Energy Economics 133 (2024) 107465

Contents lists available at ScienceDirect

Energy Economics

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

Sentiment and energy price volatility: A nonlinear high frequency analysis

Fredj Jawadi[a,*], David Bourghelle[a], Philippe Rozin[a], Abdoulkarim Idi Cheffou[b], Gazi Salah Uddin[c]

[a] Univ. Lille, ULR 4999 - LUMEN, F-59000 Lille, France
[b] ISG International Business School, Paris, France
[c] Department of Management & Engineering, Linköping University, Linköping, Sweden

ARTICLE INFO

JEL:
C22
G10
G15
Q4

Keywords:
Commodity volatility
Investor sentiment
Realized volatility
Continuous volatility
Jump
Nonlinearity

ABSTRACT

This study investigates the volatility dynamics of oil and gas prices in an environment characterized by post-coronavirus disease 2019 recovery, uncertainty, high inflation, and geopolitical tensions. Unlike previous studies, we examine a long-run series of high-frequency data on gas and oil prices from July 2007 to May 2022, which provides more than one million observations with which to analyze volatility. We compute realized volatility (RV) and decompose it into continuous volatility and jumps. We then investigate the relationship between uncertainty, investor sentiment, and RV, as well as its main components. Econometrically, we extend the heterogeneous autoregressive model of Corsi (2009) while considering not only disaggregate proxies for volatility (jumps and continuous volatility) and introducing uncertainty and heterogeneous investor sentiment, but also by allowing the model to include asymmetry, nonlinearity, and time variation according to the regime under consideration. Our results present three main findings. First, we find significant evidence of volatility decomposition, suggesting that both markets are characterized by significant jumps. Second, we show that trading volume, extra-financial news (uncertainty, investor sentiment), and jumps appear to drive commodity price volatility. Third, we find evidence of nonlinearity and threshold effects on energy price volatility. These findings are relevant for policymakers, regulators, investors, and portfolio managers, as they enable them to better characterize and forecast changes in commodity prices.

1. Introduction

In theory, the prices of most goods and services depend simultaneously on demand and supply flows and vary based on the interaction between buyers and sellers. However, the prices of commodities such as oil and gas seem to be more elastic to supply than to demand, given the strong dependence of economic agents (consumers, firms, etc.) on fossil energy (Jawadi, 2019). All things being equal, regardless of the prices of these commodities, consumers do not seem to change their habits accordingly, at least in the short term. However, an increase (cut) in the oil supply might generate an oil price decrease (decrease).

From this perspective, an increasing body of literature stipulates that besides the dependence of the commodities market on market fundamentals and oil countries, changes in commodities prices can be associated with some extra-economic drivers and behavioral factors such as economic and political uncertainty (EPU), geopolitical tensions, the level of stress of the energy industry, investor emotions, and so on (e.g., Deeney et al., 2015; Maslyuk-Escobedo et al., 2017; Qadan and Nama, 2018; Yang et al., 2019). In addition to the physical market, commodities have been largely financialized over the last two decades, and their prices have become closely dependent on investors' speculation, expectations, and positions in derivative markets. That is, accounting for these extra-economic factors is particularly helpful to better explain the ups and downs that have characterized the commodities market over the last period.

Our study analyzes the excess volatility of the oil and gas markets

Abbreviations: BV, Bi-power variation; COVID-19, Coronavirus disease 2019; CV, Continuous volatility; DW, Durbin-Watson test statistics; EMU, Equity market uncertainty; EPU, Economic and political uncertainty; GARCH, Generalized autoregressive conditional heteroscedasticity; HAR, Heterogeneous autoregressive RV; MAE, Mean absolute error; MDH, Mixture distribution hypothesis; RESET, Regression equation specification error rest; RMSE, Root mean squared error; RV, Realized volatility; STAR, Smooth transition autoregressive; TAR, Threshold autoregressive; TRMI, Thomson Reuters MarketPsych Indices; TV, Trading volume; WTI, West Texas Intermediate; WTU, World trade uncertainty.

* Corresponding author at: IAE Lille University School of Management, Campus Moulins (Office T3-05), 2 rue Mulhouse - CS 10629, 59024 Lille, France.
E-mail address: fredj.jawadi@univ-lille.fr (F. Jawadi).

https://doi.org/10.1016/j.eneco.2024.107465
Received 2 January 2024; Received in revised form 1 March 2024; Accepted 8 March 2024
Available online 22 March 2024

and identifies its main drivers. In particular, regarding these extra-economic factors, we investigate the main channels associated with the relationship between commodity volatility, uncertainty, and investor sentiment. To this end, we implement two steps.

First, we rely on high-frequency data, which enables us to compute realized volatility (RV). This is particularly relevant because we use an observed proxy for volatility rather than non-observed proxies conditioned by several parametric assumptions. In addition, unlike previous studies, which were limited by the availability of high-frequency data, our study uses a large high-frequency data sample covering >15 years and >1 million observations. Further, we divide volatility into jumps and continuous volatility to better characterize the volatility drivers.

In the second step, we test whether extra-financial and behavioral factors drive the volatility of oil and gas. Specifically, we examine the impact of uncertainty and investor sentiment on the dynamics of RV. We also model the volatility reaction using two components: jumps and continuous volatility. To capture further heterogeneity in the data, we apply Corsi's (2009) heterogeneous autoregressive (HAR) model. We extend this model to set up a nonlinear and time-varying HAR specification that is robust against asymmetry, structural breaks, and nonlinearity. This extension is a new result and provides greater flexibility in capturing different forms of lead-lag effects between investor sentiment, volatility, and other drivers.

Our current study extends the literature and offers a unique contribution by reconciling two related streams of the literature on energy finance microstructure and financial econometrics. First, unlike previous studies, we rely on a long, high-frequency time series from July 2007 to May 2022 for two important commodities (oil and gas), which enables us to handle a huge amount of information (approximately one million observations). Second, following Jawadi et al. (2019), we break volatility into continuous volatility and jumps, which allows us to model changes in commodity prices with regard to the arrival of private and public information flows. Furthermore, the use of high-frequency data enables us to rely more on concise measures related to RV than on parametric proxies of volatility that are always statistically limited (absolute value of return, conditional variance, etc.). Finally, our extension of Corsi's (2009) basic HAR model allows us to incorporate extra-financial variables related to uncertainty and investor sentiment and to correct further misspecifications related to jump modeling while extending this framework to a nonlinear context. Overall, our model shows that uncertainty and investor sentiment drive commodity volatility and improve future volatility forecasts, and that their effects enter nonlinearly.

Our findings reveal three interesting results. First, we show further evidence of volatility decomposition, suggesting that both the oil and gas markets have been characterized by significant jumps over the two last two decades. Second, we find that besides the usual fundamentals (i.e., transaction volumes), extra-financial news and behavioral factors such as uncertainty and investor's sentiments are significant drivers of volatility. Third, the intensity of this leading effect is time-varying. We do not reject the assumptions of nonlinearity and time variation in volatility dynamics; this suggests the presence of different states of volatility and that the relationship between volatility and these behavioral factors works differently according to the volatility regime or state under consideration.

The remainder of this paper is organized as follows. Section 2 briefly presents the related literature and identifies possible extensions to this study. The methodology is discussed in section 3. Section 4 discusses the main results, and Section 5 concludes the paper.

2. Literature

2.1. Brief review of commodity price volatility

The common volatility of commodities prices received much attention in the recent literature, as this volatility is relevant and offers a broad perspective when analyzing the changes in commodity prices in the physical market, valuation of futures contracts and related risk, and movement of crude oil prices relative to macroeconomic variables such as the exchange rate. Interestingly, the volatility of commodities markets increased significantly in recent decades, with recent works identifying several economic and extra-economic factors as key drivers of commodity volatility. We discuss the most closely related works while distinguishing between the two groups of studies.

The first set includes studies explaining commodity price volatility through changes in fundamentals and uncertainty. Regnier (2007) analyzes the evolution of monthly producer prices for thousands of products from January 1945 to August 2005. He compares them to the prices for crude oil, refined oil, and natural gas and reports that the volatility of oil prices exceeded the median of these prices for the first time following the 1986 oil price collapse. Kilian (2008) examines large fluctuations in US energy prices since the 1970s and investigates energy price shocks and their origins, as well as the magnitude of the energy demand response to energy price changes. Kilian (2008) advocates an endogeneity-oriented approach to energy prices and differentiates the effects of demand and supply shocks on energy markets, yielding an ongoing challenge between his work and the related studies by James Hamilton (Hamilton, 1983; Hamilton, 2003; Hamilton, 2009, etc.).

Recently, using different variants of the Generalized autoregressive conditional heteroscedasticity (GARCH) model, Wang et al. (2022a, 2022b) study the volatility dynamics of oil, natural gas, and natural resources. Their main empirical results indicate high asymmetrical volatility. Han et al. (2020) investigate the volatility link between energy and agricultural futures returns and their reactions to external macroeconomic shocks. They show a bidirectional link between energy and agricultural futures returns, which has become more pronounced recently. This link results from the co-movement effect induced by external shocks, which appears to provide effective explanatory power for the transitory volatility link. Assaf et al. (2021) study whether EPU, geopolitical risk, world trade uncertainty (WTU), and equity market uncertainty (EMU) impact the dynamics of oil, gas, and coal market returns. The authors report that WTU contributes the most to energy markets and that the oil market contributes the most to the other markets. Furthermore, they find that the total connectedness index is relatively high, coinciding with the 2008 global financial crisis and coronavirus disease 2019 (COVID-19) pandemic. Wang et al. (2022a, 2022b) examine the ability of five uncertainty indices and seven global economic conditions to predict the RV of clean energy and natural gas stock markets. They show that, when using clean energy or natural gas ETFs, the predictive information extracted from global economic conditions outperformed uncertainty indices. This suggests the need to account for real economic activities as they are more relevant for investors and policymakers than textual measures of uncertainty when analyzing clean energy and natural gas volatility.

The second set includes studies that explain commodity price volatility using extra-financial factors. For example, the oil market has distinct characteristics relative to other markets. Indeed, Cifarelli and Paladino (2010), Kaufmann (2011), and Coleman (2012) find that high speculation can explain significant changes in oil prices. This speculation is always driven by uncertainty, and various investor sentiments are marginal factors in commodity price adjustments. In general, market sentiment can take many forms: the effects of a runaway market, feelings of fear, increased stress factors, and flowing effects. Dunham and Garcia (2021) find that such global distress factors can generate individual or collective sentiments and yield widespread panic, irrational decisions, misbehaving, and mispricing. Thus, this sentiment can influence investors' judgment and impact investment decisions and prices (Gandhi et al., 2019). For example, oil prices seem to be affected by common representations of the future in an uncertain environment based on sentiment and emotion. From this perspective, as in Baker and Wurgler (2006), investors' emotions might drive them to search for opportunities that match their sentiments, which could yield negative or positive signals (optimism vs. pessimism). This would then push them to trade differently in response to this signal, which directly or indirectly impacts commodity prices (Borovkova, 2011; Deeney et al., 2015).[1]

In this context, Qadan and Nama (2018) test the impact of behavioral factors on oil price returns and volatility and find that the investor sentiment index drives changes in oil returns and volatility and that the information provided by investor sentiment can help predict oil prices. Wang et al. (2019) explain the relationship between sentiment and price formation by the fact that a rise in rumors can, for example, affect the common sentiment of investors and facilitate the effects of frenzy, hyperfocus, and ultimately investors' asset allocation choices, which can yield significant changes in prices. Using commodity-specific news sentiment data on OPEC collected by Thomson Reuters News Analytics (period 2003–2014), Gupta and Banerjee (2019) analyze the impact of foreign news sentiment on the financial performance of firms in the energy sector during periods of low or high confidence. They find that, in the US market, negative news about OPEC positively influences the stock returns of US listed firms, and that the effects of negative OPEC news dominate their results. Song et al. (2019) examine the dynamic diffusion of performance and volatility information between the fossil energy market, investor sentiment toward renewable energy, and the renewable energy market using a connectedness network approach. They demonstrate that crude oil has a greater impact on the renewable energy stock market than does investor sentiment. Li et al. (2021) analyze news about crude oil and developed a model with an investor sentiment indicator for crude oil. Their results indicate that a news sentiment shock causes significant price fluctuations in oil futures and has an asymmetric impact on the volatility of oil futures. Accordingly, incorporating a sentiment score into a deep learning model based on data decomposition demonstrates its effectiveness. Similarly, Chen et al. (2021) test whether investor sentiment has a stronger predictive power than the VIX and uncertainty indices in forecasting the RV of energy assets. They find that investor sentiment has a significant positive impact only on West Texas Intermediate (WTI) oil futures and spot contracts. Further, investor sentiment is better at forecasting oil prices, followed by the VIX, but is less significantly predictive of natural gas futures and spot contracts.

2.2. Jumps, nonlinearity, and structural breaks in commodity price volatility

Several recent studies provide evidence of nonlinearity, structural breaks, and jumps in commodity price volatility dynamics; however, their findings are inconclusive. Lee et al. (2010) focus on structural breaks in crude oil prices. They confirm the existence of permanent and transitory components in the conditional variance, with the former increasing with the occurrence of a sudden major event, and the intensity of jumps fluctuating as the latter increases in response to abnormal events. Using high-frequency data on four commodity futures markets (crude oil, heating oil, natural gas, and gasoline), Prokopczuk et al. (2016) test jumps and analyze their impact on futures volatility. They demonstrate that jumps are rare events, with intensities that vary considerably over time. In their analysis of the importance of jumps in forecasting RV over horizons ranging from 1 to 22 days, they find little difference between their benchmark model and alternative models, suggesting that explicit jump modeling does not significantly improve the accuracy of energy market volatility forecasts.

Wilmot and Mason (2013) investigate the potential presence of jumps and variable volatility in the spot prices of crude oil and futures prices related to the arrival of news. Unlike Lee et al. (2010), who use likelihood ratio tests to compare four stochastic data-generating processes, Wilmot and Mason (2013) find that allowing for both jumps and time-varying volatility improves the model's ability to explain spot and futures prices. Chan and Gray (2017) also highlight the contribution of jumps for six commodities, but find less evidence of a link between jumps and the scheduled release of economic data.[2] Recently, Jawadi et al. (2019) forecast the future dynamics of the oil and gas futures markets using high-frequency data. They decompose volatility into continuous volatility and jumps, and report evidence of nonlinearity in volatility dynamics. However, their study did not include sentiment or uncertainty factors, and relied on a limited high-frequency data sample.

