Journal of Commodity Markets 30 (2023) 100285

Contents lists available at ScienceDirect

Journal of Commodity Markets

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

Regular article

The economic impact of daily volatility persistence on energy markets

Christina Sklibosios Nikitopoulos *, Alice Carole Thomas, Jianxin Wang

University of Technology Sydney, Finance Department, UTS Business School, PO Box 123 Broadway NSW, 2007, Australia

A R T I C L E I N F O

JEL classification:
JEL
C22
C53
C58
Q40

Keywords:
realized volatility
volatility persistence
macro-economy
energy markets
HAR
forecasting

A B S T R A C T

This study examines the role of daily volatility persistence in transmitting information from macro-economy in the volatility of energy markets. In crude oil and natural gas markets, macro-economic factors, such as the VIX, the credit spread and the Baltic exchange dirty index, impact volatility, and this impact is channeled via the volatility persistence. Further, the impact of returns and variances is primarily transmitted to volatility via the daily volatility persistence. The dependence of volatility persistence on market and macro-economic conditions is termed conditional volatility persistence (CVP). The variation in daily CVP is economically significant, contributing up to 18% of future volatility and accounting for 29% of the model's explanatory power. Inclusion of the CVP in the model significantly improves volatility forecasts. Based on the utility benefits of volatility forecasts, the CVP adjusted volatility models provide up to 160 bps benefit to investors compared to the HAR models, even after accounting for transaction costs and varying trading speeds.

1. Introduction

The volatility of energy markets plays an integral role in the global economy. Information from asset classes such as macro-economy and equity has a material impact and carries significant predictive power for the future evolution of energy market volatility. In the spirit of Ross (1989), cross-market information becomes synonymous with volatility. Accordingly, empirical evidence has demonstrated the importance of this information transmission in modelling and forecasting volatility in commodity markets. Beyond the fundamentally important objective to ensure (statistical and economic) forecasting gain, this study further reveals the essential role of volatility persistence in transmitting this information in the volatility of energy markets. Although many studies find a wide range of variables affect future volatility, the information from lagged volatilities remains the most significant predictive channel (Patton and Sheppard (2015); Bollerslev et al. (2016); Bollerslev et al. (2018)). The correlation between today's and tomorrow's volatility, captured by volatility persistence, is a fundamental block of future volatility. Therefore, divulging the determinants and empirical characteristics of volatility persistence would lead to a more comprehensive understanding of how information is channeled in the volatility of energy markets.

In this paper, we quantify and gauge the importance of volatility persistence in trans-mitting information from economic market variables to the daily volatility of two key energy markets: crude oil and natural gas. We study a heterogeneous autoregressive (HAR)

* Corresponding author.
E-mail addresses: Christina.Nikitopoulos@uts.edu.au (C.S. Nikitopoulos), Alice.C.Thomas@student.uts.edu.au (A.C. Thomas), jianxin.wang@uts.edu.au (J. Wang).

https://doi.org/10.1016/j.jcomm.2022.100285
Received 27 June 2021; Received in revised form 21 August 2022; Accepted 23 August 2022
Available online 8 September 2022

model that allows volatility persistence to vary with returns, lagged volatility and economic variables. The corresponding volatility persistence is termed conditional volatility persistence (CVP). As expected, we find that returns and realized variance are important determinants of volatility persistence in energy markets, while other economic variables also affect volatility persistence. The daily CVP of crude oil is negatively related to the credit spread and the Baltic exchange dirty tanker index,[1] but positively related to the VIX.[2] Further, the impact of financial variables on the volatility persistence of natural gas markets is limited; only Treasury bills are positively related to CVP. These findings provide new evidence of the economic determinants of volatility persistence in energy markets.

Three (interrelated) results further underscore the economic significance of volatility persistence as the information-transmitting mechanism in the volatility dynamics of energy markets. First, the impact of returns is transmitted to future volatility via the volatility persistence channel. After accounting for volatility persistence, the direct impact of returns on future volatility is reduced considerably and/or becomes statistically insignificant (in natural gas). The impact of the economic variables is also transmitted to the energy volatility via the volatility persistence channel, rather than directly to the volatility level. Thus, CVP plays an important role in transmitting information from macro-economic and market conditions in future volatility. Second, the CVP determinants notably contribute to the volatility variation. The VIX, the credit spread and the Baltic exchange dirty tanker index are particularly important for the volatility persistence in crude oil markets. Collectively, the CVP-based variables explain 15%–23% of the variation in daily volatility across the two energy markets, accounting for 25%–38% of the regression R^2.[3] Thus, these variables are not only key determinants of volatility persistence but also transmit a material impact on future volatility that jointly accounts (via CVP) for about one quarter of the variation in future volatility. Third, daily volatility persistence in energy markets is considerable and has large daily variations.[4] This evidence of a statistical and economic significance of daily CVP presents a challenge to models with constant volatility persistence and to the standard practice of modelling and estimating the return impact on the volatility level, and not on its persistence.

Further, we demonstrate that models calibrating the correlation between current and future volatility and incorporating information from macro-economic variables offer enhanced predictive accuracy for energy market volatility. In the spirit of Bollerslev et al. (2018), we assess the economic value of the forecasting models by employing realized utility per unit of wealth, accounting for transaction costs and trading speed. The empirical results reveal superior statistical and economic benefits in incorporating information from macro-economic variables in the daily volatility persistence, particularly in the oil market.[5]

Our findings regarding the economic significance of volatility persistence contribute to two strands of literature. The first strand deals with modelling and analysing the empirical characteristics of volatility persistence. Following the success of the GARCH family models, a wide range of explanations have been proposed with mixed empirical support.[6] However, as Bollerslev et al. (2018) state, "[T]he economic forces behind volatility clustering per se remain poorly understood."[7] Several recent studies explored the determinants of volatility persistence in equity markets. Patton and Sheppard (2015) show that daily volatility persistence is largely driven by negative semi-variance, i.e., the sum of squared negative intraday returns. Bollerslev et al. (2016) demonstrate that the measurement errors in daily volatility reduce its information content and its impact on future volatility. Wang and Yang (2018) find strong support for CVP in the S&P 500 index and stocks in the S&P 100 index. By using a leveraged quantile HAR model, Baur and Dimpfl (2019) demonstrate that volatility persistence and asymmetry are associated with high volatility regimes. Chen and Wang (2020) show that global return and volatility are more important than their local counter-parts in determining local volatility persistence in international equity markets. We add to this literature by identifying key macro-economic determinants of daily CVP in energy markets and demonstrate that by linking volatility persistence to economic conditions, our models produce statistically and economically significant benefits relative to recent advances in modelling volatility dynamics.

The other strand of literature is associated with forecasting commodity market volatility (in particular, in the crude oil market) by conditioning on macro-economic variables. Pan, Wang, Wu, and Yin (2017) use a GARCH-MIDAS models to show that macro-economic variables improve forecasting of oil price volatility, while macroeconomic uncertainty is a strong predictor of volatility in energy markets (Bakas and Triantafyllou (2019)). Nonejad (2020) documents that the informational affinity between

[1] Credit spreads and the Baltic tanker index are positively associated with energy market volatility. An unexpected increase in credit spreads or the tanker index tends to increase the energy market volatility. This volatility reflects the gross information flow over a short period of time (e.g., a day) associated with more priced information, thus the arrival of less correlated information, meaning low volatility persistence.

[2] Ross (1989) suggests that volatility reflects information flow, thus effects transmitted between different markets reflect cross-market information flows. Correlated information, defined as information urged by the same underlying economic change, e.g., a shock in the VIX, tends to arrive within a short period of time causing volatility clustering and thus increased volatility persistence.

[3] The contribution of these determinants is assessed by the means of the Shapley R^2. The Shapley R^2 measures the marginal contributions of a set of explanatory variables to the variation of the dependent variable.

[4] The mean CVP is 54.6% in crude oil and 46.4% in natural gas, while the volatility of the CVP is 8.5% for crude oil and 7.7% for natural gas (which are larger than the standard deviations of daily volatility).

[5] We compare the forecasting performance of our model against those of the Corsi (2009) HAR model, the Patton and Sheppard (2015) HAR with semi-variance (HAR-SV) model, and the Bollerslev et al. (2016) HAR with realized quarticity (HAR-RQ) model.

[6] A partial list includes volatility regime shifts (Lamoureux and Lastrapes (1994)), persistence of information arrivals (Laux and Ng (1993); Andersen and Bollerslev (1997)), parameter uncertainty and investor learning (Brock and LeBaron (1996); Johnson (2001)), heterogeneous trading frequencies (Müller, Dacorogna, Dav'e, Olsen, Pictet, and von Weizsäcker (1997); Xue and Gençay (2012)) and investors' sensitivity to infor-mation (Liesenfeld (2001); Berger et al. (2009)).

[7] This echoes an observation by Diebold and Lopez (1995) that "a consensus economic model producing persistence in conditional variance does not exist." Goodhart and O'Hara (1997) comment that "[P]erhaps the most serious problem of GARCH modelling is that we do not yet have a good theory to explain such persistence."

macro-economic variables and monthly oil volatility is stronger after 2008, while Nguyen and Walther (2020) focus on longer-term forecasting and use MIDAS to be able to accommodate macro-economic variables at different frequencies. An emerging literature using high-frequency data analyses shor-term realized volatility forecasting in commodity markets, see Degiannakis and Filis (2017), Zhang, Ma, Shi, and Huang (2018b), Degiannakis and Filis (2018), Ma et al. (2018), Prokopczuk et al. (2019), Alam et al. (2019), Luo et al. (2020) and Bissoondoyal-Bheenick et al. (2020) and literature within. Degiannakis and Filis (2017) demonstrate that stocks, Forex, commodities and macro-economic information enhances the predictability of oil price volatility, and Degiannakis and Filis (2018) find predictive benefits in oil market volatility by using volatility and returns of financial markets. Alam et al. (2019) study sources of volatility asymmetries in oil market and show that bad volatility dominates good volatility in terms of shock transmissions. We find that the impact of economic variables is channeled to the volatility via volatility persistence. Further, when macro-economic variables matter, calibrating the variation of volatility persistence with macro-economic variables offers a significant forecasting benefit.

This study underscores the economic importance of the volatility persistence as an information transmitting mechanism in energy markets, thus it infers material implications on several aspects of energy markets dynamics. Since the financialization of commodity markets (Tang and Xiong (2012), Silvennoinen and Thorp (2013) and Cheng and Xiong (2014)), it has been argued that volatility in these markets, in particular, crude oil, has been integrated with equity markets, and volatility spillovers are evident between them (Chiang et al. (2015), Basak and Pavlova (2016) and Aromi et al., 2019). Energy market volatility has also a notable impact on the shape of futures curves, risk premiums and the asymmetric nature of the return-volatility relation which has become more pronounced following the financialization of commodity markets.[8] Further, macro-economic and financial variables, such as industrial production, term and credit spreads, the US dollar index and the VIX are key determinants of the dynamics in energy markets (in particular, oil markets) (Chiang et al. (2015), Prokopczuk et al. (2019) and Kang et al. (2020)). The magnitude of this impact differs across energy markets, thus the economic determinants of volatility persistence and their impact would potentially differ.[9] From a practical perspective, the physical energy markets have experienced significant structural changes in recent years, e.g., the expansion of the shale oil and gas markets (Kilian (2016)). Thus, investments in transportation facilities, production planning and inventory management are affected by the determinants of the volatility dynamics and their transmission channels (e.g., volatility persistence). Moreover, volatility forecasting in energy markets is critical for trading and investment performance, derivatives pricing and hedging decisions. Accounting for information flows from economic variables in energy market volatility would potentially offer forecasting gains.

The remainder of the paper is structured as follows. In Section 2, we outline the model specifications for conditional volatility persistence. Section 3 describes the data. We analyse volatility persistence in energy markets and identify its determinants in Section 4. Using forecasting considerations, the statistical and economic benefit of volatility persistence is further examined in Section 5. Section 6 concludes.

2. Modelling volatility persistence

The increasing availability of high-frequency data has improved the estimation and the forecasting of return-based realized variance (RV) measures (Bucci (2017)). The HAR model proposed by Corsi (2009) has emerged as the most popular model capturing the dynamics of daily RV. This model reproduces the empirically observed long memory of financial markets and has equal or better forecasting performance than more complicated models (Corsi (2009) and S'evi (2014)). We adapt HAR models to embed the feature of time-varying volatility persistence that may depend on market variables, such as returns, volatility and macro-economic factors. Thus, we discuss the formulation of volatility persistence in classical HAR models, and we introduce a novel approach to model conditional volatility persistence in HAR models.

2.1. Classical HAR models and volatility persistence

We denote as $RV_{t,D}$ the daily realized variance that is estimated as the sum of squared.

Intra-day returns over a day, thus given by $RV_{t,D} = \sum_{i=1}^{M} r_i^2$ where M is the number of intra-day observations. The HAR model is based

[8] See Chiarella, Kang, Nikitopoulos, and T˜o (2016), Nikitopoulos et al. (2017), Prokopczuk et al. (2017), Christoffersen and Pan (2018), Baur and Dimpfl (2018) and Prokopczuk et al. (2019) for related literature.

[9] Silvennoinen and Thorp (2013) show that macro-economic variables affect crude oil, but not natural gas. Recent empirical studies provide strong evidence of the decoupling of oil and gas prices mainly due to the shale gas revolution (Zhang and Ji (2018) and Zhang, Shi, and Shi (2018a)). Ji, Geng, and Tiwari (2018) show that oil and its refining products act as information transmitters, while natural gas markets act as information receivers.

on the notion that the unobservable variance of returns is a linear function of the lagged squared return sampled over different time horizons, reflecting the impact of an investor with varying trading frequencies (Corsi (2009)). To account for the heterogenous effects of returns on volatility, namely, the leverage effect, the lagged daily, weekly and monthly returns are also included in the model specifications.[10] Accordingly, the basic (leveraged) HAR model is estimated via the following regression:

$$RV_{t+1,D} = \alpha + \beta_D RV_{t,D} + Z_t + \varepsilon_{t+1}, \tag{1}$$

where

$$Z_t = \beta_W RV_{t,W} + \beta_M RV_{t,M} + \theta_D^+ r_{t,D}^+ + \theta_D^- r_{t,D}^- + \theta_W r_{t,W} + \theta_M r_{t,M}, \tag{2}$$

With $RV_{t,W}$ and $RV_{t,M}$ representing the (non-overlapping)[11] averages of the lagged weekly and monthly realized variances, approximated by $RV_{t,W} = \frac{1}{4}\sum_{i=1}^{4} RV_{t-i,\,D}$, and $RV_{t,M} = \frac{1}{17}\sum_{i=5}^{21} RV_{t-i,\,D}$, respectively. The negative and positive daily returns are captured by $r_{t,D}^- = r_t I_{(r_t<0)}$ and $r_{t,D}^+ = r_t I_{(r_t>0)}$, respectively, and the lagged weekly and monthly returns are constructed as $r_{t,W} = \frac{1}{4}\sum_{i=1}^{4} r_{t-i,\,D}$ and $r_{t,M} = \frac{1}{17}\sum_{i=5}^{21} r_{t-i,\,D}$, respectively. The HAR model specifications (1)–(2) infer constant daily volatility persistence, represented by β_D, and ensure that statistically and economically significant variables impacting future daily RV been used as controls including long-run dependence from weekly and monthly RV.

