 | 1.0509 | 1.0521 | 1.0035 | 1.0074 | 1.0033 |

**TABLE A4** (Continued)

| Commodity/horizon | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| S/h = 22 | 1.0102 | 1.0203 | 0.9994 | 0.9960 | 1.0009 |
| SB/h = 1 | 1.0097 | 1.0110 | 1.0092 | 0.9966 | 1.0064 |
| SB/h = 2 | 1.0066 | 1.0087 | 1.0134 | 0.9931 | 1.0086 |
| SB/h = 5 | 1.0148 | 1.0162 | 1.0025 | 0.9935 | 0.9983 |
| SB/h = 22 | 1.0064 | 1.0099 | 0.9939 | 0.9942 | 0.9940 |
| SM/h = 1 | 1.0233 | 1.0272 | 1.0138 | 1.0165 | 1.0102 |
| SM/h = 2 | 1.0379 | 1.0387 | 1.0096 | 1.0121 | 1.0091 |
| SM/h = 5 | 1.0477 | 1.0525 | 1.0059 | 1.0127 | 1.0067 |
| SM/h = 22 | 1.0252 | 1.0313 | 1.0035 | 1.0090 | 1.0047 |
| W/h = 1 | 1.0106 | 1.0158 | 1.0066 | 1.0037 | 1.0054 |
| W/h = 2 | 1.0051 | 1.0100 | 1.0051 | 1.0054 | 1.0056 |
| W/h = 5 | 1.0217 | 1.0257 | 1.0017 | 1.0050 | 1.0011 |
| W/h = 22 | 1.0022 | 1.0078 | 0.9958 | 0.9950 | 0.9938 |

*Note*: Columns: (1) The HAR-RV model versus the HAR-RV-Sentiment-MOBA-Moments model. (2) The HAR-RV-Sentiment model versus the HAR-RV-Sentiment-MOBA-Moments model. (3) The HAR-RV-Sentiment-Moments/Adj. $R^2$ model versus the HAR-RV-Sentiment-MOBA-Moments model. (4) The HAR-RV-Sentiment-Moments/BIC model versus HAR-RV-Sentiment-MOBA-Moments model. (5) The HAR-RV-Sentiment-Moments/BIC model versus the HAR-RV-Sentiment-MOBA-Moments model. (6) The HAR-RV-Sentiment-Moments/CP model versus the HAR-RV-Sentiment-MOBA-Moments model.

**TABLE A5** MAFE ratios (MOBA algorithm/$E = 500$, expanding $V$).

| Commodity/horizon | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| BO/h = 1 | 0.9863 | 1.0061 | 1.0104 | 1.0128 | 1.0127 |
| BO/h = 2 | 0.9932 | 1.0133 | 1.0004 | 1.0097 | 1.0052 |
| BO/h = 5 | 1.0089 | 1.0182 | 1.0009 | 1.0074 | 1.0019 |
| BO/h = 22 | 1.0158 | 1.0162 | 0.9957 | 1.0066 | 1.0009 |
| C/h = 1 | 1.0422 | 1.0415 | 0.9972 | 0.9856 | 0.9861 |
| C/h = 2 | 1.0229 | 1.0255 | 0.9953 | 0.9917 | 0.9956 |
| C/h = 5 | 1.0284 | 1.0326 | 0.9963 | 0.9991 | 0.9937 |
| C/h = 22 | 1.0220 | 1.0334 | 0.9955 | 1.0039 | 0.9951 |
| CC/h = 1 | 1.0143 | 1.0131 | 0.9956 | 1.0007 | 0.9963 |
| CC/h = 2 | 1.0236 | 1.0186 | 0.9936 | 1.0053 | 0.9947 |
| CC/h = 5 | 1.0060 | 1.0111 | 0.9956 | 1.0008 | 0.9969 |
| CC/h = 22 | 1.0103 | 1.0098 | 1.0025 | 1.0078 | 1.0021 |
| CT/h = 1 | 1.0077 | 1.0075 | 1.0029 | 1.0055 | 1.0052 |
| CT/h = 2 | 1.0069 | 1.0057 | 1.0037 | 1.0048 | 1.0024 |
| CT/h = 5 | 1.0103 | 1.0115 | 1.0001 | 1.0070 | 1.0017 |
| CT/h = 22 | 1.0025 | 1.0049 | 1.0030 | 0.9972 | 1.0023 |
| GF/h = 1 | 1.0104 | 1.0127 | 1.0071 | 1.0076 | 1.0079 |
| GF/h = 2 | 1.0053 | 1.0125 | 0.9986 | 0.9976 | 0.9975 |
| GF/h = 5 | 1.0035 | 1.0126 | 1.0043 | 0.9999 | 1.0041 |
| GF/h = 22 | 0.9812 | 1.0013 | 1.0069 | 1.0098 | 1.0093 |
| HE/h = 1 | 1.0281 | 1.0257 | 1.0108 | 1.0057 | 1.0071 |
| HE/h = 2 | 1.0361 | 1.0274 | 1.0054 | 1.0033 | 1.0005 |
| HE/h = 5 | 1.0433 | 1.0328 | 0.9962 | 0.9987 | 0.9937 |

