ARTICLE IN PRESS
International Journal of Forecasting xxx (xxxx) xxx

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Do oil price forecast disagreement of survey of professional forecasters predict crude oil return volatility?[☆]

Anton Hasselgren[a], Ai Jun Hou[a], Sandy Suardi[b,*], Caihong Xu[a], Xiaoxia Ye[c]

[a] Stockholm Business School, Stockholm University, Sweden
[b] School of Business, University of Wollongong, Australia
[c] University of Exeter Business School, United Kingdom

ARTICLE INFO

Keywords:
Crude oil market
GARCH-MIDAS
Professional forecasters
Disagreement
Volatility forecasting

ABSTRACT

This paper explores whether the dispersion in forecasted crude oil prices from the European Central Bank Survey of Professional Forecasters can provide insights for predicting crude oil return volatility. It is well-documented that higher disagreement among forecasters of asset price implies greater uncertainty and higher return volatility. Using several Generalized Autoregressive Conditional Heteroskedasticity with Mixed Data Sampling (GARCH-MIDAS) models, we find, based on the in-sample estimation results, the oil market experiences greater volatility when the forecasters' disagreements increase. The model that integrates both historical realized variance and forward-looking forecaster disagreement into the conditional variance, along with the model focusing solely on pure forward-looking forecaster disagreement, exhibits a much superior fit to the data compared to the model relying solely on realized variance and the models considering forward-looking forecasted mean return. The out-of-sample forecasting results unequivocally illustrate that incorporating forecaster disagreement offers valuable insights, markedly enhancing the predictive accuracy of crude oil return volatility within the GARCH-MIDAS model. Moreover, we illustrate the economic benefit of considering forecasters' disagreement when forecasting volatility, demonstrating its significance for VaR risk management.

1. Introduction

In recent years, there has been an increased focus on identifying the factors that contribute to oil price volatility, as evidenced by the works of Kilian and Murphy (2014), Van Robays (2016), and Degiannakis and Filis (2017). Notably, recent events such as the global financial crisis, the Russia-Ukraine conflict, and the post-COVID-19 surge in demand have increased oil prices and heightened interest in understanding the drivers of oil price volatility. Moreover, given that oil is a critical input in numerous production processes and a crucial factor in the economy, its price is closely monitored by businesses, market participants, and professional forecasters. Consequently, many central banks across the globe have been conducting surveys of professional forecasters to predict the price of oil.

This paper examines how surveys of professional forecasters' disagreement regarding oil prices can be utilized to predict oil return volatility. Traditionally, a greater degree of disagreement among forecasters of an economic or financial variable indicates higher uncertainty and, therefore, greater volatility of that variable, as noted

☆ We thank the editor, associate editor, and referee for their invaluable comments, which enhanced the quality of the manuscript. Any remaining errors are our own. Ai Jun Hou, Caihong Xu and Anton Hasselgren acknowledge the financial support from the Jan Wallander and Tom Hedelius Foundation of Handelsbanken, Sweden (P20-0168).
* Corresponding author.
E-mail addresses: anton.hasselgren@sbs.su.se (A. Hasselgren), aijun.hou@sbs.su.se (A.J. Hou), ssuardi@uow.edu.au (S. Suardi), caihong.xu@sbs.su.se (C. Xu), x.ye@exeter.ac.uk (X. Ye).

https://doi.org/10.1016/j.ijforecast.2024.04.005

by Chua et al. (2011) in the case of inflation and output growth. In addition, Atalla et al. (2016) demonstrate a positive correlation between the average absolute forecast errors for oil prices and the level of forecast dispersion across all horizons, suggesting that increased disagreement among forecasters reflects heightened uncertainty. Our approach complements Atalla et al. (2016), who explore the causal relationship between forecast disagreement and oil price volatility but differs significantly in our modelling framework. Moreover, our investigation is supported by theoretical models that propose heterogeneous beliefs lead to higher volatility (Banerjee & Kremer, 2010; Shalen, 2015). Additionally, empirical studies examining the link between investor disagreement and asset return volatility have found a positive association in equity markets (Antweiler & Frank, 2004; Frino et al., 2022) and mortgage-backed security markets (Carlin et al., 2014).

Why are surveys of professional forecasters of oil prices crucial for predicting oil price volatility? Patton and Timmermann (2010) examine various sources of disagreement among forecasters. One possible reason for disagreement is that forecasters may have different sets of information available during forecasting, which could be influenced by their use and weighting of oil prices in their business and models. Additionally, forecasters may differ in their opinions on which exogenous variables are most relevant or how they translate into specific price levels. These differences can arise from various methods, such as expert opinions, simple market models, or larger macroeconomic models. Disagreements may also result from strategic behaviours by some forecasters, such as attempting to manipulate the oil market or gain attention from the media. Finally, Lamont (2002) suggests that forecasters who are paid based on their relative abilities may differentiate their forecasts to stand out when making similar predictions or if there is clustering or herding behaviour among forecasters.

We focus on the impact of professional forecasters' disagreement on oil prices on long-term oil return volatility as dictated by the lower data frequency (i.e., quarterly) of the Survey of Professional Forecasters (SPF) forecasts. Investors trade different types of oil and oil-linked assets, such as oil futures and options, with profit-seeking motives. Due to oil's scarcity, demand and supply changes result in significant price fluctuations, which traders can take advantage of. Because investors have varying trading horizons and motives, their strategies for trading oil and oil-linked assets can differ significantly. For example, pension funds and other long-term investors face long-term financial risks over months and years, while day or high-frequency traders are more concerned with short-term risks on a daily or even intra-daily basis. It is thus essential to distinguish between the long- and short-term components of oil price volatility or risk. Previous research has used the GARCH-Mixed Data Sampling (MIDAS) framework to decompose equity and bond return variances and covariances into a long-term (persistent) component and a short-term (transitory) component. Asgharian et al. (2016) have demonstrated that forecasts obtained from SPF can be good predictors of long-term return variances and covariances, establishing a macro-finance link. This study aims to determine whether SPF oil price forecast disagreement can predict long-term oil return volatility, making it the first of its kind.[1]

Our study employs survey data to gauge the level of disagreement among forecasters regarding oil prices. The European Central Bank's Survey of Professional Forecasters (ECB-SPF) has been issuing quarterly forecasts for Brent crude oil prices since the first quarter of 2002. The questionnaire inquires about forecasts at short-, medium-, and long-term horizons. Our research uses both the historical price information and the forward-looking forecast information from the ECB-SPF to examine how quarterly data can be incorporated into our daily oil return model to enhance the in-sample and out-of-sample forecast of crude oil return volatility. To tackle the mixed data frequency issue, we adopt the Mixed Data Sampling (MIDAS) framework following Engle et al. (2013). This framework allows us to break down the variances of crude oil returns into a long-term (persistent, quarterly) component and a short-term (transitory, daily) component. We further investigate whether including forecasters' disagreement in the model can enhance the predictive power of long-term oil return volatility.

