Energy Economics 120 (2023) 106643

Contents lists available at ScienceDirect

Energy Economics

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

Attention to oil prices and its impact on the oil, gold and stock markets and their covariance

Piotr Fiszeder [a,b], Marcin Fałdziński [a,b], Peter Molnár [a,b,c,*]

[a] Faculty of Economic Sciences and Management, Nicolaus Copernicus University in Torun, Torun, Poland
[b] Faculty of Finance and Accounting, Prague University of Economics and Business, Prague, Czech Republic
[c] UiS Business School, University of Stavanger, Stavanger, Norway

ARTICLE INFO

Keywords:
Volatility
Google searches
Oil
Dynamic conditional correlation
High-low range
Covariance forecasting

ABSTRACT

This paper studies the impact of investor attention to oil prices on returns, volatility, and covariances of three exchange traded funds representing oil, gold, and the stock market. For this purpose, we suggest a new multivariate volatility model based on open, high, low, and closing prices that incorporates the impact of investor attention on returns, volatility, and covariances. We find that this model, which incorporates Google searches for "oil prices" as an exogeneous variable, outperforms other considered multivariate volatility models, and demonstrates that Google searches for "oil prices" can explain and forecast covariances between returns of oil, gold, and the stock market.

1. Introduction

This paper studies the impact of investor attention to oil price, as measured by Google searches for "oil prices" on returns, volatility and covariance of three exchange traded funds representing oil, gold, and the stock market. Our paper unites two strands of literature: 1) the impact of investor attention on the oil market and 2) the comovement between the oil market and other markets.

Crude oil is the world's most important commodity, and understanding the oil market and its risks is therefore important (Hung et al., 2008; Hung et al., 2011; Gkillas et al., 2020; Gkillas et al., 2021a; Tiwari et al., 2020). It has been long recognized that investor attention has an influence on financial markets, see e.g. Da et al. (2011). Attention can be measured in various ways, for example from news articles or social networks such as twitter (Gjerstad et al., 2021). One particularly useful data source to estimate attention is Google Trends. Google is the largest search engine in the world, and information regarding how much people search for a particular topic in Google is therefore a very relevant measure of attention.

It is important to emphasize that the channel between attention and financial markets is indirect. People often search for additional information before they make decisions, including trading decisions. In other words, the act of searching for information often precedes the actual action. This way, attention estimated from Google Trends might precede actual trading, and therefore might predict returns and volatility in various financial markets. In other words, it is not searching on Google that moves the markets, but information about Google searches could predict subsequent movements in the markets.

Researchers have also studied the impact of attention measured from Google Trends on the oil market. The most common applications are the impact of investor attention on oil price (Li et al., 2015; Yao et al., 2017; Elshendy et al., 2018; Li et al., 2019; Li et al., 2020; Yang et al., 2021; Chen et al., 2022), the influence on oil price volatility (Afkhami et al., 2017; Campos et al., 2017; Wang et al., 2018; Xiao and Wang, 2021; Liu et al., 2022), or the impact of investor attention on both oil price and oil price volatility (Guo and Ji, 2013; Ji and Guo, 2015; Qadan and Nama, 2018). Review of these articles is provided in the next section.

Similarly, there is a large volume of literature about comovement between oil and other commodities (Ahmadi et al., 2016; Bašta and Molnár, 2018; Behmiri et al., 2019). However, research about the impact of investor attention on the comovement (covariance) between oil price and other assets is very limited. To the best of our knowledge, the only study on this topic is Prange (2021). This novel study finds that online investor attention is a statistically significant determinant of the time-varying correlations between oil and other assets.

The main difference between Prange (2021) and our study is that Prange (2021) use a volatility model based on closing prices, while we suggest a new volatility model based on open, high, low, and closing

* Corresponding author at: University of Stavanger, UiS Business School, 4036 Stavanger, Norway.
E-mail address: peto.molnar@gmail.com (P. Molnár).

https://doi.org/10.1016/j.eneco.2023.106643
Received 3 June 2022; Received in revised form 28 February 2023; Accepted 16 March 2023
Available online 21 March 2023

prices. Overwhelmingly, multivariate volatility models are based on closing prices. Meanwhile, the incorporation of low and high prices for such models can both improve the estimation of variances and covariances and increase their forecasting accuracy (see e.g. Chou et al., 2009; Su and Wu, 2014; Fiszeder and Fałdziński, 2019; Fiszeder et al., 2019; De Nard et al., 2022; Fiszeder and Małecka, 2022; Fiszeder et al., 2023). Low and high prices are generally available with closing prices and can be easily utilized.

Similarly to Prange (2021), we use the extended dynamic conditional correlation (DCC) model of Engle (2002) to incorporate the influence of exogenous variables in the dynamic conditional correlation. There are several possible ways to introduce the influence of exogenous variables on correlations (see Schopen, 2012). Silvennoinen and Teräsvirta (2005) introduced the smooth transition conditional correlation model for which time-varying conditional correlations change smoothly between two regimes driven by an exogenous variable. Sheppard (2008) assumed that the symmetric square root of the covariance matrix is a function of exogenous variables. Chou and Liao (2008) and Vargas (2008) extended the DCC model of Engle (2002) to incorporate the influence of exogenous variables and introduced the dynamic conditional correlation with exogenous variables model (DCC-X). Min and Hwang (2012) and Kim et al. (2013) introduced the bivariate DCC-X model where the conditional correlation coefficient is determined by exogenous variables. In this paper, we explore the DCC-X model of Vargas (2008) because it is quite general (not limited to a bivariate case) but at the same time it is relatively easy to apply.

We analyze three exchange-traded funds: United States Oil Fund, SPDR Portfolio S&P 500 Growth and SPDR Gold Shares and confront the proposed model with three competing models: the DCC model of Engle (2002), the DCC-X model of Vargas (2008) and the DCC-RGARCH model of Fiszeder et al. (2019). An application of the alternative models is for benchmark purposes and used here to assess the importance of additional information from Google searches and the range-based model framework.

This study has two main contributions. The first one is a proposition of the new DCC-X-RGARCH-X model. It is an extension of the DCC-RGARCH model of Fiszeder et al. (2019) which incorporates exogenous variables in the equations of conditional means, variances and covariances. Second, we utilize this model, and demonstrate that Google searches for “oil prices” can explain and forecast covariances between the oil exchange-traded fund and exchange traded funds representing the stock market and gold. Moreover, the DCC-X-RGARCH-X model, which incorporates Google searches for “oil prices” as exogeneous variable, outperform other considered multivariate volatility models.

The rest of the paper is organized in the following way. Section 2 reviews the related literature. Section 3 presents the proposed DCC-X-RGARCH-X model. Section 4 describes the data used in the research. In Section 5 the influence of Google searches on volatility and dependence for the selected ETFs is analyzed. In Section 6, we perform a robustness check for the inclusion of the COVID-19 crisis. Section 7 provides conclusions.

2. Literature review

In this section, we review papers that studied the impact of attention on crude oil price and its volatility.

Li et al. (2015) measure attention from Google Trends for “oil price”. They utilize Granger causality and regressions and conclude that Google searches for “oil price” capture the attention of non-professional traders, and there is a feedback loop between attention and crude oil price.

Yao et al. (2017) obtain Google searches for “oil price”, “current oil prices”, “price per barrel”, “Bloomberg energy”, “oil price per barrel”, “current oil”, “crude oil chart”, “crude oil”, “current crude oil”, “current crude oil price”, and “current crude” and use principal component analysis to construct the investor attention measure. They employ the structural vector autoregression model and find that investor attention has a significant negative impact on oil prices and contributes 15% to the long-run fluctuation of the oil prices.

Elshendy et al. (2018) study whether information extracted from Twitter, Google Trends, Wikipedia, and the Global Data on Events, Location and Tone database can improve the forecasts of the crude oil price. They use autoregressive integrated moving average with explanatory variable models and find that the information combined from these platforms contains a valuable information for forecasting.

Li et al. (2019) construct the investor attention measure from Google searches for “crude oil”, “Brent” and “petroleum”. They utilize nonlinear Granger causality tests and find bidirectional causality between investor attention and crude oil returns. However, this Granger causality is stronger from investor attention to returns.