In summary, we note that studies of commodity price volatility remain inconclusive, though some recent studies indicate that both fundamentals and several behavioral factors appear to play a key role in changes in commodity prices. This lack of unanimity can be explained by the fact that volatility is not observed and is measured using relative proxies (standard deviation, absolute return squared return, conditional variance, etc). Further, while the identification of jumps in volatility dynamics is not rejected, the conclusions remain inconclusive, possibly due to jump modeling misspecifications. Our current study aims to extend these related works and fill this research gap.

3. Econometric methodology

In this section, we briefly review the tests we use to divide volatility into jumps and continuous volatility. Second, we discuss the specifications used to model the volatility-sentiment relationship.[3]

3.1. Volatility decomposition tests

First, we consider the logarithmic price process in continuous time using the following jump-diffusion model:

$$dP(t) = \mu(t)dt + \sigma(t)dW(t) + k(t)dq(t) \tag{1}$$

where P(t) is the logarithmic asset price at time t. The variable $\mu(t)$ denotes a continuous variation process. $\sigma(t)$ is stochastic volatility process with $\sigma(t) > 0$. W(t) denotes the standard Brownian equation. The variable q(t) defines a jump process characterized by a jump intensity of $\lambda(t)$ and a jump size of $\kappa(t)$.

As Jawadi et al. (2015) note, prior studies introduce different proxies for volatility (standard deviation, absolute return, squared return, proxies related to ARCH/GARCH models, proxies related to the stochastic volatility model, etc.). However, none of these are observable; they are always conditioned by parametric and restrictive assumptions related to the error distribution, among others. None of these proxies are model-free.

Andersen and Bollerslev (1998) propose an alternative measure of volatility by introducing RV, defined as

$$RV_{t+1}(\Delta) \equiv \sum_{j=1}^{1/\Delta} r^2_{t+j\Delta,\Delta} \tag{2}$$

where $r_{t,\Delta} \equiv p(t) - p(t-\Delta)$ is the discretely sampled Δ-period return, and $1/\Delta$ is the number of intraday periods, always corresponding to 5-min intervals. The main advantage of RV, defined as the sum of intraday squared returns (Eq. (2)), is that it provides an observed measure of volatility that requires no additional assumptions.

Andersen and Bollerslev (1998) break RV down into the sum of continuous volatility (δ) and jumps (k):

[1] For a nice survey on the effect of sentiment on stock markets, see Barberis et al., (1998) and Schmeling (2009), among others.

[2] Most related literature on jumps focuses on the structure of prices in the electricity market (Gudkov and Ignatieva, 2021; Hellström et al., 2012; Kostrzewski and Kostrzewska, 2021).

[3] The presentation of jump tests in this section relies on Jawadi et al. (2020) and Jawadi et al. (2015). These studies provide more detail for interested readers.

$$RV_{t+1}(\Delta)\rightarrow\int_{t-1}^{t}\delta^{2}(s)ds+\sum_{j=1}^{N_t}k_{t,j}^{2}\tag{3}$$

where N_t denotes the number of jumps occurring per day t and $k_{t,j}$is the j[th] jump size on day t.

Accordingly, Barndorff-Nielsen and Shephard (2004) suggest applying the bi-power variation (BV), as in Eq. (4), to disentangle the jump and continuous components of RV.

$$BV_{t+1}(\Delta)\equiv\mu_1^{-2}\sum_{j=2}^{1/\Delta}|r_{t+j\Delta,\Delta}|\left|r_{t+(j-1)_{\Delta,\Delta}}\right|,\text{with }\mu_1$$
$$=\sqrt{2/\pi},BV_{t+1}(\Delta)\rightarrow\int_{t-1}^{t}\delta^{2}(s)ds\tag{4}$$

Accordingly, we compute a jump as

$$Jump_{t+1}(\Delta)=RVT_{t+1}(\Delta)-BV_{t+1}(\Delta)\tag{5}$$

However, in practice, significant jumps must be distinguished from small jumps (Andersen et al., 2007a, 2007b; Jawadi et al., 2020). To this end, we employ the following test to identify significant jumps:

$$Z_{t+1}(\Delta)\equiv\frac{[RV_t(\Delta)-BV_t(\Delta)]RV_t(\Delta)^{-1}}{\left[(\mu_1^{-4}+2\mu_1^{-2}-5)max\left\{1,TQ_t(\Delta)BV_t(\Delta)^{-2}\right\}\right]^{1/2}}\tag{6}$$

where $TQ_{t+1}(\Delta)\equiv\Delta^{-1}\mu_{\frac{4}{3}}^{-3}\sum_{j=3}^{1/\Delta}|r_{t+j\Delta,\Delta}|^{4/3}|r_{t+(j-1)\Delta,\Delta}|^{4/3}|r_{t+(j-1)\Delta,\Delta}|^{4/3}$ *and* $Z_{t+1}(\Delta)$ is normally distributed.[4]

3.2. Modeling the sentiment-volatility relationship with HAR models

3.2.1. Linear specification

Financial and energy data are characterized by excess volatility and frequent and continuous price changes. This volatility excess, which is more remarkable when considering high-frequency data, is always related to the extensive presence of algorithms and automatic trading; significant flows of sales and purchases over short time intervals; reactions to market news, uncertainty, investors; and the heterogeneous behavior of traders. To adequately model the volatility of our two commodities (oil and gas), we propose the use of RV as a concise proxy for volatility. Further, our benchmark modeling is based on Corsi (2009) heterogeneous autoregressive RV (HAR-RV) model, which deals with the above characteristics by employing heterogeneous time frequencies. The underlying idea of this model is to regress RV on daily $\left(RV_t^{(d)}\right)$, weekly $\left(RV_t^{(w)}\right)$, and monthly $\left(RV_t^{(m)}\right)$ data. Formally, we set up the linear HAR-RV as

$$RV^{(d)}{}_{t+1}=\alpha_0+\alpha_1RV_t^{(d)}+\alpha_2RV_t^{(w)}+\alpha_3RV_t^{(m)}+\varepsilon_{t+1}\tag{7}$$

where $RV^{(d)}{}_{t+1}$ is the ex-post volatility estimate and ε_{t+1} denotes the error term.

Empirically, the HAR-RV can be specified as

$$RV_t=\alpha_0+\alpha_1RV_{t-1}+\alpha_2RV_{t-5}+\alpha_3RV_{t-20}+\varepsilon_t\tag{8}$$

where ε_t is an error term.

In line with the MDH theory, we augment Eq. (8) by adding trading volume (TV) as a control variable in addition to the explanatory variables. Further, we test behavioral factors (uncertainty, VIX, and sentiment) and specify the following HAR-RVX model:

$$\begin{aligned}RV_t=&\alpha_0+\alpha_1RV_{t-1}+\alpha_2RV_{t-5}+\alpha_3RV_{t-20}+\beta_1TV_t\\&+\beta_2EPU_t+\beta_3VIX_t+\beta_iX_{it}+\varepsilon_t\end{aligned}\tag{9}$$

where TV_t denotes trading volume, EPU_t denotes EPU, and VIX_t denotes the VIX index. $X_{it}\forall i=1,\dots,5$ denotes a proxy for investor sentiment.[5]

Next, to better characterize the dynamics of commodity volatility, we test whether its components (continuous volatility and jumps) drive and explain the dynamics of RV. Thus, we extend Eq. (9) and specify the HAR-CV-JX model as

$$\begin{aligned}RV_t=&\alpha_0+\alpha_1CV_{t-1}+\alpha_2CV_{t-5}+\alpha_3CV_{t-20}+\alpha'_1J_{t-1}+\alpha'_2J_{t-5}\\&+\alpha'_3J_{t-20}+\beta_1TV_t+\beta_2EPU_t+\beta_3VIX_t+\beta_iX_{it}+\varepsilon_t\end{aligned}\tag{10}$$

where CV_t denotes the continuous volatility. J_t denotes Jumps.[6]

3.2.2. Nonlinear specification

While the above specification enables us to model the dynamics of commodity volatility by considering fundamental variables (TV), volatility components, and behavioral factors, model (10) does not reproduce nonlinearity or further asymmetry in volatility dynamics. We thus extend this model to a nonlinear framework while allowing the explanatory variables to enter nonlinearly. Accordingly, we propose an on/off threshold HAR-RV model with flexible specifications.

From a theoretical perspective, this threshold specification is useful because it has the advantage of reproducing nonlinearity and multiple regimes characterizing volatility dynamics. Indeed, news does not arrive at the market either proportionally or symmetrically. Investors' reactions, and therefore the dynamics of price changes, always differ with the sign and size of news (positive vs. negative), yielding different levels and states of volatility (low, stable, or high).

Accordingly, we propose extending model (10) while allowing for multiple regime states by allowing all regressors to enter nonlinearly with the volatility regime. In line with Tsay (1989) and Hansen (1996), we set up a multiple-regime HAR-CV-JX model for which these regimes are identified, and switching between these regimes occurs when a transition variable exceeds a given threshold. The presence of different regimes allows us to capture different states of volatility such as low, normal, and high volatility, and so on. In practice, as in Tsay (1989) and Hansen (1996), this threshold is endogenously estimated, as is the number of regimes. We then estimate the threshold model following the estimation procedure developed by Tsay (1989) and Hansen (1996). We test the linearity hypothesis and number of regimes empirically through a series of threshold and linearity tests. The three-regime HAR-CV-JX model is as follows:

$$\begin{aligned}RV_T=&\alpha_{01}+\alpha_{11}CV_{t-1}+\alpha_{21}CV_{t-5}+\alpha_{31}CV_{t-20}+\alpha'_{11}J_{t-1}+\alpha'_{21}J_{t-5}\\&+\alpha_{31}CV_{t-20}+\beta_{11}TV_t+\beta_{21}EPU_t+\beta_{31}VIX_t+\beta_{i1}X_{it}+\varepsilon_t\ if\ x_t<c_1\end{aligned}\tag{11}$$

$$\begin{aligned}RV_t=&\alpha_{02}+\alpha_{12}CV_{t-1}+\alpha_{22}CV_{t-5}+\alpha_{32}CV_{t-20}+\alpha'_{12}J_{t-1}\\&+\alpha'_{22}J_{t-5}+\alpha'_{32}J_{t-20}+\beta_{12}TV_t+\beta_{22}EPU_t+\beta_{32}VIX_t+\beta_{i2}X_{it}\\&+\varepsilon_t\ if\ c_1\le x_t<c_2\end{aligned}$$

$$\begin{aligned}RV_t=&\alpha_{03}+\alpha_{13}CV_{t-1}+\alpha_{23}CV_{t-5}+\alpha_{33}CV_{t-20}+\alpha'_{13}J_{t-1}\\&+\alpha'_{23}J_{t-5}+\alpha'_{33}J_{t-20}+\beta_{13}TV_t+\beta_{23}EPU_t+\beta_{33}VIX_t+\beta_{i3}X_{it}+\varepsilon_t\ if\ x_t\ge c_2\end{aligned}$$

where c_1 and c_2 are the thresholds separating the regimes for which the transition is expected to be abrupt.

However, a generalization of this threshold HAR-CV-JX model *à la* Teräsvirta (1994) would allow for a smooth, rather abrupt, transition between volatility regimes, yielding a smooth transition HAR-CV-JX model. For example, the two-regime smooth-transition HAR-CV-JX model corresponds to

[4] See Jawadi et al. (2020) for more details about this test.

[5] We propose two proxies for sentiment to capture different aspects of investor sentiment that we describe in the data section.

[6] This extension of the basic HAR model by considering volatility components is in line with Jawadi et al. (2020).

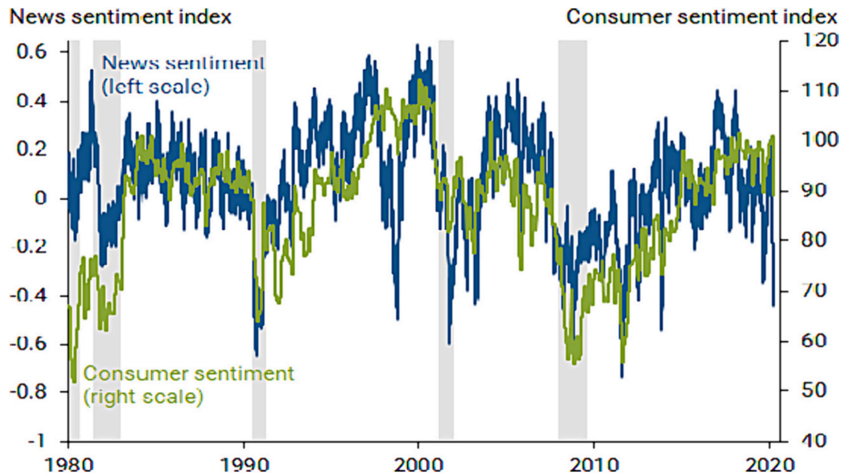


Fig. 1. Daily News Sentiment versus Monthly Consumer Sentiment.
Source: Daily News Sentiment Index and Michigan survey.

$$\begin{aligned} RV_T &= \alpha_{01} + \alpha_{11}CV_{t-1} + \alpha_{21}CV_{t-5} + \alpha_{31}CV_{t-20} + \alpha'_{11}J_{t-1} + \alpha'_{21}J_{t-5} \\ &+ \alpha_{31}CV_{t-20} + \beta_{11}TV_t + \beta_{21}EPU_t + \beta_{31}VIX_t + \beta_{i1}X_{it} + (\alpha_{02} + \alpha_{12}CV_{t-1} \\ &+ \alpha_{22}CV_{t-5} + \alpha_{32}CV_{t-20} + \alpha'_{12}J_{t-1} + \alpha'_{22}J_{t-5} + \alpha'_{32}J_{t-20} + \beta_{12}TV_t \\ &+ \beta_{22}EPU_t + \beta_{32}VIX_t + \beta_{i2}X_{it})^*F(c,\gamma) + \varepsilon_t \end{aligned} \tag{12}$$

where F is a transition function bounded between zero and one and depends on two parameters: the threshold parameter c and transition speed γ.

