What drives volatility of the U.S. oil and gas firms?

Štefan Lyócsa[a,1,*], Neda Todorova[b]

[a]*Institute of Financial Complex Systems, Masaryk University, Lipova 41a, 602 00 Brno, Czech Republic; Faculty of Management, University of Presov, Konstantinova 16, 080 01 Presov, Slovakia; Institute of Economic Research, Slovak Academy of Sciences, Sancova 56, 811 05, Slovakia*
[b]*Griffith Business School, Griffith University*
170 Kessels Road, Nathan, Queensland 4111, Australia.
Email: n.todorova@griffith.edu.au
Phone: +61 7 3735 7219, Fax: +61 7 3735 3719

Abstract

We study how the day-ahead stock price volatility of 15 firms that are S&P 500 constituents in the Oil & Gas Exploration & Production sub-industry is driven through six volatility factors represented by realized volatilities, namely, i) firms' own volatility, ii) industry market volatility, iii) local (U.S.) market volatility, iv) world equity market volatility, v) oil price volatility, and vi) natural gas price volatility. Existing studies have reported results based on analysis of one or few volatility components, but given the high dependence among volatility factors, this might bias (overestimate) the true importance of each of the volatility factors on the price fluctuation of stocks in the Oil & Gas Exploration & Production sub-industry. To take into account this inter-relatedness of volatility factors, we study all volatility factors together. Using augmented heterogeneous autoregressive (HAR) models and dynamic model averaging, our analysis shows that market volatility is most influential, followed by a stock's own volatility and industry level volatility. The role of the volatility of the oil market is of lesser importance, while the volatility of the world equity market does not appear to contain incremental information useful for predicting the volatility of firms in the Oil & Gas Exploration & Production sub-industry. The role of the natural gas market is specific. An in-sample analysis suggests a negative relationship between firm-level volatility and volatility on the natural gas market. However, in an out-of-sample framework, the volatility of the natural gas market appears to be unrelated to firm-level volatility. Dynamic model averaging further suggests that the market and industry factors are time-varying. These findings have implications for financial risk management, as we show that in an out-of-sample framework,

*Corresponding author
[1]Lyócsa acknowledges the support provided by the Czech Science Foundation (GACR) under the project number 18-05829S.

HAR models augmented with volatility factors outperform the plain HAR model by up to a 3.88% increase in volatility forecast accuracy.
Keywords: Oil & Gas sub-industry, Volatility forecasting, Volatility factors, HAR, Dynamic Model Averaging

JEL Classification: C5, C53, G17

1. Introduction

During 2010s, the United States oil and gas industry was undergoing an energy production boom, enabled by hydraulic fracturing to extract oil and gas out of shale. As a result, the United States became a net exporter of natural gas in 2017, became the world's biggest oil producer and was running at full speed to build infrastructure to meet domestic and foreign demand for shale oil and gas (Egan, 2018a; Egan, 2018b). However, in the United Nations Paris Climate Agreement signed in 2015, 195 members of the United Nations Framework Convention on Climate Change agreed to hold global warming to no more than 2 degrees above pre-industrial levels, which implies that to avoid excessive carbon emissions and meet this ambitious objective, a large portion of fossil fuel reserves must remain unburnable. Even though the U.S. withdrew from participation in the Paris Agreement in 2017 (and rejoined in 2021) and the U.S. energy industry exhibits no signs of being squeezed by the rise of electric cars and the increasing popularity of renewable energy, it is foreseeable that climate change and environmental regulations will have a significant impact on the industry in the medium to long term (Fox, 2018). Under this scenario, oil and gas producers will face prolonged low commodity prices, a tightening of climate policy with a tough budgeting of cumulative carbon emissions, and technological innovation to produce cheap substitutes for coal, oil and gas (van der Ploeg, 2016). In response to these foreseeable developments, significant academic and industry attention has been dedicated to investor implications of divesting from fossil fuels and hedging against climate risk.

Against the backdrop of these vibrant developments preceded by volatile periods driven by jobs, costs and capital spending restrictions, as well as plummeting oil prices in 2014-2015 (The Economist, 2016), and after the COVID-19 pandemic, there is an imperative for individual and institutional investors who trade with U.S. oil and gas shares to develop a strong understanding of the volatility dynamics of individual companies. This study undertakes an extensive high-frequency data-based analysis at the firm level. To the best of our knowledge, we are the first to shed light on the following questions: Is the volatility of individual oil and gas stocks impacted by the volatility of energy prices and common factors at the industry, U.S. market or global equity market levels? Can this information transmission be translated into significantly enhanced forecasting power for short-term volatility?

The analysis at hand is motivated by an extensive strand of research on the determinants of U.S. and international oil and gas company stock returns, which confirms the significant impact of energy prices and market factors.[2] Kavussanos and Marcoulis (1997) establish that the returns on the S&P 500 have a large impact on the share prices of U.S. refineries. Tjaaland et al. (2016) find that U.S. exploration and production company sub-industries have the highest exposure to both oil and gas price fluctuations. Sadorsky (2001) looks at Canadian oil and gas industry stock prices at the index level and documents that crude oil prices have large and significant impacts on stock price returns in the Canadian oil and gas industry. In a later study using firm-level data on Canadian energy companies, Boyer and Filion (2007) confirm that energy company returns are positively associated with the Canadian stock market return and with appreciations of crude oil and natural gas prices. El-Sharif et al. (2005) document a positive and highly significant relationship between the price of crude oil and equity values in the oil and gas sector using data for the United Kingdom. Li et al. (2017) study the impact of oil price shocks on the stock returns of China's listed companies in the oil industrial chain and conclude that the returns of the listed companies in the whole oil industrial chain benefit from oil price appreciation. Using firm-level monthly data from 70 countries, Gupta (2016) finds that firms located in high oil-producing countries are more sensitive to global uncertainty and oil price shocks. Nandha and Faff (2008) analyze 35 Datastream global industry indices and document that oil price increases have a negative impact on equity returns for all sectors except mining and oil and gas industries. Based on oil and gas industry data in 34 countries, Ramos and Veiga (2011) find evidence that the oil price is a globally priced factor for the oil industry, with oil and gas sectors in developed countries responding more strongly to oil price changes than those in emerging markets.

While these studies provide conclusive evidence that oil and gas companies' returns are significantly linked to energy prices and common market factors, the area of volatility has been studied less, except the role of the oil and aggregate equity market volatility, which can enable statistically and economically significant in-sample and out-of-sample forecasting improvements for equities (e.g., Feng et al., 2017; Zhang et al., 2018). The majority of existing studies focus on volatility interrelationships between oil and aggregate stock market returns (e.g., Arouri et al., 2011; Khalfaoui et al., 2015; Ewing and Malik, 2016; among others). Less attention has been dedicated to the oil and gas sector specifi-

[2] For a comprehensive review of the literature on the relationship between oil and equity markets (not limited to companies from the energy sector), refer to Smyth and Narayan (2018).

cally, and volatility forecasting aspects have not been addressed in detail. Using U.S. oil industry stock indices, Hammoudeh et al. (2003) find the oil futures market's volatility to be strongly linked to the volatility of oil exploration and production companies. Antonakakis et al. (2018) provide strong evidence of volatility co-movements and spillovers among oil prices and the stock prices of 12 of the 25 largest world oil and gas companies. Overall, while the latter study addresses volatility interrelations at the company level, the authors assess the implications of their findings more for diversification benefits and hedge effectiveness and less for volatility forecasting purposes.

This study focuses on the 15 individual S&P 500 constituents of the Oil & Gas Exploration & Production sub-industry, as this sub-industry has been identified as having the highest exposure to oil and gas prices (Tjaaland et al., 2016). Our work contributes to this strand of the literature in three important ways.

First, as far as we are aware, we are the first study to examine how day-ahead volatility of firms in the Oil & Gas Exploration & Production sub-industry is driven by six volatility factors together. Hammoudeh et al. (2003) and Antonakakis et al. (2018) have studied the role of oil prices, while Arouri et al. (2011), Khalfaoui et al. (2015), and Ewing and Malik (2016) have investigated the role of market returns, and Feng et al. (2017) and Zhang et al. (2018) have studied the role of market and oil factors. We contribute to this strand of the literature in that apart from firms' own volatility, market and oil price volatility, as additional information channels, we consider the realized volatilities of gas prices and the equity market at the industry and global levels. Specifically, we investigate whether realized volatilities from these markets enhance the in-sample explanatory power of the volatility model and the extent to which they improve out-of-sample forecasts. We use the current state-of-the-art Heterogeneous Autoregressive (HAR) model for the realized volatility of Corsi (2009). This model has enjoyed overwhelming popularity in recent literature, as it provides a simple yet powerful tool to capture and forecast volatility dynamics at various time horizons.[3] There is strong evidence that the HAR model augmented by economically relevant variables may lead to significant forecasting improvements in both equity (e.g., Todorova and Souček, 2014; Jayawardena et al., 2016;

[3]The HAR model has been extended in various ways, for example, as a continuous-jump specification (Andersen et al., 2007), including leverage effects (Corsi and Renó, 2012), a semi-volatility decomposition (Patton and Sheppard, 2015), or accounting for attenuation bias (Bollerslev et al., 2016). There is a burgeoning number of papers applying a plethora of HAR model specifications to the oil market (e.g., Sévi, 2014; Haugom et al., 2014; Prokopczuk et al., 2016; Wen et al., 2016; Ma et al., 2017; Ma et al., 2018). Recently, HAR model specifications were used to forecast value at risk on energy markets (e.g., Haugom et al., 2016).