Li et al. (2020) also use Google Trends to construct attention measure. However, they combine searches for “oil price” in the English language with equivalent searches also in Arabic, Spanish, Portuguese, Japanese, German and Persian. As methods, they use not only regressions, but also machine learning techniques such as neural networks. They find that Google Trends attention improves crude oil price predictions, and the multilingual attention measure works better than single-language attention measure.

Yang et al. (2021) utilize Google searches for 40 oil-related search terms, such as “oil price”, “oil demand” or “oil supply”, together with various economic variables. They use machine learning techniques for the dimension reduction of input variables and also for the forecasting of crude oil price. They conclude that Google Trends data is useful in oil price forecasting and that machine learning “divide and conquer” techniques work well in forecasting.

Chen et al. (2022) use Google Trends to obtain 17 oil-related search terms, such as “WTI” or “petrol price”. Attention is constructed from these using principal component analysis. They employ wavelet coherency, causality-in-quantile, and quantile-on-quantile regression methods to study the linkage between investor attention and crude oil across time and frequency domains. They find that in most cases, attention is negatively correlated with the crude oil market. However, the impact of attention on crude oil differs across quantiles, and it becomes greatest under extreme market conditions.

Afkhami et al. (2017) study volatility of not only crude oil, but also conventional gasoline, heating oil, and natural gas. They obtain 90 oil- and energy-related keywords from Google Trends and use Granger causality tests to choose relevant search terms. They use the GARCH model and find that attention improves volatility models for these energy commodities.

Campos et al. (2017) use Google searches for “oil prices” as a proxy for investor attention to oil and study whether it can predict the implied volatility of oil. They use regressions as their method and find that attention is useful in predicting the implied volatility of oil even if traditional financial and macro variables are taken into account. Moreover, attention has economic value, allowing traders of volatility-exposed portfolios to increase returns.

Wang et al. (2018) combine Google searches for “crude oil”, “oil price”, “crude oil price”, and “crude oil prices”. This variable is used in extreme machine learning models. They conclude that Google Trends is a useful tool for quantifying investor attention that can help predict volatility in the oil market.

Xiao and Wang (2021) also use Google Trends to estimate sentiment. To ensure robustness, they consider searches for “oil price”, “oil prices”, and “crude oil”. They decompose volatility into good and bad volatility, and using regressions, they find that changes in investor attention mainly affect bad volatility rather than good volatility.

Liu et al. (2022) obtain Google Trends data for 519 keywords and then use backwards regression to select the final 83 search terms. They further utilize 5-min high-frequency oil price data to construct daily volatilities. The considered models (heterogeneous autoregressive model and heterogeneous autoregressive model with Markov-switching) perform better when attention is included in these models.

Guo and Ji (2013) use Google Trends to obtain market concerns for oil price and market concerns for oil demand. They use cointegration and the exponential generalized autoregressive conditional heteroskedasticity (EGARCH) model and conclude that there is a long-term equilibrium between oil prices and long-run market concerns.

Ji and Guo (2015) utilize data from Google Trends in their study of the impact of oil-related events (hurricanes, global financial crisis, Libyan war, OPEC conference) on oil price and its volatility. They find that responses vary across events. The impact of the global financial crisis on the oil price returns was significantly negative, the impact of the Libyan war and hurricanes was significantly positive, while reactions to OPEC production announcements were inconsistent.

Qadan and Nama (2018) utilize an extensive list of variables capturing investor sentiment and find that this sentiment has a significant effect on oil prices. Moreover, volatility in these sentiment indices spills over and can explain part of the oil price volatility. Regarding the Google searches for “oil price”, “oil prices”, or “crude oil”, they find a bidirectional relationship between these searches and oil price volatility.

3. The DCC-X-RGARCH-X model

In this paper, we introduce the DCC-X-RGARCH-X model and confront it with three competing DCC specifications: the DCC model of Engle (2002), the DCC-RGARCH model of Fiszeder et al. (2019), and the DCC-X model of Vargas (2008). The DCC and DCC-RGARCH models are similar in their correlation part but differ in their specification of univariate conditional variances. The DCC model is based on the GARCH model of Bollerslev (1986), while the DCC-RGARCH model is based on the RGARCH model of Molnár (2016). Whereas the DCC-X model is an extension of the DCC model that incorporates exogenous variables that drive the time-varying conditional covariance. In this section, we introduce the new model and explain how it includes the remaining parameterisations as special cases.

Let us assume that ε_t ($N \times 1$ vector) is the innovation process for the conditional mean and can be written as:

$$\boldsymbol{\varepsilon}_t|\psi_{t-1} \sim Normal(0, cov_t), \tag{1}$$

where ψ_{t-1} is the set of all information available at time $t-1$, *Normal* is the multivariate conditional normal distribution, and $\mathbf{cov}_t$ is the $N \times N$ symmetric conditional covariance matrix.

The DCC-X(P,Q)-RGARCH-X(p,q) model can be presented as:

$$cov_t = \mathbf{D}_t \mathbf{cor}_t \mathbf{D}_t, \tag{2}$$

$$\mathbf{cor}_t = \mathbf{Q}_t^{*-1}\mathbf{Q}_t\mathbf{Q}_t^{*-1}, \tag{3}$$

$$\mathbf{Q}_t = \left(1 - \sum_{i=1}^{Q}\zeta_i - \sum_{j=1}^{P}\theta_j - \boldsymbol{K}\boldsymbol{\xi}_1'\overline{\boldsymbol{X}}\right)\mathbf{S} + \sum_{i=1}^{Q}\zeta_i(\mathbf{z}_{t-i}\mathbf{z}_{t-i}') + \sum_{j=1}^{P}\theta_j\mathbf{Q}_{t-j} + \boldsymbol{K}\boldsymbol{\xi}_1'\boldsymbol{X}_{t-1}, \tag{4}$$

where $\mathbf{D}_t =$ diag $(h_{1t}^{1/2}, h_{2t}^{1/2}, \ldots, h_{Nt}^{1/2})$, h_{kt} are conditional variances ($k = 1, 2, \ldots, N$), $\mathbf{z}_t$ is the standardized $N \times 1$ residual vector assumed to be serially independently distributed given as $\mathbf{z}_t = \mathbf{D}_t^{-1}\varepsilon_t$, $\mathbf{cor}_t$ is the time-varying $N \times N$ conditional correlation matrix of $\mathbf{z}_t$, $\mathbf{S}$ is the unconditional $N \times N$ covariance matrix of $\mathbf{z}_t$, $\mathbf{Q}_t^*$ is the diagonal $N \times N$ matrix composed of the square root of the diagonal elements of $\mathbf{Q}_t$, $\boldsymbol{K}$ is the $N \times N$ matrix that can either be an identity matrix or a matrix of ones, $\boldsymbol{\xi}_1$ is the $r \times 1$ vector of parameters, $\boldsymbol{X}_t$ is the $r \times 1$ vector of exogenous variables, $\overline{\boldsymbol{X}} = \frac{1}{n}\sum_{t=1}^{n}X_t$, and n is the number of observations used in estimation.

The parameters ζ_i (for $i = 1, 2, \ldots, Q$), θ_j (for $j = 1, 2, \ldots, P$) are nonnegative and satisfy the condition $\sum_{i=1}^{Q}\zeta_i + \sum_{j=1}^{P}\theta_j < 1$. The model implies that all conditional correlations are equally influenced by any exogenous variable.