Patton and Sheppard (2015) introduce the negative semi-variance (NSV) and the positive semi-variance (PSV) which capture the influence of negative returns and positive returns on RV, respectively. They demonstrate empirically that in equity markets negative shocks matter more to future volatility than positive shocks. The leveraged HAR model incorporating these semi-variances is denoted as the HAR-SV model, and it is estimated as

$$RV_{t+1,D} = \alpha + \beta_{PSV} PSV_{t,D} + \beta_{NSV} NSV_{t,D} + Z_t + \varepsilon_{t+1}, \tag{3}$$

where Z_t is given by (2), with $NSV_{t,D} = \sum_{i=1}^{n} r^2_{i,\,t} I_{(r_{i,t}<0)}$ and $PSV_{t,D} = \sum_{i=1}^{n} r^2_{i,\,t} I_{(r_{i,t}>0)}$. Although it is not directly inferred, the model's specifications entail a daily volatility persistence which is sensitive to the size and type of shocks - high negative shocks lead to high volatility persistence.[12]

The first explicit reference of the volatility persistence depending on market conditions was proposed by Bollerslev et al. (2016). The daily realized variance, which is computed using high frequency returns, is influenced by market microstructure noise (e.g., bid-ask bounce and tick size) and new events, see Barndorff-Nielsen and Shephard (2002) and Andersen et al. (2011). The realized quad-power quarticity (RQ) captures the variance of these measurements errors and it is estimated as $RQ_t = \frac{n}{3}\sum_{i=1}^{n} r_{i,t}^4$. High RQ_t means more noise and less information in $RV_{t,D}$, therefore less impact from $RV_{t,D}$ on $RV_{t+1,D}$. Thus, as documented by Bollerslev et al. (2016), this is a systemic variation associated with $RV_{t,D}$. The square root of RQ_t interacts with the lagged realized variance, and it is added to the HAR model to produce the (leveraged) HAR-RQ model as

$$RV_{t+1,D} = \alpha + \left(\beta_D + \beta_{RQ} RQ_t^{1/2}\right) RV_{t,D} + Z_t + \varepsilon_{t+1}, \tag{4}$$

where Z_t is determined by (2). Next, we propose a new class of HAR models with conditional volatility persistence.

2.2. HAR models with conditional volatility persistence

The dynamic nature of daily volatility persistence is impacted by market conditions and macro-economic shocks and is captured by the so-called conditional volatility persistence (CVP). Empirical evidence in equity markets suggests that daily volatility persistence depends on observed market conditions, such as returns and volatility, see Wang and Yang (2018). The net price impact of an information event, following positive and negative re-turns, triggers the arrival of correlated information the next day which leads to an increase in volatility persistence. A high-volatility market environment (high daily RV) is a signal that the market processes new information more rapidly, and there is less unpriced information thus induces a reduction in volatility persistence the next day (Andersen (1996)). Accordingly, daily volatility persistence depending on returns and volatility, can be formulated by the following HAR-CVP model:

[10] In the seminal work by Corsi and Renò (2012) and subsequent work, the leverage effect is captured by the negative (daily, weekly and monthly) returns. It is well documented that in equity markets the impact of negative returns on future volatility is more pronounced than positive returns, which motivates this representation. However, energy markets are known to react to either large negative or positive returns (Silvennoinen and Thorp (2013) and Baur and Dimpfl (2018)). This study aims to identify the impact of positive and negative returns on future RV. Accordingly, we consider a variation of the classical HAR model proposed in the literature by including the negative and positive daily returns (Wang and Yang (2018)).

[11] Non-overlapping averages allow to isolate the impact of $RV_{t,D}$ on $RV_{t+1,D}$.

[12] If we let $\theta_{t,D}^- = NSV_{t,D}/RV_{t,D}$, then model (3) can be rewritten as $RV_{t+1,D} = \alpha + [\beta_{NSV}\theta_{t,D}^- + \beta_{PSV}(1-\theta_{t,D}^-)]RV_{t,D} + Z_t + \varepsilon_{t+1}$. Then, the daily volatility persistence of the HAR-SV model is expressed as $\beta_{NSV}\theta_{t,D}^- + \beta_{PSV}(1-\theta_{t,D}^-) = \beta_{PSV} + (\beta_{NSV} - \beta_{PSV})\theta_{t,D}^-$. Based on Patton and Sheppard (2015)'s findings for equity markets, $\beta_{NSV} \gg \beta_{PSV} > 0$, thus high $\theta_{t,D}^-$ will lead to high daily volatility persistence.

$$RV_{t+1,D} = \alpha + CVP_t RV_{t,D} + Z_t + \varepsilon_{t+1}, \tag{5}$$

where

$$CVP_t = \beta_D + \beta_r^- r_{t,D}^- + \beta_r^+ r_{t,D}^+ + \beta_{RV} RV_{t,D}, \tag{6}$$

with β_r^-, β_r^+, and β_{RV} as the CVP coefficients and Z_t given again by (2). Note that the HAR-CVP model also includes non-overlapping long-run variances ($RV_{t,W}$ and $RV_{t,M}$) and returns ($r_{t,W}$ and $r_{t,M}$). Wang and Yang (2018) show that in equity markets, the coefficients $\beta_r^- < 0$, $\beta_r^+ > 0$, and $\beta_{RV} < 0$, inferring a positive relation between daily returns and volatility persistence, and a negative relation between volatility and volatility persistence.

We hypothesize that volatility persistence is conditional not only to market conditions such as returns and volatility, but also to macro-economic factors. Thus, we extend the afore-mentioned concept of the CVP to include financial indicators, such as S&P 500 returns and credit spreads, and energy sector variables, such as the Baltic exchange dirty tanker index. The information of these macro-economic factors is represented by the set of conditioning variables CV_t. To identify the information transmission channels of these macro-economic factors, we further allow the set of conditioning variables to impact the dynamics of the volatility in two ways: by their direct effect on the daily realized variance and their indirect effect via the volatility persistence. Thus, this model formulation would not only identify the macro-economic determinants of daily volatility persistence in energy markets but also disentangle the information channels of these macro-economic factors on future volatility. Accordingly, we propose the following extension of model (5)–(6) to incorporate impact from economic factors, namely the HAR-CVP-CV model, which is estimated as

$$RV_{t+1,D} = \alpha + CVP_t RV_{t,D} + Z_t + \delta_{CV} CV_t + \varepsilon_{t+1}, \tag{7}$$

where the CVP (6) is extended and estimated by

$$CVP_t = \beta_D + \beta_r^- r_{t,D}^- + \beta_r^+ r_{t,D}^+ + \beta_{RV} RV_{t,D} + \beta_{CV} CV_t, \tag{8}$$

With Z_t be represented in (2). The CVP now contains the market conditions variables $r_{t,D}^-$, $r_{t,D}^+$, $RV_{t,D}$, as well as the macro-economic conditioning variables CV_t. Thus, the corresponding regressors of CVP are $r_{t,D}^- RV_{t,D}$, $r_{t,D}^+ RV_{t,D}$, $RV_{t,D}^2$ and $CV_t RV_{t,D}$, respectively. Note that, δ_{CV} and β_{CV} represent vectors of the coefficients for the corresponding macro-economic conditioning variables in the daily realized variance dynamics and in its volatility persistence, respectively. This representation provides the flexibility to assess the direct impact of the conditioning variables on the realized variance and the contribution of the conditional volatility persistence on transmitting the (indirect) impact of the conditioning variables to the realized variance.

The HAR models can be estimated with ordinary least square (OLS), as in Corsi (2009) and Corsi and Ren'o (2012). However, daily RV exhibits frequent spikes, see Fig. 1. As pointed out by Patton and Sheppard (2015), these spikes tend to have a large influence on the estimated coefficients in the OLS estimation. Accordingly, we estimate the HAR models with weighted least squares (WLS), with weights the inverse of the fitted values of the error standard deviations, retrieved from the OLS estimation. Appendix A presents the details.

3. Data and preliminary analysis

We study the volatility persistence of two key energy markets, crude oil and natural gas, from January 2009 to August 2019.[13] We employ the corresponding energy prices and a set of conditioning variables to estimate the daily realized variances using the different specifications of volatility persistence.

3.1. Energy futures

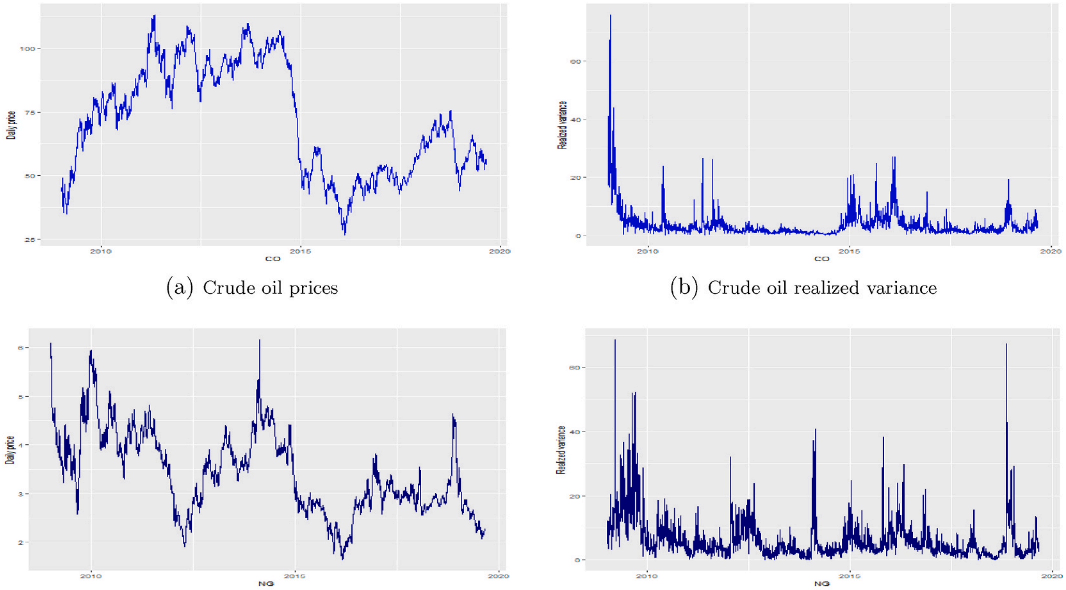
We use the prices of the nearest continuous futures contracts of the two energy commodities traded on the New York Mercantile Exchange (NYMEX). The daily RV, a measure of ex-post volatility, is constructed using the mid-quotations prices (average of.

The figure plots the time series of the daily prices and the realized variances of the crude oil and natural gas front-month futures contracts between January 2009 and August 2019. RV is scaled by 10^4.

Bid and ask prices) sampled at 5-min intervals.[14] All futures prices are collected from Thomson Reuters Tick History (TRTH). The data filter process for reducing the thin trading bias (Bollerslev et al. (2018)) is outlined in Appendix B.

[13] We decided to start our analysis from 2009 based on liquidity considerations for energy contracts, see also Bissoondoyal-Bheenick et al. (2020). The energy contracts were thinly traded (at a high-frequency level) prior to this period. Since our daily volatility measure (RV) is constructed using 5-min returns, we require that most contracts be traded over a 5-min window (with no major gaps in between). In Appendix B (Figure B1), we demonstrate that inclusion of thinly trading days can significantly overstate the average daily RV (four times more than its actual value), particularly for natural gas contracts. Similarly, during periods of low liquidity, we observe that the correlation between squared return (another proxy for daily volatility) and daily RV measure is weak. Hence, discarding the low liquidity periods removes noise in the proposed volatility proxy.

[14] In the literature, the consensus is to aggregate returns into 5-min intervals as they usually provide the best RV approximation (Liu et al. (2015)).



(a) Crude oil prices (b) Crude oil realized variance

(c) Natural gas prices (d) Natural gas realized variance

Fig. 1. Daily prices and realized variances.

We measure the daily volatility of the two energy markets by using the RV based on 5-min returns.[15] Some energy markets display seasonality and failing to account for it may distort the pricing, hedging and forecasting performance of the associated models (Suenaga et al. (2008); Mart'ınez and Torr'o (2015); Arismendi et al. (2016)). In spirit of Auer et al. (2014), we test for seasonality in the RV time series using the Kruskal and Wallis (1952) (KW) test and confirm that there is a day of the week and month of the year effect in natural gas volatility, but not in crude oil volatility.[16]

Accordingly, we deseasonalize the natural gas RV time series based on the approach used by Hameed et al. (2010) and Wang (2013). This approach ensures that the adjusted and original RV time series mean and variance remain unchanged (i.e. they share the same statistical properties). Appendix C summarises the details of the seasonality adjustments and tests.

The time series of the daily prices and the RV of the two energy markets are depicted in Fig. 1. Between 2009 and 2014, energy prices trended upward, on average, as the economic recovery from the Global Financial Crisis and increasing demand from emerging economies drove oil prices up to $125 per barrel. However, by the end of 2014, oil and natural gas prices plummeted driven by the slower economic growth of emerging countries and the rapid expansion of shale markets that led to a global oversupply of oil (oil glut). The impact of the oil glut was evident not only in the prices but also in the RV series of the crude oil market. Interestingly, the RV time series reveal that natural gas had more clusters of volatility over the years compared to the other commodities. In recent years, natural gas markets are not correlated to oil markets, which is evident from the price dynamics of these markets in Fig. 1.[17] Unlike the oil market, natural gas prices are mainly determined by supply and demand controlled by weather and production levels, with less impact from economic growth conditions. Natural gas production has increased dramatically from shale drilling, keeping natural gas prices low in recent years, with the occasional spikes driven mostly by extreme weather conditions, such as the one in November 2018.[18]

3.2. Macro-economic variables

We hypothesize that general macro-economic factors, such as financial indicators and commodity sector variables, can affect the daily volatility persistence in energy markets and transmit this impact to the volatility of these markets. To capture these effects, we use HAR models for the RV of energy markets that integrate the direct impact of macro-economic factors on volatility, but also gauge their indirect impact on volatility via (conditional) volatility persistence. We term these factors conditioning variables, and we consider two

[15] Alternative RV measures are also considered, namely, the realized quad-power quarticity (RQ), the negative semi-variance (NSV) and the positive semi-variance (PSV). See Appendix D for a discussion of their statistical properties.

[16] Crude oil markets display significant volatility spillovers and integration with equity markets, which may distort effects driven by potential oil market seasonal factors such as inventory (Cheng and Xiong (2014), Chiang et al. (2015), Basak and Pavlova (2016), Bampinas and Panagiotidis (2017), and Kang et al. (2020)).

[17] Following the liberation of natural gas prices from oil indexation, in recent years there is strong empirical evidence of the decoupling of crude oil and natural gas prices (particularly in United States) (Geng, Ji, and Fan (2016a) and Zhang and Ji (2018)). The shale gas evolution has further affected the relation betweenthe two markets (Geng, Ji, and Fan (2016b) and Caporin and Fontini (2017)).

[18] See https://www.eia.gov/todayinenergy/detail.php?id=37713.

Table 1
Descriptive statistics of conditioning variables.

	Mean	St. Dev.	Median	Skewness	Exc.Kurtosis	min	max	ADF
Financial Indicators								
SP500(%)	0.045	1.046	0.063	0	5	−6.896	6.837	14.466***
VIX	18.377	7.429	16.16	2	3	9.14	56.65	−4.495***
USDI	4.401	0.101	4.371	0	−2	4.22	4.573	−2.742
CS	1.064	0.441	0.96	3	9	0.53	3.35	−4.344***
TB	0.491	0.744	0.13	2	1	0	2.49	2.822
TS	2.019	0.918	2.01	0	0	−0.52	3.83	−2.583
Commodity sector variables								
SPGSCI	5.888	0.152	5.838	0	−1	5.551	6.245	−2.596
CRB	5.492	0.149	5.445	−1	1	5.154	5.864	−2.679
BDI	7.037	0.846	7.039	−5	40	0	8.447	−5.338***

This table details the descriptive statistics of the conditioning variables: SP500, VIX, USDI, CS, TB, TS, SPGSCI, CRB, and BDI. Based on results of the ADF test, we take the first difference of the USDI, TB, TS, SPGSCI and CRB individual time series.

groups: financial indicators and commodity sector variables. These factors are drawn by the empirical literature identifying key determinants of returns and volatility in energy markets.[19] Volatility is the main channel of volatility persistence, thus potentially they share the same determinants. However, we further aim to determine the importance of volatility persistence in transmitting this impact to volatility dynamics.