Our study explores various GARCH-MIDAS models to examine the relationship between forecasters' disagreement and the realized variance of oil price returns. Our in-sample estimation results reveal a positive correlation between forecasters' disagreement and the realized variance, indicating that the oil market becomes more volatile as disagreement increases. The model incorporating both historical realized variance and forward-looking forecaster disagreement in the conditional variance and the model including a pure forward-looking forecaster disagreement fit the data much better than the model based on realized variance only and the models incorporating forward-looking forecasted mean return. In addition, our out-of-sample forecasting results further demonstrate that forecasters' disagreement provides valuable information for predicting oil return volatility. We find that the GARCH-MIDAS model, which includes the forward-looking disagreement of oil prices amongst forecasters, yields superior out-of-sample predictions compared to the GARCH-MIDAS model that relies only on realized variance or the GARCH-MIDAS model that takes into account the mean of the forecasted oil price returns of forecasters. Importantly, we show that incorporating SPF forecast disagreement is crucial for real-world applications, such as Value-at-Risk (VaR) risk management, as it significantly improves volatility (i.e., risk) forecasting.

We contribute to the existing literature on oil return volatility prediction by demonstrating the value of incorporating forecast disagreement from the SPF survey in the GARCH-MIDAS framework. Our research complements previous studies in this area, such as Wang et al.

[1] Singleton (2014) employs monthly oil price forecasts based on Consensus Economics and demonstrates that a higher forecast dispersion is positively associated with the WTI crude oil price level and higher oil futures price volatility. However, he does not consider the mixed data frequency sampling and does not differentiate the impact of forecast dispersion on the long-run vs. short-run variance. The study also does not focus on the implications of the findings on forecasting.

(2023), who investigate the impact of extreme shocks on crude oil volatility and find that they have asymmetric effects on the short- and long-term volatility components. However, our study focuses on the impact of SPF forecast disagreement on oil return volatility, which is likely to increase with extreme shocks in the oil market. Conrad et al. (2014) examine the correlation between crude oil and stock price returns and how it is affected by the U.S. macroeconomy, but they do not consider the effect of SPF forecast disagreement on oil price volatility. Similarly, Salisu et al. (2022) investigate the predictive power of various economic activity indicators for oil price volatility, but our focus is on the SPF forecasters, who are likely to incorporate macroeconomic factors into their oil price predictions. Our findings suggest that incorporating forecast disagreement from the SPF survey can improve the prediction of long-term oil price return volatility.

The paper proceeds as follows. Section 2 outlines the GARCH-MIDAS model. Section 3 describes the data, presents descriptive statistics about ECB-SPF and provides a preliminary analysis of the relationship between the disagreement among the forecasters and the realized volatility. Section 4 reports the in-sample estimation, evaluates the out-of-sample performance of the various GARCH-MIDAS models and demonstrates the importance of our results in a real-world application. Section 5 concludes.

2. The models

2.1. The GARCH-MIDAS model

This section briefly outlines the econometric models used in the paper.

We assume that the conditional variance of the returns follows Engle et al. (2013) univariate GARCH-MIDAS framework. On day i in quarter t, assuming that the crude oil return series, $r_{i,t}$, follows the process below,

$$r_{i,t} = \mu + \sqrt{\tau_t g_{i,t}}\xi_{i,t}, \forall i = 1, \dots, N_t \\ \xi_{i,t} \big| \Phi_{i-1,t} \sim \mathcal{N}(0, 1) \tag{1}$$

where $\xi_{i,t}$ is the error term which follows a standard normal distribution, $\Phi_{i-1,t}$ is the known information up to day $(i-1)$ of quarter t, and N_t is the number of trading days in each quarter t. In Eq. (1), the variance consists of a short-run component denoted as $g_{i,t}$ and a long-run component denoted as τ_t, which only changes every quarter. The total conditional variance, $\sigma^2_{i,t}$, is defined as follows,

$$\sigma^2_{i,t} = \tau_t g_{i,t}. \tag{2}$$

The short-term component of the conditional variance will follow a GARCH(1, 1) process,

$$g_{i,t} = (1 - \alpha - \beta) + \frac{\alpha \left(r_{i-1,t} - \mu\right)^2}{\tau_t} + \beta g_{i,t-1}, \tag{3}$$

where the parameter restrictions are: $\alpha \geq 0$ and $\beta \geq 0$ and $\alpha + \beta < 1$. The long-term component of the conditional variance, τ_t, can include either backward-looking or forward-looking variables,

$$\log(\tau_t) = m + \theta \sum_{k}^{K} \phi_k(w_1, w_2) X_{t-k}, \tag{4}$$

where X_t can be realized variance (RV) computed at a lower (i.e., quarterly in this paper) frequency than the short-term component of the conditional variance, if it is a backward looking variable. K is the number of lags used to smooth the realized variance or long-term variables. When accommodating forward-looking variable, the variable X will be defined as (X^{SPF}_{t+k-1}), k will be the number of lead periods used to smooth the forward looking variables, i.e. the quarterly SPF information in this paper as we will discuss in the following section.

If it is RV, then it is computed as

$$RV_t = \sum_{i=1}^{N_t} r^2_{i,t}. \tag{5}$$

We use a fixed window for MIDAS, i.e., within a fixed number of periods t, the long-term component τ_t of the conditional variance is set to be the same. The weights used in MIDAS specified in Eq. (4) follow a beta lag polynomial process,

$$\phi_k(w_1, w_2) = \frac{(k/K)^{w_1-1}(1 - k/K)^{w_2-1}}{\sum_{j=1}^{K}(k/K)^{w_1-1}(1 - k/K)^{w_2-1}}, k = 1, \dots, K. \tag{6}$$

With this beta lag polynomial function, when setting $w_1 = 1$, the weights given to the variables always follow a decaying pattern, such that the decay speed will be controlled by w_2. The smaller w_2 is, the more weights will be given to the least distance observations in the sample.

2.2. The two-sided extension of the GARCH-MIDAS

In order to conduct our analysis of the impact of ECB-SPF on the crude oil returns volatility, we extend the GARCH-MIDAS framework with the two-sided filter. Existing studies (see e.g., Engle et al. (2013) and Asgharian et al. (2016)) find that the two-sided filter can significantly improve the predictability of GARCH-MIDAS, i.e., considering the expected future values of the macro variables when forecasting the long-term return variance.

The two-sided filter introduced by Engle et al. (2013) estimates the GARCH-MIDAS using both historical observations and the expected future values of variables, such as the forward-looking SPF data,

$$\log(\tau_t) = m + \theta_X \sum_{k=1}^{K_{\text{lag}}} \phi_k(w_1, w_2) RV_{t-k} \\ + \theta_X \sum_{k=1}^{K_{\text{lead}}} \phi_k(w_1, w_2) X^{\text{SPF}}_{t+k-1|t}, \tag{7}$$

where $\phi_k(w_1, w_2)$ is the weighting function for RV and SPF data defined as in Eq. (6). $X^{\text{SPF}}_{t+k-1|t}$ is the forecasted variables of professional forecasters. K_{lag} and K_{lead} are the number of lags and leads, which we use to smooth the realized variance and the forecasted variables of professional forecasters, respectively. In the model above, both the historical and the forecasted data share the same parameter θ_X, which implies that the observed and forecast data have the same impact on the variances. To have a clear picture of the effects of historical and SPF data on the long-term variance, we follow Asgharian et al. (2016) modifying the model above as:

$$\log(\tau_t) = m + \theta_X \sum_{k=1}^{K_{\text{lag}}} \phi_k(w_1, w_2)\, RV_{t-k} + \theta_{\text{FX}} \sum_{k=1}^{K_{\text{lead}}} \phi_k(w_1, w_2)\, X^{\text{SPF}}_{t+k-1|t}, \tag{8}$$

where K_{lag} is the number of lags of RV which can be any number of quarters, whereas the number of leads K_{lead} in a quarter can be either 2, 3 or 4 given that only four quarters forecasted data are available. The highest maximum likelihood function values in the simulation exercises decide the number of lags and leads used in the model. By fixing $w_1 = 1$, the highest weight will be given to the most recent variables. Similarly, by setting $w_2 = 1$ and estimating w_1, the highest weight will be given to the closest lead of the forecasted variables.