We assume that univariate innovation processes ε_{kt} ($k = 1, 2, \ldots, N$) are given as (to describe autocorrelation or cross-correlation of returns conditional mean equations can be extended):

$$\varepsilon_{kt} = r_{kt} - \gamma_{k0} + \gamma_{k1}' \boldsymbol{X}_{t-1}, \tag{5}$$

$$\varepsilon_{kt}|\psi_{t-1} \sim Normal(0, h_{kt}), \tag{6}$$

where r_{kt} are returns calculated as $\ln(p_{kt}/p_{kt-1})$, where p_{kt} are closing prices at time t, $\boldsymbol{\gamma}_{k1}$ is the $r \times 1$ vector of parameters, *Normal* is the univariate conditional normal distribution, and conditional variances h_{kt} are described as univariate RGARCH-X(p,q) models:

$$h_{kt} = \alpha_{k0} + \sum_{i=1}^{q}\alpha_{ki}\sigma_{P,k\ t-i}^2 + \sum_{j=1}^{p}\beta_{kj}h_{k\ t-j} + \boldsymbol{\varphi}_{k1}'\boldsymbol{X}_{t-1}, \tag{7}$$

where σ_{Pt}^2 is the Parkinson estimator of variance (Parkinson, 1980) given as $\sigma_{Pt}^2 = [ln(H_t/L_t)]^2/(4\ln 2)$ and H_t and L_t are high and low prices over a day, $\boldsymbol{\varphi}_{k1}$ is the $r \times 1$ vector of parameters, $\alpha_{k0} > 0$, $\alpha_{ki} \geq 0$, $\beta_{kj} \geq 0$ (for $k = 1, 2, \ldots, N$; $i = 1, 2, \ldots, q$; $j = 1, 2, \ldots, p$). If $\boldsymbol{X}_t \geq \mathbf{0}$, then $\boldsymbol{\varphi}_{k1} \geq \mathbf{0}$ guarantees the positivity of the conditional variance h_{kt} for all time t. In the case of negative values of the exogenous variables in $\boldsymbol{X}_t$, the positive value of the conditional variance must be verified for all t. A remedy would be to use the logarithm of h_{kt} in Eq. (7).

To ensure the positive definiteness of $\mathbf{Q}_t$, the matrix $\boldsymbol{K}$ can be assumed as an identity matrix. It is additionally specified that $\boldsymbol{\xi}'_1 = (\xi_{11}, \xi_{21}, \ldots, \xi_{r1})'$, where $\xi_{s1} = \sqrt{\xi_{s1}^{(s1)}}$, $\xi_{s1}^{(s1)} \in (0,1)$ (see Vargas, 2008). This condition is, however, restrictive because it implies that the exogenous variables only drive the conditional variances $q_{ii,t}$ but not the conditional covariances $q_{ij,t}$ ($i \neq j$), where $q_{ij,t}$ is the i, j^{th} entry of $\mathbf{Q}_t$ (nevertheless, since the conditional correlation $r_{ij,t}$ is equal to $r_{ij,t} = q_{ij,t}/(q_{ii,t}\, q_{jj,t})^{-1/2}$, it is still indirectly a function of the exogenous variables). This restriction may be relaxed by setting $\boldsymbol{K}$ as a matrix of ones instead.[1]

An additional problem is that the condition $\xi_{s1} = \sqrt{\xi_{s1}^{(s1)}}$ restricts the sign of the parameters to be non-negative. This is a limiting requirement and does not allow for the exogenous variable to have a negative influence on the matrix $\mathbf{Q}_t$. A solution would be to allow ξ_{s1} on negative values when $\boldsymbol{K}$ is an identity matrix provided that the positive definiteness of the matrix $\mathbf{Q}_t$ is not violated for all t.

All three competing DCC specifications: the DCC model of Engle (2002), the DCC-RGARCH model of Fiszeder et al. (2019), and the DCC-X model of Vargas (2008) do not contain exogenous variables in the conditional means (it means that $\gamma_{k1} = \mathbf{0}$ in Eq. (5) for all above models). Notice that for $\boldsymbol{\varphi}_{k1} = \mathbf{0}$, the RGARCH-X model (Eq. (7)) reduces to the RGARCH model of Molnár (2016). It means that exogenous variables are not present in the conditional variance. If $\boldsymbol{\varphi}_{k1} = \mathbf{0}$ and $\xi_1 = \mathbf{0}$, then the DCC-X-RGARCH-X model reduces to the DCC-RGARCH model of Fiszeder et al. (2019). Supposing $\sigma_{P,k\ t-i}^2$ in Eq. (7) are replaced by ε_{kt-i}^2 and simultaneously $\boldsymbol{\varphi}_{k1} = \mathbf{0}$, then the DCC-X-RGARCH-X model reduces to the DCC-X model of Vargas (2008).[2] If it is additionally assumed that $\xi_1 = \mathbf{0}$, then it is equivalent to the DCC model of Engle (2002).

Parameters of the DCC-X-RGARCH-X model, similarly to the parameters of the DCC model, can be estimated by the quasi-maximum likelihood method using a two-stage approach (see Engle and Sheppard, 2001). Let the parameters of the DCC-X-RGARCH-X model $\boldsymbol{\Theta}$ be written in two groups $\boldsymbol{\Theta}' = (\boldsymbol{\Theta}'_1, \boldsymbol{\Theta}'_2)$, where $\boldsymbol{\Theta}_1$ is the vector of parameters of conditional means and variances and $\boldsymbol{\Theta}_2$ is the vector of parameters of the correlation part of the model. The log-likelihood function can be presented as the sum of two parts:

[1] In the empirical part of the paper, we take this less restrictive approach and assume that $\boldsymbol{K}$ is a matrix of ones and check the positive definiteness of the $\mathbf{Q}_t$ matrix for all t.

[2] Actually, the model of Vargas (2008) contains an additional part responsible for asymmetric effects which we omit in the analysis.

$$L(\boldsymbol{\Theta}) = L_{Vol}(\boldsymbol{\Theta}_1) + L_{Corr}(\boldsymbol{\Theta}_2|\boldsymbol{\Theta}_1), \tag{8}$$

where $L_{Vol}(\boldsymbol{\Theta}_1)$ can be viewed as the volatility component:

$$L_{Vol}(\boldsymbol{\Theta}_1) = -\frac{1}{2}\sum_{t=1}^{n}\left(Nln(2\pi) + 2ln|\mathbf{D}_t| + \boldsymbol{\varepsilon}_t'\mathbf{D}_t^{-2}\boldsymbol{\varepsilon}_t\right), \tag{9}$$

while $L_{Corr}(\boldsymbol{\Theta}_2|\boldsymbol{\Theta}_1)$ represents the correlation part:

$$L_{Corr}(\boldsymbol{\Theta}_2|\boldsymbol{\Theta}_1) = -\frac{1}{2}\sum_{t=1}^{n}\left(ln|\mathbf{cor}_t| + \mathbf{z}_t'\mathbf{cor}_t^{-1}\mathbf{z}_t - \mathbf{z}_t'\mathbf{z}_t\right) \tag{10}$$

$L_{Vol}(\boldsymbol{\Theta}_1)$ can be presented as the sum of log-likelihood functions of N univariate RGARCH-X models:

$$L_{Vol}(\boldsymbol{\Theta}_1) = -\frac{1}{2}\sum_{k=1}^{N}\left(nln(2\pi) + \sum_{t=1}^{n}\left(ln(h_{kt}) + \frac{\varepsilon_{kt}^2}{h_{kt}}\right)\right) \tag{11}$$

Therefore in the first stage, the parameters of univariate RGARCH-X models can be estimated separately for each of the assets and the estimates of h_{kt} can be obtained. In the second stage, standardized by their estimated standard deviations residuals are applied to estimate the parameters of the correlation part ($\boldsymbol{\Theta}_2$) conditioning on the parameters estimated in the first stage ($\widehat{\boldsymbol{\Theta}}_1$).

The positive definiteness of the conditional covariance matrix can be achieved by the constrained quasi-maximum likelihood estimation (see Chou and Liao, 2008).

4. Data

We analyze how Google searches for the term "oil prices" influence variances and covariances for three exchange-traded funds (ETFs) listed on the New York Stock Exchange Arca, namely: United States Oil Fund (holds crude oil futures contracts and other oil-related contracts, predominantly short-term NYMEX futures contracts on WTI crude oil), SPDR Portfolio S&P 500 Growth (holds large-capitalization growth stocks selected from the S&P 500 index), and SPDR Gold Shares (holds gold bullion). They will be referred to hereinafter as oil, stocks and gold, respectively.

The dynamics of the opening jump (the difference between today's opening price and yesterday's closing price) is different from the dynamics of the trading part of the day. The overnight volatility causes the noise, so to circumvent that issue, we investigate open-to-close returns instead of close-to-close returns. We apply the percentage logarithmic returns calculated as $r_t = 100 \ln (p_{ct}/p_{ot})$, where p_{ct} and p_{ot} are closing and opening prices at time t, respectively.

