University of Groningen

Market inefficiencies associated with pricing oil stocks during shocks

Qiao, Kenan; Sun, Yuying; Wang, Shouyang

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in University of Groningen/UMCG research database](#)

Market inefficiencies associated with pricing oil stocks during shocks

Kenan Qiao [a,b,c], Yuying Sun [a,b,d,*], Shouyang Wang [a,b,d]

[a] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China
[b] Center for Forecasting Science, Chinese Academy of Sciences, China
[c] Faculty of Economics and Business, University of Groningen, Netherlands
[d] School of Economics and Management, University of Chinese Academy of Sciences, China

ARTICLE INFO

ABSTRACT

The assumption that market efficiency informs the pricing of oil stocks is critical to understanding the co-movement between stock markets and oil markets. To test this assumption in relation to various types of real oil price changes, this article proposes a two-stage analysis method that starts with a quantile regression to identify oil shocks and develop interval-valued factor pricing models. These interval-based methods, relative to traditional point-based methods, can produce more efficient parameter estimations by providing more information. The results show that oil stocks tend to be overpriced following negative oil price shocks, which partially violates the efficient market hypothesis. Yet oil stocks are efficiently priced in response to moderate changes or positive oil price shocks, such that in most cases, the market remains efficient in pricing oil stocks.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

In the effort to model and analyze oil prices (Yu et al., 2008; Zhang et al., 2008, 2009; He et al., 2010; Li et al., 2013; Lu et al., 2014; Zhao et al., 2016), studies often identify a positive co-movement between oil prices and the stock prices of oil companies (e.g., Sadorsky, 2001; Henriques and Sadorsky, 2008). Intuitively, it appears that increased oil prices enhance oil companies' profitability, and stock prices respond positively in turn. This typical, fundamental analysis predicts that oil stocks' observed prices correctly reflect their intrinsic values, in accordance with the efficient market hypothesis. As a more detailed, novel test of that assumption, this article closely examines the stock market's efficiency in pricing oil stocks by comparing the market efficiency across different types of oil price changes: negative shocks, positive shocks, and moderate price changes. The empirical evidence indicates that the efficient market hypothesis does not hold in every case.

Research already has established that oil price changes can affect the prices of oil companies' stocks: Higher oil prices should lead to the enhanced financial performance of oil firms in stock markets. Using a vector error correction model, Hammoudeh et al. (2004) find that future oil prices relate positively to U.S. oil companies' stock prices, and Sadorsky (2001) similarly finds that oil prices exert a positive effect on Canadian oil firms' stock returns. By investigating the relationship between oil prices and stock prices in China's stock market, Cong et al. (2008) find that "some important oil price shocks depress oil company stock prices." Henriques and Sadorsky (2008) also use a vector autoregressive model to show that oil price changes positively Granger cause oil stock returns. According to Boyer and Filion (2007), Canadian energy stock relates positively to crude oil and natural gas prices, and Nandha and Faff (2008) find that oil price changes have positive impacts on oil stock prices in international equity markets. Because oil companies produce oil and gas, higher oil prices tend to lead to their increased profitability, which in turn prompts positive oil stock price changes.

This explanation relies on a fundamental analysis in which oil stocks' observed prices correctly reflect their intrinsic values-that is, the efficient market hypothesis. However, relatively few studies explicitly or empirically test this assumption. Fama and French (1997) test both the capital asset pricing model (CAPM) and the Fama-French three-factor model with various stock indices for different industries and find that pricing errors for the energy industry are consistently insignificant. Similarly, Arshanapalli et al. (1998) use

a multifactor pricing model to examine the financial performance of various industry portfolios and show that the international stock market is efficient when it comes to pricing oil stocks.

Yet not all oil price changes are the same. For example, oil price shocks refer to unexpected, substantial changes in real oil prices that accompany the availability of increased information about supply and demand. Existing literature that examines the market efficiency of pricing oil stocks and confirms the efficient market hypothesis does not address these various types of oil price changes, which might include negative shocks and positive shocks, as well as moderate price changes. In this sense, previous studies provide "unconditional" tests, without confirming whether the market is consistently efficient across all these different types of oil price changes. The rare shocks in the market involve more information than do common, moderate oil price changes. Therefore, the market should be more efficient in response to moderate oil price changes than to oil price shocks, but it also is of interest to examine market efficiency for pricing oil stocks when all these different types of oil price changes occur.

In this paper, oil company's equity prices are used to study market efficiency during real oil price shocks for three reasons. First, during oil price shocks, the stock price data of oil companies better reflect market efficiency related to other industries. As the producers of crude oil, oil stock prices are expected to be strongly correlated with real oil prices. Thus, oil companies' stock prices can constitute an ideal index to assess whether the news, reflected by the oil shock, is fully digested by the financial market. Second, different from other industries, oil prices dominate oil companies' profitability. Thus, the pricing efficiency of oil stocks essentially reflects the linkage with real oil price changes and its impact on the financial market. Third, as mentioned before, the pricing efficiency of oil company stocks is the fundamental of most of the existing studies on the co-movement between oil prices and oil stock prices.