According to Terasvirta and Anderson (1992), Granger and Teräsvirta (1993), and Teräsvirta (1994), transition function F(.) can be either a logistic or exponential function. Here, Model (12) represents a logistic smooth transition HAR-CV-JX model or can become an exponential smooth transition HAR-CV-JX model. F(.) is bounded between zero and one. When $F \rightarrow 0$, Eq. (12) becomes a HAR-CV-JX model as in Eq. (10); when $F \rightarrow 1$, Eq. (12) becomes a threshold HAR-CV-JX model as in Eq. (11). Therefore, Model (12) has the advantage of capturing the linear dynamics of the energy price volatility if linearity is not rejected; it captures the dynamics of energy price for volatility in extreme volatility states when the transition is abrupt, as well as a continuum of intermediate volatility states and regimes. The specification of regimes and the choice of transition function are also specified through linearity and threshold tests (Teräsvirta, 1994), while the estimation of the smooth transition HAR-CV-JX model is carried out using the nonlinear least-squares method (Jawadi and Koubbaa, 2007; Jawadi, 2006; Teräsvirta, 1994).

Finally, we compare the five models (HAR-RV, HAR-RVX, HAR-CV-JX, threshold HAR-CV-JX, and smooth transition HAR-CV-JX) in terms of their forecasting performance for oil and gas volatility.

4. Empirical analysis

4.1. Data

The main objective of this study is to explain and forecast the dynamics of volatility of two main commodities (oil and gas). Oil and gas prices are proxied by the West Texas Intermediate (WTI) price and the Henry Hub Natural gas price, respectively. We measure the volatility of these two commodities using Andersen and Bollerslev (1998) RV model. To this end, we rely on high-frequency data for oil and gas, which consist of individual contracts available at the 5-min timeframe, enabling us to obtain continuous futures series. We also use high-frequency trading volumes in which the volume represents individual contracts. We obtain these data from the FirstRate Database for July 3, 2007 to May 27, 2022, which provides approximately 1,094,460 observations. To our knowledge, this is the first study in this area to use this important information to investigate the dynamics of volatility in oil and gas commodities. We use historical intraday market price data to compute the intraday and daily RV series (Eq. (2)). We also decompose the RV series into continuous volatility (CV) and jumps (discontinuous volatility) using Models (3), (4), and (5) to better investigate commodity price volatility dynamics.

Otherwise, as we test the contribution of behavioral factors (uncertainty, sentiment, and anxiety) to explain the dynamics of commodity price volatility, we use different behavioral time series. For example, we use the US EPU series, which is available on a daily basis, to account for further uncertainty caused by economic shocks, political shocks, and further geopolitical tensions (war in Ukraine, coordination default across OPEC countries, US-China trade war and tensions, COVID-19 outbreak, etc.). We also use the VIX, which is available on a daily basis, to capture further evidence of investor anxiety.

In addition, we consider investor and market sentiment as commodity price volatility drivers. However, sentiment is always unobserved and has been approximated differently in the related literature. Further, existing survey-based sentiment indexes are either low-frequency and therefore less useful in times of unexpected change or are available only for a short period. Rather than randomly selecting a proxy for sentiment, we propose using two proxies to capture the different aspects of investor and market sentiment. First, we use the proxy *News_Senti1*, which denotes the Daily News Sentiment Index developed by Shapiro et al. (2020). *News_Senti1* enables us to capture emotions and animal spirits related to investor confidence, as this daily sentiment index is historically highly correlated with the monthly survey-based University of Michigan Index of Consumer Sentiment and the Conference Board's Consumer Confidence Index (see Fig. 1 for an overview of this sentiment index back to 1980). Shapiro et al. (2020) use a lexical approach that relies on a predefined list of words associated with sentiments, referred to as lexicons, in three categories: negative, neutral, or positive. They construct sentiment scores for economics-related news articles using a lexical approach based on a historical archive of news articles from 16 major US newspapers compiled by the news aggregator service LexisNexis. This database covers all major regions of the country, including some with extensive national coverage, such as the *New York Times* and the *Washington Post*.[7] For this sentiment index, higher values indicate a more positive sentiment, and from 2007 onwards, we observe several uptrends and downtrends in the sentiment index (see Fig. 1).

We also use a second proxy, *ThR_sent2*, which corresponds to the

[7] For more details about the methodology to construct this sentiment index, see Shapiro et al. (2020).

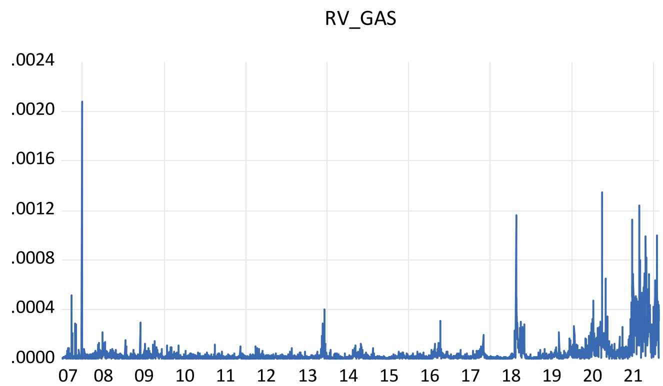


Fig. 2. Dynamics of RV for Oil.
Note: RV_WTI denotes the RV for the oil market.
We consider the negative oil price in April 2020 as an outlier and omitted it. We sincerely thank Christiane Baumeister, Tim Bollerslev, James Hamilton, and Timo Teräsvirta for their suggestions and interesting exchanges about the treatment of this negative price.

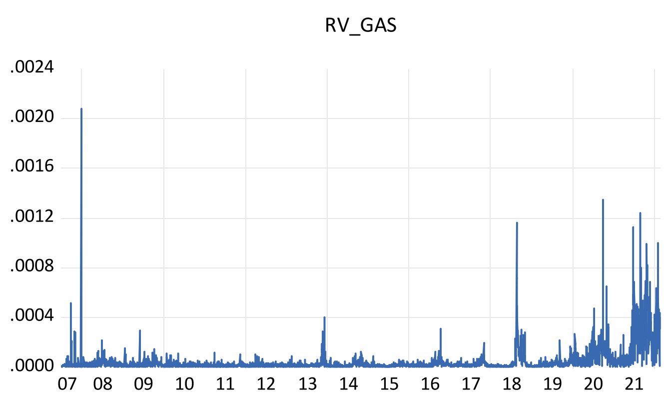


Fig. 3. Dynamics of RV for Gas.
Note: RV_Gas denotes the RV for the gas market.

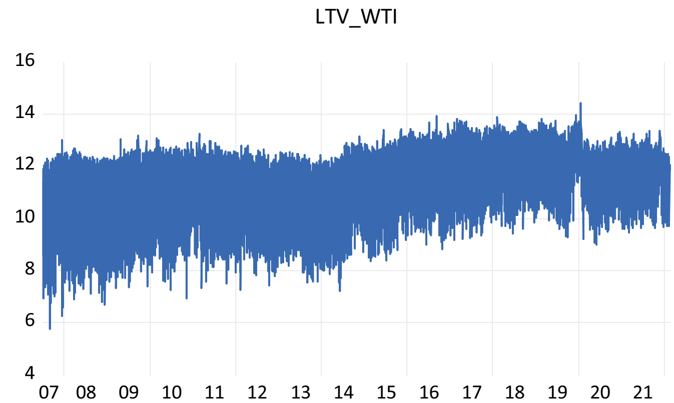


Fig. 4. Dynamics of Trading Volume for Oil.
Note: LTV_WTI denotes the oil market trading volume in logarithm.

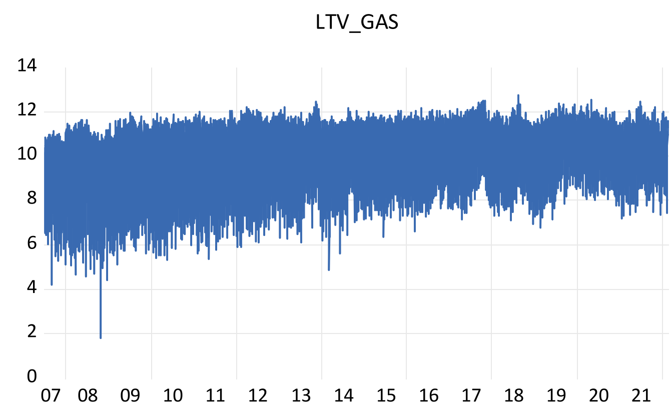


Fig. 5. Dynamics of Trading Volume for Gas.
Note: LTV_Gas denotes the gas market trading volume in logarithm.

Thomson Reuters MarketPsych Indices (TRMI) sentiment index.[8] The TRMI were constructed using machine learning and text analysis to measure key themes and sentiments that can influence price trends. This index uses several emotional and topical items from news and social media sites. Overall, we capture different dimensions of investor sentiment and emotion using these two measures.

4.2. Preliminary analysis

First, to provide an overview of the data, we report the dynamics of RV for oil and gas prices in Figs. 2 and 3, respectively, which we compute using Eq. (1). The use of RV has the advantage of dealing with an observed estimation of volatility rather than unobserved proxies for volatility, such as GARCH models and stochastic volatility models. From Figs. 2 and 3, we note that the dynamics of oil and gas volatility are characterized by several episodes of up- and down-volatility, suggesting further evidence of excess volatility for these two commodities. Indeed, the oil price volatility dynamics show some signs of volatility in 2008 (around the global financial crisis), 2014–2015 (shale revolution), and the highest increase in 2020 caused by COVID-19.[9] Volatility excess and time variation are more pronounced for gas than for oil because the gas market also shows different and important regimes of volatility. Interestingly, for both commodities, volatility reached its highest levels during crisis periods (the global financial crisis in 2008–2009, shale revolution in 2014–2015, COVID-19 pandemic in 2019–2020, and inflation episodes since 2022). In fact, for commodities, besides the physical market in which commodity prices are driven by flows of demand and supply, most commodities experienced large and impressive financialization. Accordingly, commodities are used as financial instruments that impact their exchange, and therefore, their prices. This generates more diverse transactions and market players, and thus more price volatility through commodity financialization and dependence on market news (Sabrine and Ont Aviano, 2023).

To track trading activity for the oil and gas markets, we report the trading volume series in volume in Figs. 4 and 5. Trading volumes

[8] These data are not public, and we thank and acknowledge the assistance of Thomson Reuters MarketPsych in providing access to these data for using them exclusively for this study.

[9] Indeed, the crude oil price briefly traded below 0 USD in spring 2020 because of the decline in global oil demand. In particular, on April 20, 2020, the WTI traded at a negative price because oil demand fell, and US oil inventories increased simultaneously. This is the first time the WTI futures contract fell to a negative value since trading began in 1983. On April 21, the Brent oil price, the benchmark oil price for Europe, fell to 9 USD, which is its lowest price in decades. See Appendix A for more details.

Table 1
Main Descriptive Statistics and Normality Tests.

	RV_WTI	RV_GAS
Mean	0.00013	4.70E-05
Maximum	0.0240	0.002075
Minimum	3.00E-07	3.81E-07
Std. Dev.	0.00072	9.52E-05
Skewness	24.535	6.768
Kurtosis	714.015	82.651
Jarque-Bera (p-value)	0.00	0.00
Observations	4584	4584

Note: RV_WTI and RV_GAS denote RV for oil and gas, respectively.

Table 2
Unconditional Correlation Matrix.

	RV_WTI	RV_GAS	NEWS_SENTI1	THR_SENTI2	LVIX	LUSEPU	LTV_WTI	LTV_GAS
RV_WTI	1	0.099	0.005	−0.003	0.055	0.037	0.026	−0.002
RV_GAS		1	0.033	0.042	0.093	0.093	0.194	−0.040
NEWS_SENTI1			1	−0.519	0.723	−0.131	0.143	0.023
THR_SENTI2				1	−0.399	0.005	−0.115	−0.021
LVIX					1	−0.138	0.069	0.061
LUSEPU						1	0.116	0.010
LTV_WTI							1	−0.152
LTV_GAS								1

Note: *RV_WTI* and *RV_GAS* denote RV for oil and gas, respectively. *News_Sent1* and *Thr_Sent2* denote two sentiment measures. *LVIX* and *LUSEPU* denote the logarithms of the VIX and US EPU, respectively. *LTV_WTI* and *LTV_GAS* denote the trading volumes in logarithms of oil and gas, respectively.

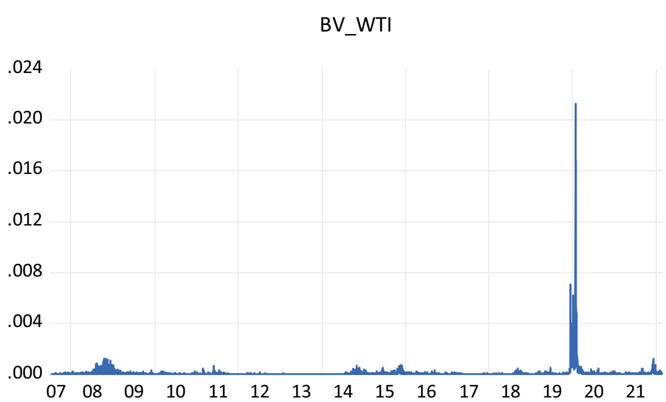


Fig. 6. Dynamics of Continuous Volatility for Oil.
Note: BV_WTI denotes the continuous volatility for Oil.

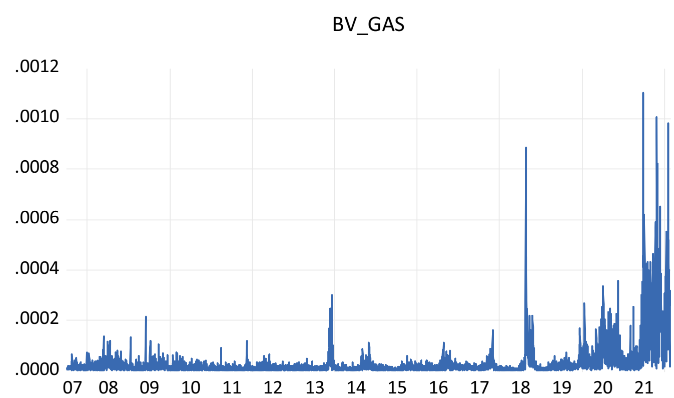


Fig. 7. Dynamics of Continuous Volatility for Gas.
Note: BV_WTI denotes the continuous volatility for gas.

appear less stable for oil than for gas; they evolve over time, but changes in volume do not appear sufficient to justify important changes in energy price volatility.