We evaluate the competing models based on daily data in the fourteen-year and two-month period from April 13, 2006, to June 11, 2020. This is a relatively large sample that comprises not only a very volatile periods, i.e. the collapse of Lehman Brothers, the global financial crisis, the European sovereign debt crisis and the financial turmoil associated with the outbreak of the COVID-19, but also tranquil periods with low volatility. Fig. 1 presents daily closing prices and returns for the three analyzed ETFs.

The use of Google searches has attracted lots of attention in the literature (see e.g. Da et al., 2011; Joseph et al., 2011; Vozlyublennaia, 2014; Bijl et al., 2016; Gwilym et al., 2016; Aalborg et al., 2019; Kim et al., 2019; among others). The Google Trends platform is used to obtain Google's search volume index (henceforth *SVI*) that has been deemed as a significant proxy for investors' attention (see Da et al., 2011). The search phrase we used is "oil prices". This phrase is directly related to oil and it also has enough searches to provide a daily volume index. Additionally, it is more likely to be used by investors and traders than other internet users because it contains the word "prices". This way, the attention coming from potential or current individual investors can be transferred into trading orders that are related to oil. The same term was used by Campos et al. (2017) for forecasting the CBOE Crude Oil Volatility Index based on Heterogeneous Autoregressive models.

We download daily data for 8 month intervals, with 3 month overlapping periods. Google Trends uses a standardized scale of 0 to 100, where 100 represents the highest query volume during a considered time period. ETFs are quoted on business days from Monday to Friday, therefore we exploit the same days for *SVI*. Raw *SVI* shouldn't be used in the analysis directly, because its value depends on the time period of downloaded data. Following Bijl et al. (2016) we standardize *SVI* to obtain abnormal *SVI* (henceforth *ASVI*). Three months of overlapping observations between every pair of adjacent windows is used to standardize the scale of the next window. The formula for $ASVI_t$ is given as:

$$ASVI_t = \frac{SVI_t - \overline{SVI}_t}{S_{SVI,t}}, \tag{12}$$

where $\overline{SVI}_t$ and $S_{SVI,\,t}$ are respectively the mean and standard deviation of SVI_t calculated from the past 3 months as follows $\overline{SVI}_t = \frac{1}{66}\sum_{i=1}^{66}SVI_{t-i}$, $S_{SVI,t} = \left(\frac{1}{66}\sum_{i=1}^{66}(SVI_{t-i} - \overline{SVI}_t)^2\right)^{1/2}$.

It is worth analyzing not only the level of $ASVI_t$ but also its changes. That is why we also investigate the first differences of $ASVI_t$ calculated as $\Delta ASVI_t = ASVI_t - ASVI_{t-1}$. The values of SVI_t and $\Delta ASVI_t$ are presented in Fig. 1.

Table 1 gives the descriptive statistics for the logarithmic returns r_t of the analyzed ETFs, levels of $ASVI_t$ and its first differences. SVI_t values are not used directly in the models, which is why we do not compare its statistics with other series. The means are positive for stocks returns and $ASVI_t$ and negative for oil returns, gold returns, and the first differences of $ASVI_t$. The highest and lowest values of standard deviation are for oil returns and gold returns, respectively. All distributions are asymmetric and all display high kurtosis. The distributions of $ASVI_t$ and $\Delta ASVI_t$ are significantly different from distributions of ETFs returns. Both of these series have much stronger asymmetry and higher leptokurtosis than returns of gold, stocks, or oil.

5. The influence of Google searches on volatility and dependence of returns

We consider four DCC models in the analysis:

1) The DCC-GARCH model by Engle (2002),
2) The DCC-X-GARCH-X model by Vargas (2008).[3] In this specification, the exogenous variable is incorporated in the DCC-GARCH model.
3) The DCC-RGARCH model by Fiszeder et al. (2019). In this specification, the RGARCH model is applied in the DCC model instead of the univariate GARCH model.
4) The proposed DCC-X-RGARCH-X model that is summarized by Eqs. (2)–(7). In this specification, the exogenous variable is incorporated in the DCC-RGARCH model.

In both the DCC-X-GARCH-X and DCC-X-RGARCH-X models the exogenous variable $\Delta ASVI_t$ is incorporated into equations of conditional means, variances and covariances. We also tried $ASVI_t$ but its influence on ETFs was much weaker. $\Delta ASVI_t$ is calculated for the search words "oil prices". For the GARCH-X and RGARCH-X models we check the positivity of the conditional variance h_{kt}, and it is met for all time t. For both the DCC-X-GARCH-X and DCC-X-RGARCH-X models, we allow ξ_{11} to take on negative values and check the positive definiteness of the matrix $\mathbf{Q}_t$. This condition is met for all t.

Parameters of the DCC-GARCH and DCC-X-GARCH-X models are estimated only based on closing prices, whereas for the estimation of

[3] Actually, the model of Vargas (2008) is the DCC-X-GARCH model, and it does not contain exogenous variable in the conditional means and variances but we add these components for a fair comparison.

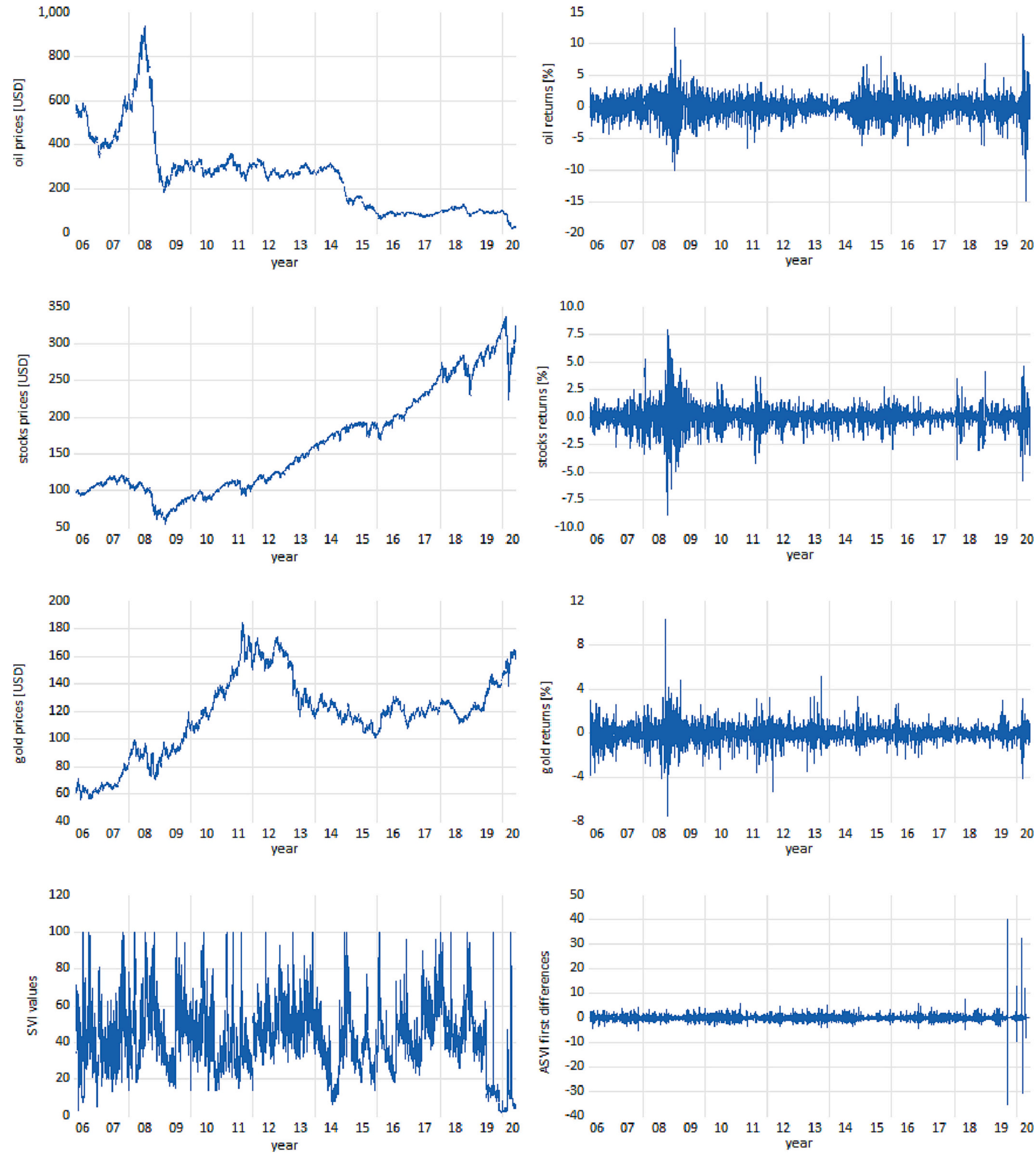

Fig. 1. Daily closing prices and returns of ETFs and values of SVI and first differences of ASVI.

both, i.e. the DCC-RGARCH and DCC-X-RGARCH-X models, both closing and low and high prices are applied.