Peng et al., 2018) and commodity markets (e.g., Lyócsa et al., 2017; Degiannakis and Filis, 2017). Our choice of volatility factors is strongly motivated by recent literature at the return level, as depicted above. As is well known, volatilities across different assets are highly inter-related (for recent evidence, see Bollerslev et al., 2018). Therefore, studying each of the volatility components separately might bias (overestimate) the true importance of each volatility factor on the price fluctuation of stock prices in the Oil & Gas sub-industry. To take into account this inter-relatedness of volatility factors, we need to study all volatility factors together, and this study is the first to do so.

Second, to provide a stronger empirical basis for our results, we perform in-sample and out-of-sample analysis. By doing this, we can judge whether in-sample results manifest into the improved out-of-sample forecasts of day-ahead volatility, which is of interest to risk and investment management. We apply a filtering technique to extract the unique risk dynamics inherent in each of the considered factors and are hence able to assess the relative importance of the companies' idiosyncratic volatility in an in-sample context.[4] We find that the price variation of firms in the Oil & Gas sub-industry is driven mostly by the development at the U.S. economy level (market factor), followed by idiosyncratic and industry factors. Although the price variation on the oil market matters as well, its magnitude is much smaller. The development on the world equity markets does not seem to contribute to the explanation of the volatility of firms in the Oil & Gas sub-industry. Finally, the effect of the natural gas market is opposite that of other markets, i.e., periods of higher volatility in the natural gas market are associated with lower firm-level volatility on the next day. This result is in line with the well-known fact that the natural gas market is unique in that it is highly volatile and exhibits specific seasonal dynamics. The out-of-sample forecasting evaluation confirmed that it is difficult to utilize information from the natural gas market to explain short-term firm-level volatility. Still, in the out-of-sample study, the benchmark HAR model was outperformed by models that utilized information from remaining markets, including the global equity market. Our results are consistent across firms and confirm that not only do market and oil volatility factors matter (as found in the literature before), but also industry factors, while even the world market volatility factor plays an important role in forecasting the day-ahead volatility of U.S. Oil

[4]By definition, idiosyncratic risk is independent of the common movement of the market (Fu, 2009). The deployed filtering technique used to obtain idiosyncratic volatility accounts not only the equity market (as captured by volatility estimates of the industry, US and the world markets) but also for risk inherent in energy price series.

& Gas sub-industry firms.

Third, we not only study the overall importance of each of the volatility factors but also show how it changes over time, using the dynamic model averaging approach. Thus, we shed light on the persistence of the role of each of the volatility factors. The results suggest that the relative importance of volatility factors can vary considerably over time. The market volatility factor appears to show the highest instability over time, while the role of the stock's own volatility is the most heterogeneous across firms.

The remainder of the paper is structured as follows. Sections 2 presents the data and methodology, respectively, Section 3 reports our empirical results, while Section 4 provides concluding remarks.

2. Data and methodology

2.1. Data source and filtering

We study which systematic market factors drive the price volatility of 15 S&P 500 companies[5] of the sub-industry 'Oil & Gas Exploration & Production' as of December 2017. We consider six systematic market volatility factors to drive the price variation of individual firms. The first is at the individual firm level, the second at the industry level, the third at the local level (U.S. market level, represented by S&P Mini index futures), the fourth at the world equity market level (MSCI world market index), and the fifth and sixth at the oil (WTI futures contracts) and natural gas (Henry Hub futures contracts) levels.

For all six factors, we estimate realized volatility based on high-frequency data obtained from the Thomson Reuters Tick History database available through the Securities Industries Research Centre of Asia Pacific (Sirca). The sample period is January 2007 to December 2017. For the futures contracts, the nearest-month contract is rolled to the next most liquid month when the daily volume of the current contract is exceeded.

We acquired data at the 1 minute calendar sampling frequency, and although the data are of high quality, prior to their use, several sanity filters were applied (see Bollerslev et al., 2018):

[5]Apache Corporation (APA), Anadarko Petroleum Corporation (APC), Chesapeake Energy Corporation (CHK), Cabot Oil & Gas Corporation (COG), ConocoPhillips (COP), Concho Resources Inc. (CXO), Devon Energy Corporation (DVN), EOG Resources Inc. (EOG), EQT Corporation (EQT), Hess Corporation (HES), Marathon Oil Corporation (MRO), Noble Energy, Inc. (NBL), Newfield Exploration Company (NFX), Pioneer Natural Resources Company (PXD), Range Resources Corporation (RRC)

- We use price data associated with least one trade.

- We use price data where the bid-ask spread was less than 1% of the bid price.

- We check if price data are (erroneously) less than zero.

- We use price data where the ask price is above the bid price.

- We use a price reversal filter that eliminates extreme price swings (for details, see Bollerslev et al., 2018, p. 2767).

Because shares of individual firms are traded at a specific time period during the day, the high-frequency data cover only that period of time (09:30 to 16:00 EST). However, for the S&P Mini index, MSCI world market index, and Oil and Natural gas data, we use the full 24-hour time period, beginning at 00:01 EST and ending at 23:59 EST.

2.2. Volatility factor estimates

2.2.1. Realized volatility

The existing literature uses various approaches to volatility estimation based on either calendar or business sampling schemes with various sampling frequencies and specific volatility estimators (e.g., the TSRV of Zhang et al., 2005; MSRV of Zhang, 2006; bi-power variation of Barndorff-Nielsen and Shephard, 2004; range-based estimators of Christensen and Podolskij, 2007; realized kernel estimators of Barndorff-Nielsen, et. al. 2008; and the MedRV and MinRV of Andersen et al., 2012). Liu et al. (2015) run a horse-race of different estimators across different assets and conclude that it is difficult to beat a 5-minute calendar sampling scheme realized volatility estimator for forecasting purposes. Given a trading day t, the realized volatility is defined as

$$RV_{t,5m}^2 = \sum_{j=1}^{N_{t,5m}} r_{t,j,5m}^2, \tag{1}$$

where $r_{t,j,5m}^2$ is the continuous returns within a 5-minute sampling frequency and $N_{t,5m}$ is the number of returns, given the 5-minute sampling frequency. However, Patton and Sheppard (2009) argue that because of the uncertainty related to the 'correct' estimator, a viable estimation strategy is to combine (e.g., average) several estimators into one. Motivated by these results, we use an averaging approach of four realized volatility estimates at the 5, 10, 15, and 30 minute sampling frequencies:[6]

$$RV_t^{2,*} = (RV_{t,5m}^2 + RV_{t,10m}^2 + RV_{t,15m}^2 + RV_{t,30m}^2)/4. \tag{2}$$

2.2.2. Overnight stock price variation

As noted before, at the firm level, high-frequency data are available only during open market times of the NYSE (9:30-16:00 EST). However, for most practical purposes (option and futures contract pricing, VaR, risk management), one is interested in the overall price variation of the asset. To calculate the overall daily volatility of a stock, we add the overnight price variation $OV_t^2 = (100(O_t - C_{t-1})/C_{t-1})^2$ (e.g., Molnár, 2012) to the estimate of the open market time realized volatility ($RV_t^{2,*}$).[7] For the U.S. market and MSCI world indices, as well as for the oil and natural gas futures prices, we have 24-hour data available, and therefore, no overnight price variation adjustment is needed. We follow an approach that is common in literature and convert all volatility estimates using the natural logarithm and we will denote the resulting volatility estimate as RV_t in the subsequent text. For example, for a given firm, the estimator is defined as

$$RV_t = ln(RV_t^2 + OV_t^2). \tag{3}$$

This transformation is known to significantly reduce excessive levels of volatility often observed at the firm level. It also makes the resulting volatility measure more suitable for modeling purposes in an autoregressive model, which tend to be highly sensitive to the presence of extreme values.

Similar to the daily estimates of volatility, the weekly and monthly estimates needed for the estimation of HAR models (as explained below) are obtained as

$$RV_t^W = \frac{1}{5}\sum_{i=0}^{4} RV_{t-i}, \text{ and } RV_t^M = \frac{1}{22}\sum_{i=0}^{21} RV_{t-i}. \tag{4}$$

[6]The $*$ denotes that the realized volatility is calculated for open market times, i.e. for stocks from 09:30 - 16:00 EST, for remaining indices it is a 24 hour window. See next sub-section for further discussion.