We use different specifications for $X^{\text{SPF}}_{t+k-1|t}$, i.e., SPF variable can be either the forecasted mean oil return (FCMean) of the crude oil or the forecasters disagreement (FCStd), proxied by the forecasted oil price standard deviation among the respondents. We also estimate the model with only $X^{\text{SPF}}_{t+k-1|t}$ considered in Eq. (4), such that, we consider five models:

RV Model: i.e., the RV model, only RV is included as specified in Eq. (4).

FCM Model: i.e., Forecasted Mean model, only forecasted mean return is considered in the variance process, as specified in Eq. (4).

FCD Model: i.e., Forecatsed Disagreement model, only the disagreement among the forecasters (measured by the standard deviation of the forecasted price) is considered in the variance process, as specified in Eq. (4).

RVFCM Model: i.e., RV+Forecasted Mean model, both RV and the forecasted mean return are considered in the variance process as in Eq. (8) .

RVFCD Model: i.e., RV+ Forecasted Disagreement model, both RV and the level of disagreement among the forecasters are considered in the variance process as in Eq. (8).

All models are estimated in MATLAB with the maximum likelihood method.

3. Data and summary statistics

3.1. Spot oil prices and ECB-SPF data

The daily spot price of Brent crude oil is collected from Refinitiv Eikon from January 2002 to November 2023.[2] Since the first quarter of 2002, the ECB has been collecting and publishing quarterly forecasts of Brent crude oil prices in its SPF survey data. According to Atalla et al. (2016), the ECB sends the survey to the participants from, i.e. the financial sector, non-financial research institutes, and employer or employee organizations to collect the average nominal spot price forecasts of Brent crude oil prices over the quarters. Currently, the SPF survey respondents provide forecasts for the current quarter and the next three quarters, referred to as horizon 0 (current quarter), 1, 2 and 3 forecasts in our paper.[3] We download the oil forecasts from the ECB-SPF database and the sample period is from 2002Q1 until 2023Q4, including 88 quarterly surveys.

Following Atalla et al. (2016), we measure the level of disagreement among the forecasters by the dispersion of the point forecasted oil prices among the survey respondents. Hence, the disagreement variable depends on how many respondents for each time and each forecast horizon, and then it is necessary to check whether the main variable of interest in this study - disagreement - is measured consistently across surveys. Panel A of Table 1 displays the number of individual forecasters specified per quarter-ahead forecast. The mean number of forecasters is 45–46 with a maximum (minimum) of 57 (32) forecasters for any given quarter. Thus, we have on average 45–46 cross-sectional variations in term of the forecasts we base on to calculate the disagreement. Panel B of Table 1 shows the fraction of forecasts any forecaster delivers for the current quarter, one, two, and three-quarters forecasts ahead. About 97 percent of forecasters deliver forecasts on all forecasting horizons. This means the disagreement measure we calculate in this study is based on information from the participants who deliver their forecasts on a continuous basis.

3.2. Preliminary analysis for the oil price, volatility and forecasters' disagreement

Table 2 presents the summary statistics of the variables used in this study. The crude oil market is very volatile during our sample period, with the lowest price of 9.12 USD/barrel, and the highest price of 143.95/barrel. The average daily return is zero, however the volatility of the daily return is 260 basis-points. Forecast is the forecasted mean oil price among the respondents, which is around 68–69 USD/barrel at all horizons and close to the mean of the daily oil price of 69.48 USD/barrel. FCStd is the standard deviation of the forecasted oil price among the respondents, measuring the disagreement among the forecasters following Singleton (2014) and Atalla et al. (2016). FCMean is the mean forecasted oil return, and shows no clear pattern over the forecast horizon. Table 2 also reveals that the FCStd increases with the forecast horizon, indicating forecasters disagree more on the oil

[2] The start date of the sample is determined by the ECB-SPF data availability.

[3] Before 2010, ECB collected forecasts for the current quarter and four quarters ahead. To make our data consistent over the whole sample period, we focus only on the forecasts of the current quarter and three quarters ahead.

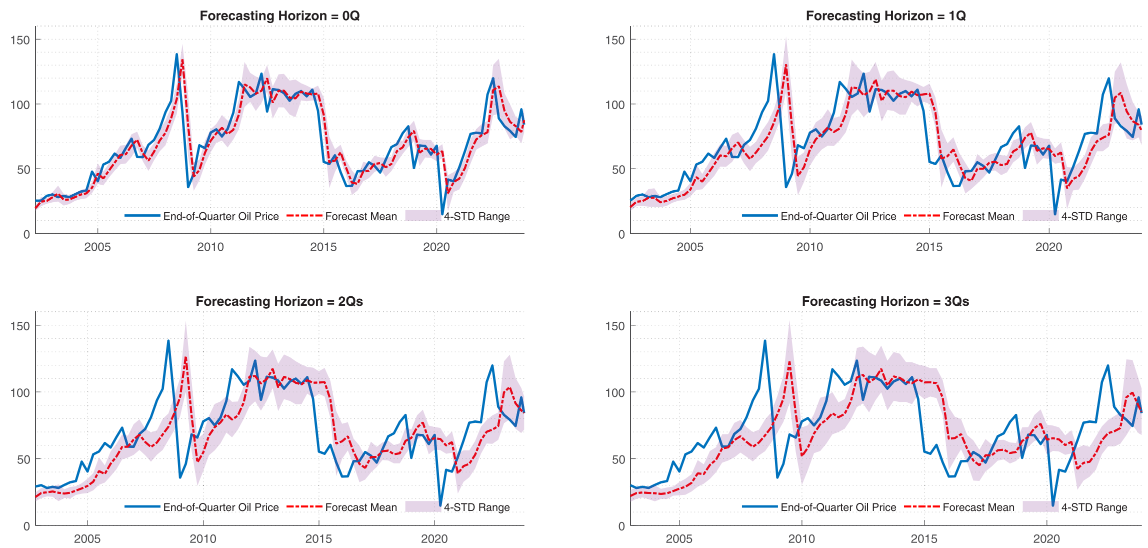

Fig. 1. The ex-ante forecasted quarterly oil price and the ex-post observed oil price.
This figure plots forecasted and observed quarterly oil price. The blue line represents ex post end-of-quarter oil price, the red dashed line represents the mean of ex ante professional forecasters' forecasted end-of-quarter oil price, and the shaded area is the four standard deviations range. Each panel plots this relationship for a particular forecast horizon, from zero, indicating the current quarter, to three quarters ahead. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Forecasters per quarter.
This table presents descriptive statistics for forecasters. Panel A presents the number of forecasts for 0, 1, 2, and 3 quarters ahead. 0 quarter ahead refers to the end of the same quarter. Panel B presents the fraction of forecasters that delivers forecasts for current, one, two, three quarters ahead, given that forecasters made any forecasts in the beginning of the current quarter.