This paper proposes a two-stage procedure to examine this market efficiency. In the first stage, quantile regression identifies oil shocks and their directions. In the second stage, a novel interval-valued factor pricing model evaluates market efficiency and produces a minimum distance estimation. The empirical study, conducted with the daily growth rates of WTI spot oil prices, produces some interesting findings. In particular, oil stocks are significantly overpriced in response to negative oil price shocks but efficiently priced for positive oil price shocks and moderate oil price changes. This result also is robust to various factor pricing models. The overpricing seemingly might arise because agents underreact to negative shocks; alternatively, agents might prefer to anchor their initial forecasts on their prior beliefs and then revise forecasts smoothly, instead of overadjusting (Capistrán and López-Moctezuma, 2014).

Compared with existing literature, this study offers three notable contributions. First, it uses quantile regression to identify oil price shocks. This method reconciles two important features of oil shocks (unpredictability and substantial changes), as revealed in previous studies, instead of just one of them. Second, the interval-valued factor pricing models that evaluate market efficiency in this study are superior to traditional point-valued factor pricing models, in that they produce more efficient estimations due to the gain in information they offer. That is, interval data contain more information (e.g., trend, volatility) than point-valued data in the same period. In turn, the proposed interval models can derive classic factor pricing models, including CAPM and the Fama-French three-factor and five-factor models. Third, the empirical results violate the efficient market hypothesis, challenging the findings of previous studies that rely on unconditional tests. Even if the market is efficient in most cases, the results reveal that the observed prices of oil stocks differ from their intrinsic values in relation to negative oil shocks.

Section 2 accordingly presents the novel methodology, including the new approach developed to identify the oil shocks and the interval-valued factor pricing models that serve to evaluate market efficiency. Section 3 describes the data analysis and empirical results. Section 4 contains robustness checks with various traditional factor pricing models, and Section 5 concludes.

2. Methodology

To establish a procedure for assessing the pricing efficiency of oil stocks in response to various types of oil price changes, two issues are relevant. First, the appropriate approach must be able to define three types of oil price changes: negative oil shocks, moderate oil price changes, and positive oil shocks. Second, an interval-valued factor regression with interval-valued dummy variables can uncover the intrinsic values of oil shocks and potential pricing errors.

2.1. Oil shocks

In this subsection, the goal is to define positive and negative oil price shocks. An oil price shock is an ambiguous concept with no generally accepted definition. Extant literature reveals two widely used measures though: net oil price increase/decrease and normalized oil price growth. First, Hamilton (1996) proposes the net oil price increase, which defines oil price shocks with quarterly data, as in the following equation:

$$NOPI_t = \max\left(0, ln\frac{p_t}{\max(p_{t-1}, p_{t-2}, p_{t-3}, p_{t-4})}\right), \tag{1}$$

where p_t denotes the oil price for quarter t. If the current oil price exceeds the maximal price over the previous year, the oil shock is equal to this percentage change; if the current oil price does not exceed the maximal oil price over the previous year, it is defined as 0. In a similar sense, a net oil price decrease is calculated as

$$NOPD_t = \min\left(0, ln\frac{p_t}{\min(p_{t-1}, p_{t-2}, p_{t-3}, p_{t-4})}\right). \tag{2}$$

This net oil increases/decrease series appears in many studies (Lee and Ni, 2002; Cunado and Gracia, 2003; Hamilton, 2003; Bernanke et al., 2004; Aloui and Jammazi, 2009; Ceylan and Berument, 2010; Engemann et al., 2014). Another widely used measure is normalized oil price growth, as proposed by Lee et al. (1995), which models the evolution of oil price growth according to an AR(n)-GARCH(p,q) process

$$r_t = \mu + \sum_{i=1}^{n} \rho_i r_{t-i} + e_t, \tag{3}$$

$$e_t = \sigma_t \epsilon_t, \tag{4}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \theta_i \sigma_{t-i}^2 + \sum_{i=1}^{q} \gamma_i e_{t-i}^2, \tag{5}$$

where $r_t = \frac{p_t}{p_{t-1}} - 1$ is the oil price growth rate. Then the normalized growth rate is simply defined as the innovation process ϵ_t. This measurement is widely adopted in existing literature (Sadorsky, 1999; Cunado and Gracia, 2003; Park and Ratti, 2008; Ceylan and Berument, 2010).