An application of the three alternative models is for benchmark purposes and for ability to assess the importance of additional information from Google searches and the range-based model framework.

Firstly, we compare the in-sample fitness of the models (Section 5.1) and then evaluate the forecasts from these models. We analyze variance (Section 5.2) and covariance forecasts (Section 5.3) separately because research regarding Google searches as the exogenous variable has already been performed for variances, whereas covariance forecasting in such a setup is virtually unknown.

5.1. In-sample analysis

Each mean equation is a constant because the sample returns of any ETF are not dependent on their own past returns nor on the past returns

Table 1
Summary statistics for daily returns of ETFs and values of search volume indices.

Assets	Mean × 10^2	Minimum	Maximum	Standard deviation	Skewness	Excess kurtosis
Exchange-traded funds						
Oil	−2.553	−14.842	12.601	1.695	−0.219	9.703
Stocks	1.359	−8.922	7.956	0.989	−0.347	13.713
Gold	−0.941	−7.483	10.328	0.805	0.231	16.899
Google searches						
ASVI	0.799	−3.747	38.869	1.539	8.570	189.892
ΔASVI	−0.021	−35.409	40.071	1.640	1.931	243.366

The sample period is from April 13, 2006, to June 11, 2020. Δ means first differences of the series.

of other ETFs. The parameters of the considered models are estimated using the quasi-maximum likelihood method with robust standard errors. The in-sample analysis is performed for the period from April 13, 2006, to April 9, 2010, and the estimates are given in Table 2. The out-of-sample forecasting results are presented in Sections 5.2 and 5.3.

For all competing models, we calculate the likelihood function, including both the volatility and correlation parts. In order to assess whether the differences between values of the likelihood function of considered models are statistically significant, we apply the Rivers and Vuong (2002) and Clarke (2007) tests for non-nested models. The results of the tests for pairs of models are given in Table 3. The value of the likelihood function is significantly higher for the DCC-X-RGARCH-X model than for the competing models, which means that the proposed model best describes the dynamics of the three ETFs. Moreover, the DCC-X-GARCH-X model is better than the standard DCC-GARCH model, which indicates that information about "oil prices" from Google searches is important in modelling the returns of oil, stocks, and gold. On the other hand, the value of the likelihood function of the DCC-RGARCH model is significantly higher than the value of the DCC-X-GARCH-X model. It supports the claim that incorporating low and high prices has greater importance than including the exogenous variable, although not conclusively.

The usage of range data significantly changes the parameters estimates. Specifically, the estimates of the parameters α_{k1} are much higher and the estimates of the parameters β_{k1} much lower in the RGARCH and RGACH-X models compared with the GARCH and GARCH-X models. It is vital for modelling and forecasting volatility because for the RGARCH and RGARCH-X models, the shocks in the previous period have a stronger influence on the current volatility than the impact which is observed for the GARCH and GARCH-X models. Thus, models formulated with range data respond more quickly to changing market conditions. This property has already been demonstrated in other studies (see e.g. Chou et al., 2009; Fiszeder and Fałdziński, 2019; Fiszeder et al., 2019). On the other hand, there are no considerable differences between the analyzed models in the estimates of parameters for the correlation component.

The estimates of the parameters γ_{k1} and φ_{k1} are not statistically different from zero. It means that Google searches for the words "oil prices" do not influence the means and variances of ETFs returns. On the other hand, the estimates of the parameter ξ_{11} are negative and statistically significant. It implies that the increase of Google searches causes the decrease of the covariance of ETFs returns.

In order to get insight into the relationship between covariances and Google searches, we also apply the τ-th linear quantile regression model (see Koenker and Bassett Jr., 1978; Koenker, 2005). It can be written as:

$$cov_{R,t} = \gamma_0(\tau) + \gamma_1(\tau) ASVI_{t-1} + \varepsilon_t(\tau) \tag{13}$$

where $cov_{R,\,t}$ is the realized covariance given as the sum of products of 5-min returns, $ASVI_{t-1}$ is the abnormal Google's search volume index for the term "oil prices".

Table 2
Estimation results of the four DCC models for the analyzed ETFs.

Parameters	DCC-GARCH		DCC-X-GARCH-X		DCC-RGARCH		DCC-X-RGARCH-X	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
γ_{10}	0.065	0.031	0.027	0.025	−0.007	0.026	−0.007	0.026
γ_{11}	–	–	0.013	0.027	–	–	−0.003	0.023
α_{10}	0.017	0.012	0.017	0.009	0.023	0.014	0.024	0.014
α_{11}	0.090	0.018	0.115	0.025	0.251	0.049	0.250	0.048
β_{11}	0.903	0.018	0.876	0.027	0.758	0.047	0.756	0.046
φ_{11}	–	–	−0.043	0.035	–	–	−0.048	0.032
γ_{20}	0.050	0.062	0.027	0.050	0.002	0.050	0.003	0.051
γ_{21}	–	–	−0.001	0.041	–	–	−0.008	0.043
α_{20}	0.044	0.025	0.041	0.021	0.048	0.029	0.046	0.029
α_{21}	0.066	0.014	0.065	0.012	0.092	0.021	0.090	0.021
β_{21}	0.927	0.015	0.923	0.014	0.890	0.025	0.893	0.026
φ_{21}	–	–	0.009	0.117	–	–	0.057	0.098
γ_{30}	0.074	0.042	−0.005	0.029	−0.001	0.028	−0.002	0.028
γ_{31}	–	–	0.003	0.023	–	–	0.001	0.024
α_{30}	0.018	0.010	0.009	0.005	0.002	0.010	0.002	0.010
α_{31}	0.052	0.015	0.064	0.019	0.109	0.032	0.108	0.031
β_{31}	0.940	0.012	0.929	0.012	0.893	0.018	0.894	0.018
φ_{31}	–	–	0.016	0.031	–	–	0.016	0.034
ζ_1	0.037	0.006	0.035	0.007	0.038	0.008	0.038	0.008
θ_1	0.952	0.009	0.950	0.012	0.945	0.013	0.944	0.013
ξ_{11}	–	–	−0.039	0.015	–	–	−0.041	0.015
ln L	−5376.4		−4592.9		−4562.4		−4558.4	

The in-sample period is from April 13, 2006, to April 9, 2010, γ_{k0}, γ_{k1} are the parameters of the conditional mean equations, α_{k0}, α_{k1}, β_{k1}, φ_{k1} are the parameters of the univariate volatility models, $k = 1, 2, 3$ for oil, stocks and gold, respectively, ζ_1, θ_1, ξ_{11} are the parameters of the correlation part. SE means standard error, ln L is the logarithm of the likelihood function.

Table 3
Results of the Rivers-Vuong and Clarke tests for model selection.

Tested models	RV statistic	RV p-value	Clarke statistic	Clarke p-value
DCC-X-GARCH-X vs DCC-GARCH	23.229	0.000	38.660	0.000
DCC-RGARCH vs DCC-GARCH	23.224	0.000	38.894	0.000
DCC-X-RGARCH-X vs DCC-GARCH	23.020	0.000	38.425	0.000
DCC-RGARCH vs DCC-X-GARCH-X	5.819	0.000	12.473	0.000
DCC-X-RGARCH-X vs DCC-X-GARCH-X	7.076	0.000	13.982	0.000
DCC-X-RGARCH-X vs DCC-RGARCH	2.213	0.013	1.777	0.038

The in-sample period is from April 13, 2006, to April 9, 2010. The RV (Rivers-Vuong) and Clarke are test statistics for model selection, where comparisons are made with the benchmark model (the second model in a pair). A low p-value means that the first model in a pair is superior to the second model.