[7]The handling of the overnight price variation might give rise to concerns regarding the proper synchronization of the data. In the subsequent analysis, we model firm-level volatility on day $t+1$, which also includes the overnight price variation for which we use the closing price on day t. To explain the volatility on day $t+1$ we will (among others) use data for the S&P Mini, MSCI index, WTI and Natural Gas index for the whole day t; thus, some of the data (after 16:00 until the end of day 23:59) are realized after the closing price at day t and used to calculate the overall firm-level volatility for the next day $t+1$. Therefore, it appears that the information sets of the explained and explanatory variables overlap. However, we argue that this does not lead to look-ahead bias, as the overnight price variation is not known until the opening price on day $t+1$ is observed.

The approach above is applied not only to data on individual stocks but also to data on the S&P Mini index and the MSCI world market index, as well as the price data from the oil and natural gas markets. It remains to define volatility at the industry level, which is simply the cross-sectional average across the 15 stocks of the Oil & Gas sub-industry.[8]

The descriptive statistics of volatility factors and daily returns are reported in Table 1. All series are subject to two stylized facts about volatility: The distribution of daily returns shows a high level of kurtosis (i.e., heavy tails), while the volatility series show a high level of persistence. Even at the 22 lag, the persistence is 0.61 for the industry factor and still 0.15 for the natural gas factor.

Please insert Table 1 around here

2.3. In-sample analysis

2.3.1. Filtration of volatility factors

The resulting series of prices and volatility factors are depicted in Figure 1. Obviously, the U.S. market and world markets seem to be correlated, and the same can be said about the prices of industry, oil and natural gas factors. Additionally, when looking at volatility, all volatility series appear to display a similar pattern. Simple calculation revealed that the average correlation among the five factors (10 correlations) is 0.54 and is always positive (exact values available upon request). While this phenomenon may be of little importance in an out-of-sample study, in the in-sample study, it may distort the true importance of each of the factors. Our in-sample analysis therefore uses filtered volatility factors.

Please insert Figure 1 around here

Let us denote a vector of endogenous realized volatilities as

$$\boldsymbol{y}_t^i = (RV_t^i, IRV_t, MRV_t, WRV_t, ORV_t, GRV_t)^T.$$

[8] We also considered factor models such as extracting the first principal component or from the approximate dynamic factor model as in Bai and Ng (2002), but because the results were similar, while the complexity of the calculations increased considerably, we opted for this simpler approach.

The components are the realized volatilities of the i-th firm, industry, U.S. market, world market, oil market and natural gas market volatility, respectively. Next, the vector

$$\boldsymbol{x}^i = (RV_t^{i,W}, IRV_t^W, MRV_t^W, WRV_t^W, ORV_t^W, GRV_t^W, RV_t^{i,M}, IRV_t^M, MRV_t^M, WRV_t^M, ORV_t^M, GRV_t^M)^T$$

has elements with weakly and monthly volatility estimates. For i-th firm, the filtering equation is a VARX(p, s) model of the form

$$\boldsymbol{y}_t^i = \boldsymbol{\Delta} + \sum_{j=1}^{p} \Phi_j \boldsymbol{y}_{t-j}^i + \sum_{j=1}^{s} \Theta_j \boldsymbol{x}_{t-j}^i + \boldsymbol{\epsilon}_t^i, \tag{5}$$

where the number of lags p was set according to the Bayesian information criterion (Schwarz, 1978) with maximum of $p = 4$. The number of lags s was set to 1. The VARX(p, s) model is estimated for each company, and in the subsequent analysis, we use the residuals $\hat{\boldsymbol{\epsilon}}_t$ standardized by each series's standard deviation as independent variables of the HAR-DMA model described in the next section.[9] For example, if we denote the elements of $\hat{\epsilon}_t^i = (\hat{\epsilon}_t^{i,F}, \hat{\epsilon}_t^{i,I}, \hat{\epsilon}_t^{i,M}, \hat{\epsilon}_t^{i,W}, \hat{\epsilon}_t^{i,O}, \hat{\epsilon}_t^{i,G})^T$, we define the filtered daily standardized volatility factors as $RV_t^{S,i,D} = \hat{\epsilon}_t^{i,F}/\sigma_{\hat{\epsilon}}^{i,F}$ (firm), $IRV_t^{S,i,D} = \hat{\epsilon}_t^{i,I}/\sigma_{\hat{\epsilon}}^{i,I}$ (industry), $MRV_t^{S,i,D} = \hat{\epsilon}_t^{i,M}/\sigma_{\hat{\epsilon}}^{i,M}$ (market), $WRV_t^{S,i,D} = \hat{\epsilon}_t^{i,W}/\sigma_{\hat{\epsilon}}^{i,W}$ (world), $ORV_t^{S,i,D} = \hat{\epsilon}_t^{i,O}/\sigma_{\hat{\epsilon}}^{i,O}$ (oil), and $GRV_t^{S,i,D} = \hat{\epsilon}_t^{i,G}/\sigma_{\hat{\epsilon}}^{i,G}$ (natural gas).[10] Note that daily standardized factors of the individual stock volatilities can be interpreted as the companies' idiosyncratic volatilities because equity and energy market common factors are filtered out of these series.

2.3.2. Volatility factor HAR-DMA: In-sample models

In the case of all volatility factors, the model to estimate is of the following form (we suppress the firm-level index i):

$$\begin{aligned} RV_{t+1} &= \beta_{0,t} + \beta_{1,t} RV_t^W + \beta_{2,t} RV_t^M + \beta_{3,t} RV_t^{S,D} + \beta_{4,t} IRV_t^{S,D} + \beta_{5,t} MRV_t^{S,D} + \\ &\quad \beta_{6,t} WRV_t^{S,D} + \beta_{7,t} ORV_t^{S,D} + \beta_{8,t} GRV_t^{S,D} + \upsilon_t,\ \upsilon_t \sim N(0, H_t). \end{aligned} \tag{6}$$

[9] Estimation was performed using the package R, Pfaff (2008).

[10] It is possible to standardize weekly and monthly volatility components as well. However, these components are not disaggregated and therefore they serve as control variables that control for potential mid- (weekly) and long-term (monthly) volatility levels. Also, by not standardizing these components, the coefficients for monthly and weekly components can be compared with the levels obtained in other studies. For example in a study of 105 individual U.S. stocks, Patton and Sheppard (2015, Table 2B) estimate the fixed effect of the weekly and monthly components to be 0.315 and 0.172, suggesting that for our sample and period of energy stocks, volatility was more persistent.

The weekly (RV_t^W) and monthly (RV_t^M) volatility estimates are included following the popular heterogeneous autoregressive (HAR) model of Corsi (2009), where the lag structure captures the long-memory property of the volatility. The remaining standardized (daily) lagged volatility factors are of interest to this research.[11] The sign and the magnitude of the coefficients allow us to assess the importance of each of the volatility factors for the next day's volatility of firms in the Oil & Gas sub-industry. However, due to the 'model' and 'parameter' uncertainty, not only the exact specification but also the value of coefficients at each time t are unknown. We therefore estimate the model by using the Dynamic Model Averaging (DMA) approach (e.g., Koop and Korobilis, 2012; Catania and Nonejad, 2016) that addresses both sources of uncertainty. We now provide a short description of the DMA model specification that follows the implementation described in Catania and Nonejad (2016).

Given a column vector $\boldsymbol{F}_{m,t}$ of possible combinations of variables as defined above and a corresponding column vector $\boldsymbol{\beta}_{m,t}$, the m-th model is given as

$$\begin{aligned} RV_{t+1} &= \boldsymbol{F}_{m,t}^T \boldsymbol{\beta}_{m,t} + \upsilon_{m,t},\ \upsilon_{m,t} \sim N(0, H_t) \\ \boldsymbol{\beta}_{m,t} &= \boldsymbol{\beta}_{m,t-1} + \eta_{m,t},\ \eta_{m,t} \sim N(0, \boldsymbol{Q}_{m,t}). \end{aligned} \tag{7}$$

The observation error $\nu_{m,t}$ and the evolution error $\eta_{m,t}$ are mutually independent variables. For all t and s ($s \neq t$), ν_t and ν_s are independent, η_t and η_s are independent, and ν_t and η_s are also independent (West and Harrison, 2006). In case of $\boldsymbol{Q}_{m,t} = \boldsymbol{0}$, this specification leads to a time-invariant regression coefficient; otherwise, the regression coefficient(s) are allowed to change over time. The parameters are estimated within the Kalman filter estimation framework. The initial prior distribution for the observational variance is $V_0|D_0 \sim IG(\frac{1}{2}, \frac{\hat{H}_0}{2})$ and for the coefficient vector $\boldsymbol{\beta}_0|D_0 \sim N(0, 100\boldsymbol{I})$, where $\boldsymbol{I}$ is the identity matrix and D_0 is the information set available at time 0.