Second, we check for the presence of a unit root in the data using the augmented Dickey and Fuller (1979, 1981) and Phillips and Perron (1988) tests. Our findings, which we omit here to save space, show that all our time series are stationary.[10]

Table 1 reports the main descriptive statistics for both energy volatilities. Table 1 shows a significant volatility excess that is more pronounced, on average, in the oil market. Furthermore, in addition to leptokurtic excess, the symmetry and normality hypotheses are strongly rejected for oil and gas. Excess kurtosis, particularly for the oil market, suggests the presence of several outliers in the distribution.

To provide an overview of the linkages between commodity price volatility, uncertainty, and sentiment, we compute an unconditional correlation matrix. We report these correlations in Table 2. First, volume and volatility are positively correlated for oil but negatively and weakly correlated for gas. This result is consistent with the mixture distribution hypothesis (MDH) particularly for the oil market. Second, we find that EPU show a positive correlation for both commodities, suggesting that high uncertainty yields greater energy price volatility. Third, the VIX is positively correlated with energy price volatility, suggesting that an increase in anxiety and fear can be a source of volatility. Finally, when considering sentiment for either the proxy *News_sent1* or *Thr_Sent2*, we find that the correlation with volatility alternates between positive and negative, suggesting that sentiment might help explain changes in volatility. This result is relevant and implies that when investors "dream" or behave with some insurance, confidence and follow their animal spirits, their action will create more volatility. This finding is in line with the theories of animal spirits (Akerloff and Shiller, 2009) and irrational exuberance (Shiller, 2015). The linkages between sentiment and energy price volatility are stronger and more pronounced when gas, rather than oil, is considered. However, these correlations are relatively weak and the framework should be extended to capture more complex and nonlinear linkages. That is, while these results provide further evidence of the linkages between sentiment, uncertainty, and volatility, we propose to improve the analysis by considering at the same time different components of volatility and other econometric tests.

4.3. Volatility decomposition

We break RV down into continuous volatility and jumps using Barndorff-Nielsen and Shephard, 2006 approach. We report in Figs. 6–7 the continuous volatility and in Figs. 8–9 the jumps for oil and gas commodities. While we find that the continuous volatility component is more important and more frequently observed than the jump component for both energy price volatilities, which is in line with previous studies

[10] These results are available upon request. We also transformed the VIX and EPU series to logarithms to stationarize their variances using the box-cox transformation.

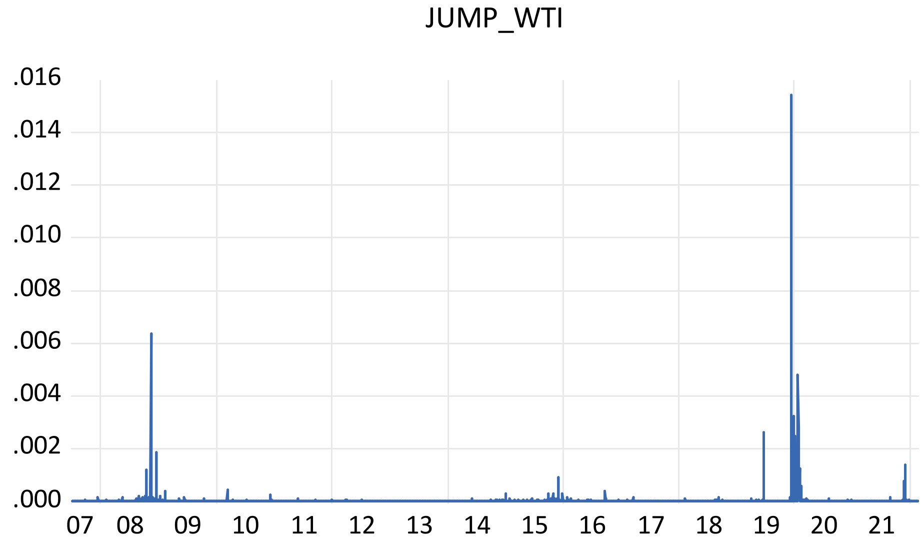


Fig. 8. Dynamics of Jumps for Oil.
Note: Jump_WTI denotes the jumps for oil.

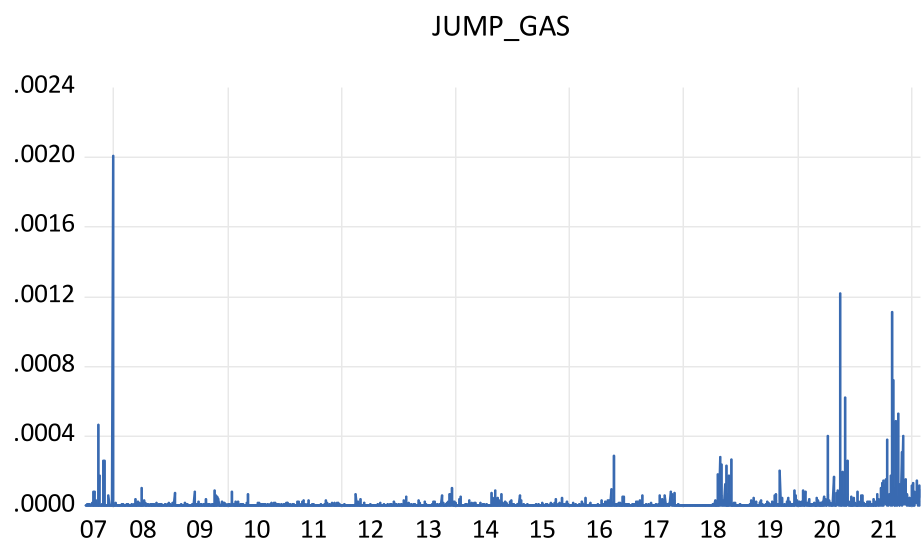


Fig. 9. Dynamics of Jumps for Gas.
Note: Jump_Gas denotes the jumps for gas.

Table 3
Main Descriptive Statistics for Continuous Volatility.

	BV_WTI	BV_GAS
Mean	0.0001	3.85E-05
Maximum	0.0212	0.0011
Minimum	1.68E-07	1.83E-07
Std. Dev.	0.0005	7.55E-05
Skewness	27.671	5.4026
Kurtosis	1003.1	44.652
Jarque-Bera (p-value)	0.00	0.00
Observations	4584	4584

Note: *BV_WTI* and *BV_GAS* denote the continuous volatility for oil and gas respectively.
Std-Dev refers to the standard deviation.

Table 4
Main Descriptive Statistics for Jumps.

	JUMP_WTI	JUMP_GAS
Mean	2.30E-05	8.93E-06
Maximum	0.0154	0.002003
Minimum	2.53E-10	1.35E-09
Std. Dev.	0.0002	4.77E-05
Skewness	42.12	24.88
Kurtosis	2155.8	845.3
Jarque-Bera (*p*-value)	0.00	0.00
Observations	4584	4584

Note: *JUMP_WTI* and *JUMP_GAS* denote the jump components of RV for oil and gas, respectively. Std-Dev refers to the standard deviation.

(Jawadi et al., 2015), we show that both commodities are characterized by the presence of significant jumps. Furthermore, jumps are more pronounced during crises and turbulent periods.

To better characterize the components of RV, we present in Tables 3 and 4, the main descriptive statistics for continuous volatility and jumps for oil and gas respectively. First, jumps in both markets are less frequent than continuous volatility. Second, we find that changes in energy prices are, on average, more pronounced for oil than for gas when considering continuous volatility or jumps. Finally, both symmetry and normality

Table 5
Basic HAR-RV Model Estimation Results.

Coefficients	Oil market	Gas Market
α_1	0.214***	0.442***
	(0.00)	(0.00)
α_2	0.506***	0.252***
	(0.01)	(0.00)
α_3	0.029	0.146***
	(0.36)	(0.00)
Adjusted R^2	0.370	0.504
DW	2.02	2.29

Note: DW denotes the statistics of the Durbin-Watson test. (***), (**), and (*) denote statistical significance at the 1%, 5%, and 10% levels, respectively. Values in (.) denote the probabilities of the t-ratios for which standard errors and covariance are corrected (Bartlett kernel and Newey-West fixed bandwidth = 10.0).

are rejected for the two components of energy price volatility. Interestingly, when considering oil, we find that jumps represent around 17.6%, while continuous volatility approaches 82.4% of the total volatility, while for gas, the quote part of jumps is around 19% of the total RV, and continuous volatility accounts for 81% of the total RV. These findings were consistent with those of previous studies (Jawadi et al., 2015; Jawadi et al., 2020). However, they are relevant as they confirm the interest in breaking down volatility into continuous volatility and jumps.

Overall, our preliminary results show a significant decomposition of RV into continuous and jump components. This finding suggests that jumps play a crucial role in explaining the energy price volatility dynamics. Further, our results point to the presence of further evidence of the dependence of energy volatility on extra-financial variables such as uncertainty and sentiment. We test hereafter these relationships linear and nonlinear frameworks.

4.4. Linear HAR model estimation results

To investigate the dynamics of energy price volatility, we first consider the basic HAR-RV (Eq. (7)), which enables us to use different frequencies (daily, weekly, and monthly) to model the dynamics of energy price commodities for oil and gas, in line with Corsi (2009). We have run different specifications and improved the model estimation to provide consistent estimators for which the estimated errors present the main "good" statistical properties. We report the main results of the

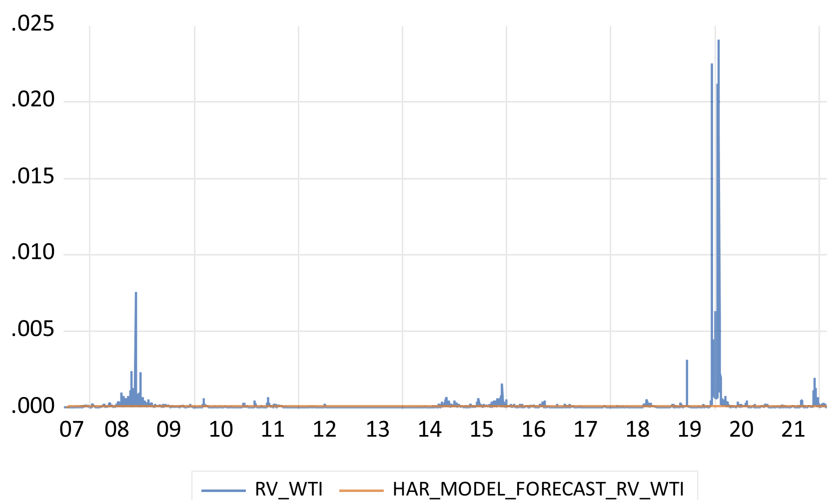


Fig. 10. Modeling Dynamics of Oil RV with the HAR_RV Model.
Note: RV_WTI denotes the observed oil RV, whereas HAR-model_forecast_RV_WTI denotes the in-sample forecast of the HAR_RV model (Eq. (8)) for oil RV.

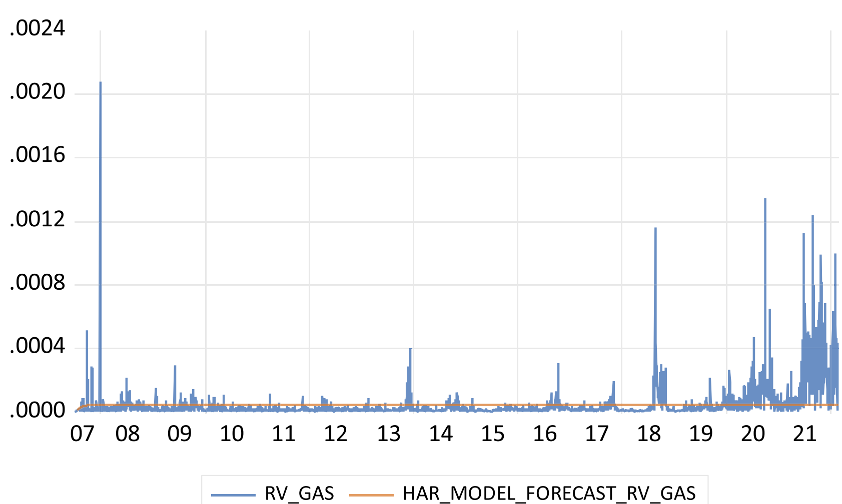


Fig. 11. Modeling Dynamics of Gas RV with the HAR_RV Model.
Note: RV_GAZ denotes the observed gas RV, whereas HAR-Model_FORECAST_RV_GAS denotes the in-sample forecast of the HAR_RV model (Eq. (8)) for the gas RV.

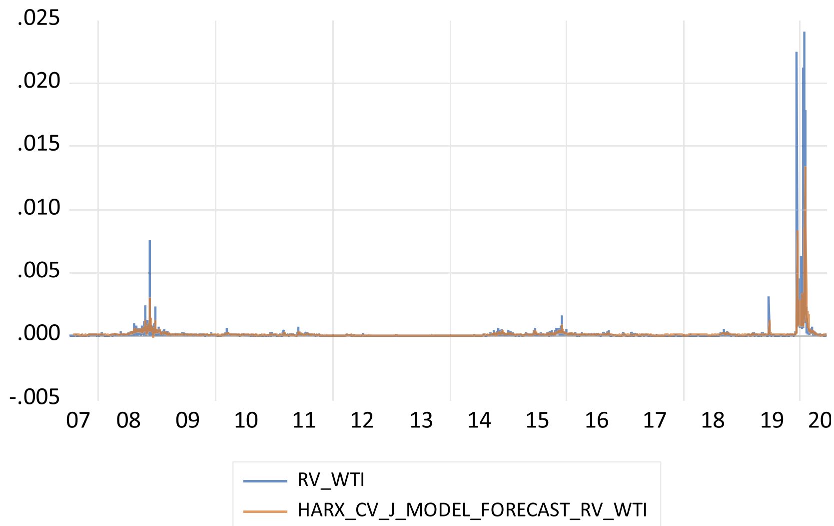


Fig. 12. Modeling Dynamics of Oil RV with the HAR_CV_J_X Model.
Note: RV_WTI denotes the observed oil RV, whereas HARX_CV_J_Model_forecast_RV_WTI denotes the in-sample forecast of the HARX_CV_J_X model (Eq. (10)) for oil RV.