The results of the parameters estimation for the 10th, 50th, and 90th quantiles are given in Table 4.

The increase of Google searches for "oil prices" causes the decrease of the covariance between oil and stock returns as well as stock and gold returns. On the other hand, the impact of $ASVI_t$ on the covariance between oil and gold returns is positive. The strongest influence of Google searches on the relationship between ETFs is observed for very high values of covariance (above the 90th quantile) between oil and stock returns, and it is negative.

5.2. Comparison of variance forecasts

In this section, we compare the forecasting performance of the four univariate models, which are used in the DCC models. The two models, i.e., GARCH and RGARCH, are based only on the historical prices of ETFs, whereas the other two, i.e., GARCH-X and RGARCH-X, also use Google searches. For the starting sample (i.e., April 13, 2006 to April 9, 2010), we estimate the parameters of the models and compute out-of-sample one-day-ahead forecasts of variance. Consecutively, we add one new observation to the estimation sample, while at the same time dropping the oldest observation, so we apply the rolling window approach. Afterwards, based on the new estimation sample, we re-estimate models and obtain forecasts. The procedure is repeated until we obtain forecasts for the ten-year and two-month period from April 12, 2010, to June 11, 2020.

The sum of squares of 5-min returns (the realized variance) is used as a proxy of the daily variance. The realized variance can be viewed as a standard approach in the literature (see e.g. Floros et al., 2020; Reschenhofer et al., 2020; Zhang et al., 2020; Kambouroudis et al., 2021). We also applied 15-min returns instead of 5-min returns, and the conclusions are very similar to those given in the paper. The forecasts are evaluated based on two measures, namely, the mean squared error (MSE) and the mean absolute error (MAE). The statistical significance of the results is evaluated based on the test of superior predictive ability (SPA) of Hansen (2005) and the model confidence set (MCS) test of Hansen et al. (2011). The SPA test verifies whether each of the models is outperformed by any of the alternatives. The MCS test provides with the best forecasting models that belong to the so-called the model confidence set with a certain probability. Table 5 presents the results from the MSE and MAE measures and the SPA and MCS tests.

According to the MSE criterion and the SPA test, there are three models: GARCH-X, RGARCH and RGARCH-X, which are not outperformed significantly by any of the alternatives. Moreover, these three models belong to the model confidence set. It means that forecasts of variance from the GARCH-X, RGARCH and RGARCH-X models are more accurate than forecasts from the standard GARCH model. At the same time, however, it is not possible to decisively point out the best forecasting model among them. According to the MAE measure, the forecasts based on two range-based models, namely the RGARCH and RGARCH-X, provide the most accurate forecasts. The obtained results do not clearly indicate whether the information from Google searches about "oil prices" increases the forecasting accuracy. On the one hand, such information improves volatility forecasts for the standard GARCH model, however, it has no significant influence on the RGARCH model. This means that including additional variable in a less-precise volatility model (GARCH) might seem to be important, but this additional variable might turn out not to be important once a more precise volatility model (RGARCH) is used.

5.3. Comparison of covariance forecasts

The main focus of the paper is the influence of Google searches on covariance forecasts, thus in this section, we compare out-of-sample one-day-ahead forecasts of covariance from the proposed DCC-X-RGARCH-X model with the three competing DCC models: DCC-GARCH, DCC-X-GARCH-X and DCC-RGARCH. We use the same estimation and forecasting samples as for the variances analysis in Section 5.2. The sum of products of 5-min returns is employed as a proxy of the daily covariance for the evaluation of the forecasts. The realized

Table 4
Quantile regression models of the realized covariance on the lagged $ASVI_t$.

Covariance of ETFs	$\gamma_0(\tau)$	s ($\gamma_0(\tau)$)	p-value	$\gamma_1(\tau)$	s ($\gamma_1(\tau)$)	p-value
10th quantile						
Oil-stocks	−0.189	0.021	0.000	−0.125	0.016	0.000
Gold-oil	0.007	0.011	0.515	0.033	0.012	0.005
Stocks-gold	−0.219	0.029	0.000	−0.025	0.021	0.243
50th quantile						
Oil-stocks	0.183	0.021	0.000	−0.097	0.018	0.000
Gold-oil	0.335	0.013	0.000	0.059	0.010	0.000
Stocks-gold	0.071	0.008	0.000	−0.020	0.007	0.003
90th quantile						
Oil-stocks	2.074	0.178	0.000	−0.445	0.073	0.000
Gold-oil	1.269	0.078	0.000	0.004	0.058	0.942
Stocks-gold	0.584	0.031	0.000	−0.059	0.025	0.019

The in-sample period is from April 13, 2006, to April 9, 2010, the realized covariance is estimated as the sum of products of 5-min returns, $\gamma_0(\tau)$, $\gamma_1(\tau)$ are the parameters of the regression (Eq. (13)), s($\gamma_0(\tau)$), s($\beta_1(\tau)$) are standard errors calculated using the Markov chain marginal bootstrap method.

Table 5
Evaluation of variance forecasts for the analyzed ETFs.

Model	Forecast evaluation criteria					
	MSE	SPA p-value	MCS p-value	MAE	SPA p-value	MCS p-value
GARCH	47.732	0.006	0.001	1.597	0.000	0.000
GARCH-X	31.321	0.126	0.237*	0.722	0.029	0.060
RGARCH	29.378	0.837	1.000*	0.690	0.676	0.858*
RGARCH-X	29.387	0.600	0.614*	0.690	0.795	1.000*

The evaluation period is April 12, 2010, to June 11, 2020, the realized variance is used as a proxy of variance and estimated as the sum of squares of 5-min returns.

* indicates that models belong to the model confidence set with a confidence level of 0.90.

covariance approach is commonly seen in the literature (see e.g. Laurent et al., 2013; Fiszeder et al., 2019; Gkillas et al., 2021b). As in the previous section, for 15-min returns we obtain similar results. In order to save space, we present only those for 5-min returns. Once again, we use the same evaluation measures and tests to evaluate forecasts as for variance. The forecasting performance results are presented in Table 6.

The SPA test results indicate that the only model which is not outperformed significantly (at the 10% significance level) by any of the alternatives is the DCC-X-RGARCH-X. According to the results of the MCS test, only the DCC-X-RGARCH-X model belongs to the model confidence set. The forecasting superiority of the proposed model does not depend on the type of loss function. Moreover, information from Google searches about "oil prices" increases the forecasting accuracy for both the standard DCC-GARCH model and the DCC-RGARCH model. The DCC-X-GARCH-X more accurately forecasts covariance than the DCC-RGARCH model. It means that information from Google searches is more important for covariance forecasting than the application of the univariate range-based volatility model. However, the best choice is to use the more precise volatility model (RGARCH) and improve it further with information from Google searches.

6. Robustness check of the COVID-19 influence

The outbreak of the COVID-19 pandemic had a huge impact on the oil market (see e.g. Bourghelle et al., 2021; Jia et al., 2021; Le et al., 2021). Lockdowns, travel restrictions, and economic turbulence led to the oil price crash in 2020. On April 20, 2020, the West Texas Intermediate crude oil price dropped to negative levels for the first time in history. Fear connected with the COVID-19 pandemic also revealed excess Google's search volume (see Lyócsa et al., 2020). The period of extreme market uncertainty coincided with the period of investors increased attention to coronavirus events. After the outbreak of the COVID-19 pandemic $\Delta ASVI_t$ for the search words "oil prices" took huge values that had never been seen before (see Fig. 1). Such outliers may cause bias on the usual maximum likelihood estimator of the parameters of GARCH models and the estimated volatilities. They can lead also to a considerable deterioration of the forecasting accuracy (Catalán and Trívez, 2007; Trucíos and Hotta, 2015). To check the robustness of our forecasting results to the COVID-19 crisis we perform the analysis for the shorter period not covering the COVID-19 pandemic. The results are presented in Tables 7 and 8 for variance and covariance forecasts, respectively.

According to both MSE and MAE criteria and both SPA and MCS tests the forecast of variance from the RGARCH and RGARCH-X models are the most accurate. The information from Google searches about "oil prices" increases the accuracy of variance forecasts but only for the standard GARCH model.