The magnitude of the random shocks that impact $\boldsymbol{\beta}_t$ is controlled by $0 < \delta \leq 1$. Let $\boldsymbol{R}_t$ denote the prediction variance of $\boldsymbol{\beta}_t$, $\boldsymbol{C_t}$ the estimated variance-covariance matrix

[11] Note that we are using filtered daily components only, not weekly and monthly. Our motivation for this simplified specification is as follows. First, for day-ahead volatility forecasts, the most relevant information is likely incorporated in the most recent price observations. Moreover, applying a filtration to weekly and monthly volatility components (see Section 2.3.1) would lead to a VARX(p, s) model where the number of parameters would increase considerably. Second, in an out-of-sample part of the study, the number of parameters would be very costly for the non-trivial estimation of the DMA, while employing a shrinkage estimator (e.g. LASSO) would not lead to time-varying parameters and would also not account for model choice uncertainty. We therefore opt for a simplified specification that is also considered in other day-ahead volatility forecasting studies (e.g. Patton and Sheppard, 2015; Bollerslev et al., 2016).

of parameters $\boldsymbol{\beta}_t$ and $\hat{H}_t$ the estimator of the observational variance. Within the Kalman recursion, the following holds $\boldsymbol{R}_t = \boldsymbol{C}_{t-1} + \boldsymbol{Q}_t = \boldsymbol{C}_{t-1} + \frac{\boldsymbol{C}_{t-1}(1-\delta)}{\delta} = \delta^{-1}\boldsymbol{C}_{t-1}$. Therefore, as δ lowers, the unpredictability of $\boldsymbol{\beta}_t$ and the time variation of the coefficients increase. In the context of linear autoregressive volatility models, Bollerslev et al. (2016) argue that when volatility is estimated with higher precision, the autoregressive coefficient(s) should be higher, and vice versa. In our setting, it might be that during turbulent time periods, a model with lower δ will be preferred to models with higher δ. The dynamics of H_t are also controlled by another forgetting factor $0 < \gamma \leq 1$, such that the estimate of H_t is given by $\hat{H}_t = (1-\gamma)\sum_{s=0}^{t-1}(\hat{H}_t - s - qe_{t-s}^S)$, where e_{t-s}^S is the standardized prediction error. Catania and Nonejad (2016) suggest that δ and γ are interconnected as both provide dynamics to the coefficients in $\boldsymbol{\beta}_{m,t}$. As a way out, one might fix one of the parameters (we fix the γ) and vary the other (δ). Furthermore, Koop and Korobilis (2012) advocate a more heuristic approach in which different values of δ are considered. We therefore consider a set of values for $\delta \in (0.90, 0.92, ..., 0.99, 1.00)$ and fix the value of $\gamma = 0.96$.

Let m_i denote a model with a given combination of predictors, where all predictor combinations are indexed as $i = 1, 2, ...2^p - 1$, where $K = 2^p - 1$ is the number of all combinations of predictors, and δ_j is a given value of the forgetting factor. Initially, the $p(m_i, \delta_j|D_0)$ is a constant that indicates that each combination of predictors and forgetting factors is equally likely. To obtain the $p(m_i|D_t)$ at time t, new observations are used and probabilities update in the following way:

$$p(m_i|D_t) = \sum_{j=1} p(m_i, \delta_j|D_t)p(\delta_j|D_t), \tag{8}$$

where

$$p(m_i, \delta_j|D_t) = \frac{p(RV_t|m_i, \delta_j, D_{t-1})p(m_i|\delta_j, D_{t-1})}{\sum_{l=1}^{K} p(RV_t|m_l, \delta_j, D_{t-1})p(m_l|\delta_j, D_{t-1})}, \tag{9}$$

$$p(m_i|\delta_j, D_{t-1}) = \frac{p(m_i|\delta_j, D_{t-1})^{\alpha}}{\sum_{l=1}^{K} p(M_l|\delta_j, D_{t-1})^{\alpha}}, \tag{10}$$

$$p(\delta_j|D_t) = \frac{p(RV_t|\delta_j, D_{t-1})p(\delta_j|D_{t-1})}{\sum_{l=1} p(RV_t|\delta_j, D_{t-1})p(\delta_l|F_{t-1})}, \tag{11}$$

$$p(\delta_j|D_{t-1}) = \frac{p(\delta_j|D_{t-1})^\alpha}{\sum_{l=1} p(\delta_l|D_{t-1})^\alpha}. \quad (12)$$

The third forgetting factor is α ($0 < \alpha \leq 1$), which is related to the given model. As α lowers, the weight to the past performance of the model lowers. As recommended by Raftery et al. (2010) and Koop and Korobilis (2012), we set α close to one, namely, to 0.99.

2.4. Out-of-sample models

2.4.1. Volatility factor HAR model specifications

In the out-of-sample part of our study, instead of using the filtered volatility factors, we use estimated realized volatilities directly. Our motivation is that the filtration procedure might induce additional uncertainty around the volatility factors which might be detrimental for forecasting purposes.[12] The benchmark model is

$$RV_{t+1} = \beta_0 + \beta_1 RV_t^D + \beta_2 RV_t^W + \beta_3 RV_t^M + \upsilon_t. \quad (13)$$

The HAR model of Corsi (2009) is popular because it captures the stylized facts of volatility well, it is easy to estimate and to enhance while it tends to provide more accurate forecasts than the GARCH model[13] (e.g., Horpestad et al., 2018). We first adjust the HAR model by simply adding one volatility factor at a time, leading to the following five models, the Industry factor HAR, IF-HAR,

$$RV_{t+1} = \beta_0 + \beta_1 RV_t^D + \beta_2 IRV_t^D + \beta_3 RV_t^W + \beta_4 RV_t^M + \upsilon_t, \quad (14)$$

the Market factor HAR, MF-HAR,

$$RV_{t+1} = \beta_0 + \beta_1 RV_t^D + \beta_2 MRV_t^D + \beta_3 RV_t^W + \beta_4 RV_t^M + \upsilon_t, \quad (15)$$

[12]Indeed, we have also tried to use the filtered volatility factors directly, but as it turned out, all models using the filtered volatility factors were less accurate than the benchmark HAR model of Corsi (2009).

[13]Promising approaches along the lines of GARCH models are those that mix the latent volatility factor model with realized volatilities, as in Hansen et al. (2012), a mixture memory GARCH model of Klein and Walther (2016) or by incorporating macroeconomics factors available at lower sampling frequencies in a GARCH-MIDAS framework as in Nguyen and Walther (2020). Combining forecasts based on such approaches with the forecasts obtained with HAR models is left to future research.

the World factor HAR, WF-HAR,

$$RV_{t+1} = \beta_0 + \beta_1 RV_t^D + \beta_2 WRV_t^D + \beta_3 RV_t^W + \beta_4 RV_t^M + \upsilon_t, \tag{16}$$

the Oil factor HAR, OF-HAR,

$$RV_{t+1} = \beta_0 + \beta_1 RV_t^D + \beta_2 ORV_t^D + \beta_3 RV_t^W + \beta_4 RV_t^M + \upsilon_t, \tag{17}$$

and the Natural gas factor HAR, GF-HAR,

$$RV_{t+1} = \beta_0 + \beta_1 RV_t^D + \beta_2 GRV_t^D + \beta_3 RV_t^W + \beta_4 RV_t^M + \upsilon_t. \tag{18}$$

Finally, we estimate HAR model with all factors, the all factor HAR, AF-HAR,

$$\begin{aligned} RV_{t+1} = \beta_0 + \beta_1 RV_t^D + \beta_2 IRV_t^D + \beta_3 MRV_t^D + \beta_4 WRV_t^D + \beta_5 ORV_t^D + \\ \beta_6 GRV_t^D + \beta_7 RV_t^W + \beta_8 RV_t^M + \upsilon_t. \end{aligned} \tag{19}$$

With respect to the estimation procedure, we rely on a weighted least squares approach in which weights are set in a way that accounts for a structural break with an unknown date (see Pesaran et al., 2013; eq. 46 - 48).[14]

2.4.2. Out-of-sample algorithm

We use a rolling window estimation forecasting scheme with an estimation window of 1000 observations. Our motivation for using a rolling-window scheme instead of the expanding window scheme is twofold. First, a rolling window is more suited in a situation where structural breaks in coefficients should be expected. Second, we use the model confidence set of Hansen et al. (2011) to statistically compare the forecast accuracy of models, and as the test requires a local stationarity of the loss series, the rolling window approach is more suitable. The motivation to use 1000 observations in the moving window is also twofold. First, the existing forecasting literature has often used a size of 1000 observations (e.g., Sévi, 2014; Lyócsa and Molnár, 2018). Second, as argued in Lyócsa and Molnár (2018) and Narayan and Gupta (2015), several years are needed to mitigate the possible effect of seasonality on the oil and natural gas market.

[14] For a recent application of this approach in the volatility forecasting literature, see Lyócsa and Molnár (2018).