The first group of (volatility and) volatility persistence determinants is financial indica-tors. We consider the following five financial indicators: the S&P 500 return (SP500), the VIX, the US dollar index (USDI), the credit spread (CS), the 3-month Treasury bill (TB) and the term spread (TS). The second group of conditioning variables is the commodity sector variables, including the S&P Goldman Sachs Commodity Index (SPGSCI), the Commodity Research Bureau (CRB) raw materials index (CRB) and the Baltic exchange dirty tanker index (BDI).[20] We report the summary statistics of the conditioning variables in Table 1. Based on results of the Augmented Dickey-Fuller (ADF) test, we take the first difference of the USDI, TB, TS, SPGSCI and CRB individual time series.

4. The role of volatility persistence in energy markets

In this section, we identify the determinants and gauge the importance of volatility persistence in energy markets. First, we identify the macro-economic determinants of volatility persistence, and we discuss the role of volatility persistence in transmitting macro-economic information in volatility. Then, we re-assess the role of two well-known determinants of volatility persistence, namely returns and variances. We also examine the economic significance of volatility persistence by evaluating the contribution of CVP determinants to volatility variation and by detecting the drivers of the CVP variation. Last, we discuss the statistical properties of the CVP.

4.1. Macro-economic information and volatility persistence

The impact of the macro-economic variables on future volatility and daily volatility persistence can be assessed by estimating the model HAR-CVP-CV (see equations (7) and (8)). Table 2 and Table 3 present the estimation results of this model in the crude oil and natural gas markets, respectively.[21]

We find that in oil markets, VIX and credit spreads are significant financial determinants of daily volatility persistence, and the Baltic exchange dirty tanker index is one of the commodity sector determinants of volatility persistence, see Table 2. Specifically, the VIX impacts future oil volatility in two ways; directly by decreasing volatility levels and indirectly by increasing volatility persistence and consequently increasing volatility. Correlated information, driven by a shock in the VIX, tends to arrive within a short period of

[19] See Morana (2013), Anzuini et al. (2015), Hitzemann (2016), Prokopczuk et al. (2019) and Kang et al. (2020).

[20] The data sources and the role of each determinant of energy markets dynamics are discussed with associated literature in Appendix E. Other macro-economic factors could have been included in this study, such as inventory, hedging pressure and industrial production. However, these factors are available on weekly or even monthly frequency. For this study, we focus on factors available in daily frequency. Inventory may have medium-to long-term effects on weekly/monthly volatility (Kogan et al. (2009), Haugom, Langeland, Moln'ar, and Westgaard (2014) and Nikitopoulos et al. (2017)). However, short-term volatility, e.g daily volatility computed from high-frequency data, tends to be more sensitive to macro-economic factors (Kang et al. (2020)). Energy inventory data are available on weekly frequency, and a price-based proxy could have been used to counteract for the difference in the frequency (Ng and Pirrong (1994); Robe and Wallen (2016); Bruno et al. (2017)). Empirical evidence though shows that even though inventories and futures interest-adjusted spreads are highly associated, their interactions with volatility may vary (Nikitopoulos et al. (2017)). Thus, a price-based estimate of inventory may not represent a robust proxy. Accordingly, we do not control for inventories (or a price-based proxy of inventories).

[21] Even though some of the control variables are highly correlated, their coefficients have the appropriate sign and plausible magnitude and are statistically significant. Accordingly, multicollinearity should not be a major concern, see page 173 of Brooks (2008). Further, the Ljung-box statistics for serial correlation in the residuals for the HAR-CVP-CV models validates that missing variables bias is not present in our model specifications, see the last two rows in Tables 2 and 3.

Table 2
The HAR-CVP-CV models for crude oil.

α	1.45E-05***	2.95E-05***	1.46E-05***	1.86E-05***	1.45E-05***	1.38E-05***	1.34E-05***	1.37E-05***	1.26E-05*	1.95E-05**
	(4.351)	(3.554)	(3.963)	(2.886)	(4.250)	(4.082)	(3.980)	(3.973)	(1.662)	(2.014)
β_D	0.436***	0.420***	0.438***	0.487***	0.435***	0.442***	0.429***	0.431***	0.562***	0.630***
	(14.176)	(11.551)	(14.339)	(14.591)	(14.084)	(14.477)	(14.126)	(14.143)	(8.073)	(9.089)
β_r^-	−0.136***	−0.143***	−0.152***	−0.153***	−0.156***	−0.134***	−0.139***	−0.142***	−0.159***	−0.117***
	(-3.974)	(-4.141)	(-3.967)	(-3.991)	(-4.272)	(-3.759)	(-4.213)	(-4.359)	(-4.178)	(-3.583)
β_r^+	0.050	0.050	0.055	0.051	0.051	0.055	0.056*	0.057	0.054	0.059*
	(1.426)	(1.547)	(1.511)	(1.528)	(1.504)	(1.609)	(1.646)	(1.633)	(1.534)	(1.948)
β_{RV}	−0.006**	−0.009***	−0.006**	−0.006**	−0.006**	−0.006**	−0.006**	−0.006**	−0.007**	−0.011***
	(-2.064)	(-2.944)	(-2.172)	(-2.023)	(-2.093)	(-2.124)	(-2.221)	(-2.328)	(-2.361)	(-3.332)
β_W	0.307***	0.291***	0.298***	0.288***	0.300***	0.305***	0.307***	0.306***	0.300***	0.292***
	(11.708)	(10.413)	(11.196)	(10.914)	(11.103)	(11.475)	(11.353)	(11.382)	(11.450)	(10.140)
β_M	0.159***	0.155***	0.160***	0.170***	0.159***	0.154***	0.163***	0.162***	0.155***	0.162***
	(6.680)	(6.165)	(6.892)	(6.816)	(6.604)	(6.552)	(6.657)	(6.665)	(6.541)	(6.846)
θ_D^-	−0.178*	−0.231**	−0.204*	−0.198*	−0.202*	−0.222**	−0.211**	−0.195*	−0.196*	−0.218**
	(-1.733)	(-2.277)	(-1.852)	(-1.853)	(-1.934)	(-2.114)	(-2.082)	(-1.928)	(-1.826)	(-2.160)
θ_D^+	0.027	0.019	0.002	0.012	0.009	0.019	0.039	0.031	0.012	0.059
	(0.280)	(0.215)	(0.016)	(0.124)	(0.101)	(0.199)	(0.418)	(0.324)	(0.128)	(0.659)
θ_W	−0.153**	−0.145**	−0.135**	−0.146**	−0.136**	−0.135**	−0.144**	−0.137**	−0.137**	−0.170**
	(-2.376)	(-2.285)	(-2.155)	(-2.361)	(-2.116)	(-2.104)	(-2.227)	(-2.120)	(-2.116)	(-2.734)
θ_M	−0.187	−0.206	−0.225	−0.134	−0.209	−0.206	−0.200	−0.196	−0.206	−0.172
	(-1.294)	(-1.474)	(-1.536)	(-0.935)	(-1.468)	(-1.465)	(-1.419)	(-1.391)	(-1.430)	(-1.265)
β_{SP500}	−1.491									−0.138
	(-1.113)									(-0.113)
β_{VIX}		0.004*								0.009***
		(1.805)								(2.837)
β_{USDI}			3.732							2.059
			(1.189)							(0.607)
β_{CS}				−0.029						−0.098**
				(-1.198)						(-2.468)
β_{TB}					0.227					−0.035
					(0.236)					(-0.036)
β_{TS}						−0.500**				−0.290
						(-2.704)				(-1.347)

(*continued on next page*)

Table 2 (*continued*)

β_{SPGSCI}							−0.028			−0.049
							(-1.460)			(-1.056)
β_{CRB}								−0.025		0.047
								(-1.187)		(0.892)
β_{BDI}									−0.017*	−0.028***
									(-1.820)	(-2.834)
δ_{SP500}	−4.96E-04									−4.25E-04
	(-1.503)									(-1.249)
δ_{VIX}		−1.28E-06**								−1.77E-06**
		(-2.091)								(-2.166)
δ_{USDI}			−1.06E-03							−1.04E-03
			(-1.445)							(-1.267)
δ_{CS}				−9.67E-06						9.72E-06
				(-1.167)						(0.904)
δ_{TB}					5.73E-05					4.69E-05
					(0.331)					(0.256)
δ_{TS}						−2.59E-05				−3.62E-05
						(-0.549)				(-0.588)
δ_{SPGSCI}							−1.84E-04			4.79E-04
							(-0.423)			(0.524)
δ_{CRB}								−2.76E-04		−1.07E-03
								(-0.618)		(-1.061)
δ_{BDI}									1.67E-07	1.02E-06
									(0.150)	(0.872)
adj R^2	0.675	0.672	0.671	0.670	0.673	0.676	0.674	0.674	0.672	0.679
AIC	−40,486	−40,491	−40,454	−40,501	−40,470	−40,493	−40,492	−40,492	−40,469	−40,542
LB(Q)*	5.04**	1.49	5.46**	7.42***	4.82**	7.01***	5.13**	4.56**	4.53**	3.41*
	(0.02)	(0.22)	(0.02)	(0.01)	(0.03)	(0.01)	(0.02)	(0.03)	(0.03)	(0.06)

This table reports the estimation results of the following regressions in the crude oil: $RV_{t+1,D} = \alpha + (\beta_D + \beta_r^- r_{t,D}^- + \beta_r^+ r_{t,D}^+ + \beta_{RV} RV_{t,D} + \beta_{CV} CV_{t,D}) RV_{t,D} + Z_t + \delta_{CV} CV_{t,D} + \varepsilon_{t+1,D}$, where Z_t is defined by (2). The t-statistic (in parentheses) is estimated using the Newey–West standard errors. AIC is the Akaike information criteria. LB(Q*) is the Ljung-Box test statistics and its p-value is reported in parentheses. *, **, *** denotes the 10%, 5% and 1% level of significance, respectively.

Table 3
The HAR-CVP-CV models for natural gas.

α	2.76E-05***	5.75E-05**	2.74E-05**	8.61E-05**	2.93E-05***	2.79E-05**	2.70E-05**	2.78E-05**	1.93E-05	7.34E-05*
	(3.022)	(2.458)	(2.763)	(2.645)	(3.153)	(2.796)	(2.697)	(2.758)	(0.492)	(1.884)
β_D	00.414***	0.346***	0.416***	0.311***	0.407***	0.414***	0.416***	0.413***	0.385***	0.281***
	(8.812)	(5.038)	(9.425)	(4.390)	(9.073)	(9.527)	(9.205)	(9.172)	(4.273)	(3.122)
β_r^-	−0.077***	−0.075***	−0.078***	−0.077***	−0.079***	−0.077***	−0.075***	−0.075***	−0.075***	−0.071***
	(-3.532)	(-3.443)	(-3.427)	(-3.559)	(-3.610)	(-3.392)	(-3.327)	(-3.308)	(-3.406)	(-3.361)
β_r^+	0.070***	0.067***	0.068***	0.067***	0.068***	0.066***	0.067***	0.067***	0.067***	0.068***
	(3.302)	(3.308)	(3.435)	(3.377)	(3.397)	(3.227)	(3.324)	(3.251)	(3.414)	(3.653)
β_{RV}	−0.003**	−0.005**	−0.003**	−0.005**	−0.003*	−0.003**	−0.003**	−0.003*	−0.003**	−0.005**
	(-2.004)	(-2.799)	(-2.341)	(-2.517)	(-1.935)	(-2.085)	(-2.011)	(-1.859)	(-2.106)	(-2.428)
β_W	0.314***	0.308***	0.313***	0.318***	0.312***	0.311***	0.311***	0.311***	0.311***	0.322***
	(8.293)	(8.381)	(8.250)	(8.124)	(8.256)	(8.430)	(8.268)	(8.289)	(8.298)	(8.365)
β_M	0.145***	0.144***	0.145***	0.137***	0.146***	0.145***	0.145***	0.145***	0.145***	0.131***
	(5.626)	(5.240)	(4.485)	(4.279)	(4.981)	(4.491)	(4.527)	(4.514)	(4.838)	(4.685)
θ_D^-	0.041	0.027	0.044	0.035	0.052	0.038	0.017	0.022	0.027	−0.021
	(0.447)	(0.295)	(0.477)	(0.392)	(0.575)	(0.423)	(0.189)	(0.244)	(0.305)	(-0.229)
θ_D^+	0.296**	0.319**	0.310**	0.304**	0.322**	0.322**	0.319**	0.323**	0.319**	0.300**
	(2.194)	(2.310)	(2.460)	(2.210)	(2.211)	(2.490)	(2.484)	(2.477)	(2.325)	(2.462)
θ_W	−0.142	−0.145	−0.148	−0.199	−0.142	−0.148	−0.138	−0.141	−0.132	−0.150
	(-1.219)	(-1.216)	(-1.184)	(-1.381)	(-1.125)	(-1.184)	(-1.113)	(-1.123)	(-1.083)	(-1.190)
θ_M	−0.122	−0.153	−0.102	−0.166	−0.099	−0.112	−0.136	−0.107	−0.091	−0.263
	(-0.589)	(-0.726)	(-0.515)	(-0.783)	(-0.489)	(-0.573)	(-0.701)	(-0.548)	(-0.423)	(-1.276)
β_{SP500}	−0.644									−0.591
	(-0.373)									(-0.368)
β_{VIX}		0.005								0.001
		(1.481)								(0.191)
β_{USDI}			−3.047							−2.789
			(-1.196)							(-0.950)
β_{CS}				0.131*						0.096
				(1.936)						(1.430)
β_{TB}					3.039***					3.162***
					(2.813)					(2.896)
β_{TS}						−0.227				−0.184
						(-0.707)				(-0.531)
β_{SPGSCI}							−0.005			0.044

(*continued on next page*)

Table 3 (*continued*)

							(-0.264)			(0.791)
β_{CRB}								−0.010		−0.056
								(-0.513)		(-0.933)
β_{BDI}									0.004	0.007
									(0.336)	(0.675)
δ_{SP500}	−1.10E-04									−3.93E-04
	(-0.128)									(-0.505)
δ_{VIX}		−2.13E-06								−8.09E-07
		(-1.463)								(-0.556)
δ_{USDI}			1.28E-03							1.20E-03
			(0.988)							(0.879)
δ_{CS}				−6.93E-05*						−4.56E-05
				(-1.911)						(-1.293)
δ_{TB}					−1.02E-03***					−1.11E-03***
					(-3.481)					(-3.548)
δ_{TS}						6.08E-05				5.70E-05
						(0.558)				(0.413)
δ_{SPGSCI}							1.21E-03**			1.94E-04
							(2.016)			(0.113)
δ_{CRB}								1.18E-03*		1.35E-03
								(1.839)		(0.728)
δ_{BDI}									1.21E-06	7.03E-07
									(0.211)	(0.176)
adj R^2	0.568	0.568	0.569	0.564	0.574	0.567	0.572	0.569	0.568	0.570
AIC	−34,892	−34,911	−34,893	−34,874	−34,923	−34,899	−34,909	−34,900	−34,901	−34,938
LB(Q)*	2.34	0.95	2.91*	0.01	2.34	2.18	2.29	2.40	2.58	1.03
	(0.13)	(0.33)	(0.09)	(0.91)	(0.13)	(0.14)	(0.13)	(0.12)	(0.11)	(0.31)

This table reports the estimation results of the following regressions for the natural gas: $RV_{t+1,D} = \alpha + (\beta_D + \beta_r^- r_{t,D}^- + \beta_r^+ r_{t,D}^+ + \beta_{RV} RV_{t,D} + \beta_{CV} CV_{t,D}) RV_{t,D} + Z_t + \delta_{CV} CV_{t,D} + \varepsilon_{t+1,D}, \neq$ where Z_t is defined by (2). The t-statistic (in parentheses), is estimated using the Newey–West standard errors. AIC is the Akaike information criteria. LB(Q*) is the Ljung-Box test statistics and its p-value is reported in parentheses. *, **, *** denotes the 10%, 5% and 1% level of significance, respectively.