Table 6
HAR-RVX Model Estimation Results.

Coefficients	Oil market	Gas Market
α_1	0.212***	0.321***
	(0.00)	(0.00)
α_2	0.507***	0.201***
	(0.01)	(0.00)
α_3	0.028	0.084*
	(0.36)	(0.06)
β_1	−0.001	−0.067***
	(0.83)	(0.00)
β_2	−0.012	0.018
	(0.45)	(0.46)
β_3	0.013	0.031*
	(0.24)	(0.08)
β_4	0.005	0.052*
	(0.62)	(0.07)
β_5	−0.008	0.033***
	(0.58)	(0.04)
Adjusted R^2	0.371	0.22
DW	2.02	2.15

Note: DW denotes the statistics of the Durbin-Watson test. (***), (**), and (*) denote statistical significance at the 1%, 5%, and 10% levels, respectively. Values in (.) denote the probabilities of the t-ratios for which standard errors and covariance are corrected (Bartlett kernel and Newey-West fixed bandwidth = 10.0).

HAR-RV estimation in Table 5.

From Table 5, we find that high-frequency information and RV (daily and weekly) mainly explain the dynamics of volatility in the oil market, whereas in the gas market, both high-frequency volatility (daily and weekly) and monthly related information have a positive and significant effect on RV. This suggests further evidence of the well-known memory/persistence effects in volatility dynamics and that the consideration of past information can help explain the changes in energy price volatility.

However, as shown in Figs. 10 and 11, HAR_RV fails to reproduce the dynamics of energy price volatility for both oil and gas, particularly during periods of turbulence and crisis times (2008–2009, 2014–2015, 2020).

In other words, there is room for improvement in specifications. In this context, we propose to estimate an HAR_RVX model (Eq. (9)) that extends the basic HAR_RV model while introducing trading volume, EPU, the VIX, and two proxies for sentiment (*News_Sent 1* and *Thr_Senti2*).[11]

We summarize the main results in Table 6. Considering the oil market, we observe that HAR_RVX performs less than the basic HAR_RV model in explaining dynamic volatility. Indeed, the inclusion of exogenous variables does not improve the model, and neither the trading volume nor the behavioral factors (uncertainty, VIX, and sentiment) are statistically significant, notably for oil. Only past information related to oil RV appeared to be statistically significant in the basic HAR_RV model. Regarding the gas market, daily, weekly, and monthly news significantly drove RV dynamics. We also find that the VIX and sentiment have a positive and significant but low effect on volatility. EPU do not have a significant effect, whereas trading volume has a negative and significant effect.

Next, to improve this specification and better explain the dynamics of RV and gas, we propose considering the HAR_CV_J_X specification (model (10)), for which we consider the volatility components (continuous volatility and jumps) among the potential drivers of RV. We report the main results in Table 7. Accordingly, we find that the dynamics of RV in the oil market are driven mainly and significantly by continuous volatility (daily and weekly), while jumps and behavioral factors do not show any significant effects. Regarding the gas market, we find that the RV depends on both continuous volatility and jumps. We also find that the trading volume still has a negative effect on RV, while both the VIX and sentiment appear useful in explaining changes in gas prices. Interestingly, when looking at the performance of this HAR_CV_J_X model, we note that the consideration of RV decomposition into continuous volatility and jumps improved the analysis, as this specification enables us to reproduce the abrupt changes in oil and gas prices during turbulent times (2008–2009 global financial crisis, 2014–2015 shale revolution, COVID-19 outbreak, etc.).[12]

Overall, the above linear specifications confirm the usefulness of volatility components and high-frequency data for improving the

[11] We recall that *News_sent1* refers to the sentiment index proposed by Shapiro et al. (2020), while *ThR_sent2* is the sentiment index proposed by Thomson Reuters.

[12] See Figs. 12 and 13 for more details.

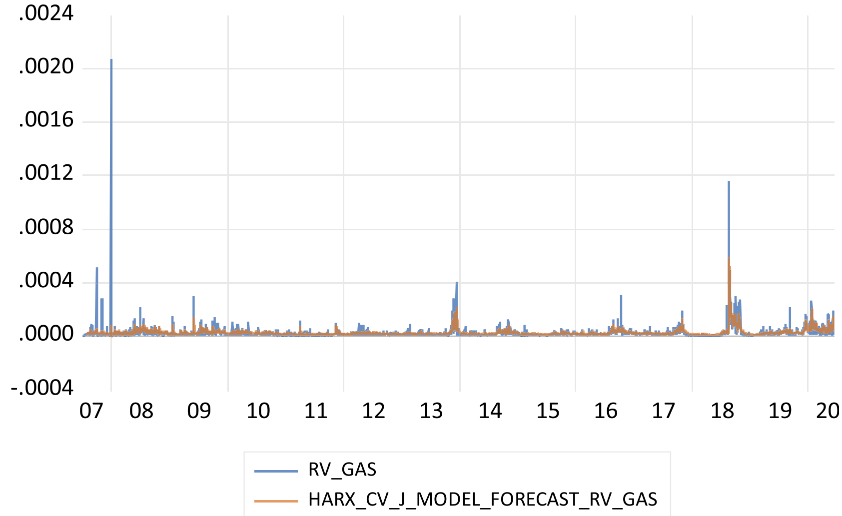


Fig. 13. Modeling Dynamics of Gas RV with the HAR_CV_J_X Model.
Note: RV_WTI denotes the observed gas RV, whereas HARX_CV_J_Model_forecast_RV_gas denotes the in-sample forecast of the HARX_CV_J_X model (Eq. (10)) for the gas RV.

Table 7
Basic HAR-CV_J_X Model Estimation Results.

Coefficients	Oil market	Gas Market
α_1	0.231***	0.386***
	(0.00)	(0.00)
α_2	0.329***	0.199***
	(0.00)	(0.00)
α_3	0.060	0.045***
	(0.23)	(0.01)
α'_1	0.142	0.012
	(0.35)	(0.63)
α'_2	0.045	−0.019**
	(0.68)	(0.03)
α'_3	−0.023*	−0.008
	(0.09)	(0.25)
β_1	0.001	−0.059***
	(0.87)	(0.00)
β_2	−5.49E-06	0.003
	(0.99)	(0.85)
β_3	0.011	0.024*
	(0.28)	(0.07)
β_4	0.013	0.034*
	(0.35)	(0.08)
β_5	−0.012	0.017*
	(0.45)	(0.10)
Adjusted R^2	0.311	0.32
DW	2.16	2.06

Note: DW denotes the statistics of the Durbin-Watson test. (***), (**), and (*) denote statistical significance at the 1%, 5%, and 10% levels, respectively. Values in (.) denote the probabilities of the t-ratios for which standard errors and covariance are corrected (Bartlett kernel and Newey-West fixed bandwidth = 10.0).

analysis of energy price volatility. We also find evidence of the significant impact of behavioral factors (market anxiety (VIX) and investor sentiment) on the volatility of energy prices. However, this modeling imposes restrictive assumptions such as normality, linearity, and symmetry, which cannot hold in practice, as it is most likely that investors react differently when the market is in its down state versus in its up state. Therefore, this econometric framework might be less suitable for appropriately reproducing the dynamics of energy price volatility, and in particular, for capturing the effects of behavioral factors on volatility. To extend this framework to a more general and nonlinear context, we apply different structural break and linearity tests. Next, we discuss the main results.

Table 8
Ramsey RESET Test Results.

	Oil RV	Gas RV
F-statistic	19.056	14.80
(*p*-value)	(0.00)	(0.00)

4.5. Linearity tests

First, we considered Brock et al.'s (1987) BDS test, which detects nonlinear serial dependence in time series. The BDS test tests the null hypothesis of independent identical distribution (i.i.d.) against the alternative hypothesis of linear or nonlinear dependence. Accordingly, the BDS test can be considered as an ad-hoc diagnostic test for detecting nonlinearities in a time series. When applying the BDS to the series of RV for oil and gas, we found that the null hypothesis (i.i.d.) could not be accepted for either series. Thus, either a linear or nonlinear structure characterizes the dynamics of the RV of the two energy prices.

Second, to check the validity of the above linear HAR_CV_J_X model, we apply the Ramsey (1969) regression equation specification error test (RESET), which checks the misspecification of the underlined linear model. Indeed, its basic idea is to check whether nonlinear combinations of explanatory variables have the potential to explain the dependent variable of the linear model (the RV of the energy price in our case). If the null hypothesis of the RESET test is rejected, it indicates that the linear model is misspecified, and that the dynamics of the dependent variable might exhibit nonlinearity. We reported the main results of this test in Table 8. For both energy markets, the null hypothesis is rejected, suggesting that the linear HAR_CV_J_X model is misspecified, and that further nonlinear combinations of the explanatory variables might help improve the modeling of RV dynamics.

To better explore the nonlinearity option, Bai and Perron (2003) applied structural break tests to test for multiple breakpoints. The Bai-Perron test tests the null hypothesis of L + 1 versus L sequentially determined breaks and presents the advantage of identifying further breaks endogenously.[13] From Table 9, we find that the null hypothesis of zero breaks is rejected for both commodities and that at least one break

[13] The break test options are trimming = 0.15, max breaks = 8, and the significance level is 5%.

Table 9
Bai-Perron Test Results.

Break test (number of breaks)	Oil RV	Gas RV
F-statistic (1)	57.77*	43.97*
F-statistic (2)	4.67	24.97

Note: The critical value for the test (0 breaks versus 1 break) is 27.03, whereas that for the test (1 vs. 2 breaks) is 29.24. (*) denotes a significant break at the 5% level.

Table 10
Linearity Test Results.

Linearity Test (*p*-value)	Oil RV	Gas RV
Tsay (1989) Test	0.00***	0.00***
Hansen (1996) test	0.00***	0.00***
Teräsvirta (1994) sequential tests		
H_3	0.00***	0.00***
H_2	0.00***	0.00***
H_1	0.00***	0.00***
Escribano and Jorda (2001) tests		
H_{0L}	0.00***	0.00***
H_{0E}	0.00***	0.00***

Note: (***) denotes the rejection of the null hypothesis of linearity at the 5% level. All Teräsvirta sequential tests are based on the 3rd order Taylor expansion, whereas all Escribano-Jorda tests are based on the 4th order Taylor expansion. L and E denote Logistic and Exponential transition functions, respectively.
See Teräsvirta (1994) for more details on the properties of these functions.

is statistically significant. This result is in line with the results of the Ramsey RESET test, as they point to the presence of structural breaks and nonlinearity in the RV dynamics. This finding is particularly relevant as it suggests that energy price volatility is characterized by structural breaks and nonlinearity; therefore, the drivers of volatility might enter nonlinearly.

To test the linearity hypothesis more explicitly and double-check for the presence of threshold effects in RV dynamics, we ran two classes of linearity tests. Tsay (1989) and Hansen (1996) test the null hypothesis of linearity against the alternative hypothesis of nonlinearity given by threshold autoregressive (TAR) models, whereas Teräsvirta (1994) tests linearity against the nonlinearity given by smooth transition autoregressive (STAR) models. The former check for nonlinearity and multiple regimes under the assumption of an abrupt transition between regimes, whereas the latter check for nonlinearity with a smooth transition between regimes. We also apply the Escribano-Jorda (2001) linearity test, which specifies the transition function (exponential or logistic) that conducts the transition across regimes. In other words, these two classes of linearity tests were required to identify the threshold variable. To this end, we apply these tests to different threshold variables, and when linearity is rejected, the optimal threshold variable was that for which the null hypothesis of linearity is most strongly rejected. Our findings show that linearity is strongly rejected for both commodities: oil and gas and the linearity was most strongly rejected when considering "Jump" as the threshold variable. The main results of the linearity tests are listed in Table 10.

From Table 10, we note that, regardless of the threshold test under consideration, linearity was strongly rejected at the 1% level. This result is relevant because it suggests that the RV dynamics for oil and gas exhibit nonlinearity and threshold effects, which point to a time variation related to the drivers of energy price volatility. Indeed, the reaction of energy price volatility to these drivers may vary, depending on the volatility regime under consideration. Furthermore, when analyzing the intensity of the rejection of linearity, we find that Teräsvirta (1994) sequential tests and Escribano-Jorda's (2001) linearity tests recommend a first-order logistic function with a nonzero threshold to the exponential function for the gas market, and the inverse is true for the oil market. This means that, when considering the smooth transition model, the transition between volatility regimes is carried out by a first-order logistic function (exponential function) for gas (oil), for which the transition speed and threshold were estimated endogenously.

Table 11
Threshold HAR-CV_J_X Model Results with Two Regimes for Oil.

Regime 1		Regime 2	
Coefficients	Estimators	Coefficients	Estimators
α_{11}	0.627*** (0.00)	α_{12}	0.057 (0.69)
α_{21}	0.479*** (0.00)	α_{22}	0.299*** (0.00)
α_{31}	0.008 (0.17)	α_{32}	0.183 (0.16)
α'_{11}	−0.065 (0.14)	α'_{12}	0.997*** (0.02)
α'_{21}	−0.376 (0.44)	α'_{22}	0.027 (0.68)
α'_{31}	−0.073* (0.06)	α'_{32}	−0.048* (0.07)
β_{11}	0.001 (0.85)	β_{12}	−0.044 (0.30)
β_{21}	−0.019 (0.11)	β_{22}	0.077 (0.21)
β_{31}	0.001 (0.96)	β_{32}	0.001 (0.96)
β_{41}	0.006 (0.28)	β_{42}	0.077* (0.07)
β_{51}	0.0005 (0.95)	β_{52}	−0.024 (0.70)
n_1	3455	n_2	651
Threshold (c)	5.162E-05		
Adjusted R^2	0.418		
DW	2.10		

Note: DW denotes the statistics of the Durbin-Watson test. (***), (**), and (*) denote statistical significance at the 1%, 5%, and 10% levels, respectively. Values in (.) denote the probabilities of the t-ratios for which standard errors and covariance are corrected (Bartlett kernel and Newey-West fixed bandwidth = 10.0). n_1 and n_2 denote the number of observations in Regimes 1 and 2, respectively. Threshold (c) denotes the estimated threshold separating the two regimes.