Panel A: Number of forecasters per quarter

	Mean	Median	SD	Min	Max
#Forecast0	46.25	45	4.94	32	57
#Forecast1	46.08	45	4.93	32	57
#Forecast2	45.75	45	5.18	32	57
#Forecast3	45.30	45	5.25	32	57

Panel B: Fraction of forecasters per quarter

	Fraction
Zero	0.993
One	0.989
Two	0.982
Three	0.972

price at longer horizons. The mean quarterly realized variance (RV) is around 4.4%.

In Table 3 the correlation matrix for the relevant variables is presented. The SPF forecasts (Forecast) are positively correlated with Oil prices (from 0.86 to 0.87 depending on the horizon), and negatively correlated with Oil Returns (from −0.23 to −0.25 depending on the horizon). Furthermore, the SPF forecasts have no significant correlation with FCMean at any horizon except for the current quarter and next quarter. But the SPF forecasts are positively correlated with the forecasters disagreement (FCStd) at all horizons (0.58 to 0.70). The disagreement of forecasts are positively correlated with Oil Price (0.45 to 0.52 depending on the horizon), and negatively correlated with Oil Returns (−0.15 to −0.18, depending on the horizon). Furthermore, the disagreement of forecasters is positively correlated with realized variance at the current quarter and the next quarter. This is an indication that the disagreement of forecasters is related to underlying uncertainty, and thus volatility.

As a further preliminary check of whether useful information is contained in the ECB-SPF data, we plot the time series movements for the observed oil price and the average forecasted oil price from SPF in Fig. 1. The blue line represents the end-of-quarter oil price, the red dashed line represents the forecasted mean, and the shaded area is the four standard deviation range. Each panel represents a forecast horizon (from zero, indicating the current quarter, to three quarters ahead). The figure shows that the oil prices at any horizons lead the forecasted mean, indicating that forecasters consistently underestimate price fluctuations. The differences between the forecasted mean and observed oil prices increase with forecast horizons getting longer.

Fig. 2 shows how the realized variance and the disagreement among the forecasted prices evolve. A similar observation can be seen in Fig. 2 as in Fig. 1. Consistent with Atalla et al. (2016), we find a positive and significant correlation between the ex-post realized variance and the ex-ante disagreement among the oil market participants over the short horizons but less in the longer horizons. This suggests that we should consider the disagreement information among the forecasters when we model and forecast the oil return volatility and assign higher weights

Table 2
Summary statistics.
This table presents the summary statistics for the data used. The sample period is from January 2002 to November 2023. The data of the oil prices returns are in daily frequency, while the other variables are in quarterly frequency. 1, 2, and 3 indicate the forecasting horizons in quarters and 0 means forecasts for the end of the same quarter. FCMean refers to the mean forecasted return, and FCStd refers to the standard deviation of price forecasts. RV refers to quarterly realized variance.

	Mean	Median	SD	Skew	Kurt	Min	Max
Daily Oil Price	69.483	66.535	27.925	0.251	−0.848	9.120	143.950
Daily Oil Return	0.000	0.001	0.026	−2.104	85.047	−0.644	0.412
Forecast0	69.049	64.481	27.637	0.228	−0.875	19.475	134.447
Forecast1	68.822	65.084	27.135	0.178	−0.871	20.363	130.573
Forecast2	68.675	64.996	26.682	0.130	−0.839	21.258	126.828
Forecast3	68.925	65.368	26.456	0.091	−0.816	22.020	122.815
FCMean0	0.007	−0.007	0.096	5.172	36.245	−0.118	0.739
FCMean1	−0.002	−0.004	0.036	0.228	1.079	−0.109	0.117
FCMean2	−0.001	−0.006	0.033	0.042	2.288	−0.123	0.106
FCMean3	0.005	0.001	0.026	0.852	1.495	−0.047	0.093
FCStd0	4.305	3.945	2.060	1.311	2.157	1.081	11.887
FCStd1	5.231	4.659	2.409	0.993	0.834	1.495	12.711
FCStd2	5.798	5.292	2.647	0.804	0.467	1.548	13.299
FCStd3	6.321	5.643	2.897	0.912	0.800	1.729	15.511
RV	0.044	0.025	0.106	7.611	62.082	0.004	0.963

Table 3
Correlation matrix.
This table presents the correlation matrix for the data used. The sample period is January 2002 to November 2023. The data of the oil prices and returns are converted from daily to quarterly to match the quarterly frequency. 1, 2, and 3 indicate the forecasting horizons in quarters and 0 refers to forecasts for the end of the same quarter. FCMean refers to the mean forecasted return, and FCStd refers to the standard deviation of price forecasts. RV refers to quarterly realized variance. ***, **, and * indicate the significance levels at 1%, 5%, and 10%, respectively.

	Oil Price	Oil Return	Forecast0	Forecast1	Forecast2	Forecast3	FCMean0	FCMean1	FCMean2	FCMean3	FCStd0	FCStd1	FCStd2	FCStd3
Oil Return	0.18*													
Forecast0	0.87***	−0.25**												
Forecast1	0.87***	−0.25**	1.00***											
Forecast2	0.86***	−0.23**	0.99***	1.00***										
Forecast3	0.86***	−0.23**	0.98***	0.99***	1.00***									
FCMean0	−0.05	0.54***	−0.15	−0.12	−0.10	−0.09								
FCMean1	−0.11	0.21*	−0.12	−0.06	−0.01	0.03	0.46***							
FCMean2	−0.11	0.24**	−0.16	−0.11	−0.05	−0.01	0.42***	0.88***						
FCMean3	−0.18*	0.23**	−0.23**	−0.18*	−0.13	−0.07	0.47***	0.78***	0.87***					
FCstd0	0.45***	−0.15	0.58***	0.59***	0.59***	0.59***	0.13	0.10	0.08	0.10				
FCstd1	0.47***	−0.16	0.63***	0.63***	0.64***	0.64***	0.16	0.13	0.10	0.12	0.94***			
FCstd2	0.51***	−0.18*	0.67***	0.68***	0.69***	0.70***	0.12	0.15	0.15	0.15	0.90***	0.97***		
FCstd3	0.52***	−0.15	0.67***	0.67***	0.68***	0.69***	0.13	0.14	0.15	0.16	0.86***	0.94***	0.98***	
RV	−0.24**	0.17	−0.19*	−0.18*	−0.17	−0.16	0.77***	0.34***	0.36***	0.38***	0.18*	0.18*	0.12	0.11

to the information contained in the SPF disagreement over the shorter horizon, as specified in the model section.