Table 4
Contribution of the CVP-CV determinants to volatility variation (shapley R^2).

	SP500	VIX	USDI	CS	TB	TS	SPGSCI	CRB	BDI
Panel A: HAR-CVP-CV model for Crude Oil									
RV_t	18.10%	16.43%	18.94%	17.02%	19.12%	18.35%	18.40%	18.32%	15.63%
$r^-_{t,D}RV_{t,D}$	2.68%	3.37%	3.74%	3.63%	3.95%	3.39%	3.33%	3.19%	3.38%
$r^+_{t,D}RV_{t,D}$	3.47%	1.93%	2.61%	1.89%	2.37%	2.90%	2.65%	2.64%	2.41%
$RV^2_{t,D}$	9.56%	8.65%	10.36%	8.53%	10.46%	9.70%	10.00%	9.89%	10.09%
$CV_{t,D}RV_{t,D}$	2.73%	7.47%	0.45%	7.28%	0.11%	2.78%	1.52%	1.86%	7.06%
$\widehat{CVP}_{t,D}$	18.44%	21.42%	17.16%	21.32%	16.90%	18.76%	17.49%	17.57%	22.94%
$CV_{t,D}$	1.03%	0.96%	0.04%	0.77%	0.05%	0.52%	0.24%	0.34%	1.11%
$RV_{t,W} + RV_{t,M}$	25.99%	24.27%	26.82%	23.98%	26.94%	25.08%	26.80%	26.58%	24.32%
$r^-_{t,D} + r^+_{t,D}$	3.93%	3.49%	3.63%	3.38%	3.77%	4.20%	3.81%	4.13%	2.86%
$r_{t,W} + r_{t,M}$	0.00%	0.59%	0.50%	0.57%	0.47%	0.67%	0.62%	0.48%	0.34%
R^2	67.50%	67.16%	67.08%	67.03%	67.26%	67.59%	67.36%	67.42%	67.20%
$\widehat{CVP}_{t,D}/R^2$	27.33%	31.89%	25.58%	31.81%	25.13%	27.76%	25.97%	26.05%	34.13%
Panel B: HAR-CVP-CV model for Natural Gas									
RV_t	17.06%	13.17%	17.07%	13.01%	16.89%	17.02%	17.14%	17.00%	12.57%
$r^-_{t,D}RV_{t,D}$	2.65%	2.14%	2.69%	1.92%	2.71%	2.68%	2.58%	2.36%	2.17%
$r^+_{t,D}RV_{t,D}$	3.15%	2.56%	3.17%	2.71%	3.20%	3.16%	3.12%	2.98%	2.69%
$RV^2_{t,D}$	9.41%	7.28%	9.44%	7.34%	9.43%	9.42%	9.34%	9.33%	8.39%
$CV_{t,D}RV_{t,D}$	0.04%	7.83%	0.08%	7.56%	0.34%	0.02%	0.38%	0.23%	8.20%
$\widehat{CVP}_{t,D}$	15.25%	19.80%	15.38%	19.53%	15.68%	15.28%	15.41%	14.91%	21.45%
$CV_{t,D}$	0.04%	0.65%	0.01%	0.11%	0.17%	0.01%	0.27%	0.01%	0.92%
$RV_{t,W} + RV_{t,M}$	22.09%	20.89%	22.00%	21.27%	22.06%	21.95%	21.92%	22.77%	19.60%
$r^-_{t,D} + r^+_{t,D}$	2.31%	1.73%	2.34%	2.00%	2.47%	2.34%	2.33%	2.08%	2.20%
$r_{t,W} + r_{t,M}$	0.04%	0.54%	0.06%	0.51%	0.10%	0.06%	0.08%	0.15%	0.03%
R^2	56.79%	56.77%	56.86%	56.43%	57.37%	56.67%	57.15%	56.91%	56.77%
$\widehat{CVP}_{t,D}/R^2$	26.85%	34.88%	27.04%	34.61%	27.33%	26.97%	26.96%	26.20%	37.78%

This table reports the Shapley decomposition of the regression R^2 in the HAR-CVP-CV models in the crude oil and natural gas market: $RV_{t+1,D} = \alpha + (\beta_D + \beta_r^- r^-_{t,D} + \beta_r^+ r^+_{t,D} + \beta_{RV}RV_{t,D} + \beta_{CV}CV_{t,D})RV_{t,D} + Z_t + \delta_{CV}CV_{t,D} + \varepsilon_{t+1,D}$, with $\widehat{CVP}_{t,D} = r^-_{t,D}RV_{t,D} + r^+_{t,D}RV_{t,D} + RV^2_{t,D} + CV_{t,D}RV_{t,D}$, and Z_t is defined by (2).

time triggering volatility clustering, and thus increased volatility persistence. As the impact of the indirect channel via volatility persistence is far more pronounced than the direct one, VIX would positively affect oil volatility. Following the financialization of commodity markets, volatility spillovers and co-movements between oil and equity markets are well documented (Tang and Xiong (2012); Basak and Pavlova (2016)) providing compelling evidence of a positive relation between VIX and oil market volatility (Silvennoinen and Thorp (2013), Kang et al. (2020) and Wang et al. (2020)). Credit spreads have also been identified as an important predictor of crude oil volatility (Hitzemann (2016) and Prokopczuk et al. (2019)). Worsening of credit conditions in the economy implies higher inflation compensation and volatility (Chen et al. (2014)). As a sudden increase in credit spreads would increase oil market volatility, this gross information flow over a short period of time (e.g., a day) is associated with more priced and thus less correlated information, reflecting a drop in volatility persistence (Hasbrouck (1995) and Andersen (1996)). This justifies the negative relation between credit spreads and oil volatility persistence. The Baltic exchange dirty tanker index - which represents the physical factor that influences the supply of energy commodities - tends to decrease volatility persistence. Empirical literature suggests that BDI and future volatility are positively related (Breitenfellner et al. (2009), Kilian (2009) and Fan and Xu (2011)), and this impact is transmitted to volatility via a reduction in the oil market's volatility persistence.

The Treasury bill, a proxy for monetary policy, is the only financial factor affecting volatility persistence in the natural gas market. Indeed, the indirect impact of the Treasury bill via the volatility persistence channel implies that the Treasury bill increases volatility persistence, i.e., the CVP_t variable, which subsequently indicates that the impact of volatility persistence on realized volatility is positive. There is also a direct negative impact of Treasury bills on volatility levels. Typically, low interest rates lead to high demand for inventories, which among other economic channels, puts upward pressure on future prices and potentially, volatility in energy markets (Arora and Tanner (2013); Frankel (2014); Cheng, Nikitopoulos, and Schlögl (2018); Kang et al. (2020)). Based on the argument that high volatility reflects better priced and thus less correlated information, a reduction in volatility persistence is stipulated. Another financial variable that directly affects the daily volatility levels in natural gas is the credit spreads, a well-accepted predictor of volatility in energy markets (Table 3). Conversely to the oil markets, commodity sector variables such as the S&P GSCI Non-Energy index and CRB Raw Materials Index have a direct influence on the daily volatility in the natural gas market.

Empirical studies demonstrate that financial indicators and commodity sector variables play an important role in determining volatility in energy markets (Robe and Wallen (2016), Prokopczuk et al. (2019) and Kang et al. (2020)). We find that the impact of

Table 5
The HAR and HAR-CVP models.

	Crude Oil		Natural Gas	
	HAR	HAR-CVP	HAR	HAR-CVP
Panel A: Model estimation				
α	9.79E-06***	1.45E-05***	1.37E-05*	2.76E-05**
	(3.265)	(4.276)	(1.957)	(2.716)
β_D	0.435***	0.436***	0.453***	0.415***
	(18.234)	(13.970)	(12.439)	(9.584)
β_r^-		−0.156***		−0.077***
		(-4.331)		(-3.400)
β_r^+		0.052		0.067***
		(1.531)		(3.309)
β_{RV}		−0.006**		−0.003*
		(-2.124)		(-2.208)
β_W	0.301***	0.299***	0.309***	0.311***
	(11.161)	(10.759)	(8.511)	(8.452)
β_M	0.159***	0.159***	0.140***	0.145***
	(6.608)	(6.455)	(4.809)	(4.485)
θ_D^-	−0.665***	−0.198*	−0.260**	0.040
	(-7.453)	(-1.910)	(-2.952)	(0.442)
θ_D^+	0.155*	0.009	0.592***	0.317*
	(1.705)	(0.094)	(5.780)	(2.499)
θ_W	−0.116*	−0.135**	−0.127	−0.145
	(-1.787)	(-2.078)	(-1.055)	(-1.161)
θ_M	−0.189	−0.208	−0.101	−0.110
	(-1.304)	(-1.464)	(-0.500)	(-0.565)
R^2	0.668	0.672	0.563	0.568
AIC	−40,448	−40,465	−34,885	−34,898
Panel B: Shapley R^2				
$RV_{t,D}$	25.50%	19.13%	25.34%	17.07%
Δ%		−24.98%		−32.62%
$r_{t,D}^- RV_{t,D}$		2.34%		2.68%
$r_{t,D}^+ RV_{t,D}$		3.99%		3.17%
$RV_{t,D}^2$		10.49%		9.43%
$\widehat{CVP}_{t,D}$		16.83%		15.28%
$RV_{t,W} + RV_{t,M}$	33.62%	26.99%	27.40%	22.01%
Δ%		−19.73%		−19.68%
$r_{t,D}^- + r_{t,D}^+$	6.24%	3.73%	3.51%	2.35%
Δ%		−40.19%		−32.97%
$r_{t,W} + r_{t,M}$	1.43%	0.48%	0.08%	0.06%
Δ%		−66.54%		−21.44%
R^2	66.80%	67.16%	56.33%	56.77%
$\widehat{CVP}_{t,D}/R^2$		25.05%		26.91%

This table reports the estimation results of the following regressions in the crude oil and natural gas markets: **HAR** : $RV_{t+1,D} = \alpha + \beta_D RV_{t,D} + Z_t + \varepsilon_{t+1,D}$, **HAR − CVP** : $RV_{t+1,D} = \alpha + (\beta_D + \beta_r^- r_{t,D}^- + \beta_r^+ r_{t,D}^+ + \beta_{RV} RV_{t,D}) RV_{t,D} + Z_t + \varepsilon_{t+1,D}$, with $\widehat{CVP}_{t,D} = r_{t,D}^- RV_{t,D} + r_{t,D}^+ RV_{t,D} + RV_{t,D}^2$, and Z_t is defined by (2). Panel A details the estimation results for two regression models in the crude oil and natural markets, outlined above. The t-statistic (in parentheses) is estimated using the Newey–West standard errors. AIC is the Akaike information criteria. *, **, *** denotes the 10%, 5% and 1% level of significance, respectively. Panel B reports the Shapley decomposition of the regression R^2 in the HAR and HAR-CVP models. Δ% is the percentage change of the Shapley R^2 in the HAR-CVP model compared to the HAR model.

these economic determinants is transmitted to future oil market volatility via its volatility persistence channel (and not directly to the volatility levels). Natural gas is less integrated with the oil market, and volatility is mostly driven by fundamentals,[22] thus the impact of financial indicators on volatility is moderate (only via credit spreads and Treasury bills). Furthermore, most of the macro-economic factors impact volatility directly in natural gas markets.

[22] Geng et al. (2016a) find that supply and demand are the main determinants of natural gas prices in United States, while oil prices play a key role in determining natural gas prices in Europe and Japan.

Table 6
CVP variance decomposition.

	HAR-CVP		
	$w(r^-_{t,D})$	$w(r^+_{t,D})$	$w(RV_{t,D})$
Crude Oil			
Mean	87.1%	5.3%	7.5%
Median	86.7%	5.5%	7.8%
Min	76.0%	0.0%	0.1%
Max	99.9%	11.1%	15.4%
Natural Gas			
Mean	52.5%	38.4%	9.0%
Median	52.7%	38.6%	8.8%ss
Min	29.4%	15.1%	0.1%
Max	75.2%	61.3%	18.2%

This table reports the summary statistics of the weights $w(r^-_{t,D})$, $w(r^+_{t,D})$ and $w(RV_{t,D})$ in the HAR-CVP model across six permutations in the orthogonalization process, respectively. The weights represent the percentages of the variance of $CVP_{t,D}$ decompose into the variances of its orthogonalized components $r^-_{t,D}$, $r^+_{t,D}$ and $RV_{t,D}$ in the HAR-CVP model.

4.2. Impact of returns and past volatility

Two well-known determinants of volatility include returns and past RV; thus, we re-assess the impact of these two market conditions on volatility and the role of volatility persistence in transmitting this impact. We consider firstly the classical HAR model specifications (that infer constant volatility persistence (represented by β_D)) and then compare these with the conditional volatility persistence HAR models, namely HAR-CVP and HAR-CVP-CV models (see Section 2.2), in which the volatility persistence is modeled via the CVP that depends on the RV, (positive and negative) returns and macro-economic variables. Reflecting an in-sample estimation, Panel A of Table 5 displays the impact of returns and RV on the next day's volatility for the classical HAR (see equations (1) and (2)) and the HAR-CVP models. Tables 2 and 3 present the in-sample estimation results of the HAR-CVP-CV models in the crude oil and natural gas markets, respectively. We find two main results. First, we confirm that daily returns and RV are important drivers of future volatility. Daily RV increases the next day's volatility, and daily returns decrease future volatility in the HAR models. Second and most importantly, the conditional volatility persistence HAR models estimation reveals that daily returns and RV are also significant determinants of volatility persistence and play a prominent role in transferring the impact of returns and RV to future volatility.

When using HAR models (see the first and third columns of Table 5), crude oil future volatility is determined by weekly returns and (positive and negative) daily returns, with the impact of negative daily returns higher than the positive daily returns and the weekly returns. Past (positive and negative) daily returns are also key determinants of future volatility in natural gas markets, yet the impact of positive daily returns is more pronounced (compared to negative daily returns), a behavior reflecting the inverse leverage effect. This is consistent with the theory of storage and the impact of fundamental commodity supply and demand factors (Ng and Pirrong (1994) and Geng et al. (2016a)).[23] Nevertheless, in the last decade, energy markets have been very actively traded markets and popular investment vehicles, justifying the similarity of the behavior of the crude oil market to the equity markets (Chiang et al. (2015) and Basak and Pavlova (2016)). Further, in the energy markets, past (daily, weekly and monthly) RV is highly significant and positively related to the next day's volatility, with the impact of short-term volatilities more pronounced compared to the impact of monthly volatilities. Accordingly, short-term trading in energy markets seems to be more influential on future daily volatility compared to longer-term trading (Ma et al. (2018)).