Both measures reflect important features of oil price shocks. First, the net oil price increase/decrease implies that oil price shocks are rare, such that only dramatic oil price changes can be regarded as shocks. Relatively insignificant oil price changes are excluded from this definition. Second, the normalized oil price growth measure implies that only unexpected oil price changes can be called shocks. Thus, rarity and unexpectedness mutually and essentially define oil

price shocks; in turn, for this study, oil price shocks are formally defined as substantial and unpredictable changes in oil prices.

Using conditional quantiles of oil price growth as thresholds can distinguish the types of oil price changes. Suppose $\mathcal{F}_t$ is an information set observed at time t. The lower τth and upper τth quantiles of r_{t+1}, conditional on $\mathcal{F}_t$, can be calculated as

$$q_{t+1}^{\tau} = \sup\{q : Prob(r_{t+1} < q|\mathcal{F}_t) < \tau\}, \tag{6}$$

$$q_{t+1}^{1-\tau} = \inf\{q : Prob(r_{t+1} > q|\mathcal{F}_t) < \tau\}, \tag{7}$$

where τ is a relatively small number that can be set subjectively, at 1%, 5%, or 10% for example. Then oil price change r_{t+1} is a negative shock at time $t + 1$ if its realized value is smaller than q_{t+1}^{τ} or a positive shock at time $t + 1$ if its realized value is greater than $q_{t+1}^{1-\tau}$; otherwise, r_{t+1} is a moderate oil price change. Unlike the net oil price increase/decrease series and normalized oil price growth, this quantile-based identification considers both attributions of oil price shocks, rather than focusing on one or the other. The conditional quantiles q_{t+1}^{τ} and $q_{t+1}^{1-\tau}$ are derived from the information set $\mathcal{F}_t$, such that the predictable components of r_{t+1} have been incorporated into these quantiles. By using lower and upper quantiles of r_{t+1} as thresholds, this method also can identify relatively insignificant or small shifts in oil prices as moderate oil price changes. Only sharp increases or decreases are identified as shocks, according to both of their crucial characteristics.

Quantile regression in turn can provide the conditional quantiles of r_t. Suppose $x_t^1, x_t^2, \ldots, x_t^n$ are n $\mathcal{F}_{t-1}$-adopted random variables. Then a quantile regression for r_t can be written as

$$q_t^{\tau} = Q^{\tau}\left(r_t|x_t^1, x_t^2, \ldots, x_t^n\right) = f\left(x_t^1, x_t^2, \ldots, x_t^n; \beta^{\tau}\right), \tag{8}$$

where f is a predetermined function, and β^{τ} is the vector of unknown parameters. With some regularity conditions, β^{τ} can be estimated according to the following equation:

$$\hat{\beta}^{\tau} = \arg\min_{\beta^{\tau}}\{(\tau - 1)\sum_{r_t < q_t^{\tau}}(r_t - q_t^{\tau}) + \tau\sum_{r_t > q_t^{\tau}}(r_t - q_t^{\tau})\}. \tag{9}$$

Thus, the lower and upper τth quantiles of r_t can be calculated as $\hat{q}_t^{\tau} = f\left(x_t^1, \ldots, x_t^n; \hat{\beta}^{\tau}\right)$ and $\hat{q}_t^{1-\tau} = f\left(x_t^1, \ldots, x_t^n; \hat{\beta}^{1-\tau}\right)$, respectively. Quantile regressions offer three main advantages over GARCH-based approaches (Lee et al., 1995). First, they impose few restrictions on the data-generating process for r_t. No economic theory exists to suggest a GARCH model for oil price growth rates, such that the choice of a GARCH model setting is arbitrary. Second, quantile regression establishes a general framework for various choices of function $f(*)$ and regressors $x_t^1, \ldots, x_t^n$. Third, without the assumption of normally distributed error terms, quantile regression performs well in fitting r_t's tail distribution. Empirical experience reveals that normal distribution often achieves only poor fit with the tail distribution of many financial and economic variables, despite its strong performance in fitting most observations except extreme values. In most cases, this gap is not a serious issue, but for this study, it becomes crucial, because the focus is the extreme values of r_t.

To specify the quantile regression, τ is set to 5%. Technically, τ can be set arbitrarily anywhere between 0 and 1. But 1%, 5%, and 10% are three frequently used values to recognize low probability events in practice, suggesting 5% as an appropriate choice. In addition, f is assumed to be a linear function of x_t, such that

$$f\left(x_t^1, x_t^2, \ldots, x_t^n; \beta^{\tau}\right) = \alpha^{\tau} + \beta_1^{\tau}x_t^1 + \beta_2^{\tau}x_t^2 + \ldots + \beta_n^{\tau}x_t^n. \tag{10}$$