4.6. Nonlinear modeling

In this subsection, we analyze nonlinear or threshold models. The null hypothesis of linearity was rejected against nonlinearity when the alternative was provided by either an abrupt threshold model or a smooth transition model. Rather than randomly selecting a nonlinear or threshold model, we propose estimating both nonlinear models to compare their results and identify the model that fits the data better.

We begin with the oil market. First, we estimate a threshold HAR_CV_J_X model with two regimes in line with the model or Eq. (11). We report the main results of this threshold model in Table 11 and note the interesting results. First, the dynamics of oil price volatility exhibit nonlinearity, time- threshold effects, and are time-varying. The drivers of oil price volatility vary across regimes. In the first regime, only the information provided by daily and weekly continuous volatility is relevant for explaining changes in oil prices. In the second regime, we find that both continuous volatility and jumps have a nonlinear effect on the RV. Furthermore, interestingly, we find a positive effect of sentiment on the upper regime of volatility.

This finding is relevant, and it indicates that in the lower regime of oil price volatility, when the threshold volatility defined by jump is less than the threshold (5.162E-05), the oil price RV is driven only by daily and weekly continuous volatility, suggesting that only public information accounts in this regime explain and forecast oil price volatility. However, in the upper volatility regime, when the oil jump exceeds this threshold, the oil RV becomes more driven by jumps and sentiment, in addition to weekly continuous volatility. This suggests that when oil

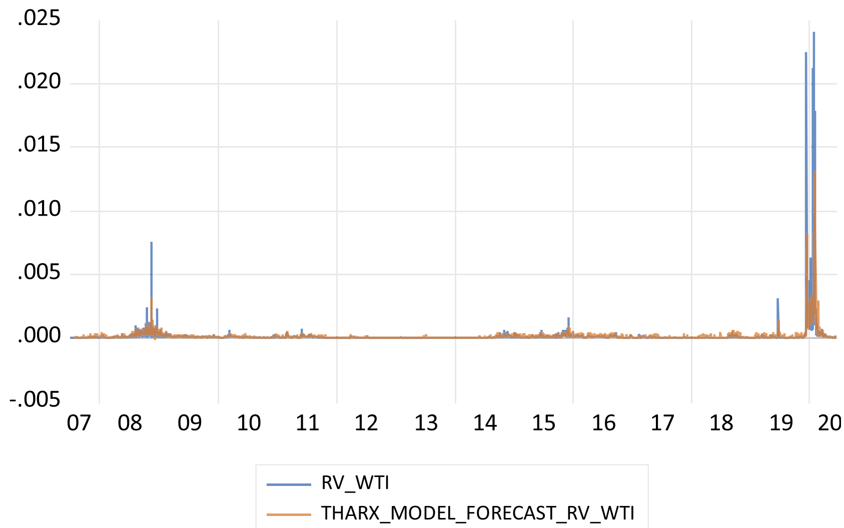


Fig. 14. Modeling Dynamics of Oil RV with the Two-regime Threshold HAR_CV_J_X Model.
Note: RV_WTI denotes the observed gas RV, whereas HARX_CV_J_Model_forecast_RV_gas denotes the in-sample forecast of the HARX_CV_J_X model (Eq. (11)) for the gas RV.

Table 12
Exponential Smooth Transition HAR-CV_J_X Model Estimation Results for Oil.

Linear Regime		Nonlinear Regime	
Coefficients	Estimators	Coefficients	Estimators
α_{11}	−0.014 (0.89)	α_{12}	0.243* (0.08)
α_{21}	−0.352*** (0.00)	α_{22}	0.844*** (0.00)
α_{31}	0.277 (0.19)	α_{32}	−0.239 (0.22)
α'_{11}	0.907** (0.04)	α'_{12}	−0.886** (0.05)
α'_{21}	0.284*** (0.00)	α'_{22}	−0.294*** (0.00)
α'_{31}	−0.083 (0.45)	α'_{32}	−0.083 (0.45)
β_{11}	−0.019* (0.06)	β_{12}	0.013 (0.50)
β_{21}	−0.017 (0.63)	β_{22}	−0.004 (0.92)
β_{31}	3.29E-06 (0.87)	β_{32}	−0.00- (0.86)
β_{41}	−0.028 (0.35)	β_{42}	0.046*** (0.01)
β_{51}	−0.076** (0.03)	β_{52}	0.074** (0.05)
γ	274.03*** (0.00)		
Threshold (c)	−0.107*** (0.00)		
Adjusted R^2	0.531		
DW	1.99		

Note: DW denotes the statistics of the Durbin-Watson test. (***), (**), and (*) denote statistical significance at the 1%, 5%, and 10% levels, respectively. Values in (.) denote the probabilities of the t-ratios for which standard errors and covariance are corrected (Bartlett kernel and Newey-West fixed bandwidth = 10.0). n_1 and n_2 denote the number of observations in Regimes 1 and 2, respectively. Threshold (c) denotes the estimated threshold separating the two regimes.

price volatility exceeds a certain threshold, investing in oil-related assets becomes more attractive and retains more investor attention, which is why both investor sentiment (proxy related to investor confidence, animal spirits) and private information (jump) are more active in driving oil price volatility.

Overall, the two-regime threshold HAR-CV_J_X model provides interesting results and improves the modeling of oil price volatility provided by the above linear models. Indeed, from Fig. 14, we find that this threshold model captures the most important changes in oil prices, but only partially captures those of 2008 and 2020.

However, it is worth recalling that while the threshold model captures further nonlinearity, asymmetry and multiple regimes, it imposes an abrupt transition across regimes and therefore it only reproduces extreme states of volatility. To extend this nonlinear framework and simultaneously reproduce at the same time both extreme states of oil price volatility and a continuum of intermediate states, we propose allowing the transition across regimes to occur smoothly rather than abruptly, which is the most realistic hypothesis to consider investor heterogeneity. Accordingly, and while taking into account the above results of linearity and threshold tests, we estimated an exponential smooth transition HAR_CV_J_X model, and we reported the main results in Table 12.

From Table 12, we note interesting results. First, we confirm that the oil price RV exhibits nonlinearity, threshold effects, and time variation, and that its drivers vary per regime. Indeed, in the first regime, oil price volatility is driven by trading activities, such as trading volume, weekly continuous volatility, and jumps. Further, sentiment shows a negative and significant effect on volatility, suggesting that below a certain threshold, when volatility is low and changes in oil prices are relatively weak, an increase in sentiment yields some feedback or correction that attenuates oil price volatility. Indeed, in this regime, which is also identified by a threshold variable related to sentiment, oil price variations and jumps are not high enough to guarantee a high expected return, justifying that sentiment appeals to more precaution and attenuates oil price volatility. However, in the second regime, when the sentiment exceeds a given threshold ($c = -0.107$ that is endogenously estimated by the model), the oil price volatility will be driven by both continuous volatility and jumps. However, whereas public information related to continuous volatility amplifies volatility, private news related to jumps may attenuate it. Furthermore, the role of sentiment becomes positive and more significant, which amplifies and drives oil price volatility in this regime. Indeed, investors would follow these high oil price variations and would aim to benefit from this oil price valuation. They activate more rapidly and significantly their *animal spirits* and this *irrational exuberance à la Shiller*, drives oil price volatility to unprecedented levels. Interestingly, though sentiment shows a negative effect in the first regime and a positive effect in the second regime, the overall

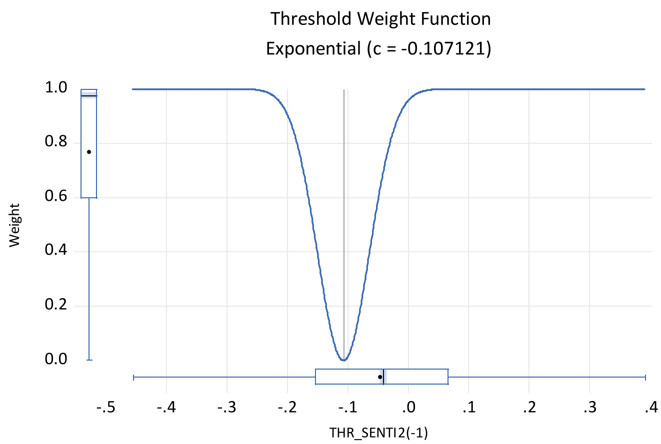


Fig. 15. Estimated Exponential Transition Function for Oil.

impact of sentiment is rather positive, suggesting that such behavioral factors amplify oil price volatility and can help explain its excess.

To better illustrate the importance of this nonlinearity in oil price volatility and the predominance of this upper volatility regime, characterized by an important volatility excess, we report the estimated exponential transition function in Fig. 15. This function reaches several times and persists in the upper regime and around the value of 1, suggesting that the nonlinear regime dominates the dynamics and, therefore, the importance of the drivers identified in the nonlinear regime. This finding is particularly relevant, as it highlights that sentiment plays a key role in driving oil price volatility and maintaining the upper regime.

In Fig. 16, we report the observed oil price RV against the oil price RV estimated by the Exponential Smooth Transition HAR-CV_J_X Model. We find that the Exponential Smooth Transition HAR-CV_J_X Model faithfully reproduces the dynamics of oil price volatility, as it reproduces most changes in the oil price. In particular, it seems that the combination of jumps/continuous volatility decomposition, nonlinearity, multiple regimes, and smooth transition hypotheses significantly improves the previous specifications. Interestingly, when we run the main diagnostic tests on the estimated residuals of this model, we can reject the hypothesis of remaining nonlinearity and accept the hypothesis of parameter constancy, suggesting that the estimated residuals do not present further evidence of nonlinearity and that the tour model fits the data well.

We consider the gas market in which the linearity hypothesis is also significantly rejected. Accordingly, we specify and tested different threshold HAR specifications, and the best specification that fit the data corresponded well to a threshold HAR-CV_J_X model with two regimes that we retained, and we report its results in Table 13. From Table 13, for the oil market, we note that the dynamics of the gas RV are nonlinear and time-varying according to the regime under consideration. In the first regime, gas volatility is driven only by daily, weekly, and monthly continuous volatilities, jumps, and trading volumes. In the second regime, when the gas RV exceeds a threshold of 2.20E-05, gas volatility becomes more strongly and significantly driven by continuous volatility and sentiment. In this second regime, sentiment has a positive and significant effect on gas volatility, confirming that an increase in sentiment dealing with further evidence of overconfidence and irrational exuberance might amplify gas volatility. This suggests that in the upper regime, several behavioral factors might impact and drive this volatility to the highest levels. Overall, this three-regime threshold model provides interesting results and improves on different linear modeling for gas volatility, as shown in Fig. 17.

Finally, from the perspective of improving the threshold HAR modeling for gas volatility while trying to capture different intermediate states of volatility besides these extreme regimes or states, we also allowed the transition across regimes of gas volatility to occur smoothly rather than abruptly, and we estimated a Logistic Smooth Transition HAR_CV_J_X Model, also supported by the above linearity tests. We reported the main results in Table 14.

Thus, we find that gas price volatility is explained and driven differently by the regime. In the low-volatility regime, continuous volatility (daily, weekly, and monthly) and jumps (weekly) drive significant changes in gas prices. This sign is positive for continuous volatility (indicating the amplification effect of public information) and negative for jumps, suggesting that private information attenuates gas volatility. Additionally, trading volume has a negative and significant effect on gas volatility, for which the negative sign is not in line with the MDH. Furthermore, neither sentiment nor uncertainty has a significant effect on gas volatility in this linear regime.

When considering the second nonlinear regime, we find that gas volatility is more significantly driven and amplified by daily and sentiment, and less by the other factors. Interestingly, for the oil market, but

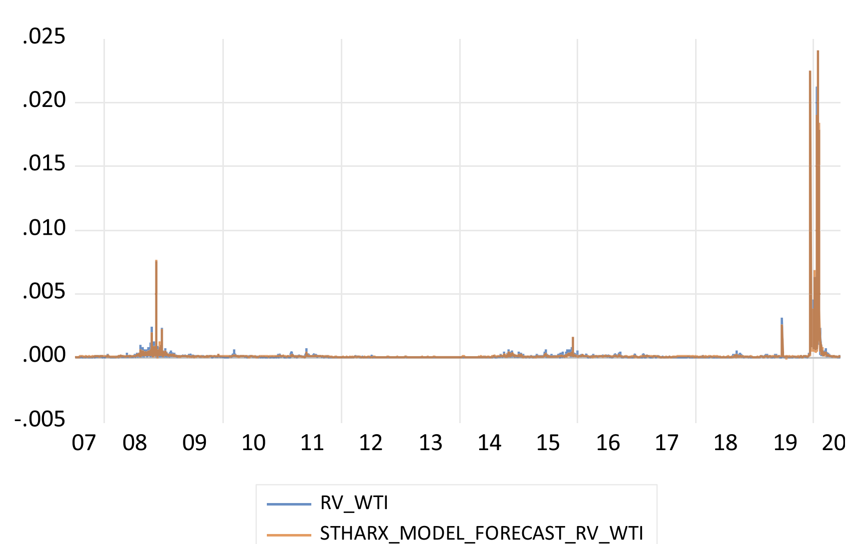


Fig. 16. Modeling Dynamics of Oil RV with the Exponential Smooth Transition HAR_CV_J_X Model.
Note: RV_WTI denotes the observed oil RV, whereas STHARX_Model_Forecast_RV_WTI denotes the in-sample forecast of the exponential STHARX_Model (Eq. (12)) for oil RV.

Table 13
Threshold HAR-CV_J_X Model Estimation Results with Two Regimes for Gas.