The effects on volatility persistence are determined from the (in-sample) estimation of the HAR-CVP models reported in Panel A of Table 5 (see the second and forth columns). Daily returns and RV are important determinants of volatility persistence in energy markets. Negative returns affect volatility persistence in crude oil market, and negative returns increase future oil volatility directly and indirectly via the volatility persistence channel. In the natural gas market, positive[24] and negative returns determine volatility persistence, and their impact is transmitted to future volatility entirely via volatility persistence. A negative return of 1% leads to an average increase in daily volatility persistence of 28.6% in the crude oil market and 16.6% in the natural gas market.[25] Note that, the impact of positive returns on natural gas volatility persistence is marginally higher compared to the impact of negative returns. The RV

[23] In principle, low inventory (among other reasons and commodity shortage) drives commodity prices and volatility up, implying a positive relation between returns and volatility.

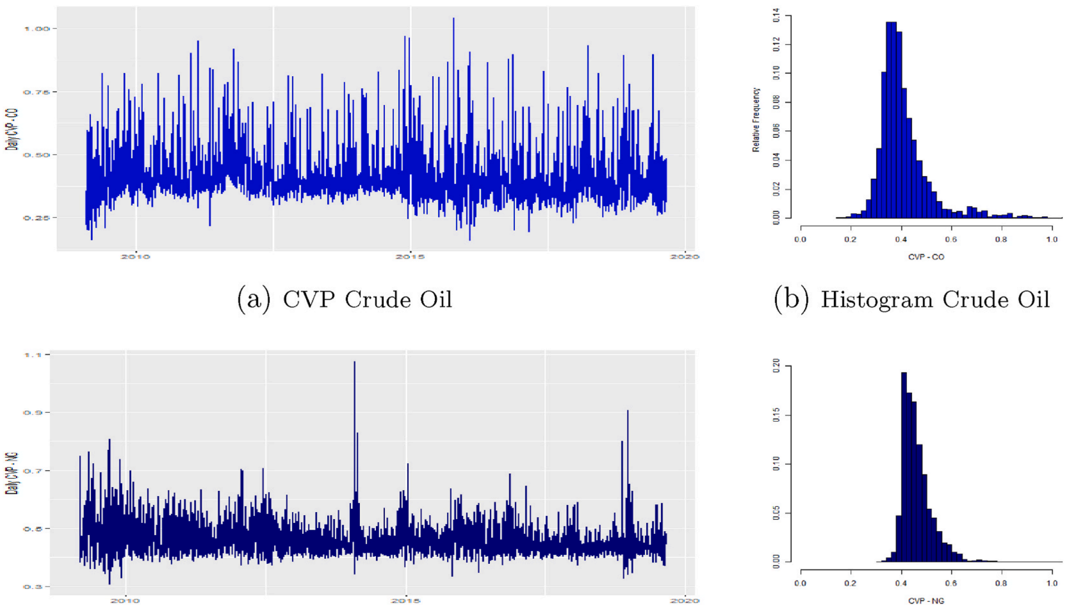
[24] The short-term effect of positive returns on natural gas volatility is consistent with the inverse leverage effect, thus for natural gas, fundamentals dominate in the short run.

[25] In Table 5, $\beta^-_r = -0.156$ for crude oil and -0.077 for natural gas, while in Table 7, the mean level of CVP is 0.546 for crude oil and 0.464 for natural gas. Accordingly, the average increase in daily volatility persistence is $0.156/0.546 = 28.6\%$ for crude oil and $0.077/0.464 = 16.6\%$ for natural gas.

Table 7
Statistical properties of CVP.

Panel A: Descriptive Statistics								
	Mean	St. Dev.	Median	Skewness	Exc. Kurtosis	Min	Max	$LB(1)$
Crude oil	0.546	0.085	0.521	2.071	7.384	0.373	1.336	400.111
Natural Gas	0.464	0.077	0.455	0.926	4.520	0.155	1.073	4.541
Panel B: Correlation								
	r_D^+	r_D^-	RV_D	VIX_D	CS_D	TB_D	BDI_D	
Crude oil	0.074***	−0.753***	0.429***	0.506***	0.160***	–	−0.238***	
Natural gas	0.349***	−0.488***	0.146***	–	–	0.617***	–	

The CVP is calculated as: $CVP_{t,D} = \beta_D + \beta_t^- r_{t,D}^- + \beta_t^+ r_{t,D}^+ + \beta_{RV} RV_{t,D} + \beta_{CV} CV_{t,D}$.This table reports the statistical properties of CVP in the crude oil and natural gas market. Panel A is the descriptive statistics of CVP in the energy markets. $LB(1)$ is the Ljung-Box test statistics at 1 lag. Panel B is the correlation between the CVP and individual components of the CVP. We only consider the conditional variables having a significant impact on future RV to construct $CVP_{t,D}$. *, **, *** denotes the 10%, 5% and 1% level of significance, respectively.



(a) CVP Crude Oil (b) Histogram Crude Oil

(c) CVP Natural Gas (d) Histogram Natural Gas

Fig. 2. Daily conditional volatility persistence (CVP) and histogram of CVP with conditioning variables.

matters to the volatility persistence of crude oil and natural gas and is negatively associated with the CVP. On an average volatility day, volatility persistence is expected to decline by 4.4% in crude oil and 4.0% in natural gas.[26] Similarly to the HAR models, past (daily, weekly and monthly) RV remains highly significant and positively related to the next day's volatility. However, some of the RV impact is transmitted to future volatility via the CVP, and it has an inverse effect on volatility persistence. This can be justified by the argument that high volatility allows more information to be priced, resulting in lower volatility persistence (Andersen (1996)). Finally, results from Tables 2 and 3 for the HAR-CVP-CV models reveal that the addition of the macro-economic variables does not affect the statistical significance of the other predictors of future volatility and daily volatility persistence, such as positive and negative returns and RV. Therefore, we conclude that volatility persistence plays an important role in transmitting the impact of returns and RV to volatility of energy markets.[27]

[26] This is computed by combining information from Table 5, Table 7 and Table D2 from Appendix D. For crude oil is computed as $0.006 \times 4.035/0.546 = 4.4\%$ and for natural gas, is $0.003 \times 6.206/0.464 = 4.0\%$, accordingly.

[27] The HAR-CVP model can be extended to include variations of the RV measure: the semi-variances (SV) of Patton and Sheppard (2015) and the Bollerslev et al. (2016) realized quad-power quarticity (RQ). We use these extensions of the HAR-CVP model, namely, HAR-CVP-SV and HAR-CVP-RQ, as robustness tests and we find that the results are robust for these HAR-CVP model extensions, see Appendix F for details.

Table 8
Out-of sample Forecasting Performance.

Panel A: Loss functions comparison

	HAR	HAR-SV	HAR-RQ	HAR-CVP	HAR-CVP-CV
Crude Oil					
MSE-ln					
Mean	0.236	0.294	0.211	0.185	0.155
Median	0.073	0.074	0.073	0.061	0.057
St. Dev.	0.558	1.054	0.419	0.420	0.333
QLIKE					
Mean	0.035	0.036	0.036	0.030	0.028
Median	0.037	0.034	0.037	0.033	0.032
St. Dev.	0.590	3.278	0.326	0.412	0.265
Natural Gas					
MSE-ln					
Mean	0.287	0.310	0.291	0.287	0.289
Median	0.077	0.088	0.077	0.076	0.079
St. Dev.	0.810	0.894	0.803	0.778	0.773
QLIKE					
Mean	0.125	0.131	0.127	0.126	0.127
Median	0.039	0.042	0.039	0.038	0.040
St. Dev.	0.280	0.283	0.282	0.282	0.279

Panel B: DM tests

	Crude Oil			Natural Gas		
HAR-CVP						
	HAR	HAR-SV	HAR-RQ	HAR	HAR-SV	HAR-RQ
MSE-ln	−3.666***	−4.254***	−2.628**	−0.128	−3.991***	−1.777*
QLIKE	−2.537**	−2.532**	−1.261	0.968	−1.601	−1.071
HAR-CVP-CV						
	HAR	HAR-SV	HAR-RQ	HAR	HAR-SV	HAR-RQ
MSE-ln	−6.294***	−5.470***	−5.853***	0.383	−3.614***	−0.851
QLIKE	−4.303***	−2.819**	−3.819***	1.441	−1.281	−0.231

Panel A reports a comparison of loss functions, namely MSE-ln and QLIKE loss functions, across different HAR models for crude oil and natural gas market. Panel B reports the DM test statistics of the HAR_i against the $HAR-CVP$ and the $HAR-CVP-CV$ model, where i is the HAR, HAR-SV and HAR-RQ model. A negative value means that the $HAR-CVP$ and/or $HAR-CVP-CV$ model has lower loss values compared to the competing models. *, **, *** denotes the 10%, 5% and 1% level of significance, respectively.

4.3. Economic significance of daily CVP

To further appreciate the contribution of the daily CVP to future volatility, we assess the economic significance of the daily CVP. We gauge the contribution of CVP determinants to the volatility variation and then identify the drivers of the CVP variation.

4.3.1. Contribution of CVP determinants to the volatility variation

We have identified returns, RV and macro-economic variables such as the VIX, credit spreads and Treasury bills as significant determinants of CVP in the oil and natural gas markets. We use the Shapley-Owen decomposition of the regression R^2 to measure the marginal contribution (via their explanatory power) of each variables to the volatility variation captured by the conditional volatility persistence HAR models.[28]

Panel B of Table 5 presents the decomposition of the CVP regressors, namely, $r^+_{t,D}RV_{t,D}$, $r^-_{t,D}RV_{t,D}$, and $RV^2_{t,D}$ in the HAR-CVP models. We find that the CVP regressors explain 16.83% of future variations in crude oil's RV and 15.28% of future variations in natural gas's RV. Further, the CVP accounts for 25.05% of the model's explanatory power in crude oil and 26.91% of the model's explanatory power in natural gas. Among the CVP regressors, RV^2 makes the stronger contribution (higher explanatory power) with a Shapley R^2 of

[28] We follow the Lahaye and Neely (2018) and Wang and Yang (2018) approach to estimate the Shapley R^2. Henceforth, the total R^2 is the sum of the estimated Shapley R^2 for each variable. The Shapley R^2 decomposition helps to better assess the contribution of the CVP regresssors (as well as other regressors) to the HAR model. One advantage of the Shapley R^2 regression over the general linear regression technique is that it takes care of multicollinearity.

Table 9
Realized Utility

Panel A: Realized utility with varying coefficient of risk aversion

		HAR	HAR-SV	HAR-RQ	HAR-CVP	HAR-CVP-CV	HAR	HAR-SV	HAR-RQ	HAR-CVP	HAR-CVP-CV
Crude Oil											
T.Costs	Spread	$\gamma = 2$					$\gamma = 4$				
Zero		8.24%	7.04%	8.44%	8.49%	8.61%	4.12%	3.52%	4.22%	4.24%	4.31%
Full	Full	8.19%	6.97%	8.38%	8.44%	8.57%	4.10%	3.49%	4.19%	4.22%	4.29%
Full	Half	8.22%	7.00%	8.41%	8.46%	8.59%	4.11%	3.50%	4.20%	4.23%	4.30%
Gradual	Full	8.24%	7.03%	8.43%	8.48%	8.60%	4.12%	3.51%	4.21%	4.24%	4.30%
Gradual	Half	8.24%	7.03%	8.43%	8.48%	8.61%	4.12%	3.52%	4.22%	4.24%	4.30%
Natural Gas											
T.Costs	Spread	$\gamma = 2$					$\gamma = 4$				
Zero		11.52%	11.49%	11.50%	11.51%	11.50%	5.76%	5.75%	5.75%	5.75%	5.75%
Full	Full	11.42%	11.42%	11.40%	11.41%	11.39%	5.71%	5.71%	5.70%	5.70%	5.70%
Full	Half	11.47%	11.46%	11.45%	11.46%	11.45%	5.74%	5.73%	5.72%	5.73%	5.72%
Gradual	Full	11.50%	11.48%	11.48%	11.49%	11.48%	5.75%	5.74%	5.74%	5.75%	5.74%
Gradual	Half	11.51%	11.49%	11.49%	11.50%	11.49%	5.76%	5.74%	5.75%	5.75%	5.75%

Panel B: Differential in realized utility and DM tests

		HAR-CVP			HAR-CVP-CV			HAR-CVP			HAR-CVP-CV		
		HAR	HAR-SV	HAR-RQ	HAR	HAR-SV	HAR-RQ	HAR	HAR-SV	HAR-RQ	HAR	HAR-SV	HAR-RQ
Crude Oil													
T.Costs	Spread	$\gamma = 2$						$\gamma = 4$					
Zero		24.321**	144.892**	5.147	36.663***	157.234**	17.489***	12.160**	72.446**	2.574	18.331***	78.617**	8.744***
		(2.275)	(2.363)	(0.880)	(3.858)	(2.575)	(3.244)	(2.275)	(2.363)	(0.880)	(3.858)	(2.575)	(3.244)
Full	Full	24.929**	146.872**	5.727	38.448***	160.392**	19.246***	12.464**	73.436**	2.864	19.224***	80.196**	9.623***
		(2.270)	(2.358)	(0.867)	(3.836)	(2.568)	(3.201)	(2.270)	(2.358)	(0.867)	(3.836)	(2.568)	(3.201)
Full	Half	24.632**	145.926**	5.438	37.566***	158.860**	18.373***	12.316**	72.963**	2.719	18.783***	79.430**	9.186***
		(2.286)	(2.371)	(0.911)	(3.906)	(2.590)	(3.341)	(2.286)	(2.371)	(0.911)	(3.906)	(2.590)	(3.341)
Gradual	Full	24.424**	145.264**	5.236	36.949***	157.788**	17.761***	12.212**	72.632**	2.618	18.475***	78.894**	8.881***
		(2.301)	(2.383)	(0.953)	(3.973)	(2.611)	(3.478)	(2.301)	(2.383)	(0.953)	(3.973)	(2.611)	(3.478)
Gradual	Half	24.380**	145.122**	5.193	36.817***	157.559**	17.630***	12.190**	72.561**	2.597	18.409***	78.779**	8.815***
		(2.273)	(2.361)	(0.874)	(3.848)	(2.572)	(3.223)	(2.273)	(2.361)	(0.874)	(3.848)	(2.572)	(3.223)
Natural Gas													
T.Costs	Spread	$\gamma = 2$						$\gamma = 4$					
Zero		−1.144	1.378	0.647	−1.760	0.762	0.030	−0.572	0.689	0.323	−0.880	0.381	0.015
		(-1.216)	(0.595)	(0.708)	(-1.612)	(0.330)	(0.028)	(-1.216)	(0.595)	(0.708)	(-1.612)	(0.330)	(0.028)
Full	Full	−1.812*	−1.554	0.689	−3.216**	−2.958	−0.715	−0.906*	−0.777	0.345	−1.608**	−1.479	−0.357
		(-1.864)	(-0.664)	(0.721)	(-2.781)	(-1.264)	(-0.611)	(-1.864)	(-0.664)	(0.721)	(-2.781)	(-1.264)	(-0.611)
Full	Half	−1.478	−0.086	0.668	−2.489**	−1.097	−0.342	−0.739	−0.043	0.334	−1.244**	−0.549	−0.171
		(-1.550)	(-0.037)	(0.717)	(-2.222)	(-0.473)	(-0.304)	(-1.550)	(-0.037)	(0.717)	(-2.222)	(-0.473)	(-0.304)
Gradual	Full	−1.244	0.941	0.654	−1.980*	0.206	−0.082	−0.622	0.470	0.327	−0.990*	0.103	−0.041
		(-1.318)	(0.406)	(0.711)	(-1.800)	(0.089)	(-0.074)	(-1.318)	(0.406)	(0.711)	(-1.800)	(0.089)	(-0.074)
Gradual	Half	−1.194	1.161	0.651	−1.871*	0.485	−0.026	−0.597	0.581	0.325	−0.935*	0.242	−0.013
		(-1.267)	(0.501)	(0.710)	(-1.707)	(0.210)	(-0.024)	(-1.267)	(0.501)	(0.710)	(-1.707)	(0.210)	(-0.024)

Panel A report the average realized utility (UoW) with varying coefficient of risk aversion. The realized utility is estimated under five scenarios: no transaction costs ("Zero"), with transaction costs equal to the average full ("Full") and half ("Half") spreads with investment positions fully rebalanced at the close of each business day ("Full") and with transaction costs equal to the average full ("Full") and half ("Half") spreads with investment positions rebalanced gradually ("'Gradual"). The full and half spreads are the difference between the ask and bid prices divided by the midquote and half the full spread over the past nine months, respectively. The full spread is equal to 1.90 and 5.20 bps (basis points) in the crude oil and natural gas markets. The half spread stands at 0.95 and 2.60 bps in the respective markets. Panel B reports the reports the differential in realized utility between the classical HAR models and the HAR-CVP models and their respective DM tests (in parentheses). The differential in utility is reported in bps. A positive DM value means that the HAR-CVP and/or HAR-CVP-CV model has higher realized utility compared to the classical HAR models. *, **, *** denotes the 10%, 5% and 1% level of significance, respectively.