Although it is still possible to specify a nonlinear function for f, a linear specification for quantile regressions is common in previous studies. Two classes of variables could represent the choice of regressors x_t: the AR(n)-GARCH(p,q)-based conditional expectation ($\bar{r}_t$) and variance (σ_t^2) of r_t, or else the empirical moments of r_t in an n-period rolling window $[t - n, t - 1]$. Because the conditional expected values and volatility of r_t largely determine its quantiles, this study follows previous literature and uses AR(n)-GARCH(p,q) to estimate them. In addition, the empirical moments of r_t reveal more information about r_t's conditional distribution, especially its third- and fourth-order moments. Thus, it is possible to calculate the sample mean, variance, skewness, and kurtosis of r_t in the n-period rolling window, then include them in the quantile regression as well. The skewness and kurtosis are calculated as

$$skew_t = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{r_{t-i} - ave_t}{vol_t}\right)^3, kurt_t = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{r_{t-i} - ave_t}{vol_t}\right)^4, \tag{11}$$

where $aver_t$ and vol_t are sample mean and standard deviation of r_t in rolling window $[t - n, t - 1]$.

Finally, the procedure for identifying oil price shocks is as follows: Initially, the construction of the time series for the AR(n)-GARCH(p,q)-based conditional expectation ($\bar{r}_t$) and variance (σ_t^2) of r_t, as in Eqs. (3) to (5), relies on the sample mean ($aver_t$), variance (vol_t), skewness ($skew_t$), and kurtosis ($kurt_t$) of r_t in a rolling window $[t - n, t - 1]$. Following the suggestions of Bollerslev et al. (1992), this study uses a lower-order model, namely, AR(1)-GARCH(1,1). The chosen length of the rolling window is 21 trading days, or one calendar month. Then, the quantile regressions can be obtained as follows:

$$q_t^{0.05} = \alpha^{0.05} + \beta_{\bar{r}}^{0.05}\bar{r}_t + \beta_{\sigma^2}^{0.05}\sigma_t^2 + \beta_{ave}^{0.05}ave_t + \beta_{vol^2}^{0.05}vol_t^2$$
$$+ \beta_{skew}^{0.05}skew_t + \beta_{kurt}^{0.05}kurt_t, \tag{12}$$

and

$$q_t^{0.95} = \alpha^{0.95} + \beta_{\bar{r}}^{0.95}\bar{r}_t + \beta_{\sigma^2}^{0.95}\sigma_t^2 + \beta_{ave}^{0.95}ave_t + \beta_{vol^2}^{0.95}vol_t^2$$
$$+ \beta_{skew}^{0.95}skew_t + \beta_{kurt}^{0.95}kurt_t. \tag{13}$$

Next, the process calculates the lower and upper 0.05th quantiles, $q_t^{0.05}$ and $q_t^{0.95}$. Finally, if $r_t < q_t^{0.05}$, r_t is identified as a negative oil price shock; if $r_t > q_t^{0.95}$, r_t is identified as a positive oil price shock; and otherwise, r_t is identified as a moderate oil price change.

2.2. Interval-valued factor models with interval dummy

A popular definition of an efficient market indicates that the observed prices of financial assets fully reflect available information. Thus, an extreme efficient market implies that the asset prices equal their intrinsic value, associated with future cash flows (Fama, 1970, 1991; Timmermann and Granger, 2004). Suppose P_t and P_t^* are the observed price and intrinsic value of an asset at time t. For this study, the asset is a portfolio that consists of some stocks issued by oil companies, so in an efficient market, P_t and P_t^* must be identical, $P_t = P_t^*$.

Tests of market efficiency inevitably run into the problem of joint tests of the efficient market hypothesis and pricing models. That is, to compare the observed price P_t with the intrinsic value P_t^*, it is necessary first to find a value of P_t^*. In essence, P_t^* must be determined by a general equilibrium (pricing) model. When the asset is mispriced

though, it is never possible to determine whether the market is inefficient or the chosen pricing model has been misspecified. Thus, all tests represent joint tests of efficient market and pricing models.

As an alternative, interval-valued factor models with interval-valued dummy variables can examine market efficiency with three key advantages. First, these models can derive traditional point-valued factor pricing models, such as CAPM, the Fama-French three-factor model, the Carhart four-factor model, or the Fama-French five-factor model (e.g., Fama and French, 1993, 2015). Because tests of market efficiency suffer from the problem of joint tests, it becomes critical to choose a reliable pricing model with strong empirical performance, and these proposed interval-valued factor pricing models offer ideal options for enriching existing pricing models. Second, interval data contain more information than point data for the same period, and this informational gain should produce more efficient parameter estimates and statistical inferences for the interval-based models. Third, in the spirit of classic point-valued factor pricing models, interval-based models provide a flexible framework that can incorporate various choices of pricing factors. Different specifications of interval-valued pricing factors with interval-valued dummy variables also provide good robustness checks.