Regime 1		Regime 2	
Coefficients	Estimators	Coefficients	Estimators
α_{11}	0.175*** (0.00)	α_{12}	0.407*** (0.00)
α_{21}	0.267*** (0.00)	α_{22}	0.155*** (0.00)
α_{31}	0.065*** (0.00)	α_{32}	0.018 (0.40)
α'_{11}	0.311*** (0.00)	α'_{12}	0.007 (0.72)
α'_{21}	−0.015*** (0.00)	α'_{22}	−0.092* (0.10)
α'_{31}	−0.040** (0.04)	α'_{32}	0.001 (0.87)
β_{11}	−0.064* (0.00)	β_{12}	−0.082* (0.09)
β_{21}	−0.026** (0.03)	β_{22}	0.110 (0.15)
β_{31}	0.026 (0.11)	β_{32}	0.053 (0.22)
β_{41}	−0.002 (0.77)	β_{42}	0.212*** (0.01)
β_{51}	0.006 (0.52)	β_{52}	0.093** (0.03)
n_1	3462	n_2	644
Threshold 1 (c)	2.203E-05		
Adjusted R^2	0.33		
DW	2.06		
LL	−5213.99		

Note: DW denotes the statistics of the Durbin-Watson test. (***), (**), and (*) denote statistical significance at the 1%, 5%, and 10% levels, respectively. Values in (.) denote the probabilities of the t-ratios for which standard errors and covariance are corrected (Bartlett kernel and Newey-West fixed bandwidth = 10.0). n_1, and n_2 denote the number of observations in Regimes 1 and 2, respectively. Threshold (c) denotes the estimated threshold parameter separating the two regimes under consideration.

Table 14
Logistic Smooth Transition HAR-CV_J_X Model Estimation Results for Gas.

Linear Regime		Nonlinear Regime	
Coefficients	Estimators	Coefficients	Estimators
α_{11}	0.328*** (0.00)	α_{12}	0.021 (0.60)
α_{21}	0.230*** (0.00)	α_{22}	−0.099** (0.02)
α_{31}	0.048*** (0.00)	α_{32}	0.019 (0.63)
α'_{11}	−0.005 (0.49)	α'_{12}	0.246*** (0.00)
α'_{21}	−0.017*** (0.00)	α'_{22}	−0.057 (0.38)
α'_{31}	−0.001 (0.75)	α'_{32}	−0.184 (0.22)
β_{11}	−0.062*** (0.00)	β_{12}	0.175 (0.36)
β_{21}	−0.014 (0.31)	β_{22}	0.107 (0.41)
β_{31}	0.020 (0.13)	β_{32}	0.096 (0.45)
β_{41}	0.006 (0.55)	β_{42}	0.425** (0.03)
β_{51}	0.002 (0.76)	β_{52}	0.290*** (0.00)
γ	277.68*** (0.00)		
Threshold (c)	0.0004*** (0.00)		
Adjusted R^2	0.34		
DW	2.04		

Note: DW denotes the statistics of the Durbin-Watson test. (***), (**), and (*) denote statistical significance at the 1%, 5%, and 10% levels, respectively. Values in (.) denote the probabilities of the t-ratios for which standard errors and covariance are corrected (Bartlett kernel and Newey-West fixed bandwidth = 10.0). Threshold (c) denotes the estimated threshold separating the two regimes, and γ is the adjustment speed.

more profoundly for gas than for oil, the impact of sentiment on driving volatility is more relevant in the nonlinear regime than in the linear regime. Furthermore, when considering the nonlinear estimation as a whole, we obtain interesting results. On the one hand, the global effect of EPU on gas price volatility is around +9.7%, suggesting that an increase in uncertainty by 1% provokes an increase in gas price volatility. The global sentiment effect (taking into account the two sentiment proxies) on gas price volatility is positive and is around 71%, suggesting that a 1% increase in sentiment (illustrated by overconfidence, optimism excess, irrational exuberance, etc.) would imply a 71% increase in gas price volatility. This finding is relevant and in line with Akerlof and Shiller (2010) (2015), suggesting that behavioral factors (uncertainty and sentiment) and animal spirits are relevant in explaining the dynamics of gas price volatility. On the other hand, regarding fundamentals, while the MDH hypothesis is not rejected, the global impact of trading volumes on gas price volatility is negative and significant only in

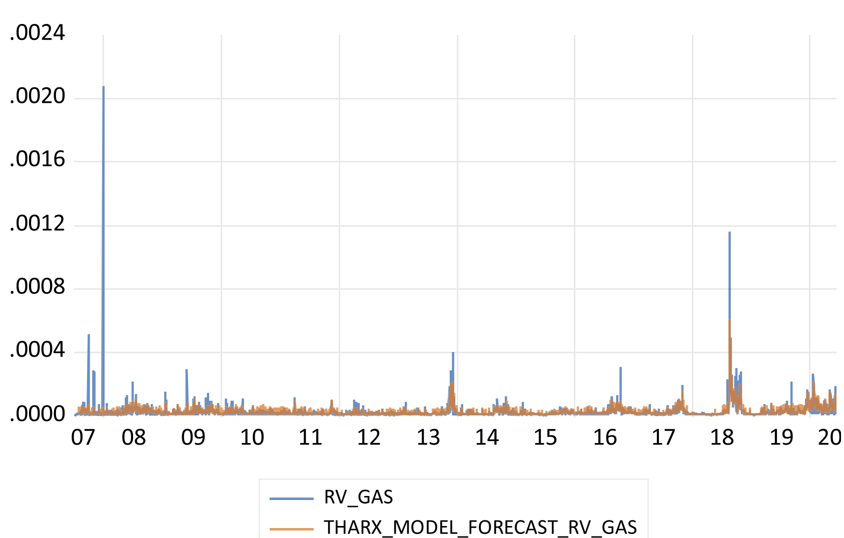


Fig. 17. Modeling Dynamics of Gas RV using the Threshold HAR-CV_J_X Model with Two Regimes.

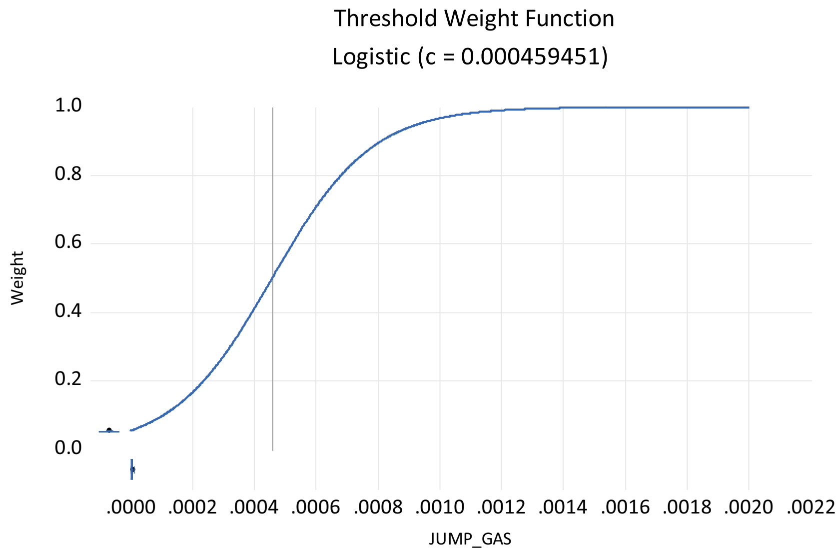


Fig. 18. Estimated Logistic Transition Function for Gas.

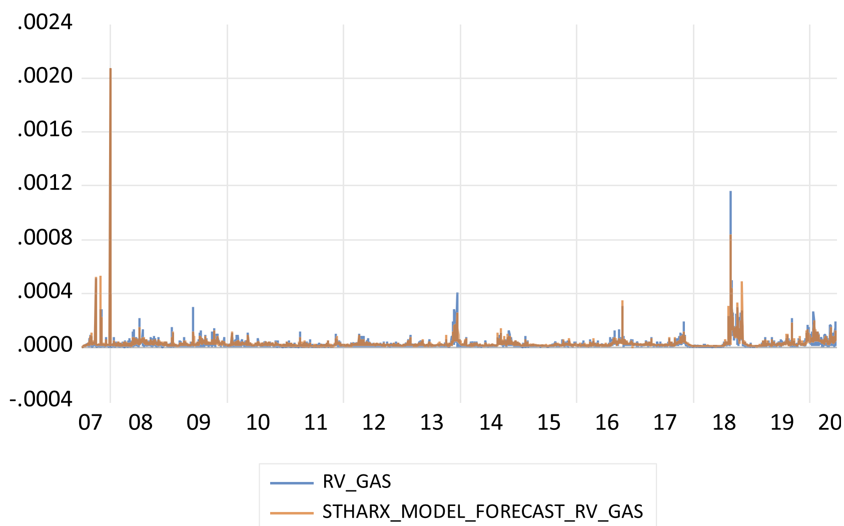


Fig. 19. Modeling Dynamics of Gas RV using the Logistic Smooth Transition HAR-CV_J_X Model.
Note: RV_Gas denotes the observed gas RV, whereas STHARX_Model_Forecast_RV_Gas denotes the in-sample forecast of the logistic STHARX_Model (Eq. (12)) for gas RV.

the first regime (approximately 6.2%). This finding confirms the dominance of behavioral factors driving volatility.

Otherwise, switching between the two regimes occurs when the jump exceeds the threshold value (0.00045) and seems abrupt, as the slope parameter (γ) is significantly high, suggesting a highly rapid transition between the two regimes, as illustrated by the estimated transition logistic function in Fig. 18. For the oil market, the upper gas price regime appears to dominate gas price volatility and provides further evidence of high volatility amplified or driven by behavioral factors rather than by fundamentals.

Overall, the smooth logistic transition HAR-CV_J_X model provides good results and reproduces the main properties of gas price volatility (Fig. 19). Indeed, unlike the previous specifications, it is the only model that captures the main changes in gas price volatility. Diagnostic tests also largely validated the model and showed that the estimated residuals present good statistical properties.

4.7. Forecasting

We estimated different econometric specifications (basic HAR_RV, HAR_RVX, HAR_CV_JX, threshold HAR_CV_JX, and Smooth Transition HAR_CV_JX) to try to model the dynamics of price volatility and assess changes in oil and gas prices. While all models helped improve the analysis of energy price volatility and identifying its main drivers, nonlinear models appear more relevant. To compare these specifications, we propose hereafter to analyze their forecasting performance. Therefore, we retained the basic HAR_RV specification as the benchmark model. Further, we re-estimate these models over the period of July 3, 2007 to May 27, 2021; we analyze the forecasting performance of these models over the period of May 28, 2021 to May 27, 2022.

In practice, we use two loss functions, the root mean squared error (RMSE) and mean absolute error (MAE), as well as the Theil U_2 coefficient, to compare the out-of-sample forecasting performance of our models with that of a naive approach. The lower the values of MAE and RMSE, the better the forecasting performance of the model. For Theil's U_2 coefficient, when $U_2 < 1$ (resp. $U_2 > 1$), indicates that the forecasting

Table 15a
Forecasting Evaluation for Oil Price Volatility.

Model	Theil U_2	MAE Ratio	RMSE Ratio
Basic HAR_RV	0.999	–	–
HAR_RVX	1.03	1.028	1.047
HAR_CV_JX	0.882	0.978	0.972
Threshold HAR_CV_JX	0.821	0.927	0.962
Smooth Transition HAR-CV_JX	0.807	0.870	0.825

Table 15b
Forecasting Evaluation for Gas Price Volatility.

Model	Theil U_2	MAE Ratio	RMSE Ratio
Basic HAR_RV	1.00	–	–
HAR_RVX	0.825	0.979	0.953
HAR_CV_JX	0.725	0.973	0.945
Threshold HAR_CV_JX	0.767	0.937	
Smooth Transition HAR-CV_JX	0.696	0.879	0.857

accuracy of the model under consideration was greater (resp. less) than the naive approach.

We report the main results in Tables 15a and 15b for oil and gas, respectively. Accordingly, we find that the smooth transition HAR_CV_JX model outperforms the other models in forecasting the future dynamics of volatility for both oil and gas. Furthermore, we conclude that the forecasts of the two nonlinear models and the HAR_CV_JX model are more accurate than the forecasts provided by the naïve approach.

Our findings are relevant for three reasons. First, they show an interest in breaking down volatility into continuous volatility and jumps. Second, they point to the importance of considering behavioral factors such as sentiment and uncertainty to explain the dynamics of energy price volatility and to explain its dynamics. Finally, the consideration of nonlinearity, asymmetry, and switching regime assumptions is helpful and very useful for improving the modeling and forecasting of oil and gas price volatility.

5. Conclusion

This study investigates the volatility dynamics of two major commodities–oil and gas–in an environment characterized by post-COVID-19 uncertainty, high inflation, and geopolitical tensions. We analyze the excess volatility of these markets and identify their main fundamental and psychological drivers. Thus, we test the contribution of behavioral factors (uncertainty, sentiment, and anxiety) in explaining the dynamics of commodity price volatility. We use the West Texas Intermediate (WTI) price and Henry Hub Natural gas and measure the RV model of Andersen and Bollerslev (1998) with 5-min timeframe of high-frequency data. We investigate the dynamics of commodity price volatility by decomposing the RV series into continuous volatility (CV) and jumps (discontinuous volatility). We also test the contribution of behavioral factors using the US EPU and VIX indexes to capture investor anxiety, the Daily News Sentiment Index developed by Shapiro et al. (2020), and the TRMI to capture investor sentiment.

Our study contributes to the literature by extending Corsi (2009) HAR model to establish a nonlinear and time-varying specification robust to asymmetry, structural breaks, and nonlinearity. This nonlinear framework enables us to capture the different forms of lead-lag effects between investor sentiment and volatility. Empirically, we show that both oil and gas markets were characterized by significant jumps over the last two decades. Second, we find that extra-financial news such as uncertainty and investor sentiment are significant drivers of volatility, but the intensity of this leading effect is time-varying, and the reaction of volatility to behavioral factors is more pronounced in the upper regime when volatility reaches high levels. Accordingly, we find that the relationship between volatility and its drivers differs according to the volatility regime or state under consideration. While fundamentals and factors related to public news drive volatility in the first regime, or a state of low volatility, behavioral factors dominate and drive energy price volatility in the upper regime of volatility. A further extension of this work would be to assess the volatility spillovers of these two commodities using high-frequency data and different components of volatility. To this end, an identified structural vector autoregressive model can be set up to analyze these linkages across oil and gas markets.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Fredj Jawadi: Writing – review & editing, Writing – original draft, Supervision, Software, Project administration, Methodology, Conceptualization. **David Bourghelle:** Writing – review & editing, Data curation. **Philippe Rozin:** Writing – review & editing, Resources. **Abdoulkarim Idi Cheffou:** Writing – review & editing, Resources, Investigation. **Gazi Salah Uddin:** Writing – review & editing, Resources, Data curation.