10.49% for crude oil and 9.43% for natural gas.[29] Further, the direct influence of past short- and long-term returns and RV on future RV is significantly reduced in the two markets (with the effect more pronounced in crude oil market) for the HAR-CVP models (compared to the HAR models). Moreover, in the oil market large negative returns have greater impact on future volatility than large positive returns, a feature shared with equity markets, providing support for the notion of volatility spillovers between energy and equity market.[30]

The Shapley decomposition of the regression R^2 for the HAR-CVP-CV models is presented in Table 4 for crude oil and natural gas, in Panel A and Panel B, respectively. The CVP regressors, i.e., $r^+_{t,D}RV_{t,D}$, $r^-_{t,D}RV_{t,D}$, $RV^2_{t,D}$ and $CV_{t,D}RV_{t,D}$, explain up to 18% of future variations in RV and account for up to 29% of the models' explanatory power. RV^2 generally, has the most explanatory power (with a Shapley R^2 up to 10%), but the $CV_{t,D}RV_{t,D}$ reaches similar explanatory power for the models associated with key determinants of CVP, such as the VIX (7.47%), credit spreads (7.28%) and the Baltic exchange dirty tanker index (7.06%). This underscores the economic significance of the conditioning variables in explaining variation in future volatility, an effect that is channeled via volatility persistence. In fact, the indirect impact of each conditioning variable on future volatility (via CVP) reaches a Shapley R^2 of up to 7.5%, while comparatively, the direct impact of the same variables on future RV is much lower (reaching only 1%). Similar to the WLS regression results, the Shapley R^2 of the CVP regressors representing the conditioning variables is higher in the crude oil market compared to the natural gas market. Thus, the contribution of the conditioning variables is stronger to volatility persistence than the direct impact on volatility levels. This result is consistent with our previous findings which demonstrate that these variables are key determinants of volatility persistence and transmit their impact to future volatility via the volatility persistence channel. This impact is considerable and jointly accounts, via the CVP, for more than a quarter of the variation in future volatility.

4.3.2. Determinants of CVP variation

Beyond the determinants of the CVP, we seek to also identify the determinants of CVP variation. To this end, we decompose the variance of $CVP_{t,D}$ into the variances of its orthogonalized components $r^-_{t,D}$, $r^+_{t,D}$, and $RV_{t,D}$ The weights $w(r^-_{t,D})$, $w(r^+_{t,D})$, and $w(RV_{t,D})$ denote the contributions of $r^-_{t,D}$, $r^+_{t,D}$, and $RV_{t,D}$ to the variance of $CVP_{t,D}$, respectively.[31]

The contribution of CVP variables ($r^-_{t,D}$, $r^+_{t,D}$, and $RV_{t,D}$) to the variance of CVP for the HAR-CVP models is reported in Table 6. Negative returns account for 87.1% and 52.5% of the CVP variance in the HAR-CVP models for crude oil and natural gas, respectively. The volatility level (RV) makes up approximately 7.5% of the CVP variance in the crude oil market, and 9.0% of the CVP variance in the natural gas market. The variance decomposition of the CVP variables reveals that negative return, rather than positive return, is the strongest driver of the variance of volatility persistence in the crude oil market. However, in natural gas markets, positive returns are the strongest driver of variation in volatility persistence. The combined effect of positive and negative returns accounts for 90–92% of the CVP variation across all markets.

4.4. Statistical properties of daily CVP

The daily CVP for the two energy markets are estimated by equation (7) and their statistical properties are reported in Table 7.[32] The time series and histogram of the daily CVP for each market are displayed in Fig. 2.

Volatility persistence in energy markets fluctuates over time and exhibits significant variability. The mean of the daily volatility persistence reaches 54.6% in the crude oil market and 46.4% in the natural gas market, while its variation is 8.5% in the crude oil.

The figure plots the daily conditional volatility persistence and the corresponding histogram for crude oil and natural gas front-month futures contracts between January 2009 and August 2019.

Market and 7.7% in the natural gas markets. The histograms in Fig. 2 show that the mean CVP is larger than the median (long right-hand tail) in both energy markets. Specifically, the crude oil market has a higher proportion of positive outliers (with 0.11% and 0.08% of the sample days having a CVP greater or equal to one, respectively). The cross-correlation between CVP and its components (r^-_D, r^+_D,

[29] Empirical evidence in equity markets shows that $r^+_{t,D}RV_{t,D}$ has the most explanatory power for future RV and equity markets exhibit lower RV and volatility of RV than the energy markets (Wang and Yang (2018)).

[30] The explanatory power of CVP has also been considered for the HAR-SV and HAR-RQ models, and we found similar results, see Appendix F for details. Thus, these results are robust under the different HAR and HAR-CVP model specifications.

[31] We follow the Wang and Yang (2018) variance decomposition approach to assess the marginal contribution of return and volatility level on the CVP. We assume that $y = ax_1 + bx_2 + cx_3$. y represents the CVP, and x_i represents the three CVP regressors ($r^-_{t,D}RV_{t,D}$, $r^+_{t,D}RV_{t,D}$, $RV^2_{t,D}$). As there are three regressors, $y = ax_1 + bx_2 + cx_3$ can be rewritten in 6 ways (3! = 6 permutations). We extract the residuals u_{21} and u_{31} from the following equations: (1) $x_2 = \alpha_0 + \alpha_1 x_1 + u_{21}$ and (2) $x_3 = \beta_0 + \beta_1 x_1 + u_{31}$. Then, $\widehat{u}_{31} = \lambda\widehat{u}_{21} + u_{32}$ is estimated, and u_{32} is retrieved. With the previously estimated coefficients and regressors, y can be further decomposed as $y = ax_1 + b(\widehat{\alpha}_0 + \widehat{\alpha}_1 x_1 + \widehat{u}_{21}) + c(\widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{u}_{31})$. Thus, $y = (a + b\widehat{\alpha}_1 + c\widehat{\beta}_1)x_1 + (b + c\widehat{\lambda})\widehat{u}_{21} + c\widehat{u}_{32} + constant$, which simplifies to $= Ax_1 + B\widehat{u}_{21} + c\widehat{u}_{32} + constant$. The variance of y is given by $var(y) = A^2var(x_1) + B^2var(\widehat{u}_{21}) + c^2var(\widehat{u}_{32})$. The weights of each regressor are estimated as $w(x_1) \equiv \frac{A^2var(x_1)}{var(y)}$, $w(x_2) \equiv \frac{B^2var(x_2)}{var(y)}$ and $w(x_3) \equiv \frac{c^2var(x_3)}{var(y)}$. Ultimately, $RV_{t,D}$ is replaced by $RQ^{1/2}_{t,D}$ as a CVP regressor in the HAR-CVP-RQ model.

[32] We identify the conditioning variables that have a significant impact (directly and indirectly) on future RV from the last column of Tables 2 and 3 for the two energy commodities. The HAR-CVP-CV regression model is re-estimated, and their coefficients are extracted to compute $CVP_{t,D} = \beta_D + \beta^-_t r^-_{t,D} + \beta^+_t r^+_{t,D} + \beta_{RV}RV_{t,D} + \beta_{CV}CV_{t,D}$.

RV_D and $CV_{t,D}$) is also presented in Panel B of Table 7. As expected, in the crude oil markets, the correlation between volatility persistence and negative returns is much stronger compared to the positive returns, while the reverse holds in the natural gas market. Daily volatility persistence is positively related (averaging 0.29) to RV in all markets.[33] The VIX and Treasury bills are positively correlated to volatility persistence, while credit spreads and the Baltic exchange dirty tanker index are negatively correlated with volatility persistence. These results are consistent with the results of the in-sample estimation of the HAR-CVP-CV models (see the last column of Tables 2 and 3).

5. Forecasting performance

We assess next the out-of-sample forecasting performance of the conditional volatility persistence HAR models in predicting daily volatility, compared to classical HAR models. Beyond the statistical significance of forecasting daily energy market volatility, we also con-sider a utility-based methodology to evaluate the economic significance of the forecasts.

5.1. Models and loss functions

Aiming to gauge the role of embedding macro-economic information in improving predictive accuracy, we compare the out-of-sample performance of three classical HAR models (HAR, HAR-SV, HAR-RQ) and two conditional volatility persistence HAR models (HAR- CVP and HAR-CVP-CV), see Section 2 for models description. We use the fixed rolling window out-of-sample forecast (Patton and Sheppard (2015) and Bollerslev et al. (2016)), where the rolling window is set to 1, 000 days representing approximately 4 years.[34] The HAR models are estimated using the WLS regressions. We use two loss functions (measuring the prediction error between two competing volatility models): the log of the mean squared errors (*MSE*-ln) and the quasi-likelihood (*QLIKE*) loss functions defined as

$$L\left(RV_{t+1,D}, \widehat{RV}_{t+1,D}\right) = \begin{cases} \left(\ln RV_{t+1,D} - \ln \widehat{RV}_{t+1,D}\right)^2; for MSE - \ln, \\ \dfrac{RV_{t+1,D}}{\widehat{RV}_{t+1,D}} - RV_{t+1,D} \ln \dfrac{RV_{t+1,D}}{\widehat{RV}_{t+1,D}}; for QLIKE, \end{cases} \tag{9}$$

where $\widehat{RV}_{t+1,D}$ is the forecast of $RV_{t+1,D}$. Derived from the mean squared errors loss function, the *MSE*-ln reduces the impact of large forecast errors ($RV_{t+1,D} - \widehat{RV}_{t+1,D}$). Alternatively, *QLIKE* measures the difference in the ratio of $RV_{t+1,D}/ \widehat{RV}_{t+1,D}$. The *MSE*-ln and *QLIKE* have been used extensively in the equity related literature (Patton (2011)) and commodity (Byun and Cho (2013), Li and Li (2015) and Zhu et al. (2017)). Using the loss functions, the forecasting accuracy is also assessed with the Diebold and Mariano (1995) (DM) test.[35] The pair-wise loss difference ($d_{t+1,D}(HAR_k)$) against the two HAR-CVP models is defined as:

$$d_{t+1,D}(HAR_k) = L\left(RV_{t+1,D}, \widehat{RV}_{t+1,D}; HAR - CVP\right) - L\left(RV_t, \widehat{RV}_{t+1,D}; HAR_k\right),$$

where k represents the three classical HAR models (HAR, HAR-SV and HAR-RQ). The DM test statistic, $DM[d_{t+1,D}(HAR_k)]$, is estimated via the following specification:

$$DM\left[\bar{d}_{t+1,D}\left(HAR_k\right)\right] = \frac{\bar{d}_{t+1,D}(HAR_k)}{\sqrt{\left(Var\left[d_{t+1,D}(HAR_k)\right]/T\right)}},$$

where $\bar{d}_{t+1,D}(HAR_k)$ is the mean value of $d_{t+1,D}(HAR_k)$, $Var[d_{t+1,D}(HAR_k)]$ is the variance of $d_{t+1,D}(HAR_k)$ and T is the number of forecasts.

[33] We rely on the Ljung-Box test to measure persistence over a day, see Table 7.

[34] We use a fixed rolling window estimation that ensures that the rolling window and out-of-sample size stay constant. As we do not impose any restrictions on the parameters in the rolling window estimation, the forecasts can potentially be negative. Although negative forecasts rarely occur in this analysis, we apply an "insanity filter". This ensures that the negative forecasts are replaced by the minimum positive RV within each rolling window, see Patton and Sheppard (2015). We also replace the RV forecasts that exceed (fall behind) the maximum (minimum) observed RV in the rolling window with the mean observed RV in each rolling window (Swanson and White (1995), Bollerslev et al. (2016) and Bollerslev et al. (2018)). While the latter affects only 0.36% of the RV forecasts in the crude oil market, the natural gas market is unaffected by this. We reassess the out-of-sample forecasting performance of the HAR models without the Swanson and White (1995) filtering process or an alternative "insanity filter" where the "insane" forecasts are replaced with the dumb forecasts (Hyndman and Koehler (2006) and Lux and Kaizoji (2007)) and the results obtained are quantitatively similar.

[35] The DM test serves as a measure for comparing the forecasting accuracy of two competing models. It is a well-recognised forecasting evaluation statistic that has been widely used in empirical studies in equity markets (Han et al. (2015), Sharma and Vipul. (2016) and Wang and Yang (2018)) and commodity markets (Jiang et al. (2015), Herrera et al. (2018) and Gong and Lin (2018)).

5.2. Forecast comparison

We find that the conditional volatility persistence HAR models, on average, outperform the classical HAR models in forecasting accuracy (see Panel A of Table 8). The improvement in crude oil market is significant (with the lowest mean and median loss values for both loss functions (MSE-ln and QLIKE)), while the improvement in natural gas is marginal. The DM test statistics, reported in Panel B of Table 8, corroborate the loss function results in Panel A. These results confirm significant benefits in out-of-sample daily forecasting, when information from market and macro-economic variables is integrated in the model (Degiannakis and Filis (2017) and Luo et al. (2020)). In the crude oil market, the impact of macro-economic variables, such as the VIX and credit spreads, is substantial, and it is channeled in the oil price volatility via its persistence. Accounting for this information transmission channel brings significant forecasting gains. However, in the natural gas market, most of these macro-economic variables do not affect volatility persistence, thus the proposed HAR-CVP- CV models would not improve forecasting performance. Prices and volatility in the natural gas market are predominantly driven by commodity sector factors, such as demand and inventory considerations and the recent shale gas expansion (Geng et al. (2016a) and Caporin and Fontini (2017)). This underscores the need for more robust forecasting models, which account for a wider range of macro-economic conditions that have a measurable impact on price dynamics of the oil and natural gas markets, such as demand and inventory (Geng et al. (2016b) and Kang et al. (2020)).[36]

Although conditional volatility persistence HAR models provide statistically stronger forecasts, from a practical perspective, the models should also outperform when these volatility forecasts are implemented in risk management investment strategies, as discussed next.