Therefore, the interval dummy variable to assess the pricing efficiency of oil shocks under different types can be defined as follows:

Definition 1. The interval dummy variable for positive and negative shocks consists of a pair of interval-valued dummy variables denoted as ID_t and Id_t, namely,

$$ID_t = \begin{cases} \left[-\frac{1}{2}, \frac{1}{2}\right], & \text{positive shock occurs} \\ [0,0], & \text{otherwise,} \end{cases}$$

and

$$Id_t = \begin{cases} \left[-\frac{1}{2}, \frac{1}{2}\right], & \text{negative shock occurs} \\ [0,0], & \text{otherwise.} \end{cases}$$

These interval-valued dummy variables can indicate the three types of oil price changes (negative oil shocks, moderate oil price changes, and positive oil shocks) but do not attempt to measure their sizes, which is pertinent for two reasons. First, this study is dedicated to determining whether the stock market is consistently efficient in response to different types of oil price changes, rather than exploring the quantitative relationship between oil price shocks and pricing errors. Dummy variables provide a more direct way to address this research question. Second, there is no empirical or theoretical guideline for specifying a functional relationship between oil price shocks and pricing errors. A misspecified model might disturb subsequent tests and lead to incorrect conclusions. Therefore, it is reasonable to indicate the types of oil price changes by using dummy variables. Specifically, if a negative oil price shock occurs, the dummy variable Id_t is set at a unit interval $\left[-\frac{1}{2}, \frac{1}{2}\right]$; if a positive oil price shock occurs, the dummy variable ID_t is set at a unit interval $\left[-\frac{1}{2}, \frac{1}{2}\right]$; and otherwise, Id_t and ID_t are both interval-valued zeros [0,0].

Denote R_{ft} as the return on the risk-free asset and R_t as the return on a risky asset. Let $\{Y_t | Y_t = [R_{ft}, R_t]\}$ and $\left\{X_t^k | X_t^k = \left[X_{Lt}^k, X_{Rt}^k\right]\right\}, k = 1, \cdots, n$ be a D_K weakly stationary interval process, and Y_t and X_t^k are defined by their left bounds (i.e., R_{ft} or X_{Lt}^k) and right bounds (i.e., R_t or X_{Rt}^k), respectively (see Yang et al. (2016), Sun et al. (2018)). The D_K weak stationarity is defined in Theorem 2.1 of Han et al. (2016). Specifically, if $\{Y_t\}$ is a D_K weakly stationary interval process $\{Y_t\}$, this satisfies that $E(Y_t) = \mu$, $E\left\langle s_{Y_t} - s_\mu, s_{Y_{t-j}} - s_\mu\right\rangle_K = \gamma(j)$ and $\sum_{j=-\infty}^\infty |\gamma(j)| < \infty$ for all t and j. The main difference from the traditional weakly stationary is the autocovariance function with respect

to D_K-distance. Each interval process is an inseparable set of ordered numbers, which can be referred to as an extended random interval, following the definition of a generalized interval in Kaucher (1980), to cover the reverse order of interval bounds, such as when the left bound is larger than the right bound.

Following the spirit of Sun et al. (2018), we propose the general form of the conditional interval-valued factor pricing model:

$$Y_t = \alpha_0 + \alpha I_0 + \alpha^d Id_{t-1} + \alpha^D ID_{t-1} + \sum_{k=1}^n \beta^k X_t^k + u_t = Z_t'\theta + u_t, \quad (14)$$

$$E_{t-1}\left(u_t \otimes \left(1, Id_t, ID_t, X_t^1, \ldots, X_t^n\right)\right) = 0_{1\times(n+3)}. \quad (15)$$

where Y_t is the interval-valued return constructed by the risk-free rate and asset return; X_t^k is the kth interval-valued pricing factor; $E_{t-1}(\cdot)$ refers to the conditional expectation; $0_{1\times(n+3)}$ is a one-by-$(n+3)$ zero matrix; $Z_t = ([1,1], I_0, Id_{t-1}, ID_{t-1}, X_t^1, \cdots, X_t^n)$; $I_0 = [-1/2, 1/2]$ is a constant, unit interval; and $\theta = \left(\alpha_0, \alpha, \alpha^d, \alpha^D, \beta^1, \cdots, \beta^n\right)'$. Eq. (15) is based on the assumption that $u_t = [u_{Lt}, u_{Rt}]$ is an interval martingale difference sequence with respect to the information set $I_{t-1} = \{Id_0, \cdots, Id_{t-1}, ID_0, \cdots, ID_{t-1}, X_1, \cdots, X_t^n\}$, $\otimes$ is defined to reflect the Hadamard product for matrices based on the support function s_A (see Section 2.4). In turn, $E_{t-1}\left\langle s_{u_t}, s_{X_t^j}\right\rangle_K = 0, j = 1, \cdots, n$, $E_{t-1}\left\langle s_{u_t}, s_{Id_{t-1}}\right\rangle_K = 0$ and $E_{t-1}\left\langle s_{u_t}, s_{ID_{t-1}}\right\rangle_K = 0$. For simplicity, the bounds of all interval-valued pricing factors are assumed to be the returns of some assets. Then the ranges of these interval-valued variables represent the difference between the returns of two assets. Thus, the interval-based model can derive a traditional point-valued factor pricing model as follows:

$$R_t - R_{ft} = \alpha + \alpha^d d_{t-1} + \alpha^D D_{t-1} + \sum_{k=1}^n \beta^k X_{rt}^k + \epsilon_t, \quad (16)$$

where $X_{rt} = X_{Rt} - X_{Lt}$, $d_{t-1} = 1$ if a negative oil price shock occurs; $D_{t-1} = 1$ if a positive oil price shock occurs; and otherwise, both are zero. In turn, it is possible to define $Id_t = \left[-\frac{1}{2}d_t, \frac{1}{2}d_t\right]$ and $ID_t = \left[-\frac{1}{2}D_t, \frac{1}{2}D_t\right]$, as a special case of traditional factor models under state-based pricing errors (Ferson and Korajczyk, 1995).

Furthermore, α is the pricing error for moderate oil price changes. When ID_t and Id_t are both [0,0], interval-valued factor models can produce traditional point-valued factor pricing models, including CAPM and the Fama-French three- or five-factor models. In particular, the test asset is an aggregated price index of oil stocks. If the classic factor pricing model is correctly specified with appropriate factors, the asset is efficiently priced if and only if the constant α is zero. If $\alpha > 0$, the asset has a greater expected return than the fair value required to take systematic risks. In this case, the observed price of the asset is too "cheap" relative to its intrinsic value. That is, a positive price error implies the asset is underpriced. If instead $\alpha < 0$, the asset is overpriced.

The analysis of pricing error under different types of oil shocks relies on Eq. (14). If α^d is the pricing error under negative oil price shocks, and α^D is the pricing error under positive oil price shocks. Using $\alpha = 0$ as a benchmark, Table 1 summarizes the economic implications of α, α^d and α^D. It is worth noting that a significant α^d or α^D cannot be interpreted as a causal relation between oil price shocks and market inefficiency. That is, oil prices are determined by supply and demand, so they typically are endogenous variables. Meanwhile, oil stock prices inevitably are driven by other economic variables too. Thus, both oil prices and oil stock prices are endogenous and could be driven by some other common factors. In turn, it is impossible to find

Table 1
Economic implications of pricing errors under different types of oil price changes.

	Negative oil shock	Moderate change	Positive oil shock
$\alpha^d < 0$	Over-priced		
$\alpha^d > 0$	Under-priced		
$\alpha < 0$		Over-priced	
$\alpha > 0$		Under-priced	
$\alpha^D < 0$			Over-priced
$\alpha^D > 0$			Under-priced

Note: This table shows the economic implications of the pricing errors under different types of oil price changes. The pricing errors α, α^d and α^D are derived by

$$R^e_{t+1} = \alpha + \alpha^d d_t + \alpha^D D_t + \sum_{k=1}^n \beta^k F^k_{t+1} + \epsilon_{t+1},$$

$$E(\epsilon_{t+1}) = E(\epsilon_{t+1} d_t) = E(\epsilon_{t+1} D_t) = E(\epsilon_{t+1} F^1_{t+1}) = \ldots = E(\epsilon_{t+1} F^n_{t+1}) = 0.$$

If there is a positive oil price shock at time t, set $d = 1$; If there is a negative oil price shock at time t, set $D = 1$; otherwise, $d_t = D_t = 0$.

any conclusive causal relation between oil price shocks and market inefficiency with Eq. (14) alone, even if α^d or α^D were significant.

2.3. Special cases

The general form of the conditional interval-valued factor pricing model (Eq. (14)) also can produce some special cases, including an interval-valued CAPM, interval-valued Fama-French three-factor model, and interval-valued Fama-French five-factor model. The interval version of CAPM is:

$$Y_t = \alpha_0 + \alpha I_0 + \beta X_t + \alpha^d Id_{t-1} + \alpha^D ID_{t-1} + u_t, \tag{17}$$

where $Y_t = [R_{ft}, R_t]$, $X_t = [R_{ft}, R_{mt}]$, R_{mt} is the market return and $E(u_t|I_{t-1}) = [0,0]$, α_0, α and β are scalar-valued unknown parameters, $I_0 = [-1/2, 1/2]$ is a constant, unit interval; $u_t = [u_{Lt}, u_{Rt}]$ is an interval martingale difference sequence with respect to the information set I_{t-1}, i.e., $E(u_t|I_{t-1}) = [0,0]$ a.s. From our interval CAMP, it is possible to derive a point-valued model, reflecting the difference between the right and left bounds of the interval. Therefore, a classic CAPM model with point-valued dummy variables results:

$$R_t - R_{ft} = \alpha + \beta(R_{mt} - R_{ft}) + \alpha^d d_{t-1} + \alpha^D D_{t-1} + \epsilon_t, t = 1, \cdots, T, \tag{18}$$

where $E(\epsilon_t|I_{t-1}) = E(u_{Rt} - u_{Lt}|I_{t-1}) = 0$. The CAPM beta is measured by the slope coefficient β from Eq. (18). The main advantage of this interval modeling approach is that it estimates the model using interval data, which contain more information than range data. Thus, more efficient estimates result, even if the focus ultimately is on the range model.

Following a similar approach, the interval-valued three-factor model, constructed by 2-by-3 Fama-French portfolios formed on the basis of size and book-to-market ratios, can be written as:

$$Y_t = \alpha_0 + \alpha I_0 + \alpha^d Id_{t-1} + \alpha^D ID_{t-1} + \beta_1 X^1_t + \beta_2 X^2_t + \beta_3 X^3_t + u_t, \tag{19}$$

where

$$X^1_t = [R_{ft}, R_{mt}], \quad X^2_t = \left[\frac{1}{3}(B/L_t + B/M_t + B/H_t), \frac{1}{3}(S/L_t + S/M_t + S/H_t)\right],$$

$$X^3_t = \left[\frac{1}{2}(S/L_t + B/L_t), \frac{1}{2}(S/H_t + B/H_t)\right], \frac{1}{3}(B/L_t + B/M_t + B/H_t)$$

is the return on the portfolio of big market value firms, $\frac{1}{3}(S/L_t + S/M_t + S/H_t)$ is the return on the portfolio of small market value firms, $\frac{1}{2}(S/L_t + B/L_t)$ is the return on the portfolio of low book-to-market ratio firms, and $\frac{1}{2}(S/H_t + B/H_t)$ is the return on the portfolio

of high book-to-market ratio firms (for portfolio construction details, see Fama and French (1993)). Each interval-valued factor is well defined, according to the concept of an extended interval. By taking the difference between the right and left bounds of the three-factor interval CAMP, it is possible to derive a point-valued three-factor model:

$$R_t - R_{ft} = \alpha + \alpha^D D_{t-1} + \alpha^d d_{t-1}$$
$$+ \beta_1(R_{mt} - R_{ft}) + \beta_2 SMB_t + \beta_3 HML_t + \epsilon_t, t \cdots, T, \tag{20}$$

where *SMB* and *HML* are the differences between the returns on the high and low book-to-market value portfolios and on the small minus big firm portfolios, respectively. The Fama-French three-factor beta is measured by the coefficient $\beta_{it}, i = 1, 2, 3$ for each factor. Small firms usually have relatively large factor loadings β_2, and high book-to-market ratio firms tend to have relatively large β_3.

Similarly, the interval-valued five-factor model is constructed by 2-by-3 Fama-French portfolios formed on size and book-to-market ratio, size and operating profitability, and size and investments, as follows:

$$Y_t = \alpha_0 + \alpha + \alpha^d Id_{t-1} + \alpha^D ID_{t-1}$$
$$+ I_0 + \beta_1 X^1_t + \beta_2 X^2_t + \beta_3 X^3_t + \beta_4 X^4_t + \beta_5 X^5_t + u_t, \tag{21}$$

where

$$X^4_t = \left[\frac{1}{2}(S/W + B/W), \frac{1}{2}(S/R + B/R)\right], X^5_t = \left[\frac{1}{2}(S/A + B/A), \frac{1}{2}(S/C + B/C)\right], \frac{1}{2}(S/W + B/W)$$

is the return on diversified portfolios of stocks with weak profitability, $\frac{1}{2}(S/R + B/R)$ is the return on diversified portfolios of stocks with robust profitability, $\frac{1}{2}(S/A + B/A)$ is the return on the two aggressive investment portfolios, $\frac{1}{2}(S/C + B/C)$ is the return on the two conservative investment portfolios. Then the classic Fama-French five-factor model is derived as follows:

$$R_t - R_{ft} = \alpha + \beta_1(R_{mt} - R_{ft}) + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 RMW_t$$
$$+ \beta_5 CMA_t + \alpha^d d_{t-1} + \alpha^D D_{t-1} + \epsilon_t, \tag{22}$$

where RMW_t is the difference between the returns on diversified portfolios of stocks with robust and weak profitability, and CMA_t is the difference between the returns on diversified portfolios of stocks of conservative and aggressive investment firms (Fama and French, 2015). Firms with robust profitability have relatively large factor loadings β_4, but conservative firms have relatively large β_5.

Table 2
Data description of main variables.