According to both MSE and MAE criteria and both SPA and MCS tests, the covariance forecasts from the DCC-X-RGARCH-X model are the most precise. The information from Google searches about "oil prices" increases the accuracy of covariance forecasts for both the standard DCC-GARCH model and the DCC-RGARCH model. It means that the main conclusions of the research remain unchanged, irrespective of whether the COVID-19 crisis is included in the analysis or not.

Table 6
Evaluation of covariance forecasts for the analyzed ETFs.

Model	Forecast evaluation criteria					
	MSE	SPA p-value	MCS p-value	MAE	SPA p-value	MCS p-value
DCC-GARCH	1.427	0.026	0.010	0.395	0.000	0.000
DCC-X-GARCH-X	0.452	0.059	0.041	0.220	0.001	0.000
DCC-RGARCH	0.659	0.001	0.010	0.329	0.000	0.000
DCC-X-RGARCH-X	0.381	0.942	1.000*	0.212	0.527	1.000*

The evaluation period is April 12, 2010, to June 11, 2020, the realized covariance is used as a proxy of covariance and estimated as the sum of products of 5-min returns.

* indicates that models belong to the model confidence set with a confidence level of 0.90.

Table 7
Evaluation of variance forecasts for the period without the COVID-19 turmoil.

Model	Forecast evaluation criteria					
	MSE	SPA p-value	MCS p-value	MAE	SPA p-value	MCS p-value
GARCH	3.107	0.000	0.000	1.127	0.000	0.000
GARCH-X	0.987	0.002	0.002	0.496	0.000	0.000
RGARCH	0.883	0.386	0.709*	0.462	0.278	0.566*
RGARCH-X	0.882	0.614	1.000*	0.462	0.614	1.000*

The evaluation period is April 12, 2010, to December 31, 2019, the realized variance is used as a proxy of variance and estimated as the sum of squares of 5-min returns.

* indicates that models belong to the model confidence set with a confidence level of 0.90.

Table 8
Evaluation of covariance forecasts for the period without the COVID-19 turmoil.

Model	Forecast evaluation criteria					
	MSE	SPA p-value	MCS p-value	MAE	SPA *p*-value	MCS p-value
DCC-GARCH	0.285	0.000	0.000	0.327	0.000	0.000
DCC-X-GARCH-X	0.131	0.007	0.005	0.179	0.000	0.000
DCC-RGARCH	0.228	0.000	0.000	0.281	0.000	0.000
DCC-X-RGARCH-X	0.123	0.536	1.000*	0.172	0.513	1.000*

The evaluation period is April 12, 2010, to December 31, 2019, the realized covariance is used as a proxy of covariance and estimated as the sum of products of 5-min returns.

* indicates that models belong to the model confidence set with a confidence level of 0.90.

7. Conclusions

The relation between investor attention and financial markets has recently attracted a lot of interest. Google's search volume index is considered an important proxy for the attention and sentiment of investors. Such information is important not only for the stock market but also for the commodity market, especially for the oil market. We analyze the influence of attention to oil prices, measured by Google searches for the term "oil prices", on variances and covariances of three exchange-traded funds representing oil, the stock market, and gold.

For this purpose, we introduce a new DCC-X-RGARCH-X model. It is an extension of the DCC-RGARCH model of Fiszeder et al. (2019) which incorporates exogenous variables in the equations of conditional means, variances, and covariances. We show in the in-sample analysis that attention to oil prices does not influence the return and volatility of oil, gold, and the stock market but does have a significant impact on the covariance of their returns. The rise of interest in the phrase "oil prices" induces *the opposite reaction of stocks to oil and gold.* On the other hand, the increase of Google searches causes similar changes between oil and gold returns. The proposed DCC-X-RGARCH-X model describes the dynamics of oil, gold and the stock market better than competing models.

In the out-of-sample analysis, the influence of Google searches for the phrase "oil prices" on the accuracy of variance forecasts depends on the considered volatility model. Such information improves volatility forecasts for the standard GARCH model, however, it has no significant influence on the more precise RGARCH model. On the other hand, the models which incorporate Google searches for the phrase "oil prices" better forecast covariance than the standard models based exclusively on the past returns of series. The DCC-X- RGARCH-X model outperforms other considered multivariate volatility models, and this forecasting superiority is robust to the type of loss function and the inclusion of the COVID-19 crisis.

Altogether, our results show the importance of investor attention to oil prices as a factor influencing the covariance of oil returns with the stock market returns and with gold returns. In addition, it also improves the standard GARCH model for oil return volatility.

CRediT authorship contribution statement

Piotr Fiszeder: Conceptualization, Methodology, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Marcin Fałdziński:** Conceptualization, Methodology, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Peter Molnár:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing.

Acknowledgements

Financial support from the Czech Science Foundation under grant no. 20-16786S is acknowledged.

References

Aalborg, H.A., Molnár, P., de Vries, J.E., 2019. What can explain the price, volatility and trading volume of bitcoin? Financ. Res. Lett. 29, 255–265.

Afkhami, M., Cormack, L., Ghoddusi, H., 2017. Google search keywords that best predict energy price volatility. Energy Econ. 67, 17–27.

Ahmadi, M., Behmiri, N.B., Manera, M., 2016. How is volatility in commodity markets linked to oil price shocks? Energy Econ. 59, 11–23.

Bašta, M., Molnár, P., 2018. Oil market volatility and stock market volatility. Financ. Res. Lett. 26, 204–214.

Behmiri, N.B., Manera, M., Nicolini, M., 2019. Understanding dynamic conditional correlations between oil, natural gas and non-energy commodity futures markets. Energy J. 40 (2).

Bijl, L., Kringhaug, G., Molnár, P., Sandvik, E., 2016. Google searches and stock returns. Int. Rev. Financ. Anal. 45, 150–156.

Bollerslev, T., 1986. Generalised autoregressive conditional heteroskedasticity. J. Econ. 31, 307–327.

Bourghelle, D., Jawadi, F., Rozin, P., 2021. Oil price volatility in the context of Covid-19. Int. Econ. 167, 39–49.

Campos, I., Cortazar, G., Reyes, T., 2017. Modeling and predicting oil VIX: internet search volume versus traditional mariables. Energy Econ. 66, 194–204.

Catalán, B., Trívez, F.J., 2007. Forecasting volatility in GARCH models with additive outliers. Quant. Financ. 7 (6), 591–596.

Chen, Q., Zhu, H., Yu, D., Hau, L., 2022. How does investor attention matter for crude oil prices and returns? Evidence from time-frequency quantile causality analysis. N. Am. J. Econ. Financ. 59, 101581.

Chou, R.Y., Liao, W.-Y., 2008. Explaining the Great Decoupling of the Equity-Bond Linkage with a Modified Dynamic Conditional Correlation Model, Institute of Economics, Academia Sinica Taipei, Taiwan Working Paper.

Chou, R.Y., Wu, C.C., Liu, N., 2009. Forecasting time-varying covariance with a range-based dynamic conditional correlation model. Rev. Quant. Finan. Acc. 33 (4), 327–345.

Clarke, K.A., 2007. A simple distribution-free test for nonnested model selection. Polit. Anal. 15 (3), 347–363.

Da, Z., Engelberg, J., Gao, P., 2011. In search of attention. J. Financ. 66, 1461–1499.

De Nard, G., Engle, R.F., Ledoit, O., Wolf, M., 2022. Large dynamic covariance matrices: enhancements based on intraday data. J. Bank. Financ. 138, 106426.

Elshendy, M., Colladon, A.F., Battistoni, E., Gloor, P.A., 2018. Using four different online media sources to forecast the crude oil price. J. Inf. Sci. 44 (3), 408–421.

Engle, R.F., 2002. Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. J. Bus. Econ. Stat. 20 (3), 339–350.

Engle, R.F., Sheppard, K., 2001. Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH, NBER Working Paper No. 8554.

Fiszeder, P., Fałdziński, M., 2019. Improving forecasts with the co-range dynamic conditional correlation model. J. Econ. Dyn. Control. 108, 103736.