5.3. Economic significance of the forecasts

Motivated by Bollerslev et al. (2018), we employ a realized utility-based approach based on volatility forecasts to assess the economic benefit of implementing the proposed HAR models.[37] This approach assumes that rational investors[38] trade in a risky asset (i.e. energy asset) with time-varying volatility and earn a constant risk-adjusted return or Sharpe ratio (SR).[39] Investors seek to keep a constant level of risk and adjust their optimal portfolio size accordingly. Therefore, the risk/volatility target (RT) is set to reflect the annualized volatility in the respective market.[40] Bollerslev et al. (2018) assume that the coefficient of risk aversion, γ (metric of investors' degree of risk aversion), is the same across all asset classes. Gauging the actual degree of investors' risk aversion in different energy markets is not trivial, thus we assume two levels of risk aversion; $\gamma = 2$ and 4. Because the RT is estimated by dividing the annualized SR by γ, the SR is retrieved accordingly. Moreover, we use various combinations of the SR in the respective markets to derive the optimal γ.[41]

The expected return of an investor's strategy, optimal targeted position (OTP), is estimated by multiplying the SR by the RT. Because of the disutility of risk, as pointed out in Bollerslev et al. (2018), the net optimal targeted position (NOTP) is halved. Under a perfect risk model, the value of a risky asset is worth NOTP% of wealth. For instance, if $\gamma = 2$, then the NOTP is equal to 9% and 13% in the crude oil and natural gas markets, respectively.[42] Given that investors choose an initial investment position of $x^{\theta}_{t+1,D} = RT/\sqrt{E_t(RV^{\theta}_{t+1,D})}$, the average realized utility per unit of wealth (UoW^{θ}) associated with each HAR model[43] is estimated for the crude oil and natural gas markets as follows:

[36] We recognize that inventory is an important driver of commodity markets price and volatility dynamics. The theory of storage induces a monotonic relation between inventories and volatility where low inventory levels are associated with high volatility due to the increased risk of inventory exhaustion, see Working (1949) and Brennan (1958). However, recent empirical studies reveal that inventory holds an asymmetric V-shaped relation with oil price volatility as high inventory levels reflect inelastic supply and limited inventory adjustments that cause an increase in volatility (Kogan et al. (2009), Haugom et al. (2014) and Nikitopoulos et al. (2017)). Empirical evidence further suggests that the dearth of inventories and storage capacity affect prices non-linearly (Büyükcahin et al., 2013), while this non-linearity extends also to the second moment, particularly implied volatilities (Robe and Wallen (2016)),) which can also affect commodity arbitraging (Ederington et al. (2021)).

[37] Commodity (with energy a commodity subclass) volatility is the highest among the different asset classes (such as equity or bond), and thus, is prone to more frequent shocks. Therefore, we use a dynamic approach to portfolio rebalancing and assume it happens daily. As Bollerslev et al. (2018) examine the aggregate volatility of different asset classes, they assume that portfolio rebalancing occurs monthly (less frequently).

[38] We assume that investors have mean-variance preferences, i.e., they want low risk and high return.

[39] In this setting, the SR is measured by $SR = E_t(r^e_{t+1,D})/\sqrt{E_t(RV_{t+1,D})}$, where the excess return ($r^e_{t+1,D}$) is equal to the return on the risky asset ($r_{t+1,D}$) less the return on a risk-free asset ($r^f_{t,D}$).

[40] In the crude oil and natural markets, the average annualized RV (over the sample period) stands at 30% and 36%, respectively.

[41] The estimated realized utility using varying SR in each market is reported in Appendix G.

[42] Similarly, for a SR of 0.6 and 0.7 and $\gamma = 4$, the NOTP is 5% and 6% in the crude oil and natural gas markets, respectively.

[43] θ represents the HAR, HAR-SV, HAR-RQ, HAR-CVP and HAR-CVP-CV models accordingly.

$$UoW^{\theta}_{CO} = \frac{1}{T}\sum_{t,D}^{T}\left(18\%\frac{\sqrt{RV^{\theta}_{t+1,D}}}{\sqrt{E_t\left(RV^{\theta}_{t+1,D}\right)}} - 9\%\frac{RV^{\theta}_{t+1,D}}{E_t\left(RV^{\theta}_{t+1,D}\right)}\right),$$

$$UoW^{\theta}_{NG} = \frac{1}{T}\sum_{t,D}^{T}\left(25\%\frac{\sqrt{RV^{\theta}_{t+1,D}}}{\sqrt{E_t\left(RV^{\theta}_{t+1,D}\right)}} - 13\%\frac{RV^{\theta}_{t+1,D}}{E_t\left(RV^{\theta}_{t+1,D}\right)}\right),$$

where $RV^{\theta}_{t+1,D}$ and $E_t(RV^{\theta}_{t+1,D})$ represent the actual RV and the forecasted RV, respectively. The economic intuition behind the utility-based framework is that models with better forecasting accuracy provide a higher level of realized utility (economic benefit).

We also examine the effect of the transaction costs on the realized utility by adopting an approach similar to that of Bollerslev et al. (2018). Transaction costs can have a damaging effect on investors' position, particularly for less informed (naïve) investors (Palczewski et al. (2015)). Taking into account the actual cost of implementing risk-targeted positions causes a loss in utility. Transaction costs are estimated by using the "full-spread" (median bid-ask spread)[44] and the "half-spread" (half of the median bid-ask spread) over the previous nine months. Under this approach, the linear trading cost may vary with the absolute magnitude of the change in the positions, $\left|x^{\theta}_{t,D} - x^{\theta}_{t+1,D}\right|$. We emulate the trading strategy of Gârleanu and Pedersen (2013), Gârleanu and Pedersen (2016) and Bollerslev et al. (2018) and allow the investment positions to be rebalanced gradually. Only 15% of the positions are traded toward the zero-cost optimal target every day. This adjustment makes up for the fact that trading is done partially, and therefore, the loss in utility is not heavily penalised. The results of the average realized utility under five scenarios of transaction costs are reported in Table 9. Panel A shows the estimated realized utility with varying coefficients of risk aversion, and Panel B reports the differential in the realized utility between the classical HAR models and the conditional volatility persistence HAR models and their respective DM tests.

We find that the conditional volatility persistence HAR models consistently outperform the classical HAR models in the crude oil market. When $\gamma = 2$, the utility benefits of HAR-CVP and HAR-CVP-CV are 8.486% and 8.610%, and these levels of utility benefits are 24.3 bps and 36.7 bps above the benefits of the basic HAR model (respectively and in the no transaction cost case), see Panel A of Table 9. To put this in an economic perspective, the utility benefits of the model incorporating information from macro-economic variables (HAR-CVP-CV) imply that investors are willing to pay 36.7 bps to use the HAR-CVP-CV model for risk management rather than to use the basic HAR models. This is a comparative significant level of utility benefits for risk management which is marginally below the institutional fees typically required for active asset management (Bollerslev et al. (2018)). The improvement the conditional volatility persistence HAR models offer compared to the HAR with semi-variances is more than fivefold, reaching 144.9 bps for HAR-CVP and 157.2 bps for HAR-CVP-CV (no transaction cost and for $\gamma = 2$). Interestingly, the HAR with realized quarticity and the HAR-CVP provide very similar benefits to investors, potentially because both models accommodate time-varying volatility persistence that depends on market conditions. However, including the additional information from macro-economic variables (by using the HARR-CVP-CV model) improves the realized utility for crude oil by 17.5 bps under no transaction costs and even more than that when transaction costs are involved. Furthermore, when $\gamma = 4$, the utility benefits of HAR-CVP and HAR-CVP-CV are 4.243% and 4.305%, respectively (in the no transaction cost case), which is 12.2 bps and 18.4 bps above the utility benefits of the basic HAR model. Although these benefits are lower than the benefits of a less risk-averse investor ($\gamma = 2$), this still represents a substantial level of utility benefits in the crude oil market. The results are quantitatively the same for the four scenarios concerning transaction costs and trading speed. Using the full-spread (as a proxy for the transaction costs) causes the largest reduction in realized utility in all markets. However, the utility benefit differential reaches 38.5 bps for the HAR-CVP-CV model (over the basic HAR) when $\gamma = 2$ and 19.2 bps when $\gamma = 4$. The "Gradual" trading has a marginal impact on the utility benefits for all models. The DM tests, presented in Panel B in Table 9 confirm further the validity of the results above. These findings underscore the substantial economic benefits of embedding information from macro-economic factors in volatility forecasting applications, information that is transmitted in volatility via its persistence.

In the natural gas market, the benefits are marginal (and not statistically significant) between the models. Assuming different degrees of risk aversion and no transaction costs, the utility benefits are up to 1.4 bps higher for the conditional volatility persistence HAR models (compared to the HAR-SV and HAR-RQ models). At the different levels of trading speed, the utility benefits of the HAR-CVP or HAR-CVP-CV models over competing HAR model is (almost completely) lost, especially for the "full" transaction costs scenarios where trading is the slowest. These results are mainly driven by the substantial impact of the relatively high transaction costs occurring in the natural gas market. The estimated transaction costs in the natural gas market (5.2 bps) is almost threefold higher than in the crude oil market (1.9 bps), see Bollerslev et al. (2018).[45] Furthermore, the impact of the macro-economic variables (considered in this study) in the volatility of natural gas market is marginal and mostly direct (as it is not channeled via the volatility persistence), thus using models that incorporate information from the underlying market variables would not offer any benefits in forecasting

[44] We calculate the spread by using the difference between the ask and bid prices divided by the mid-quote.

[45] We re-estimate the realized utility in the natural gas market while assuming the same level of transaction costs as in the crude oil market, i.e. 1.9 bps, and we find that the transaction costs are no longer penalizing the realized utility to the same extent. These results are available upon request from the authors.

performance.[46] When information from the macro-economy matters in the volatility dynamics and it is transmitted to the volatility via its persistence, as it happens in the crude oil market, then the conditional volatility persistence HAR models would provide substantial utility benefits. Thus, recognising volatility persistence as an important information transmission channel in volatility forecasting brings statistically significant forecasting performance and substantial utility gains in risk management strategies.[47]

6. Conclusion

We explore the role of daily volatility persistence in shaping the dynamics of future volatility in two energy markets, crude oil and natural gas. By allowing the daily volatility persistence to be time-varying in the HAR model, we identify the determinants of volatility persistence, and we analyse its contribution to predicting future volatility. We further hypothesize that macro-economic variables impact daily volatility persistence, and we investigate the role of volatility persistence in transmitting the impact of these variables to future volatility. The ability of these models to forecast short-term volatility in energy markets is examined, and their benefits for investment strategies are accordingly evaluated.

Returns, volatility and macro-economic variables are key determinants of volatility persistence. Generally, negative returns impact volatility persistence in crude oil market, while positive returns matter more in volatility persistence in natural gas market. Furthermore, the Baltic exchange dirty index and financial indicators, including the VIX, the credit spreads and the 3-month Treasury bill rate, affect volatility persistence in energy markets, in particular, the crude oil market. This impact is transmitted to the volatility dynamics mostly indirectly via the volatility persistence channel, rather than directly to the volatility levels. In crude oil market, volatility persistence plays a dominant role in diffusing the impact of returns, RV and macro-economic conditions in future volatility. The statistical properties of daily CVP in the energy markets reveal that negative (positive) returns have a greater impact on oil (natural gas) volatility persistence, and that the mean volatility persistence reaches 50.5%. The out-of-sample forecasting analysis demonstrates that the conditional volatility persistence HAR models economically and statistically outperform the classical HAR models. The utility benefit can reach up to 160 bps (subject to the Sharpe ratio and the risk target) for the models accommodating the information from macro-economic variables.

Several practical implications have emerged from this study. The volatility of energy markets displays distinct characteristics, but the energy markets are becoming more integrated with equity markets. The contribution of volatility persistence in forecasting daily realized variance (ex-post volatility) is measurable and indicative of volatility persistence being a priced risk factor. Thus, the modelling consideration of volatility persistence is important and provides useful insights to the energy market participants, from institutional investors to energy producers and in particular, to short-term traders. Based on the superior forecasting performance and economic gains achieved by the proposed daily CVP models, they offer a robust approach for assessing and managing short-term risk exposures relevant to daily and momentum trading strategies and dynamic hedging applications. Further, the COVID-19 pandemic underscores the role of energy markets in the stability of global economies, when we witnessed negative oil prices at the climax of the outbreak.[48] As the aftermath of the COVID-19 pandemic is unfolding, the volatility of energy markets (along with the volatility of financial markets) is the main concern. Consequently, effective modelling and accurate forecasting of volatility are extremely important. These findings have also opened new directions for research including the effects of volatility persistence on longer-term volatility forecasting and expanding to embed the effects of a wider range of macro-economic volatility determinants in energy markets.

Author contribution statement

Thomas, Nikitopoulos, Wang (Conceptualization); **Thomas** (Data curation); **Thomas, Wang** (Formal analysis); **Thomas, Nikitopoulos, Wang** (Investigation); **Thomas, Nikitopoulos, Wang** (Methodology); Nikitopoulos (Project administration); **Thomas, Nikitopoulos** (Resources); Thomas (Software); **Thomas, Nikitopoulos, Wang** (Validation); Thomas, Nikitopoulos (Visualization); **Thomas, Nikitopoulos** (Writing - original draft); **Thomas, Nikitopoulos, Wang** (Writing - review & editing).

Financial support from the UTS Business Research Support funding scheme is acknowledged.

Acknowledgements

We thank Marcel Prokopczuk, an associate editor, two anonymous reviewers, Susan Thorp, Laura Ryan, Talis Putnins and the participants of the UTS Finance Research Show-case 2020 for fruitful discussions and helpful suggestions at different stages of this project.

[46] While oil markets act as information transmitters, natural gas markets act as information receivers (Ji et al. (2018)).

[47] To account for structural breaks, we split the data sample into two subperiods: pre-2015 (January 2009 to December 2014) and post-2015 (January 2015 to August 2019) and find quantitatively similar results (Asai et al. (2020) and Dahl et al. (2020)). The detailed results and discussions of this analysis are presented in Appendix H.

[48] On 20 April 2020, for first time in history the US oil benchmark dropped into deep negative territory as a combined result of the sinking oil demand in a system with limited storage capacity and inability to take delivery of the long position of (deliverable) oil futures contracts.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jcomm.2022.100285.