4. Empirical analysis

In this section, we present the empirical results. Our analysis starts with the in-sample analysis where we use the GARCH-MIDAS model to estimate the oil return variances using the five models described in the Model section. We then evaluate the performance of these models in the out-of-sample, and finally, we show the importance of our results with a real world application.

4.1. The in-sample results

We use 8 lags for RV and 4 leads for X^{SPF} in all models. The highest loglikelihood funtion suggests the 8 historical lags, while the longest possible forecast horizons decide the leads of 4. Table 4 shows the results from estimating different GARCH-MIDAS models for the estimation period from January 2002 to November 2023. The parameter of the short-run variance component β in the GARCH model (Eq. (3)) is highly significant and from 0.84 to 0.92 across the five models, indicating a high clustering pattern in the short-run oil return variance. For the long-run variance, the parameter of RV, θ_{RV}, is significant for the RV and RVFCD models indicating RV is important and helpful in modelling the total variance of oil returns, but it is insignificant in RVFCM. The estimate of w_{RV} represents that the degree of smoothing varies over different horizons of the realized variance variable (RV). A smaller w_{RV} leads to a higher degree of smoothing over the lagged observations. We find that w_{RV} is highly significant in RV, RVFCM and RVFCD models.

The forward-looking SPF variables, X^{SPF}, is either FCMean or FCStd and its impact on the variance of oil return is captured by θ_{FX} in each model. We can see θ_{FX}

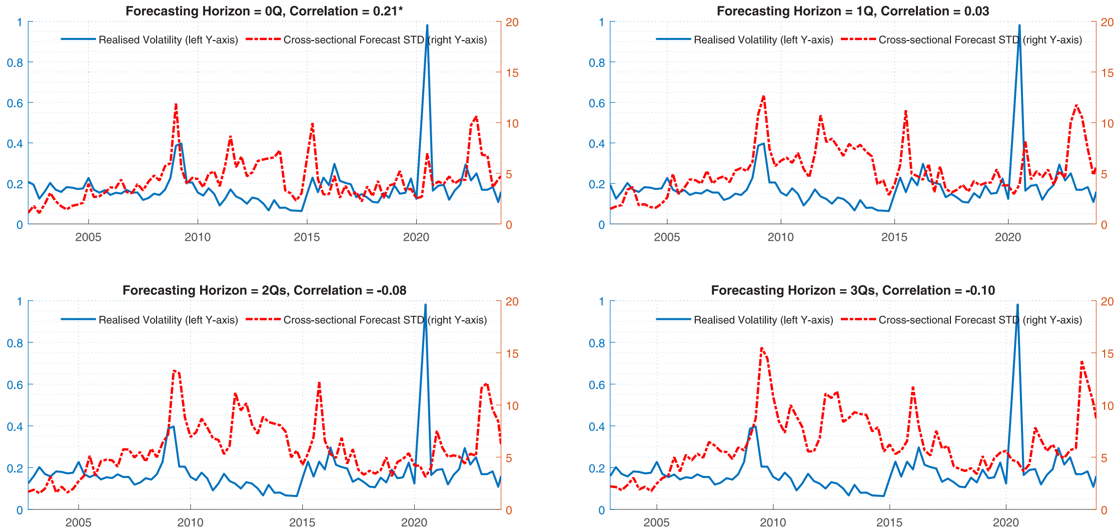

Fig. 2. The ex-ante forecasters disagreement and the ex-post realised oil price volatility.
This figure plots the professional forecasters disagreement and the realised oil price volatility. The blue line represents ex post realised oil price volatility (left Y-axis) and the dashed red line represents the ex ante cross-sectional professional forecasters' end-of-quarter forecast disagreement expressed as standard deviation (right Y-axis). Each panel plots this relationship for a particular forecast horizon, from zero, indicating the current quarter, to three quarters ahead. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
In Sample Estimation results for different models.
This table reports the results of different GARCH-MIDAS model for estimating the variance of the oil returns. These models are described in Eq. (8),
$\log(\tau_t) = m + \theta_X \sum_{k=1}^{K_{\text{lag}}} \phi_k(w_1, w_2)\, RV_{t-k} + \theta_{\text{FX}} \sum_{k=1}^{K_{\text{lead}}} \phi_k(w_1, w_2)\, X_{t+k-1|t}^{\text{SPF}}$,
The weighting function is specified as in Eq. (6),
$\phi_k(w_1, w_2) = \frac{(k/K)^{w_1-1}(1-k/K)^{w_2-1}}{\sum_{j=1}^{K}(k/K)^{w_1-1}(1-k/K)^{w_2-1}}, k = 1, \dots, K.$
The RV model reports the results of the GARCH-MIDAS model which only includes the realized variance in the variance process. The FCM/FCD includes only the forecasted mean/forecasters disagreement in the model. The RVFCM/RVFCD includes both the Realized Variance and the forecasters mean prices/forecasters disagreement in the model. LLF is the value of the maximized log-likelihood function. AIC is the Akaike information criterion and BIC the Bayesian information criterion. ***, ** and * indicate significance at the 1%, 5% and 10% levels, respectively.

Model	μ	α	β	m	θ_{RV}	θ_{FX}	w_{RV}	w_{FX}	*LLF*	*AIC*	*BIC*
RV	0.013***	0.133***	0.840***	−2.009***	1.599***		1.000**		3360.087	−6708.174	−6668.513
FCM	0.022***	0.072***	0.915***	−1.999***		0.999		1.000	3366.009	−6720.018	−6680.357
FCD	0.009**	0.085***	0.909***	−1.818***		0.065***		3.100	3383.509	−6755.018	−6715.357
RVFCM	0.010**	0.085***	0.903***	−2.011***	0.900	0.013	1.000***	1.001***	3382.088	−6748.176	−6695.295
RVFCD	0.010**	0.086***	0.901***	−2.357***	2.081***	0.061***	1.001**	1.354	3386.562	−6757.125	−6704.243

is significantly positive for the FCD model and RVFCD model at 1% level. This indicates that the disagreement among the forecasters (FCStd) significantly contributes to explaining the variance of the oil return, and the impact is positive, which means that a higher degree of disagreement among the forecasters is associated with a higher oil return volatility. The fact that θ_{FX} is also significantly positive in the FCD model, where the disagreement among the forecasters (FCStd) is used as the main solo variable without having the RV variable in the model, provides supportive evidence of the robustness of informativeness of the disagreement variable in modelling the oil return volatility. In FCM and RVFCM models, we report a positive impact of the forecasted mean return on the variance of oil returns. However, the impact is insignificant in both FCM model and RVFCM models. The estimate of w_{FX} represents that the degree of smoothing varies over different horizons of the forward-looking SPF variable. Here a smaller w_{FX} leads to a higher degree of smoothing over the lead observations.