(Continues)

**TABLE A5** (Continued)

| Commodity/horizon | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| HE/h = 22 | 1.0359 | 1.0233 | 1.0000 | 0.9983 | 0.9989 |
| KC/h = 1 | 0.9960 | 1.0021 | 1.0020 | 1.0045 | 1.0012 |
| KC/h = 2 | 1.0043 | 1.0097 | 1.0077 | 1.0090 | 1.0094 |
| KC/h = 5 | 0.9964 | 1.0017 | 1.0033 | 0.9976 | 1.0013 |
| KC/h = 22 | 1.0075 | 1.0064 | 0.9986 | 0.9959 | 0.9990 |
| LE/h = 1 | 1.0185 | 1.0203 | 0.9993 | 0.9974 | 0.9989 |
| LE/h = 2 | 1.0034 | 1.0058 | 0.9979 | 0.9967 | 0.9977 |
| LE/h = 5 | 1.0108 | 1.0147 | 0.9972 | 1.0046 | 0.9941 |
| LE/h = 22 | 1.0052 | 1.0129 | 0.9990 | 0.9967 | 0.9983 |
| OJ/h = 1 | 0.9900 | 0.9907 | 1.0066 | 1.0026 | 1.0061 |
| OJ/h = 2 | 0.9936 | 0.9944 | 1.0127 | 1.0063 | 1.0093 |
| OJ/h = 5 | 1.0014 | 1.0020 | 1.0048 | 1.0034 | 1.0025 |
| OJ/h = 22 | 0.9979 | 0.9972 | 1.0019 | 0.9986 | 1.0007 |
| RR/h = 1 | 1.0101 | 1.0100 | 0.9972 | 0.9903 | 0.9968 |
| RR/h = 2 | 1.0145 | 1.0143 | 1.0030 | 1.0019 | 1.0031 |
| RR/h = 5 | 1.0196 | 1.0148 | 0.9850 | 0.9937 | 0.9891 |
| RR/h = 22 | 1.0132 | 1.0125 | 0.9997 | 0.9940 | 1.0019 |
| S/h = 1 | 1.0257 | 1.0275 | 1.0017 | 1.0071 | 1.0009 |
| S/h = 2 | 1.0354 | 1.0382 | 1.0088 | 1.0214 | 1.0092 |
| S/h = 5 | 1.0337 | 1.0373 | 1.0095 | 1.0189 | 1.0084 |
| S/h = 22 | 1.0235 | 1.0271 | 0.9970 | 0.9935 | 0.9916 |
| SB/h = 1 | 1.0235 | 1.0253 | 0.9959 | 1.0002 | 0.9965 |
| SB/h = 2 | 1.0191 | 1.0208 | 0.9967 | 1.0049 | 1.0007 |
| SB/h = 5 | 1.0231 | 1.0250 | 0.9964 | 0.9993 | 0.9956 |
| SB/h = 22 | 1.0222 | 1.0213 | 0.9905 | 1.0036 | 0.9928 |
| SM/h = 1 | 1.0291 | 1.0287 | 1.0043 | 1.0020 | 1.0050 |
| SM/h = 2 | 1.0398 | 1.0405 | 0.9977 | 1.0030 | 0.9979 |
| SM/h = 5 | 1.0348 | 1.0352 | 1.0087 | 1.0069 | 1.0087 |
| SM/h = 22 | 1.0017 | 1.0000 | 1.0097 | 1.0047 | 1.0025 |
| W/h = 1 | 1.0146 | 1.0230 | 0.9938 | 0.9936 | 0.9925 |
| W/h = 2 | 1.0151 | 1.0211 | 0.9976 | 0.9973 | 0.9988 |
| W/h = 5 | 1.0310 | 1.0359 | 0.9921 | 0.9953 | 0.9922 |
| W/h = 22 | 0.9980 | 1.0100 | 0.9905 | 0.9976 | 0.9926 |