2.4.3. Forecast evaluation

We evaluate forecasts using two popular loss functions, namely, the square forecast error (SFE) and the QLIKE, which are both advocated by Patton (2011) as leading to consistent model rankings even in the presence of a noisy proxy. While SFE is symmetric, giving equal loss to forecast under- and over-estimations, the QLIKE penalizes forecast under-estimations more. Let RV_t denote the observed realized volatility (our proxy for the true volatility) and RV_t^F the predicted volatility. The SFE and QLIKE are defined as

$$SFE_t = (RV_t - RV_t^F)^2 \tag{20}$$

and

$$QLIKE_t = \frac{RV_t}{RV_t^F} - ln\left(\frac{RV_t}{RV_t^F}\right) - 1, \tag{21}$$

respectively. The overall model performance is evaluated using the time average of the SFE_t (denoted as MSFE) and $QLIKE_t$. Statistical forecast evaluation is conducted using the model confidence set of Hansen et al. (2011), where we test which models (for each stock separately) belong to the set of superior models with an 90% confidence, a value usually employed in the literature (e.g., Hansen et al., 2011; Wang et al., 2016; Meng et al., 2018). We use the t_{max} version of the test and the routines provided by Bernardi and Catania (2018).

3. Results

3.1. Volatility drivers: in-sample analysis

We estimate the volatility factor HAR model (see Section 2.3.2) for each stock separately. However, as results across stocks are very similar, instead of reporting individual stock-level results, we aggregate the results by averaging coefficients first across time (for each stock separately) and then across stocks (see Table 2). By averaging across time, we are interested in the expected effects of each component over time, while subsequent averaging across stocks stocks allows us to observe expected effects of components on the next day's volatility of Oil & Gas sub-industry stocks. Time-variation of coefficients can be observed in Figure 2, while variation across individual stocks is presented in Table A1 (Appendix). As the volatility factors were first filtrated using the VARX(p, s) model and consequently standardized to have a unit variance and zero mean, the magnitude of coefficients is comparable across volatility factors.

Please insert Table 2 around here

We initially expected that the most important volatility driver will be the stock's own volatility, followed by industry, market and world factors. However, the results in Table 2 suggest that the price variation of firms in the Oil & Gas sub-industry is driven mostly by the development at the whole economy level (market factor, 0.040) followed by stock (0.036) and industry factors (0.031) that are similar in magnitude. Interestingly, although the price variation on the oil (0.001) and natural gas (-0.015) markets matters as well, the magnitude is much smaller, while in the case of the Natural gas market, the association is negative, i.e., higher levels of volatility in the natural gas market tend to precede smaller firm-level volatility. Finally, the results for the world factor are mixed, as for 47% of firms, the coefficient was positive, and overall, the magnitude of the effect was small. The importance of the market volatility is in line with existing studies (e.g., Feng et al., 2017; Zhang et al., 2018), as well as the role of the oil price fluctuation (e.g., Hammoudeh et al., 2003; Antonakakis et al., 2018). However, the role of the price fluctuations on the world equity, industry and natural gas markets is new to the literature and shows that firm-level volatility is driven not only by market and oil factors but also by industry and stock idiosyncratic factors, while it is rather weakly linked to the development of the natural gas and global equity markets. Moreover, our results allow us to compare the relative importance of these factors with respect to one another.

Please insert Figure 2 around here

The HAR-DMA model allows us to observe the possible time-varying nature of volatility drivers. In Figure 2, we plot the time-varying posterior inclusion probabilities (PIP), which are averaged across firms. Three additional interesting observations are made. First, the industry and, particularly, the market factor show large variation in PIPs. Therefore, it appears that although the market factor has the strongest impact on firm-level volatility, this influence is surely not constant over time, as the market factor's PIP ranges from around 0.20 to 0.75. The industry factor appears to play a more prominent role during the financial crisis, and the stock and industry factors also appear to drive firm-level volatility during 2015, when oil prices started to decline. Second, we find that the oil and world factors appear to be less volatile over time, as the PIPs vary little.

This result may be explained by the fact that we study daily levels of firm volatility, while energy prices tend to influence firms in the Oil & Gas sub-industry only if they remain relatively (compared to past values) high or small over a longer period of time. Due to the variability of commodity prices, contracts made by firms may have fixed prices in advance, reducing the variation in the future cash flows of such firms and, consequently, the stock price variation. Third, even the natural gas factor might be of considerable importance, suggesting that to model the volatility of firms in the Oil & Gas sub-industry, one needs to take into account model uncertainty, as different volatility factors drive volatility over different time periods. All these results are consistent across firms, as the variability of PIPs across firms seems to be small (note that the shaded area shows the variability of PIP across firms), although the idiosyncratic volatility factor clearly shows the highest variation across firms. This finding is expected, as by definition, idiosyncratic volatility should be more unique to each firm, while common volatility factors are not.

3.2. Volatility drivers: out-of-sample analysis

The in-sample results reported above give an indication about which volatility factors are important drivers of the price variation of Oil & Gas sub-industry firms. However, firm-level (unfiltered) realized volatilities might also subsume information from other markets. In that case, the simple HAR model should out-perform models that also use realized volatilities from other markets, as the need to estimate additional parameters without adding new information usually increases the forecast error. On the other hand, if realized volatility in the industry, economy, world, oil and natural gas markets contains additional information, augmented HAR models might outperform the benchmark HAR model.

Note that in contrast to those used in the in-sample study, the HAR models used in the out-of-sample analysis include non-filtered (raw) realized volatilities (see Section 2.5.1 for details). Results from the out-of-sample analysis are aggregated across firms in Table 3.[15]

Several interesting observations are made. First, regardless of the loss function employed, except of the GS-HAR (natural gas factor) model, all models outperformed the benchmark HAR model. This result is consistent across firms and further cements our empirical evidence that not only the market volatility factor (Feng et al., 2017 and

[15]Results for each firm separately are available in Appendix 2.

Zhang et al., 2018) but also the industry, world and oil market volatility factors matter.

Please insert Table 3 around here

Second, we find that when a statistically valid comparison is undertaken, then only the IF-HAR, MF-HAR, and AF-HAR models tend to systematically out-perform the benchmark HAR model. The most accurate forecasts are provided using the AF-HAR model, i.e., when all volatility factors are used together. This result provides strong statistical evidence in an out-of-sample framework that not only the market volatility but also other volatility factors contain useful information with regard to a firm's next-day volatility.

Third, comparing the results of the MSE and QLIKE loss function, we see that forecast improvements are higher for QLIKE losses. The average improvements ranged from -0.52% (WF-HAR, World factor model) to -3.72% (AF-HAR, All volatility factor model) for the MSE loss function and were even slightly better when volatility underestimation was of higher concern, as is the case of the QLIKE loss function, ranging from -0.77% (WF-HAR) to -3.88% (AF-HAR). These results suggest that volatility factors are of particular interest to risk managers, as models with volatility factors improve forecasts when it matters more, i.e., when volatility underestimation is of concern.

4. Conclusion

Market and oil prices play an important role in the development of share prices of firms in the Oil & Gas sub-industry (e.g., Kavussanos and Marcoulis, 1997; Tjaaland et al., 2016; Sadorsky, 2001; Boyer and Filion, 2007; Li et al., 2017; Nandha and Faff, 2008; Ramos and Veiga, 2011). Much less is known about the role of volatility on equity and oil markets with respect to the volatility of share prices in the Oil & Gas sub-industry (e.g., Hammoudeh et al., 2003; Feng et al., 2017; Zhang et al., 2018; Antonakakis et al., 2018). We contribute to this strand in the literature by studying how the day-ahead volatility of 15 firms that are S&P 500 constituents of the Oil & Gas sub-industry depends on six volatility factors represented by realized volatilities, namely, i) a firm's own volatility, ii) industry market volatility, iii) local (U.S.) market volatility, iv) world equity market volatility, v) oil price volatility, and vi) natural gas price volatility.

Using the augmented HAR model of Corsi (2009) and the DMA framework, we found that with respect to a firm's day-ahead volatility, the market volatility factor appears to be the most important, followed by a firm's own and industry-level volatilities. Contrary to previous studies, we found that the role of the fluctuations on the oil market is of lesser importance. We further confirm that the role of the market and industry volatility factors exhibits considerable variation over time. On the other hand, the role of the oil price fluctuations is rather constant. Finally, we show that these results can be utilized to improve volatility forecasts in an out-of-sample framework. A HAR model that uses all volatility factors decreases forecast error upon a benchmark HAR model of Corsi (2009) by up to -3.88%. Moreover, our analysis reveals that volatility factors are more useful when volatility underestimation is of higher concern, i.e., our results have important implications for risk and investment management practices.

References

References

Antonakakis, N., Cunado, J., Filis, G., Gabauer, D., Perez de Gracia, F., 2018. Oil volatility, oil and gas firms and portfolio diversification. Energy Economics 70, 499–515.