References

Alam, M.S., Shahzad, S.J.H., Ferrer, R., 2019. Causal flows between oil and forex markets using high- frequency data: asymmetries from good and bad volatility. Energy Econ. 84, 104513.
Andersen, T., 1996. Return volatility and trading volume: an information flow interpretation of stochastic volatility. J. Finance 51 (1), 169–204.
Andersen, T.G., Bollerslev, T., 1997. Heterogeneous information arrivals and return volatility dynamics: uncovering the long-run in high frequency returns. J. Finance 52 (3), 975–1005.
Andersen, T.G., Bollerslev, T., Meddahi, N., 2011. Realized volatility forecasting and market microstructure noise. J. Econom. 160 (1), 220–234.
Anzuini, A., Pagano, P., Pisani, M., 2015. Macroeconomic effects of precautionary demand for oil. J. Appl. Econom. 30 (6), 968–986.
Arismendi, J.C., Back, J., Prokopczuk, M., Paschke, R., Rudolf, M., 2016. Seasonal stochastic volatility: implications for the pricing of commodity options. J. Bank. Finance 66, 53–65.
Aromi, D., Clements, A., 2019. Spillovers between the oil sector and the S&P500: the impact of information flow about crude oil. Energy Econ. 81, 187–196.
Arora, V., Tanner, M., 2013. Do oil prices respond to real interest rates? Energy Econ. 36, 546–555.
Asai, M., Gupta, R., McAleer, M., 2020. Forecasting volatility and co-volatility of crude oil and gold futures: effects of leverage, jumps, spillovers, and geopolitical risks. Int. J. Forecast. 36 (3), 933–948.
Auer, B.R., 2014. Daily seasonality in crude oil returns and volatilities. Energy Econ. 43, 82–88.
Bakas, D., Triantafyllou, A., 2019. Volatility forecasting in commodity markets using macro uncertainty. Energy Econ. 81, 79–94.
Bampinas, G., Panagiotidis, T., 2017. Oil and stock markets before and after financial crises: a local Gaussian correlation approach. J. Futures Mark. 37 (12), 1179–1204.
Barndorff-Nielsen, O.E., Shephard, N., 2002. Econometric analysis of relized volatility and its use in esti- mating stochastic volatility models. J. Roy. Stat. Soc. 64 (2), 253–280.
Basak, S., Pavlova, A., 2016. A model of financialization of commodities. J. Finance 71 (4), 1511–1556.
Baur, D.G., Dimpfl, T., 2018. The asymmetric return-volatility relationship of commodity prices. Energy Econ. 76, 378–387.
Baur, D.G., Dimpfl, T., 2019. Think again: volatility asymmetry and volatility persistence. Stud. Nonlinear Dynam. Econom. 23 (1), 20170020.
Berger, D., Chaboud, A., Hjalmarsson, E., 2009. What drives volatility persistence in the foreign exchnage market? J. Financ. Econ. 94 (2), 192–213.
Bissoondoyal-Bheenick, E., Brooks, R., Do, H.X., Smyth, R., 2020. Exploiting the heteroskedasticity in measurement error to improve volatility predictions in oil and biofuel feedstock markets. Energy Econ. 86, 104689.
Bollerslev, T., Hood, B., Huss, J., Pedersen, L.H., 2018. Risk everywhere: modeling and managing volatility. Rev. Financ. Stud. 31 (7), 2729–2773.
Bollerslev, T., Patton, A., Quaedvlieg, R., 2016. Exploiting the errors: a simple approach for improved volatility forecasting. J. Econom. 192 (1), 1–18.
Breitenfellner, A., Cuaresma, J.C., Keppel, C., 2009. Determinants of crude oil prices: supply, demand, cartel or speculation? Monetary Policy and the Economy. Oesterreichische Nationalbank (Austrian Cen- tral Bank) (4), 111–136. URL. https://ideas.repec.org/a/onb/oenbmp/y2009i4b6.html.
Brennan, M.J., 1958. The supply of storage. Am. Econ. Rev. 48, 50–72.
Brock, W., LeBaron, B., 1996. A dynamic structural model for stock return volatility and trading volume. Rev. Econ. Stat. 78, 94–110.
Brooks, C., 2008. Introductory Econometrics for Finance, second ed. Edition. Cambridge University Press.
Bruno, V.G., Bykahin, B., Robe, M.A., 2017. The financialization of food? Am. J. Agric. Econ. 99 (1), 243–264.
Bucci, A., 2017. Forecasting realized volatility: a review. Journal of Advanced Studies in Finance 8 (2), 94–138.
Büyükcahin, B., Lee, T.K., Moser, J.T., Robe, M.A., 2013. Physical markets, paper markets and the WTI-Brent spread. Energy J. 34 (3), 129–151.
Byun, S.J., Cho, H., 2013. Forecasting carbon futures volatility using GARCH models with energy volatili- ties. Energy Econ. 40, 207–221.
Caporin, M., Fontini, F., 2017. The long-run oil-natural gas price relationship and the shale gas revolution. Energy Econ. 64, 511–519.
Chen, S., Wang, J., 2020. On the Common Determinants of Volatility Persistence and Asymmetry. Available at SSRN 3335572.
Chen, Y., Turnovsky, S., Zivot, E., 2014. Forecasting inflation using commodity price aggregates. J. Econom. 183, 117–134.
Cheng, B., Nikitopoulos, C.S., Schlögl, E., 2018. Pricing of long-dated commodity derivatives: do stochastic interest rates matter? J. Bank. Finance 95, 148–166.
Cheng, I.-H., Xiong, W., 2014. Financialization of commodity markets. Annual Review of Financial Eco- nomics 6 (1), 419–441.
Chiang, I., Hughen, W.K., Sagi, J.S., 2015. Estimating oil risk factors using information from equity and derivatives markets. J. Finance 70, 769–804.
Chiarella, C., Kang, B., Nikitopoulos, C.S., Tˆo, T., 2016. The return-volatility relation in commodity futures markets. J. Futures Mark. 36 (2), 127–152.
Christoffersen, P., Pan, X., 2018. Oil volatility risk and expected stock returns. J. Bank. Finance 95, 5–26.
Corsi, F., 2009. A simple approximate long-memory model of realized volatility. Journal of Financial Econo- metrics 7 (2), 174–196.
Corsi, F., Ren′o, R., 2012. Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. J. Bus. Econ. Stat. 30 (3), 368–380.
Dahl, R.E., Oglend, A., Yahya, M., 2020. Dynamics of volatility spillover in commodity markets: linking crude oil to agriculture. Journal of Commodity Markets 20, 100111.
Degiannakis, S., Filis, G., 2017. Forecasting oil prices realized volatility using information channels from other asset classes. J. Int. Money Finance 76, 28–49.
Degiannakis, S., Filis, G., 2018. Forecasting oil prices: high-frequency financial data are indeed useful. Energy Econ. 76, 388–402.
Diebold, F.X., Lopez, J.A., 1995. Modeling volatility dynamics. In: K D, H. (Ed.), Macroeconometrics. Recent Economic Thought Series, vol. 46. Springer, Dordrecht.
Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. J. Bus. Econ. Stat. 13, 253–265.
Ederington, L.H., Fernando, C.S., Holland, K., Lee, T.C., Linn, S., 2021. Dynamics of arbitrage. J. Financ. Quant. Anal. 56 (4), 1350–1380.
Fan, Y., Xu, J.H., 2011. What has been driven oil prices since 2000? A structural change perspective. Energy Econ. 33, 1082–1094.
Frankel, J., 2014. Effects of speculation and interest rates in a "carry trade" model of commodity prices. J. Int. Money Finance 42, 88–112.
Gˆarleanu, N., Pedersen, L., 2013. Dynamic trading with predictable returns and transaction costs. J. Finance 68 (6), 2309–2340.
Gˆarleanu, N., Pedersen, L., 2016. Dynamic portfolio choice with frictions. J. Econ. Theor. 165, 487–516.
Geng, J.B., Ji, Q., Fan, Y., 2016a. The behavior mechanism analysis or regional natural gas prices: a multi-scale perspective. Energy 101, 266–277.
Geng, J.B., Ji, Q., Fan, Y., 2016b. The impact of the North American shale gas revolution on regional natural gas markets: evidence from the regime-switching model. Energy Pol. 96, 167–178.
Gong, X., Lin, B., 2018. Structural breaks and volatility forecasting in the copper futures market. J. Futures Mark. 38 (3), 290–339.
Goodhart, C.A.E., O'Hara, M., 1997. High frequency data in financial markets: issues and applications. J. Empir. Finance 4 (2–3), 73–114.
Hameed, A., Kang, W., Viswanathan, S., 2010. Stock market declines and liquidity. J. Finance 65 (1), 257–293.
Han, H., Park, M.D., Zhang, S., 2015. A multiplicative error model with heterogeneous components for forecasting realized volatility. J. Forecast. 34 (3), 209–219.
Hasbrouck, J., 1995. One security, many markets: determining the contributions to price discovery. J. Finance 50 (4), 1175–1199.
Haugom, E., Langeland, H., Moln′ar, P., Westgaard, S., 2014. Forecasting volatility of the U.S. oil market. J. Bank. Finance 47, 1–14.
Herrera, A.M., Hu, L., Pastor, D., 2018. Forecasting crude oil price volatility. Int. J. Forecast. 34 (4), 622–635.
Hitzemann, S., 2016. Macroeconomic Fluctuations, Oil Supply Shocks, and Equilibrium Oil Futures Prices. Working paper.

Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. Int. J. Forecast. 22 (4), 679–688.
Ji, Q., Geng, J.B., Tiwari, A.K., 2018. Information spillovers and connectedness networks in the oil and gas markets. Energy Econ. 75, 71–84.
Jiang, Y., Liu, X., Ye, W., 2015. Do intraday data contain more information for volatility forecasting? Evidence from the Chinese commodity futures market. Appl. Econ. Lett. 22 (3), 218–222.
Johnson, T.C., 2001. Return dynamics when persistence is unobservable. Math. Finance 11 (4), 415–445.
Kang, B., Nikitopoulos, C.S., Prokopczuk, M., 2020. Economic determinants of oil futures volatility: a term structure perspective. Energy Econ. 88, 104743.
Kilian, L., 2009. Not all oil price shocks are alike: disentangling demand and supply shocks in the crude oil market. Am. Econ. Rev. 99 (3), 1053–1069.
Kilian, L., 2016. The impact of the shale oil revolution on U.S. oil and gasoline prices. Rev. Environ. Econ. Pol. 10 (2), 185–205.
Kogan, L., Livdan, D., Yaron, A., 2009. Oil futures prices in a production economy with investment con- straints. J. Finance 64 (3), 1345–1375.
Kruskal, W.H., Wallis, W.A., 1952. Use of ranks in one-criterion variance analysis. J. Am. Stat. Assoc. 47 (260), 583–621.
Lahaye, J., Neely, C., 2018. The role of jumps in volatility spillovers in foreign exchange markets: meteor shower and heat waves revisited. J. Bus. Econ. Stat. 38 (2), 410–427.
Lamoureux, C.G., Lastrapes, W.D., 1994. Endogenous trading volume and momentum in stock-return volatility. J. Bus. Econ. Stat. 12 (2), 253–260.
Laux, P.A., Ng, L.K., 1993. The sources of GARCH: empirical evidence from an intraday returns model incorporating systematic and unique risks. J. Int. Money Finance 12 (5), 543–560.
Li, G., Li, Y., 2015. Forecasting copper futures volatility under model uncertainty. Resour. Pol. 46, 167–176.
Liesenfeld, R., 2001. A generalized bivariate mixture model for stock price volatility and trading volume. J. Econom. 104 (1), 141–178.
Liu, L.Y., Patton, A.J., Sheppard, K., 2015. Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. J. Econom. 187 (1), 293–311.
Luo, J., Ji, Q., Klein, T., Todorova, N., Zhang, D., 2020. On realized volatility of crude oil futures markets: forecasting with exogenous predictors under structural breaks. Energy Econ. 89, 104781.
Lux, T., Kaizoji, T., 2007. Forecasting volatility and volume in the tokyo stock market. J. Econ. Dynam. Control 31 (6), 1808–1843.
Ma, F., Zhang, Y., Huang, D., Lai, X., 2018. Forecasting oil futures price volatility: new evidence from realized range-based volatility. Energy Econ. 75, 400–409.
Mart′ınez, B., Torr′o, H., 2015. European natural gas seasonal effects on futures hedging. Energy Econ. 50, 154–168.
Morana, C., 2013. Oil price dynamics, macro-finance interactions and the role of financial speculation. J. Bank. Finance 37, 206–226.
Müller, U.A., Dacorogna, M.M., Dav′e, R.D., Olsen, R.B., Pictet, O.V., von Weizsäcker, J.E., 1997. Volatilities of different time resolutions - analyzing the dynamics of market components. J. Empir. Finance 4 (2–3), 213–239.
Ng, V.K., Pirrong, S.C., 1994. Fundamentals and volatility: storage, spreads, and the dynamics of metals prices. J. Bus. 67 (2), 203–230.
Nguyen, D.K., Walther, T., 2020. Modeling and forecasting commodity market volatility with long-term economic and financial variables. J. Forecast. 39, 126–142.
Nikitopoulos, C.S., Squires, M., Thorp, S., Yeung, D., 2017. Determinants of the crude oil futures curve: inventory, consumption and volatility. J. Bank. Finance 84, 53–67.
Nonejad, N., 2020. A detailed look at crude oil price volatility prediction using macroeconomic variables. J. Forecast. https://doi.org/10.1002/for.2679.
Palczewski, J., Poulsen, R., Schenk-Hopp′e, K., Wang, H., 2015. Dynamic portfolio optimization with trans- action costs and state-dependent drift. Eur. J. Oper. Res. 243 (3), 921–931.
Pan, Z., Wang, Y., Wu, C., Yin, L., 2017. Oil price volatility and macroeconomic fundamentals: a regime switching GARCH-MIDAS model. J. Empir. Finance 43, 130–142.
Patton, A., 2011. Volatility forecast comparison using imperfect volatility proxies. J. Econom. 160 (1), 246–256.
Patton, A.J., Sheppard, K., 2015. Good volatility, bad volatility: signed jumps and the persistence of volatility. Rev. Econ. Stat. 97, 683–697.
Prokopczuk, M., Stancu, A., Symeonidis, L., 2019. The economic drivers of commodity market volatility. J. Int. Money Finance 98, 102063.
Prokopczuk, M., Symeonidis, L., Simen, C.W., 2017. Variance risk in commodity markets. J. Bank. Finance 81, 136–149.
Robe, M.A., Wallen, J., 2016. Fundamentals, derivatives market information and oil price volatility. J. Futures Mark. 36 (4), 317–344.
Ross, S.A., 1989. Information and volatility: the no-arbitrage martingale approach to timing and resolution irrelevancy. J. Finance 44 (1), 1–17.
S′evi, B., 2014. Forecasting the volatility of crude oil futures using intraday data. Eur. J. Oper. Res. 235 (3), 643–659.
Sharma, P., Vipul, 2016. Forecasting stock market volatility using realized GARCH model: international evidence. Q. Rev. Econ. Finance 59, 222–230.
Silvennoinen, A., Thorp, S., 2013. Financialization, crisis and commodity correlation dynamics. J. Int. Financ. Mark. Inst. Money 24, 42–65.
Suenaga, H., Smith, A., Williams, J., 2008. Volatility dynamics of NYMEX natural gas futures prices. J. Futures Mark. 28 (5), 438–463.
Swanson, N.R., White, H., 1995. A model-selection approach to assessing the information in the term structure using linear models and artificial neural networks. J. Bus. Econ. Stat. 13 (3), 265–275.
Tang, K., Xiong, W., 2012. Index investment and the financialization of commodities. Financ. Anal. J. 68 (6), 54–74.
Wang, J., 2013. Liquidity commonality among Asian equity markets. Pac. Basin Finance J. 21 (1), 1209–1231.
Wang, J., Huang, Y., Ma, F., Chevallier, J., 2020. Does high-frequency crude oil futures data contain useful information for predicting volatility in the U.S. stock market? New evidence. Energy Econ. 91, 104897.
Wang, J.-X., Yang, M., 2018. Conditional volatility persistence. Available at SSRN 3080693. SSRN 3080693. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3080693.
Working, H., 1949. The theory of the price of storage. Am. Econ. Rev. 39, 1254–1262.
Xue, Y., Gençay, R., 2012. Trading frequency and volatility clustering. J. Bank. Finance 36 (3), 760–773.
Zhang, D., Ji, Q., 2018. Further evidence on the debate of oil-gas price decoupling: a long memory approach. Energy Pol. 113, 68–75.
Zhang, D., Shi, M., Shi, X., 2018a. Oil indexation, market fundamentals, and natural gas prices: an inves- tigation of the Asian premium in natural gas trade. Energy Econ. 69, 33–41.
Zhang, Y., Ma, F., Shi, B., Huang, D., 2018b. Forecasting the prices of crude oil: an iterated combination approach. Energy Econ. 70, 472–483.
Zhu, X., Zhang, H., Zhong, M., 2017. Volatility forecasting using high frequency data: the role of after-hours information and leverage effects. Resour. Pol. 54, 58–70.