*Note*: Columns: (1) The HAR-RV model versus the HAR-RV-Sentiment-MOBA-Moments model. (2) The HAR-RV-Sentiment model versus the HAR-RV-Sentiment-MOBA-Moments model. (3) The HAR-RV-Sentiment-Moments/Adj. $R^2$ model versus the HAR-RV-Sentiment-MOBA-Moments model. (4) The HAR-RV-Sentiment-Moments/BIC model versus HAR-RV-Sentiment-MOBA-Moments model. (5) The HAR-RV-Sentiment-Moments/BIC model versus the HAR-RV-Sentiment-MOBA-Moments model. (6) The HAR-RV-Sentiment-Moments/CP model versus the HAR-RV-Sentiment-MOBA-Moments model.

TABLE A6 Clark–West tests for alternative measures of sentiment (forward stepwise predictor selection/recursive-estimation window).

| Commodity/horizon | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| BO/h = 1 | 0.0001 | 0.0051 | 0.0001 | 0.0001 | 0.0035 | 0.0001 |
| BO/h = 2 | 0.0001 | 0.0010 | 0.0001 | 0.0001 | 0.0005 | 0.0001 |
| BO/h = 5 | 0.0000 | 0.0001 | 0.0001 | 0.0000 | 0.0000 | 0.0000 |
| BO/h = 22 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| C/h = 1 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| C/h = 2 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| C/h = 5 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| C/h = 22 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| CC/h = 1 | 0.0000 | 0.0001 | 0.0000 | 0.0000 | 0.0001 | 0.0000 |
| CC/h = 2 | 0.0000 | 0.0047 | 0.0001 | 0.0000 | 0.0025 | 0.0000 |
| CC/h = 5 | 0.0000 | 0.0042 | 0.0000 | 0.0000 | 0.0086 | 0.0000 |
| CC/h = 22 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| CT/h = 1 | 0.0046 | 0.0040 | 0.0034 | 0.0045 | 0.0040 | 0.0033 |
| CT/h = 2 | 0.0026 | 0.0039 | 0.0028 | 0.0026 | 0.0040 | 0.0030 |
| CT/h = 5 | 0.0057 | 0.0101 | 0.0061 | 0.0058 | 0.0095 | 0.0057 |
| CT/h = 22 | 0.0062 | 0.0121 | 0.0057 | 0.0072 | 0.0125 | 0.0065 |
| HE/h = 1 | 0.0000 | 0.0001 | 0.0000 | 0.0000 | 0.0003 | 0.0000 |
| HE/h = 2 | 0.0000 | 0.0001 | 0.0000 | 0.0000 | 0.0004 | 0.0001 |
| HE/h = 5 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| HE/h = 22 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| KC/h = 1 | 0.0002 | 0.0008 | 0.0002 | 0.0003 | 0.0012 | 0.0003 |
| KC/h = 2 | 0.0001 | 0.0001 | 0.0001 | 0.0002 | 0.0003 | 0.0003 |
| KC/h = 5 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0000 | 0.0000 |
| KC/h = 22 | 0.1127 | 0.0233 | 0.0933 | 0.1780 | 0.0397 | 0.1344 |
| LE/h = 1 | 0.0056 | 0.0198 | 0.0079 | 0.0081 | 0.0264 | 0.0113 |
| LE/h = 2 | 0.0047 | 0.0104 | 0.0066 | 0.0053 | 0.0104 | 0.0071 |
| LE/h = 5 | 0.0001 | 0.0003 | 0.0002 | 0.0001 | 0.0003 | 0.0002 |
| LE/h = 22 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 |
| OJ/h = 1 | 0.0354 | 0.0793 | 0.0452 | 0.0377 | 0.1211 | 0.0506 |
| OJ/h = 2 | 0.0235 | 0.2167 | 0.0256 | 0.0310 | 0.2105 | 0.0305 |
| OJ/h = 5 | 0.0236 | 0.0305 | 0.0271 | 0.0193 | 0.0208 | 0.0185 |
| OJ/h = 22 | 0.1512 | 0.3117 | 0.2349 | 0.0770 | 0.1327 | 0.0977 |
| RR/h = 1 | 0.0499 | 0.1509 | 0.0716 | 0.0474 | 0.1280 | 0.0640 |
| RR/h = 2 | 0.0495 | 0.2458 | 0.0662 | 0.0503 | 0.2339 | 0.0631 |
| RR/h = 5 | 0.0514 | 0.3324 | 0.0117 | 0.0579 | 0.2793 | 0.0722 |
| RR/h = 22 | 0.0004 | 0.1224 | 0.0004 | 0.0277 | 0.0931 | 0.0004 |
| S/h = 1 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| S/h = 2 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| S/h = 5 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| S/h = 22 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| SB/h = 1 | 0.0243 | 0.0411 | 0.0279 | 0.0217 | 0.0374 | 0.0247 |
| SB/h = 2 | 0.0197 | 0.0281 | 0.0234 | 0.0172 | 0.0230 | 0.0201 |
| SB/h = 5 | 0.0167 | 0.0306 | 0.0151 | 0.0136 | 0.0246 | 0.0129 |