Andersen, T. G., Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review 39, 885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X., 2007. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. Review of Economics and Statisics 89, 701–720.

Andersen, T. G., Dobrev, D., Schaumburg, E., 2012. Jump-robust volatility estimation using nearest neighbor truncation. Journal of Econometrics, 169, 75–93.

Arouri, M. E. H., Jouini, J., Nguyen, D. K., 2011. Volatility spillovers between oil prices and stock sector returns: implications for portfolio management. Journal of International Money and Finance 30, 1387–1405.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191–221.

Barndorff-Nielsen, O. E., Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps. Journal of Financial Econometrics 2, 1–37.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., Shephard, N., 2008. Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. Econometrica 76, 1481–1536.

Bernardi, M., Catania, L., 2018. The model confidence set package for R. International Journal of Computational Economics and Econometrics 8, 144-158.

Bollerslev, T., Patton, A. J., Quaedvlieg, R., 2016. Exploiting the errors: A simple approach for improved volatility forecasting. Journal of Econometrics 192, 1–18.

Bollerslev, T., Hood, B., Huss, J., Pedersen, L. H., 2018. Risk everywhere: Modeling and managing volatility. The Review of Financial Studies 31, 2729-2773.

Boyer, M. M., Filion, D., 2007. Common and fundamental factors in stock returns of Canadian oil and gas companies. Energy Economics 29, 428–453.

Bubák, V., Kočenda, E., Žikeš, F., 2011. Volatility transmission in emerging European foreign exchange markets. Journal of Banking and Finance 35, 2829–2841.

Catania, L., Nonejad, N., 2016. Dynamic model averaging for practitioners in economics and finance: The eDMA package. arXiv preprint arXiv:1606.05656.

Christensen, K., Podolskij, M., 2007. Realized range-based estimation of integrated variance. Journal of Econometrics, 141, 323–349.

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics 7, 174–196.

Corsi, F., Renó, R., 2012. Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. Journal of Business & Economic Statistics, 368–380.

Degiannakis, S., Filis, G., 2017. Forecasting oil price realized volatility using information channels from other asset classes. Journal of International Money and Finance 79, 28–49.

Egan, M., 2018a. Texas Gulf Coast exports more oil than it imports for the first time. CNN Money. Retrieved from: https://money.cnn.com/2018/08/23/investing/oil-exports-texas-houston/index.html.

Egan, M., 2018b. American oil refineries are working harder than ever before. CNN Money. Retrieved from https://money.cnn.com/2018/08/14/investing/oil-refinery-gasoline-demand/index.html.

El-Sharif, I., Brown, D., Nixon, B., Russell, A., 2005. Evidence on the nature and extent of the relationship between oil prices and equity values in the UK. Energy Economics, 27, 819–830.

Ewing, B. T., Malik, F., 2016. Volatility spillovers between oil prices and the stock market under structural breaks. Global Finance Journal 29, 12–23.

Feng, J., Wang, Y., Yin, L., 2017. Oil volatility risk and stock market volatility predictability: Evidence from G7 countries. Energy Economics 68, 240–254.

Fox, J., 2018. A dark side of the shale boom. Bloomberg Opinion. Retrieved from https://www.bloomberg.com/view/articles/2018-03-08/there-s-a-dark-side-to-the-american-shale-boom.

Fu, F., 2018. Idiosyncratic risk and the cross-section of expected stock returns. Journal of Financial Economics 91, 24–37.

Gupta, K., 2016. Oil price shocks, competition and oil & gas stock returns – global evidence. Energy Economics 57, 140–153.

Hammoudeh, D., Dibooglu, S., Aleisa, E., 2003. Relationships among U.S. oil prices and oil industry equity indices. International Review of Economics & Finance 15, 1–29.

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

Hansen, P. R., Huang, Z., Shek H. H., 2012. Realized GARCH: a joint model for returns and realized measures of volatility. Journal of Applied Econometrics 27, 877–906.

Haugom, E., Langeland, H., Molnàr, P., Westgaard, S., 2014. Forecasting volatility of the US oil market. Journal of Banking & Finance, 47, 1–14.

Haugom, E., Ray, R., Ullrich, C. J., Veka, S., Westgaard, S., 2016. A parsimonious quantile regression model to forecast day-ahead value-at-risk. Finance Research Letters, 16, 196–207.

Horpestad, J. B., Lyócsa, S., Molnár, P., Olsen, T. B., 2018. Asymmetric volatility in equity markets around the world. The North American Journal of Economics and Finance.

Jayawardena, N. I., Todorova, N., Li, B., Su, J.-J., 2016. Forecasting stock volatility using after-hour information: Evidence from the Australian Stock Exchange. Economic Modelling 52, 592–608.

Kavussanos, M. G., Marcoulis, S. N., 1997. The stock market perception of industry risk and microeconomic factors: The case of the US water transportation industry versus other transport industries. Transportation Research. Part E, Logistics and Transportation Review 33, 147–158.

Khalfaoui, R., Boutahar, M., Boubaker, H., 2015. Analyzing volatility spillovers and hedging between oil and stock markets: evidence from wavelet analysis. Energy Economics 49, 540–549.

Klein, T., Walther, T., 2016. Oil price volatility forecast with mixture memory GARCH. Energy Economics, 58, 46–58.

Koop, G., Korobilis, D., 2012. Forecasting inflation using dynamic model averaging. International Economic Review, 53, 867–886.

Li, Q., Cheng, K., Yang, Z., 2017. Response pattern of stock returns to international oil price shocks from the perspective of China's oil industrial chain. Applied Energy 185, 1821–1831.

Liu, Ch., Maheu, J. M., 2009. Forecasting realized volatility: a Bayesian model-averaging approach. Journal of Applied Econometrics 24, 709–733.

Liu, L. Y., Patton, A. J., Sheppard, K., 2015. Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. Journal of Econometrics, 187, 293–311.

Lyócsa, Š., Molnàr, P., Todorova, N., 2017. Volatility forecasting of non-ferrous metal futures: Covariances, covariates or combinations? Journal of International Financial Markets, Institutions & Money 51, 228–247.

Lyócsa, Š., Molnár, P., 2018. Exploiting dependence: Day-ahead volatility forecasting for crude oil and natural gas exchange-traded funds. Energy 155, 462–473.

Ma, F., Wahab, M., Huang, D., Xu W., 2017. Forecasting the realized volatility of the oil futures market: A regime switching approach. Energy Economics 67, 136–145.

Ma, F., Wei, Y., Liu, L., Huang D., 2018. Forecasting realized volatility of oil futures market: A new insight. Journal of Forecasting 37, 419–436.

Meng, X., Taylor, J. W., 2018. An approximate long-memory range-based approach for value at risk estimation. International Journal of Forecasting 34, 377–388.

Molnár, P. 2012. Properties of range-based volatility estimators. International Review of Financial Analysis, 23, 20–29.

Nandha, M., Faff, R., 2008. Does oil move equity prices? A global view. Energy Economics 30, 986–997.

Narayan, P. K., Gupta, R., 2015. Has oil price predicted stock returns for over a century?. Energy Economics, 48, 18–23.

Nguyen, D. K., Walther, T., 2020. Modeling and forecasting commodity market volatility with long-term economic and financial variables. Journal of Forecasting 39, 126–142.

Patton, A. J., Sheppard, K., 2009. Optimal combinations of realised volatility estimators. International Journal of Forecasting, 25, 218–238.

Patton, A. J., 2011. Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics, 160, 246–256.

Patton, A. J., Sheppard, K., 2015. Good volatility, bad volatility: Signed jumps and the persistence of volatility. Review of Economics and Statistics 97, 683–697.

Peng H., Chen R., Mei, D., Diao X., 2018. Forecasting the realized volatility of the Chinese stock market: Do the G7 stock markets help? Physica A 501, 78–85.

Pesaran, M. H., Pick, A., Pranovich, M., 2013. Optimal forecasts in the presence of structural breaks. Journal of Econometrics, 177, 134–152.

Pfaff, B., 2008. VAR, SVAR and SVEC Models: Implementation Within R: package vars. Journal of Statistical Software, 27.

Prokopczuk, M., Symeonidis, L., Simen, C. W., 2016. Do jumps matter for volatility forecasting? Evidence from energy markets. Journal of Futures Markets 36, 758–792.

Raftery, A. E., Kárný, M., Ettler, P., 2010. Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. Technometrics, 52, 52–66.

Ramos, S. B., Veiga, H., 2011. Risk factors in oil and gas industry returns: international evidence. Energy Economics 33, 525–542.

Sadorsky, P., 2001. Risk factors in stock returns of Canadian oil and gas companies. Energy Economics 23, 17–28.

Sévi, B., 2014. Forecasting the volatility of crude oil futures using intraday data. European Journal of Operational Research 235, 643–659.

Schwarz, G., 1978. Estimating the dimension of a model. The Annals of Statistics, 6, 461–464.