Fig. 3 plots the realized variance and long-run components from the five in-sample GARCH-MIDAS models (i.e., RV, FCM, RVFCM, FCD, and RVFCD) presented in the model section. The blue line represents the short-run variance and the red line represents the long-run variance. The RV model reports the results of the GARCH-MIDAS model which only includes the realized variance in the variance process. The FCM/FCD includes only model's forecasted mean/forecasters disagreement. The RVFCM/RVFCD includes both the realized variance and the forecasters mean prices/forecasters disagreement in the model. The estimated long-term variances are generally fairly

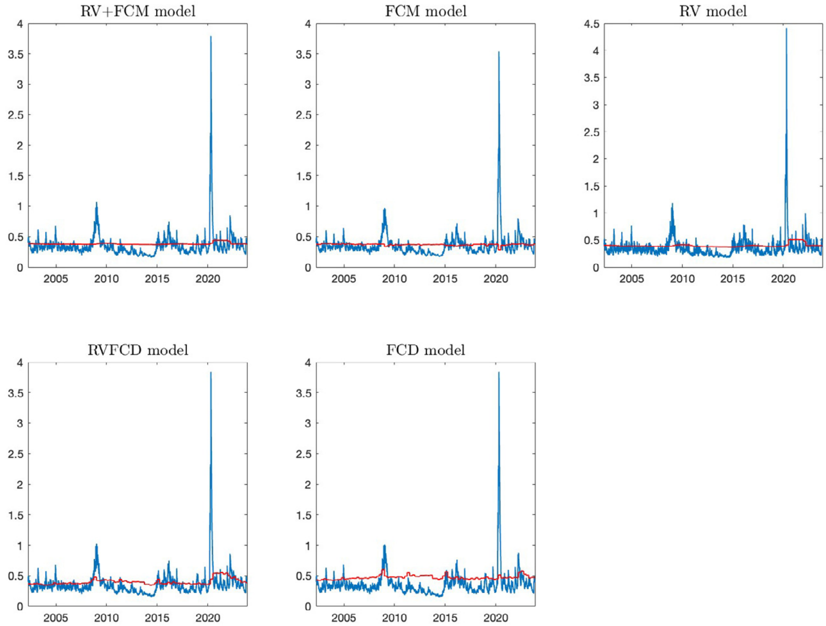

Fig. 3. Estimated long-term variance and total variance: in sample.
This figure plots the realized variances and long-run components from the five in-sample GARCH-MIDAS models (RV, FCM, RVFCM, FCD, and RVFCD). The blue line represents the short-run variance and the red line represents the long-run variance. Each panel plots this relationship for a particular model, which are further detailed in the model section. The RV model reports the results of the GARCH-MIDAS model which only includes the realized variance in the variance process. The FCM/FCD includes only the forecasted mean/forecasters disagreement in the model. The RVFCM/RVFCD includes both RV and the forecasters mean/disagreement in the model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

stable, except for apparent increases during the global financial crisis in 2007–2008, the recent pandemic period and the Russian-Ukraine conflict. The long-run variance obtained from FCD and RVFCD models seem to capture the movements of the realized variances better than the others.

We compute AIC and BIC and the maximized log-likelihood (LLF) value to further compare these models' overall fit to the data. The last three columns of Table 4 report the AIC and BIC along with the LLF for the five models: RV, FCM, FCD, RVFCM, and RVFCD. As is shown in Table 4, the RVFCD model has the highest LLF, the lowest AIC and the second lowest BIC whereas the FCD model has the second highest LLF, the second low AIC and the lowest BIC. Thus, the RVFCD model including both the historical information and the forward-looking disagreement and the FCD model including a pure forward-looking disagreement work well in capturing the process, much better than the models including forecasted mean model and the RV model. Adding the mean forecasted oil return (either FCM or RCFCM) does not improve the model fit as do the models with forecasters' disagreement. In addition, we conduct the likelihood ratio (LR) tests to see if the RVFCM and RVFCD models (non-restricted) are superior to the RV model (restricted). The LR test results, displayed in Table 5, confirm that the null hypothesis that the RV is better than RVFCM or RVFCD is rejected at a 1% significance level. Hence, the in-sample results show that including the disagreement among forecasters of the ECB-SPF improves the in-sample fit of the oil return volatility.

4.2. Out-of-sample results

We perform out-of-sample analyses for these five models to further evaluate the model's forecasting accuracy on oil return volatility. The parameters are obtained using rolling- window estimations with a 10-year window. Then, the estimated parameters are used to predict the variance in the subsequent year.

We calculate their loss functions, i.e., the mean squared errors (MSE), the mean absolute errors (MAE), and QLIKE for each of the five models for the total variance from daily variance predictions and for the long-run variance from quarterly prediction. The MSE and MAE are defined as,

$$MSE = \frac{1}{T}\sum_{t=1}^{T}(\sigma_t^2 - \hat{\sigma}_t^2)^2 \tag{9}$$

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|\sigma_t^2 - \hat{\sigma}_t^2| \tag{10}$$

where ${\sigma_t}^2$ is the actual variance, and $\hat{\sigma}_t^2$ is the forecasted variance from the model. Compared with MAE measure, MSE is a quadratic loss function that usually gives a higher weight to large prediction errors. Thus,

Table 5
The likelihood ratio tests.
This table reports the results of the Likelihood Ratio (LR) test for the RVFCM and RVFCD model against the RV model (restricted). The LR test statistic is calculated as $LR = -2(\log L_0 - \log L)$, where $\log L_0$ and $\log L$ are the log likelihood of the restricted and unrestricted model, respectively. The null hypothesis H0 is: the restricted model, i.e., the RV model is the better model.***,**,* denotes for significance level 1%, 5% and 10% respectively.

	RVFCD	RVFCM
Test Statistic	52.950***	44.002***
p-value	0.000	0.000
df	2	2

according to Brooks and Persand (2003), MSE is more appropriate for large errors.

The QLIKE which corresponds to the loss implied by a Gaussian likelihood is given by,

$$QLIKE = \frac{1}{T}\sum_{t=1}^{T}(ln(\hat{\sigma}_t^2) + \frac{\sigma_t^2}{\hat{\sigma}_t^2}) \tag{11}$$

where ${\sigma_t}^2$ is the actual variance, and $\hat{\sigma}_t^2$ is the forecasted variance from the model.

To further confirm whether the difference in the out-of-sample MSE is significant among the models, we adopt the DM test in Diebold and Mariano (1995) to compare the forecasting accuracy of two competing models under general assumptions (e.g., corrected for autocorrelation).

$$DM = \frac{E(d_t)}{\sqrt{Var(d_t)}} \sim N(0, 1) \tag{12}$$
$$d_t = e_{1,t}^2 - e_{2,t}^2$$

where $e_{1,t}$ and $e_{2,t}$ are prediction errors from two competing models 1 and 2, respectively, and $E(d_t)$ and $Var(d_t)$ are the mean and the variance of d_t. In addition to these measures, we also consider the modified DM test due to Harvey et al. (1997) for the total variance and long-run variance prediction comparison of all models, as follows,

$$\text{Modified DM} = k \times E(d_t)/\sqrt{Var(d_t)/T}, \tag{13}$$

where $k = \sqrt{(T + 1 - 2h + h(h-1)/T)/T}$ and h is the forecast horizon.