(Continues)

**TABLE A6** (Continued)

| Commodity/horizon | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| SB/h = 22 | 0.0029 | 0.0117 | 0.0025 | 0.0024 | 0.0091 | 0.0022 |

*Note*: The table summarizes *p* values (based on robust standard errors) of the Clark–West test. Columns: (1) The HAR-RV-BUZZ model versus the HAR-RV-BUZZ-Moments model (selected by means of the adjusted $R^2$ statistic). (2) The HAR-RV-BUZZ model versus the HAR-RV-BUZZ-Moments model (selected by means of the BIC). (3) The HAR-RV-BUZZ model versus the HAR-RV-BUZZ-Moments model (selected by means of the CP criterion). (4) The HAR-RV-PCA model versus the HAR-RV-PCA-Moments model (selected by means of the adjusted $R^2$ statistic). (5) The HAR-RV-PCA model versus the HAR-RV-PCA-Moments model (selected by means of the BIC). (6) The HAR-RV-PCA model versus the HAR-RV-PCA-Moments model (selected by means of the CP criterion). PCA: First three principal components.

**TABLE A7** MAFE and RMSFE ratios for a HAR-RV-sentiment-subindices model versus a HAR-RV-moments model (forward stepwise predictor selection/recursive-estimation window).

| Commodity/horizon | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| BO/h = 1 | 1.0154 | 0.9954 | 1.0151 | 1.0174 | 1.0022 | 1.0179 |
| BO/h = 2 | 1.0156 | 0.9974 | 1.0115 | 1.0236 | 1.0083 | 1.0190 |
| BO/h = 5 | 1.0276 | 1.0073 | 1.0233 | 1.0383 | 1.0189 | 1.0350 |
| BO/h = 22 | 1.0207 | 1.0165 | 1.0175 | 1.0275 | 1.0183 | 1.0244 |
| C/h = 1 | 1.0399 | 1.0278 | 1.0411 | 1.0344 | 1.0168 | 1.0291 |
| C/h = 2 | 1.0411 | 1.0251 | 1.0336 | 1.0391 | 1.0226 | 1.0326 |
| C/h = 5 | 1.0447 | 1.0237 | 1.0337 | 1.0425 | 1.0272 | 1.0360 |
| C/h = 22 | 1.0225 | 1.0290 | 1.0281 | 1.0381 | 1.0389 | 1.0414 |
| CC/h = 1 | 1.0198 | 1.0126 | 1.0177 | 1.0179 | 1.0111 | 1.0162 |
| CC/h = 2 | 1.0219 | 1.0111 | 1.0185 | 1.0193 | 1.0082 | 1.0178 |
| CC/h = 5 | 1.0340 | 1.0087 | 1.0274 | 1.0292 | 1.0101 | 1.0258 |
| CC/h = 22 | 1.0289 | 1.0172 | 1.0258 | 1.0329 | 1.0212 | 1.0310 |
| CT/h = 1 | 1.0214 | 1.0077 | 1.0144 | 1.0264 | 1.0166 | 1.0222 |
| CT/h = 2 | 1.0173 | 1.0032 | 1.0114 | 1.0272 | 1.0150 | 1.0230 |
| CT/h = 5 | 1.0340 | 1.0118 | 1.0322 | 1.0350 | 1.0199 | 1.0332 |
| CT/h = 22 | 1.0368 | 1.0163 | 1.0318 | 1.0632 | 1.0353 | 1.0580 |
| HE/h = 1 | 1.0805 | 1.0580 | 1.0733 | 1.0404 | 1.0417 | 1.0395 |