Smyth, R., Narayan, P. K., 2018. What do we know about oil prices and stock returns? International Review of Financial Analysis 57, 148–156.

The Economist, 2016. Oil Companies: In the Dark Ages. Retrieved from: http://tinyurl.com/zu8nrlm.

Tjaaland, SH., Westgaard, S., Osmundsen, P., Frydenberg, S., 2016. Oil and gas risk factor sensitivities for U.S. energy companies. The Journal of Energy and Development 41, 135–173.

Todorova, N., Souček, M., 2014. Overnight information flow and realised volatility forecasting. Finance Research Letters 11, 420–428.

van der Ploeg, F., 2016. Fossil fuel producers under threat. Oxford Review of Economic Policy 32, 206–222.

Wang, Y., Wu, Ch., Yang, L., 2016. Forecasting crude oil market volatility: A Markov switching multifractal volatility approach. International Journal of Forecasting 32, 1–9.

Wen, F., Gong, X., Cai, S., 2016. Forecasting the volatility of crude oil futures using HAR-type models with structural breaks. Energy Economics 59, 400–413.

West, M., Harrison, J., 2006. Bayesian forecasting and dynamic models. Springer Science and Business Media.

Zhang, L., Mykland, P. A., Ait-Sahalia, Y., 2005. A tale of two time scales determining integrated volatility with noisy high-frequency data. Journal of the American Statistical Association 100, 1394–1411.

Zhang, L., 2006. Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. Bernoulli 12, 1019–1043.

Zhang, Y., Ma, F., Liao, Y., Shi, B., 2018. Forecasting global equity market volatilities. Working Paper. Available on Research Gate.

Tables and figures

Figure 1: Prices and volatility factors

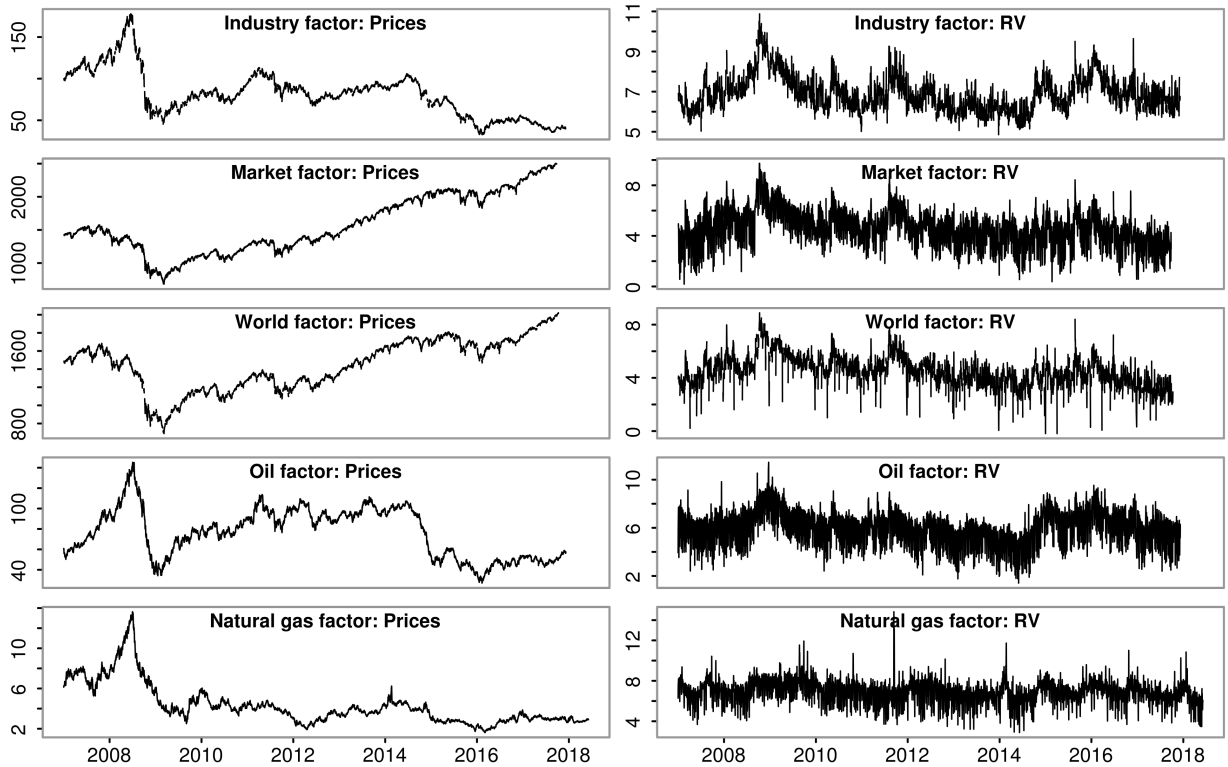
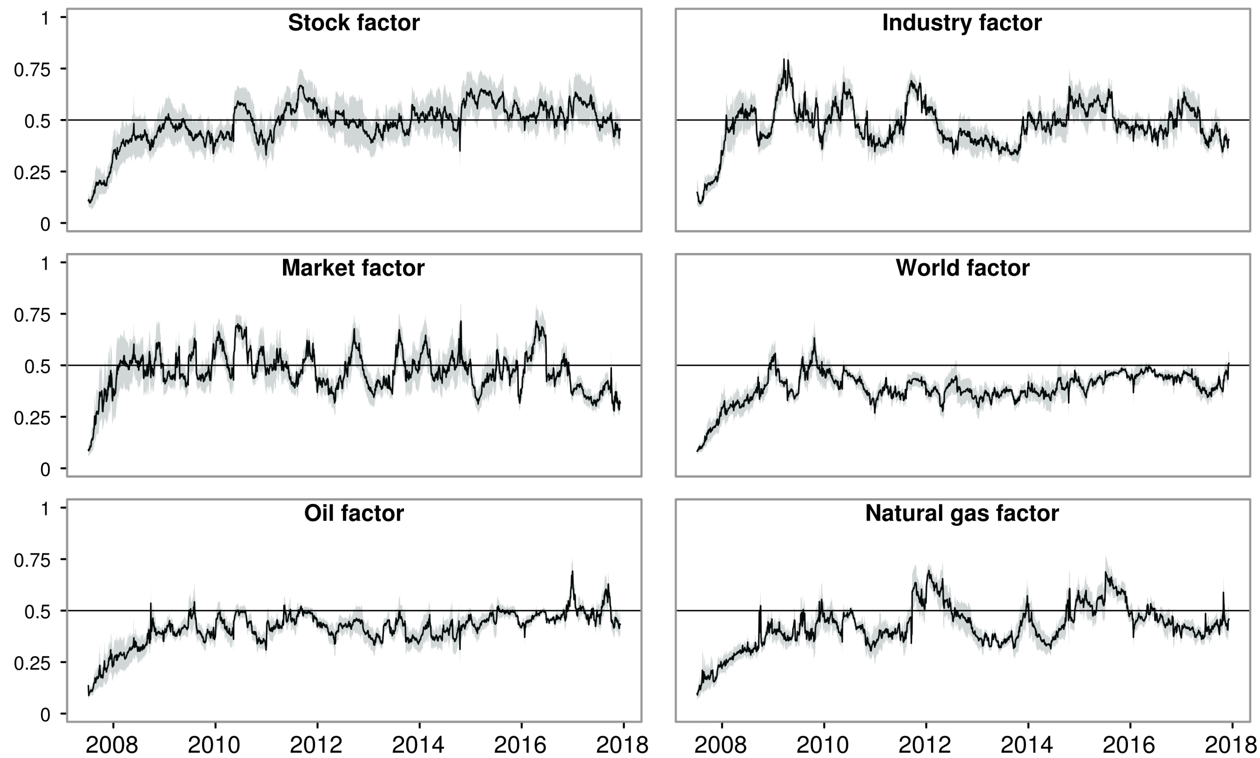


Figure 2: Posterior inclusion probabilities: Averages across firms



Note: PIP are on the y-axis, time on the x-axis. Shaded area highlight spans $\pm 2 \times SD$ calculated from 15 coefficients across firms.

Table 1: Summary statistics of realized volatilities and daily returns

	Mean	SD	Skew	Kurt	AC1	AC5	AC22
Panel A: Realized volatility							
Industry	6.88	0.86	0.88	4.03	0.80	0.73	0.61
Market	4.83	1.35	-0.07	3.50	0.59	0.52	0.36
World	4.47	1.15	0.13	3.75	0.75	0.64	0.49
Oil	6.31	1.29	-0.50	3.80	0.42	0.39	0.29
Natgas	7.01	0.99	-0.24	6.12	0.33	0.31	0.15
Panel B: Daily returns							
Industry	-0.0003	0.02	-0.83	12.42	-0.05	-0.03	0.03
Market	0.0002	0.01	-0.49	17.05	-0.06	0.02	-0.02
World	0.0001	0.01	-0.47	12.20	0.10	-0.05	-0.01
Oil	0.0000	0.02	0.26	9.64	-0.04	0.05	0.00
Natgas	-0.0002	0.03	0.41	7.57	-0.03	0.00	-0.03

Note: SD denotes standard deviation, AC1, AC5, and AC22 the first, fifth and twenty-second autocorrelation coefficients, respectively.