When calculating MSE and MAE, we use the quarterly realized variance based on daily returns within that quarter to proxy for the true long-term variance and the daily squared return to proxy for the true total variance. The calculated MSE and MAE of all models are presented in Table 6. Panel A reports MSE, MAE, and QLIKE for the total variance, and Panel B reports MSE, MAE, and QLIKE for the long-run variance. The MSEs reported in Panel A show that the FCD model has the smallest MSE and MAE followed by the RVFCD model for the total variance prediction. For the long-term variance prediction in Panel B, the RVFCD model has the smallest MSE and MAE, followed by the FCD model. The QLIKE for total variance in Panel A shows the RV model has the smallest loss function whereas in Panel B for the long-term variance the RVFCD model has the smallest loss function. To sum up, the out-of-sample results are in general consistent with the in-sample results. Thus, we conclude that considering the disagreement among the professional forecasters improve the in-sample and out-of-sample of both the long-term and total variance prediction.

The DM and the modified DM tests results are presented in Table 7 for the total variance prediction at a daily frequency and Table 8 for the quarterly long-run variance prediction. From Tables 7 and 8, we can see both the DM and the modified DM tests give consistent conclusions on the significance of the out-of-sample forecasting accuracy differences. Specifically, in Table 7, we find that the FCD model significantly outperforms RV, FCM and RVFCM models, but the difference between the FCD model and the RVFCD model for the total variance is insignificant. As is shown in Table 8, the RVFCD significantly outperforms all the other four models for the long-run variance prediction at the 1% significance level, indicated by both the DM test and modified DM tests. The FCD model is the second best for the long term variance prediction. This means that FCD's and RVFCD's out-of-sample forecasting accuracy is substantially better than the other models, indicating that including forecasters' disagreement (FCStd) improves performance. The results also show that the RVFCD model forecasts better than the RVFCM model, suggesting that the disgreement among the forecasters, not the average forecasted return, contains useful information about the future oil return volatility. Overall, our out-of-sample forecasting exercise provides further evidence supporting the informativeness and practical usefulness of the disagreement variable (FCStd) in isolation and when combined with RV.

4.3. Real-world example: the implications of our results for risk management

Our in-sample and out-of-sample analysis results are important for portfolio selection, asset pricing and risk management. For example, as our estimation implies, the FCD and RVFCD models provide a more accurate prediction for the total variance and long-term variance than the RV, FCM and RVFCM models. This may lead to a significant difference in risk management, portfolio selection, and hedging strategies.

To further demonstrate the importance of considering the expectation errors of investors in the volatility prediction, we apply these five models to the forecasts of the value-at-risk (VaR). In particular, we first estimate the out-of-sample forecasted total variance. Then we use the out-of-sample forecasted total volatility to build up a 90% interval for oil returns. The 90% intervals of the oil returns

Table 6
MSE, MAE and QLIKE of out-of-sample forecasts.
This table presents the mean square error (MSE), the mean absolute error (MAE), and the QLIKE of the out-of-sample forecast using the five models: RV, FCM, FCD, RVFCM, and RVFCD. Panel A reports the MSE, MAE and QLIKE for total variance, and Panel B for the long-term variance.

	RV	FCM	FCD	RVFCM	RVFCD
Panel A: Total variance					
MSE	0.071	0.074	0.065	0.072	0.066
MAE	0.013	0.035	0.001	0.012	0.010
QLIKE	−6.789	−5.471	−6.819	−6.792	−6.833
Panle B: Long-term variance					
MSE	0.082	0.081	0.074	0.082	0.072
MAE	0.098	0.093	0.074	0.098	0.073
QLIKE	158.602	17.405	−1.168	175.475	−1.372

Table 7
The DM test: total variance.
This table presents the test statistics from the DM test of Diebold and Mariano (1995) and the modified DM test of Harvey et al. (1997) for different models for the prediction of total variance. The true total variance is proxied by the squared daily returns. The test statistic compares the model in the column to the model in the row. Negative (positive) statistics indicate that the model in the column outperforms (underperforms) the model in the row. ***,**,* denotes for significance level 1%, 5% and 10% respectively.

Model	DM Test		Modified DM Test	
	FCD	RVFCD	FCD	RVFCD
RV	−2.342***	−1.551	−2.342***	−1.551
FCM	−3.462***	−4.597***	−3.461***	−4.596***
RVFCM	−3.223***	−3.643***	−3.223***	−3.643***
FCD	–	0.839	–	0.839

Table 8
The DM test: the long-term variance (τ).
This table presents the test statistics from the DM test of Diebold and Mariano (1995) and the modified DM test of Harvey et al. (1997) for different models for the prediction of total variance. The true long term variance is proxied by the realized variance by summing up the daily squared returns within the same quarter. The test statistic compares the model in the column to the model in the row. Negative (positive) statistics indicate that the model in the column outperforms (underperforms) the model in the row.***,**,* denotes for significance level 1%, 5% and 10% respectively.

Model	DM Test		Modified DM Test	
	FCD	RVFCD	FCD	RVFCD
RV	−2.402***	−2.504***	−2.379***	−2.480***
FCM	−2.292***	−2.420***	−2.270***	−2.396***
RVFCM	−2.100***	−2.259***	−2.079**	−2.237***
FCD	–	−2.516***	–	−2.492***

are computed as,

$$\hat{r}_t = \hat{\mu}_t + q_k\sqrt{\hat{\sigma}_t}$$

where q_k is the quantile of the normally distributed errors. $\hat{r}_t$ and $\hat{\sigma}_t$ are the out-of-sample predicted mean and volatility, respectively. It is interesting to note that the lower bound of the 90% return interval will be the 5% daily value-at-risk (VaR) measure given that the initial investment is 1 US dollar. Fig. 4 plots the 90% intervals of the forecasted returns based on the out of sample predicted volatility from the RV, FCM, FCD, RVFCM, and RVFCD models. The return intervals built upon the out-of-sample forecasted conditional mean and variance of various models (besides the FCM model) do not differ that much during the relatively stable periods. However, when extreme shocks enter the market, the models that do not consider the forecasters' disagreement provide much wider return interval than those considering the forecasters' disagreement. One example is the sudden oil price drops in during 2020 (pandemic). We can see that the RV, FCM, RVFCM models overestimate the oil returns of the upper bound and underestimate the oil returns of the lower bound. Because the 5% daily VaR measure is the lower bound of the interval for a 1 US dollar initial investment, and as shown in the previous section, both the FCD and RVFCD model can provide more accurate in-sample and out-of-sample volatility prediction than the other models, hence, during the extremely volatile period, the 5% VaR based on these models without considering the forecasters' disagreement is overestimated.