| HE/h = 2 | 1.0509 | 1.0374 | 1.0435 | 1.0316 | 1.0182 | 1.0276 |
| HE/h = 5 | 1.0627 | 1.0442 | 1.0532 | 1.0363 | 1.0310 | 1.0321 |
| HE/h = 22 | 1.0448 | 1.0368 | 1.0452 | 1.0398 | 1.0360 | 1.0388 |
| KC/h = 1 | 1.0160 | 1.0039 | 1.0109 | 1.0153 | 1.0035 | 1.0101 |
| KC/h = 2 | 1.0232 | 1.0054 | 1.0197 | 1.0195 | 1.0044 | 1.0151 |
| KC/h = 5 | 1.0345 | 1.0165 | 1.0325 | 1.0235 | 1.0162 | 1.0217 |
| KC/h = 22 | 1.0277 | 1.0109 | 1.0232 | 1.0241 | 1.0115 | 1.0201 |
| LE/h = 1 | 1.0162 | 1.0144 | 1.0148 | 1.0162 | 1.0121 | 1.0148 |
| LE/h = 2 | 1.0133 | 1.0045 | 1.0110 | 1.0234 | 1.0163 | 1.0220 |
| LE/h = 5 | 1.0238 | 1.0116 | 1.0219 | 1.0344 | 1.0292 | 1.0334 |
| LE/h = 22 | 1.0366 | 1.0224 | 1.0316 | 1.0288 | 1.0231 | 1.0264 |
| OJ/h = 1 | 1.0059 | 1.0029 | 1.0002 | 1.0207 | 1.0109 | 1.0173 |
| OJ/h = 2 | 1.0167 | 0.9990 | 1.0110 | 1.0197 | 1.0002 | 1.0143 |
| OJ/h = 5 | 1.0250 | 0.9994 | 1.0082 | 1.0280 | 1.0052 | 1.0110 |
| OJ/h = 22 | 1.0347 | 1.0075 | 1.0252 | 1.0457 | 1.0093 | 1.0384 |
| RR/h = 1 | 1.0166 | 1.0120 | 1.0158 | 1.0160 | 1.0041 | 1.0120 |

TABLE A7 (Continued)

| Commodity/horizon | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| RR/h = 2 | 1.0272 | 1.0111 | 1.0214 | 1.0216 | 1.0043 | 1.0184 |
| RR/h = 5 | 1.0241 | 1.0076 | 1.0137 | 1.0268 | 1.0040 | 1.0119 |
| RR/h = 22 | 1.0174 | 1.0124 | 1.0155 | 1.0263 | 1.0082 | 1.0182 |
| S/h = 1 | 1.0629 | 1.0396 | 1.0600 | 1.0321 | 1.0245 | 1.0310 |
| S/h = 2 | 1.0598 | 1.0515 | 1.0618 | 1.0540 | 1.0425 | 1.0548 |
| S/h = 5 | 1.0481 | 1.0425 | 1.0471 | 1.0644 | 1.0505 | 1.0642 |
| S/h = 22 | 0.9856 | 0.9985 | 0.9894 | 1.0223 | 1.0234 | 1.0234 |
| SB/h = 1 | 1.0251 | 1.0153 | 1.0182 | 1.0252 | 1.0248 | 1.0235 |
| SB/h = 2 | 1.0358 | 1.0301 | 1.0365 | 1.0394 | 1.0361 | 1.0398 |
| SB/h = 5 | 1.0403 | 1.0324 | 1.0403 | 1.0517 | 1.0460 | 1.0532 |
| SB/h = 22 | 1.0364 | 1.0379 | 1.0416 | 1.0560 | 1.0514 | 1.0602 |

*Note*: Columns (1)–(3): MAFE ratios for the HAR-RV-Sentiment-Subindices model versus HAR-RV-Moments model (both selected by means of the adjusted $R^2$ statistic, BIC, and CP criterion). Columns (4)–(6) RMSFE ratio for the HAR-RV-Sentiment-Subindices model versus HAR-RV-Moments model (both selected by means of the adjusted $R^2$ statistic, BIC, and CP criterion).