Fiszeder, P., Małecka, M., 2022. Forecasting volatility during the outbreak of Russian invasion of Ukraine: application to commodities, stock indices, currencies, and cryptocurrencies. Equilibr. Q. J. Econ. Econ. Policy 17 (4), 939–967.

Fiszeder, P., Fałdziński, M., Molnár, P., 2019. Range-based DCC models for covariance and value-at-risk forecasting. J. Empir. Financ. 54, 58–76.

Fiszeder, P., Fałdziński, M., Molnár, P., 2023. Modeling and forecasting dynamic conditional correlation with opening, high, low and closing prices. J. Empir. Financ. 70, 308–321.

Floros, C., Gkillas, K., Konstantatos, C., Tsagkanos, A., 2020. Realized measures to explain volatility changes over time. J. Risk Financ. Manag. 13 (6), 125.

Gjerstad, P., Meyn, P.F., Molnár, P., Næss, T.D., 2021. Do president trump's tweets affect financial markets? Decis. Support. Syst. 147, 113577.

Gkillas, K., Gupta, R., Pierdzioch, C., 2020. Forecasting realized oil-price volatility: the role of financial stress and asymmetric loss. J. Int. Money Financ. 104, 102137.

Gkillas, K., Gupta, R., Pierdzioch, C., Yoon, S.M., 2021a. OPEC news and jumps in the oil market. Energy Econ. 96, 105096.

Gkillas, K., Konstantatos, C., Siriopoulos, C., 2021b. Uncertainty due to infectious diseases and stock-bond correlation. Econometrics 9 (2), 17.

Guo, J.F., Ji, Q., 2013. How does market concern derived from the internet affect oil prices? Appl. Energy 112, 1536–1543.

Gwilym, O.A., Hasan, I., Wang, Q., Xie, R., 2016. In search of concepts: the effects of speculative demand on stock returns: in search of concepts. Eur. Financ. Manag. 22, 427–449.

Hansen, P.R., 2005. A test for superior predictive ability. J. Bus. Econ. Stat. 23 (4), 365–380.

Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79, 453–497.

Hung, J.C., Lee, M.C., Liu, H.C., 2008. Estimation of value-at-risk for energy commodities via fat-tailed GARCH models. Energy Econ. 30 (3), 1173–1191.

Hung, J.C., Wang, Y.H., Chang, M.C., Shih, K.H., Kao, H.H., 2011. Minimum variance hedging with bivariate regime-switching model for WTI crude oil. Energy 36 (5), 3050–3057.

Ji, Q., Guo, J.F., 2015. Oil price volatility and oil-related events: an internet concern study perspective. Appl. Energy 137, 256–264.

Jia, Z., Wen, S., Lin, B., 2021. The effects and reacts of COVID-19 pandemic and international oil price on energy, economy, and environment in China. Appl. Energy 302, 117612.

Joseph, K., Babajide Wintoki, M., Zhang, Z., 2011. Forecasting abnormal stock returns and trading volume using investor sentiment: evidence from online search. Int. J. Forecast. 27, 1116–1127.

Kambouroudis, D.S., McMillan, D.G., Tsakou, K., 2021. Forecasting realized volatility: the role of implied volatility, leverage effect, overnight returns, and volatility of realized volatility. J. Futur. Mark. 41, 1618–1639.

Kim, B.-H., Kim, H., Min, H.-G., 2013. Reassessing the link between Japanese yen and emerging Asian currencies. J. Int. Money Financ. 33, 306–326.

Kim, N., Lučivjanská, K., Molnár, P., Villa, R., 2019. Google searches and stock market activity: evidence from Norway. Financ. Res. Lett. 28, 208–220.

Koenker, R., 2005. Quantile Regression. Cambridge University Press.

Koenker, R., Bassett Jr., G., 1978. Regression quantiles. Econometrica 46 (1), 33–50.

Laurent, S., Rombouts, J.V.K., Violante, F., 2013. On loss functions and ranking forecasting performances of multivariate volatility models. J. Econ. 173 (1), 1–10, 2013.

Le, T.-H., Le, A.T., Le, H.-C., 2021. The historic oil price fluctuation during the Covid-19 pandemic: what are the causes? Res. Int. Bus. Financ. 58, 101489.

Li, X., Ma, J., Wang, S., Zhang, X., 2015. How does Google search affect trader positions and crude oil prices? Econ. Model. 49, 162–171.

Li, S., Zhang, H., Yuan, D., 2019. Investor attention and crude oil prices: evidence from nonlinear granger causality tests. Energy Econ. 84, 104494.

Li, J., Tang, L., Wang, S., 2020. Forecasting crude oil price with multilingual search engine data. Phys. A: Stat. Mech. Appl. 551, 124178.

Liu, Y., Niu, Z., Suleman, M.T., Yin, L., Zhang, H., 2022. Forecasting the volatility of crude oil futures: the role of oil investor attention and its regime switching characteristics under a high-frequency framework. Energy 238, 121779.

Lyócsa, Š., Baumöhl, E., Výrost, T., Molnár, P., 2020. Fear of the coronavirus and the stock markets. Financ. Res. Lett. 36, 101735.

Min, H.-G., Hwang, Y.-S., 2012. Dynamic correlation analysis of US financial crisis and contagion: evidence from four OECD countries. Appl. Financ. Econ. 22, 2063–2074.

Molnár, P., 2016. High-low range in GARCH models of stock return volatility. Appl. Econ. 48 (51), 4977–4991.

Parkinson, M., 1980. The extreme value method for estimating the variance of the rate of return. J. Bus. 53 (1), 61–65.

Prange, P., 2021. Does online investor attention drive the co-movement of stock-, commodity-, and energy markets? Insights from Google searches. Energy Econ. 99, 105282.

Qadan, M., Nama, H., 2018. Investor sentiment and the price of oil. Energy Econ. 69, 42–58.

Reschenhofer, E., Mangat, M.K., Stark, T., 2020. Volatility forecasts, proxies and loss functions. J. Empir. Financ. 59, 133–153.

Rivers, D., Vuong, Q., 2002. Model selection tests for nonlinear dynamic models. Econ. J. 5 (1), 1–39.

Schopen, J.-H., 2012. Exogenous Variables in Dynamic Conditional Correlation Models for Financial Markets, PhD Thesis. Bremen University.

Sheppard, K., 2008. Economic Factors and the Covariance of Equity Returns, Working Paper. University of Oxford.

Silvennoinen, A., Teräsvirta, T., 2005. Multivariate Autoregressive Conditional Heteroskedasticity with Smooth Transitions in Conditional Correlations, SSE/EFI Working Paper Series in Economics and Finance No. 577.

Su, Y.K., Wu, C.C., 2014. A new range-based regime-switching dynamic conditional correlation model for minimum-variance hedging. J. Math. Financ. 4 (3), 207–219.

Tiwari, A.K., Aye, G.C., Gupta, R., Gkillas, K., 2020. Gold-oil dependence dynamics and the role of geopolitical risks: evidence from a Markov-switching time-varying copula model. Energy Econ. 88, 104748.

Trucíos, C., Hotta, L.K., 2015. Bootstrap prediction in univariate volatility models with leverage effect. Math. Comput. Simul. 120, 91–103.

Vargas, G.A., 2008. What Drives the Dynamic Conditional Correlation of Foreign Exchange and Equity Returns?, MPRA Paper No. 7174.

Vozlyublennaia, N., 2014. Investor attention, index performance, and return predictability. J. Bank. Financ. 41, 17–35.

Wang, J., Athanasopoulos, G., Hyndman, R.J., Wang, S., 2018. Crude oil price forecasting based on internet concern using an extreme learning machine. Int. J. Forecast. 34 (4), 665–677.

Xiao, J., Wang, Y., 2021. Investor attention and oil market volatility: does economic policy uncertainty matter? Energy Econ. 97, 105180.

Yang, Y., Guo, J.E., Sun, S., Li, Y., 2021. Forecasting crude oil price with a new hybrid approach and multi-source data. Eng. Appl. Artif. Intell. 101, 104217.

Yao, T., Zhang, Y.J., Ma, C.Q., 2017. How does investor attention affect international crude oil prices? Appl. Energy 205, 336–344.

Zhang, Y., Ma, F., Liao, Y., 2020. Forecasting global equity market volatilities. Int. J. Forecast. 36 (4), 1454–1475.