Declaration of competing interest

None.

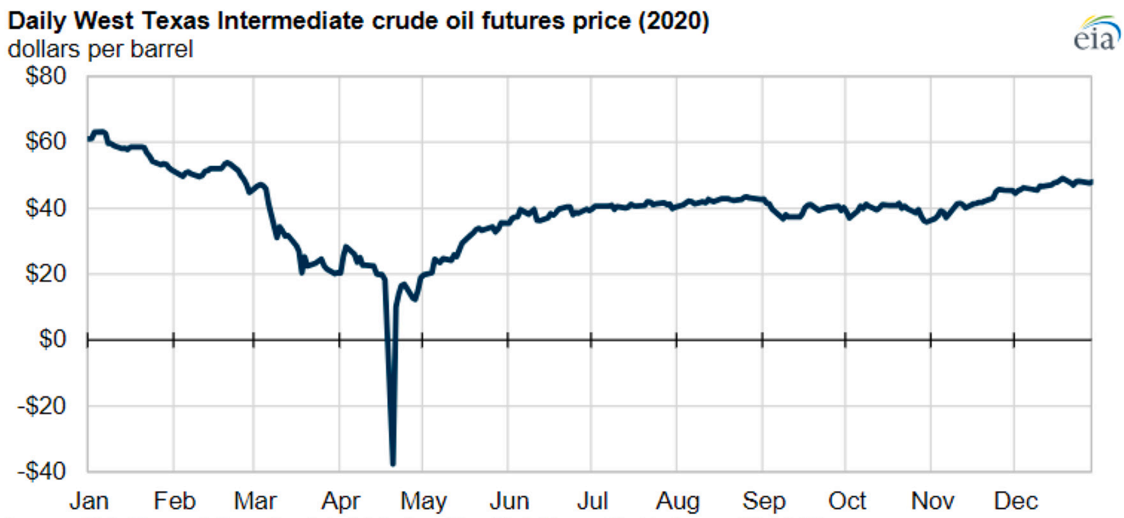
Acknowledgments

The authors thank Christiane Baumeister, Tim Bollerslev, James Hamilton, and Timo Teräsvirta for very useful and interesting discussions that helped us to improve the paper. We would like also to thank Cheng Hsiao and the participants of the Econometric Seminar of the Economic Department at the University of Southern California for very helpful comments and feedback.

Appendix A. Negative oil price

JANUARY 5, 2021

Crude oil prices briefly traded below $0 in spring 2020 but have since been mostly flat



Source: U.S. Energy Information Administration, New York Mercantile Exchange (NYMEX)
Note: Prices reflect front-month (contract with the earliest delivery date) futures prices.

In the first half of 2020, responses to the COVID-19 pandemic led to steep declines in global petroleum demand and to volatile crude oil markets. The second half of the year was characterized by relatively stable prices as demand began to recover. As petroleum demand fell and U.S. crude oil inventories increased, West Texas Intermediate (WTI) crude oil traded at negative prices on April 20, the first time the price for the WTI futures contract fell to less than zero since trading began in 1983. The next day, Brent crude oil, another global crude oil price benchmark, fell to $9.12 per barrel (b), its lowest daily price in decades.

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eneco.2024.107465.

References

Akerlof, G.A., Shiller, R.J., 2010. Animal Spirits: How Human Psychology Drives the Economy, and why it Matters for Global Capitalism. Princeton university press.

Andersen, T., Bollerslev, T., 1998. Deutsche mark-dollar volatility: intraday volatility patterns, macroeconomic announcements and longer run dependencies. J. Financ. 1, 219–265.

Andersen, T.G., Bollerslev, T., Diebold, F.X., 2007a. Roughing it up: including jump components in the measurement, modeling and forecasting of return volatility. Rev. Econ. Stat. 89, 701–720. https://doi.org/10.1162/rest.89.4.701.

Andersen, T.G., Bollerslev, T., Dobrev, D., 2007b. No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: theory and testable distributional implications. J. Econ. 138, 125–180. https://doi.org/10.1016/j.jeconom.2006.05.018.

Assaf, A., Charif, H., Mokni, K., 2021. Dynamic connectedness between uncertainty and energy markets: Do investor sentiments matter? Resources Policy 72, 102112.

Bai, J., Perron, P., 2003. Computation and analysis of multiple structural change models. J. Appl. Econ. 18, 1–22. https://doi.org/10.1002/jae.659.

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross section of stock returns. J. Financ. 61, 1645–1680. https://doi.org/10.1111/j.1540-6261.2006.00885.x.

Barndorff-Nielsen, O.E., Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps. J. Financ. Econ. 2, 1–37. https://doi.org/10.1093/jjfinec/nbh001.

Barndorff-Nielsen, O.E., Shephard, N., 2006. Econometrics of testing for jumps in financial economics using bipower variation. J. Financ. Econ. 4, 1–30. https://doi.org/10.1093/jjfinec/nbi022.

Borovkova, S., 2011. News analytics for energy futures. (Available at SSRN 1719582). SSRN J. https://doi.org/10.2139/ssrn.1719582.

Chan, K.F., Gray, P., 2017. Do scheduled macroeconomic announcements influence energy price jumps? J. Futur. Mark. 37, 71–89. https://doi.org/10.1002/fut.21796.

Chen, Z., Liang, C., Umar, M., 2021. Is investor sentiment stronger than VIX and uncertainty indices in predicting energy volatility? Res. Policy 74, 102391. https://doi.org/10.1016/j.resourpol.2021.102391.

Cifarelli, G., Paladino, G., 2010. Oil price dynamics and speculation. Energy Econ. 32, 363–372. https://doi.org/10.1016/j.eneco.2009.08.014.

Coleman, L., 2012. Explaining crude oil prices using fundamental measures. Energy Policy 40, 318–324. https://doi.org/10.1016/j.enpol.2011.10.012.

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. J. Financ. Econ. 7, 174–196. https://doi.org/10.1093/jjfinec/nbp001.

Deeney, P., Cummins, M., Dowling, M., Bermingham, A., 2015. Sentiment in oil markets. Int. Rev. Financ. Anal. 39, 179–185. https://doi.org/10.1016/j.irfa.2015.01.005.

Dickey, D.A., Fuller, W.A., 1979. Distribution of the Estimators for Autoregressive Time Series. Journal of the American Statistical Association 74, 427–431.

Dickey, D.A., Fuller, W.A., 1981. Likelihood Ratio Tests for Autoregressive Time Series with a Unit Root. Econometrica 49, 1057–1072.

Dunham, L.M., Garcia, J., 2021. Measuring the effect of investor sentiment on financial distress. Manag. Financ. 47, 1834–1852. https://doi.org/10.1108/MF-02-2021-0056.

Escribano, A., Jordá, O., 2001. Testing nonlinearity: Decision rules for selecting between logistic and exponential STAR models. Spanish Economic Review 3, 193–209.

Gandhi, P., Loughran, T., McDonald, B., 2019. Using annual report sentiment as a proxy for financial distress in US banks. J. Behav. Financ. 20, 424–436. https://doi.org/10.1080/15427560.2019.1553176.

Granger, C.W., Teräsvirta, T., 1993. Modelling Nonlinear Economic Relationships. Oxford University Press.

Gudkov, N., Ignatieva, K., 2021. Electricity price modelling with stochastic volatility and jumps: an empirical investigation. Energy Econ. 98, 105260 https://doi.org/10.1016/j.eneco.2021.105260.

Gupta, K., Banerjee, R., 2019. Does OPEC news sentiment influence stock returns of energy firms in the United States? Energy Econ. 77, 34–45. https://doi.org/10.1016/j.eneco.2018.03.017.

Hamilton, J.D., 1983. Oil and the macroeconomy since World War II. J. Polit. Econ. 91, 228–248. https://doi.org/10.1086/261140.

Hamilton, J.D., 2003. What is an oil shock? J. Econ. 113, 363–398. https://doi.org/10.1016/S0304-4076(02)00207-5.

Hamilton, J.D., 2009. Causes and consequences of the oil shock of 2007-08. SSRN J. 215–261 https://doi.org/10.2139/ssrn.2583456.

Han, L., Jin, J., Wu, L., Zeng, H., 2020. The volatility linkage between energy and agricultural futures markets with external shocks. Int. Rev. Financ. Anal. 68, 101317 https://doi.org/10.1016/j.irfa.2019.01.011.

Hansen, B.E., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. Econometrica 64, 413–430. https://doi.org/10.2307/2171789.

Hellström, J., Lundgren, J., Yu, H., 2012. Why do electricity prices jump? Empirical evidence from the Nordic electricity market. Energy Econ. 34, 1774–1781. https://doi.org/10.1016/j.eneco.2012.07.006.

Jawadi, F., 2006. Ajustements non linéaires des cours boursiers dans les pays du G7: essai de modélisation. Doctoral Dissertation. Univ. Paris 10-Nanterre.

Jawadi, F., 2019. Understanding oil price dynamics and their effects over recent decades: an interview with James Hamilton. Energy J. 40, 1–14. https://doi.org/10.1177/01956574201940s21.

Jawadi, F., Louhichi, W., Idi Cheffou, A., 2015. Testing and modeling jump contagion across international stock markets: a nonparametric intraday approach. J. Financ. Mark. 26 (November), 64–84. https://doi.org/10.1016/j.finmar.2015.09.004.

Jawadi, F., Koubbaa, Y., 2007. Dynamique non-linéaire des marchés boursiers du G7: une application des modèles STAR. Finance 28 (1), 29–74.

Jawadi, F., Ftiti, Z., Louhichi, W., 2020. Forecasting energy futures volatility with threshold augmented heterogeneous autoregressive jump models. Econ. Rev. 39, 54–70. https://doi.org/10.1080/07474938.2019.1690190.

Kaufmann, R.K., 2011. The role of market fundamentals and speculation in recent price changes for crude oil. Energy Policy 39, 105–115. https://doi.org/10.1016/j.enpol.2010.09.018.

Kilian, L., 2008. The economic effects of energy price shocks. J. Econ. Lit. 46, 871–909. https://doi.org/10.1257/jel.46.4.871.

Kostrzewski, M., Kostrzewska, J., 2021. The impact of forecasting jumps on forecasting electricity prices. Energies 14, 336. https://doi.org/10.3390/en14020336.

Lee, Y.H., Hu, H.N., Chiou, J.S., 2010. Jump dynamics with structural breaks for crude oil prices. Energy Econ. 32, 343–350. https://doi.org/10.1016/j.eneco.2009.08.006.

Li, Y., Jiang, S., Li, X., Wang, S., 2021. The role of news sentiment in oil futures returns and volatility forecasting: data-decomposition based deep learning approach. Energy Econ. 95, 105140 https://doi.org/10.1016/j.eneco.2021.105140.

Maslyuk-Escobedo, S., Rotaru, K., Dokumentov, A., 2017. News sentiment and jumps in energy spot and futures markets. Pacific-Basin Fin. J. 45, 186–210. https://doi.org/10.1016/j.pacfin.2016.07.001.

Phillips, P.C.B., Perron, P., 1988. Testing for a unit root in time series regression. Biometrika 75 (2), 335–346.

Prokopczuk, M., Symeonidis, L., Wese Simen, C., 2016. Do jumps matter for volatility forecasting? Evidence from energy markets. J. Futur. Mark. 36, 758–792. https://doi.org/10.1002/fut.21759.

Qadan, M., Nama, H., 2018. Investor sentiment and the price of oil. Energy Econ. 69, 42–58. https://doi.org/10.1016/j.eneco.2017.10.035.

Ramsey, D.D., 1969. On the social rate of discount: Comment. The American Economic Review 59 (5), 919–924.

Regnier, E., 2007. Oil and energy price volatility. Energy Econ. 29, 405–427. https://doi.org/10.1016/j.eneco.2005.11.003.

Shapiro, A.H., Sudhof, M., Wilson, D.J., 2020. Measuring News Sentiment. FRBSF Working Paper, pp. 2017–2001. https://doi.org/10.1016/j.jeconom.2020.07.053.

Shiller, R.J., 2015. Irrational exuberance. In: Irrational Exuberance. Princeton University Press, 392 pages.

Song, Y., Ji, Q., Du, Y.J., Geng, J.B., 2019. The dynamic dependence of fossil energy, investor sentiment and renewable energy stock markets. Energy Econ. 84, 104564 https://doi.org/10.1016/j.eneco.2019.104564.

Teräsvirta, T., 1994. Specification, estimation, and evaluation of smooth transition autoregressive models. J. Am. Stat. Assoc. 89, 208–218. https://doi.org/10.1080/01621459.1994.10476462.

Terasvirta, T., Anderson, H.M., 1992. Characterizing nonlinearities in business cycles using smooth transition autoregressive models. J. Appl. Econ. 7 (S1), S119–S136. https://doi.org/10.1002/jae.3950070509.

Tsay, R.S., 1989. Testing and modeling threshold autoregressive processes. J. Am. Stat. Assoc. 84, 231–240. https://doi.org/10.1080/01621459.1989.10478760.

Wang, J., Xie, Z., Li, Q., Tan, J., Xing, R., Chen, Y., Wu, F., 2019. Effect of digitalized rumor clarification on stock markets. Emerg. Markets Fin. Trade. 55, 450–474. https://doi.org/10.1080/1540496X.2018.1534683.

Wang, Y., Li, H., Altuntaş, M., 2022a. Volatility in natural resources commodity prices: evaluating volatility in oil and gas rents. Res. Policy 77, 102766. https://doi.org/10.1016/j.resourpol.2022.102766.

Wang, J., Ma, F., Bouri, E., Zhong, J., 2022b. Volatility of clean energy and natural gas, uncertainty indices, and global economic conditions. Energy Econ. 108, 105904 https://doi.org/10.1016/j.eneco.2022.105904.

Wilmot, N.A., Mason, C.F., 2013. Jump processes in the market for crude oil. Energy J. 34, 33–48. https://doi.org/10.5547/01956574.34.1.2.

Yang, C., Gong, X., Zhang, H., 2019. Volatility forecasting of crude oil futures: the role of investor sentiment and leverage effect. Res. Policy 61, 548–563. https://doi.org/10.1016/j.resourpol.2018.05.012.