Table 2: Average estimated coefficients from the HAR model with (filtered) volatility factors

	Coef	SE	T-ratio	%positive
Panel A: Standard HAR model variables				
Intercept	2.351	0.140	16.770	100.0%
Weekly volatility	0.448	0.015	29.860	100.0%
Monthly volatility	0.189	0.016	11.790	100.0%
Panel B: Daily lagged volatility factors (unit variance standardized variables)				
Stock	0.036	0.004	9.280	100%
Industry	0.031	0.002	17.110	100%
Market	0.040	0.004	11.510	100%
World	0.000	0.001	-0.310	47%
Oil	0.001	0.002	0.480	67%
Natural gas	-0.011	0.001	-12.250	0%

Note: Values in the column 'Coef' are two-times averages: first, for each stock, we obtained an average value of the filtered estimate of $\theta_{t,i}$ across the time dimension (t); second, we took the averages across stocks (i). Column 'SE' denotes the standard error calculated from averaged coefficients across stocks. 'T-ratio' is the corresponding ratio of the column 'Coef' and 'SE' ,and '% positive' is the proportion of positive coefficients across stocks (after averaging for each stock across the time dimension).

Table 3: Out-of-sample forecast evaluation for standard HAR models aggregated across firms

		Industry	Market	World	Oil	Natgas	All factors
	HAR	IF-HAR	MF-HAR	WF-HAR	OF-HAR	GF-HAR	AF-HAR
	A	B	C	D	E	F	G
Panel A: MSFE							
Proportion [in %] of cases in which the column model outperformed the benchmark HAR model							
		100	100	66.7	100	20	100
Average forecast improvement (negative value) in % of the column model against the benchmark HAR model							
		-3.02	-2.28	-0.52	-1.46	0.11	-3.72
Proportion [in %] of cases in which the column model was part of the superior set of models							
	0	100	80	0	46.7	0	93.3
Panel B: QLIKE							
Proportion [in %] of cases in which the column model outperformed the benchmark HAR model							
		100	100	93.3	100	20	100
Average forecast improvement (negative value) in % of the column model against the benchmark HAR model							
		-3.16	2.52	0.77	1.38	0.16	3.88
Proportion [in %] of cases in which the column model was part of the superior set of models							
	0	100	86.7	6.7	40	0	93.3

Note: XXX

Table A1: In-sample level averages across the time domain of the estimated coefficients from HAR model with (filtered) volatility factors

	APA	APC	CHK	COG	COP	CXO	DVN	EOG	EQT	HES	MRO	NBL	NFX	PXD	RRC
Panel A: Standard HAR model variables															
Intercept	1.3241	1.2747	1.4413	1.2290	1.1396	1.3948	1.7696	1.1412	1.2482	1.5022	1.3419	1.4071	1.3116	1.3774	1.3591
Weekly volatility	0.6551	0.6724	0.5878	0.5963	0.6651	0.5159	0.5777	0.6723	0.5964	0.5753	0.6366	0.6392	0.5672	0.6358	0.5828
Monthly volatility	0.1230	0.1169	0.1828	0.2078	0.1286	0.2749	0.1215	0.1383	0.1860	0.1772	0.1435	0.1222	0.2244	0.1401	0.2025
Panel B: Daily lagged volatility factors (unit variance standardized variables)															
Stock	0.0348	0.0356	0.0504	0.0392	0.0257	0.0274	0.0256	0.0274	0.0429	0.0283	0.0619	0.0137	0.0384	0.0397	0.0218
Industry	0.0145	0.0109	0.0254	0.0215	0.0182	0.0256	0.0188	0.0230	0.0184	0.0195	0.0116	0.0226	0.0230	0.0187	0.0266
Market	0.0317	0.0270	0.0221	0.0374	0.0509	0.0170	0.0271	0.0374	0.0385	0.0399	0.0289	0.0399	0.0253	0.0294	0.0473
World	0.0054	0.0105	0.0087	0.0157	0.0094	0.0124	0.0017	0.0026	0.0098	0.0048	0.0043	0.0126	0.0053	0.0095	0.0104
Oil	0.0057	0.0080	0.0049	0.0072	0.0080	0.0009	-0.0008	0.0053	0.0025	-0.0086	0.0125	-0.0021	0.0037	-0.0071	0.0040
Natural gas	-0.0071	-0.0071	0.0010	-0.0048	-0.0072	-0.0006	-0.0032	-0.0071	-0.0101	0.0042	-0.0024	-0.0040	-0.0039	0.0012	-0.0063

Note: Values in the table are averages obtained as an average value of the filtered estimate of $\theta_{t,i}$ across the time dimension (t).

Table A2: Out-of-sample forecast evaluation of standard models at the firm level

	Code	APA	APC	CHK	COG	COP	CXO	DVN	EOG	EQT	HES	MRO	NBL	NFX	PXD	RRC
Panel A: MSFE																
HAR	A	0.406	0.406	0.474	0.476	0.443	0.385	0.390	0.450	0.393	0.439	0.429	0.439	0.410	0.421	0.362
IF-HAR	B	-2.4%†	-1.7%†	-2.9%†	-5.1%†	-1.6%	-1.8%	-1.9%	-5.4%†	-1.7%	-2.5%†	-4.0%†	-5.7%†	-2.9%†	-2.6%	-3.2%†
MF-HAR	C	-3.2%†	-2.1%†	-1.8%†	-4.9%†	-3.8%†	-3.1%†	-3.2%†	-5.5%†	-3.3%†	-3.2%†	-4.0%†	-3.4%†	-2.8%†	-4.0%†	-2.6%†
WF-HAR	D	-0.7%	-0.5%	-0.7%	-2.9%	-0.4%	-0.8%	-0.4%	-2.6%	-0.4%	-0.8%	-1.6%	-1.9%	-1.2%	-1.4%	-0.9%
OF-HAR	E	-2.5%†	-2.5%†	-1.4%†	-4.5%†	-2.6%†	-1.5%	-1.4%	-4.2	-0.7%	-2.1%†	-3.4%†	-3.0%†	-2.8%†	-2.1%†	-2.0%†
GF-HAR	F	0.2%	0.2%	0.3%	-0.3%	0.3%	0.2%	0.3%	0.0%	0.4%	0.3%	0.2%	0.2%	0.3%	0.3%	0.3%
AF-HAR	G	-4.6%†	-3.3%†	-2.2%†	-6.6%†	-4.6%†	-3.6%†	-3.9%†	-7.9%†	-3.2%†	-3.9%†	-6.2%†	-6.7%†	-3.8%†	-4.4%†	-3.3%†
Panel B: QLIKE																
HAR	A	0.459	0.442	0.421	0.456	0.592	0.397	0.434	0.484	0.452	0.481	0.449	0.471	0.382	0.428	0.364
IF-HAR	B	-2.9%†	-1.5%†	-3.5%†	-4.8%†	-2.0%	-1.6%	-2.7%†	-4.9%	-2.3%	-3.0%	-4.3%	-5.3%	-3.2%†	-2.5%	-3.7%†
MF-HAR	C	-4.1%†	-2.6%†	-2.6%†	-5.3%†	-4.5%†	-3.6%†	-4.1%†	-6.4%†	-4.9%†	-4.3%†	-4.9%	-4.0%	-3.6%†	-4.8%†	-3.3%†
WF-HAR	D	-1.4%	-0.8%	-1.3%	-3.3%	-0.9%	-1.2%	-0.9%	-3.2%	-1.2%	-1.6%	-2.2%	-2.2%	-1.9%	-1.9%	-1.3%
OF-HAR	E	-3.0%†	-3.2%†	-1.9%†	-4.9%†	-3.4%†	-1.6%	-1.8%	-3.6%	-1.%	-2.5%	-3.9%	-3.0%	-3.3%†	-2.1%	-2.6%†
GF-HAR	F	0.2%	0.2%	0.4%	-0.4%	0.5%	0.3%	0.4%	0.0%	0.4%	0.4%	0.3%	0.3%	0.3%	0.3%	0.2%
AF-HAR	G	-5.3%†	-3.8%†	-2.8%†	-6.9%†	-5.6%†	-3.7%†	-4.5%†	-7.8%†	-4.3%†	-4.7%†	-6.9%†	-6.7%†	-4.4%†	-4.6%†	-4.4%†

Note: For the benchmark model 'A', we report the actual average value of the loss function. For the remaining models, we report the relative (in %) forecast improvements (negative values) with respect to the value of the benchmark model. The superscript † denotes whether the given model was in the set of superior forecasting models for the given firm.