In general, the FCD and the RVFCD models perform much better than the RV, FCM and RVFCM models when forcasting the volatility and capturing the rise and fall movements of the crude oil returns volatility. Because the FCM and FRVCM models are prone to overestimate the volatility of the oil returns in the highly volatile periods and hence cause much larger estimation errors, they are less appropriate for predicting the crude oil return

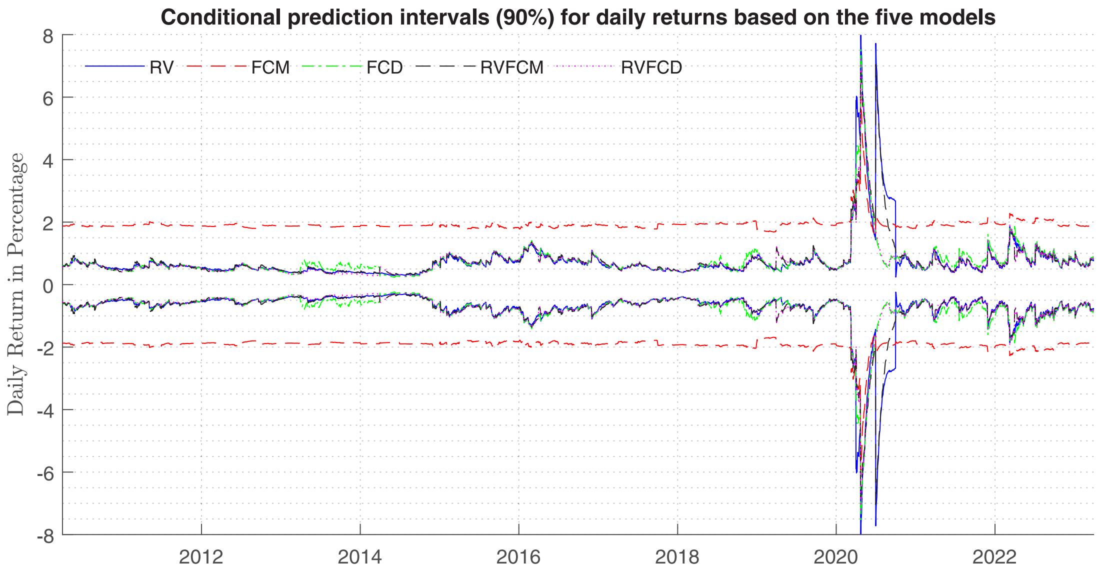

Fig. 4. The 90% conditional prediction interval for daily returns (percentage).
This figure plots the 90% conditional prediction interval for daily returns of the oil prices. The predicted intervals are calculated based on the out-of-sample variance prediction. The out-of-sample forecasts start from 2010.

volatility than the models such as FCD and RVFCD. The poor volatility prediction of such models will significantly affect the investor's portfolio selection, risk management, and dynamic hedging strategy.

5. Conclusion

Understanding how the disagreement among participants affects the oil market volatility is important for investors and firms to accurately forecast the volatility of crude oil returns when designing their investment and hedging strategies and for economic policy makers to make effective regulation policies.

This paper studies how the disagreement among forecasters about the future oil price may affect the oil return volatility. One empirical challenge to carrying out our research is that data on disagreement in expectations and forecasts of the market participants are not readily available. We use the data from ECB-SPF to measure the disagreement among the forecasters. The disagreement measure of ECB-SPF is only available quarterly and the oil returns are calculated daily. We tackle this problem by applying the MIDAS framework of Engle et al. (2013), which enables us to decompose the variances of crude oil returns into a long-run (persistent, quarterly) component and a short-run (transitory, daily) component, and to analyze if adding the forecasters disagreement to the variance process can improve the forecasting abilities of the short-run and long-run oil return volatility.

Our results indicate that including the forecasters' agreement into the model is important when predicting the crude oil return volatility. In particular, we find that GARCH-MIDAS, including the forward-looking disagreement in oil price among the forecasters, improves upon the GARCH-MIDAS model based on realized volatility only and upon the GARCH-MIDAS including the mean forecasted return among the forecasters in terms of in-sample and out-of-sample performance. In addition, we demonstrate that the improved crude oil forecasts have important implications in real-world applications, such as VaR risk management.

CRediT authorship contribution statement

Anton Hasselgren: Conceptualization, Methodology, Formal analysis, Data curation. **Ai Jun Hou:** Conceptualization, Methodology, Investigation, Formal analysis, Validation, Funding acquisition, Writing – original draft. **Sandy Suardi:** Conceptualization, Formal analysis, Writing – review & editing. **Caihong Xu:** Conceptualization, Methodology, Formal analysis, Data curation, Funding acquisition, Investigation, Validation, Writing – original draft, Writing – review & editing. **Xiaoxia Ye:** Conceptualization, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance*, *59*, 1259–1294.

Asgharian, H., Christiansen, C., & Hou, A. J. (2016). Macro-finance determinants of the long-run stock–bond correlation: The DCC-MIDAS specification. *Journal of Financial Econometrics*, *14*, 617–642.

Atalla, T. k., Joutz, F., & Pierru, A. (2016). Does disagreement among oil price forecasters reflect volatility? Evidence from the ECB surveys. *International Journal of Forecasting*, *32*, 1178–1192.

Banerjee, S., & Kremer, I. (2010). Disagreement and learning: Dynamic patterns of trade. *The Journal of Finance*, *65*, 1269–1302.

Brooks, C., & Persand, G. (2003). The effect of asymmetries on stock index return value at risk estimates. *Journal of Risk Finance*, *4*, 29–42.

Carlin, B. I., Longstaff, F. A., & Matoba, K. (2014). Disagreement and asset prices. *Journal of Financial Economics*, *114*, 226–238.

Chua, C. L., Kim, D., & Suardi, S. (2011). Are empirical measures of economic uncertainty alike? *Journal of Economic Surveys*, *25*, 807–821.

Conrad, C., Loch, K., & Rittler, D. (2014). On the macroeconomic determinants of long-term volatilities and correlations in U.S. stock and crude oil markets. *Journal of Empirical Finance*, *29*, 26–40.

Degiannakis, S., & Filis, G. (2017). Forecasting oil price realized volatility using information channels from other asset classes. *Journal of International Money and Finance*, *75*, 28–49.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*, 253–265.

Engle, R. F., Ghysels, E., & Sohn, B. (2013). Stock market volatility and macroeconomic fundamentals. *The Review of Economics and Statistics*, *95*, 776–797.

Frino, A., Xu, C., & Zhou, Z. I. (2022). Are option traders more informed than Twitter users? A PVAR analysis. *Journal of Futures Markets*, *42*, 1755–1771.

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, *13*, 281–291.

Kilian, L., & Murphy, D. P. (2014). The role of inventories and speculative trading in the global market for crude oil. *Journal of Applied Econometrics*, *29*, 454–478.

Lamont, O. A. (2002). Macroeconomic forecasts and microeconomic forecasters. *Journal of Economic Behaviour and Organization*, *48*, 265–280.

Patton, A. J., & Timmermann, A. (2010). Why do forecasters disagree? Lessons from the term structure of cross-sectional dispersion. *Journal of Monetary Economics*, *57*, 803–820.

Salisu, A. A., Gupta, R., Bouri, E., & Ji, Q. (2022). Mixed-frequency forecasting of crude oil volatility based on the information content of global economic conditions. *Journal of Forecasting*, *41*, 134–157.

Shalen, C. T. (2015). Volume, volatility, and the dispersion of beliefs. *The Review of Financial Studies*, *6*, 405–434.

Singleton, K. J. (2014). Investor flows and the 2008 boom/bust in oil prices. *Management Science*, *60*, 300–318.

Van Robays, I. (2016). Macroeconomic uncertainty and oil price volatility. *Oxford Bulletin of Economics and Statistics*, *78*, 671–693.

Wang, L., Ma, F., Liu, G., & Lang, Q. (2023). Do extreme shocks help forecast oil price volatility? *International Journal of Forecasting*, *28*, 2056–2073.