	# of obs.	Mean (%)	S.D. (%)
Growth rate of oil price (r_t)	8001	0.039	2.509
T-bill rate (rf_t)	8026	0.013	0.010
Excess return on oil stock index (R^e_t)	8026	0.037	1.472
Excess return on market portfolio ($Mktrf_t$)	8026	0.034	1.109
Size factor (SMB_t)	8026	0.002	0.586
Value factor (HML_t)	8026	0.011	0.573
Operating profitability factor (RMW_t)	8026	0.017	0.439
Investment factor (CMA_t)	8026	0.012	0.401

Note: This table reports the data description of main variables. Daily growth rate of oil price (r_t) is calculated as $r_t = \frac{p_t - p_{t-1}}{p_{t-1}}$, where p_t is the WTI oil price which is downloaded from FRED Economic Data; daily T-bill rate (rf_t), Excess return on oil stock index (R^e_t), Excess return on market portfolio ($Mktrf_t$), Size factor (SMB_t), Value factor (HML_t), Operating profitability factor (RMW_t) and Investment factor (CMA_t) are all collected from Kenneth R. French's website. The excess return on Oil Stock Index (R^e_t) is calculated as $R_t - rf_t$, where R_t is the net return on oil stock index. The sample is from January 2nd, 1986 to October 31st, 2017.

Table 3
Estimation of AR(1)-GARCH(1,1) and descriptive statistics of oil price growth rates' conditional moments.

Panel A: quasi-maximum likelihood estimation of AR(1)-GARCH(1,1) model					
	Conditional expectation		Conditional variance		
	μ (%)	ρ	ω	θ	γ
Coef.	0.046**	−0.024**	0.000***	0.891***	0.103***
	(2.410)	(−2.080)	(8.320)	(191.120)	(24.440)

Panel B: descriptive statistics of oil price growth rates' conditional moments				
	AR(1)-GARCH(1,1) based		Rolling window based	
	Mean	S.D.	Mean	S.D.
Expectation(%)	0.045	0.060	0.040	0.491
Variance(%)	0.066	0.081	0.060	0.082
Skewness			−0.055	0.678
Kurtosis			3.283	1.409

Note: Panel A shows the quasi maximum-likelihood estimation of AR(1)-GARCH(1,1) model for the growth rates of WTI oil prices. The growth rates of oil prices are calculated as $r_t = \frac{p_t - p_{t-1}}{p_{t-1}}$, where p_t is the WTI oil price. Panel B presents the descriptive statistics of conditional moments of r_t, including AR(1)-GARCH(1,1) based conditional expectation and variance, and 21-day rolling window based expectation, variance, skewness and kurtosis. Skewness and kurtosis are calculated as

$$skew_t = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{r_{t-i} - ave_t}{vol_t} \right)^3, kurt_t = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{r_{t-i} - ave_t}{vol_t} \right)^4,$$

where ave_t and vol_t are sample mean and standard deviation of r_t in rolling window $[t - 21, t - 1]$. The WTI crude oil prices are downloaded from FRED Economic Data. The sample period is from January 2nd, 1986 to October 31st, 2017. *, ** and *** denote 10%, 5% and 1% significance respectively. The z-statistics of coefficients' estimators are presented in the brackets.

2.4. Estimation

Interval information supports estimates of the coefficients θ. The challenge arises in the specification of the objective function to measure the sum of the squared distance between the observed interval-valued sets and the interval models. Following Nather (1997) and Nather (2000), this study seeks to measure the squared D_K distance between set-valued intervals Y_t and its fitted value $Z_t'\theta$. Specifically, the D_K metric is:

$$D_K^2(Y_t, Z_t'\beta) = \int_{(u,v) \in s^0} \left[s_{Y_t}(u), s_{Z_t'\theta}(u) \right] \left[s_{Y_t}(v), s_{Z_t'\theta}(v) \right] dK(u,v),$$

$$= \left\langle s_{Y_t - Z_t'\theta}, s_{Y_t - Z_t'\theta} \right\rangle_K = ||Y_t - Z_t'\theta||_K^2 = ||u_t||_K^2, \qquad (23)$$

where the unit space $s^0 = \{u \in R^1, |u| = 1\} = \{1, -1\}$, $K(u,v)$ is a symmetric positive definite weighing function on s^0 to ensure that $D_K(Y_t, Z_t'\theta)$ is a metric for extended intervals, and $\langle \cdot, \cdot, \rangle$ indicates the inner product in s^0 with respect to kernel $K(u,v)$. By minimizing the sum of squared errors $\sum_{t=1}^{T} D_K^2(Y_t, Z_t'\theta)$, it is possible to obtain the estimator as follows:

$$\hat{\theta} = \left(\sum_{t=1}^{T} \left\langle s_{Z_t}, s_{Z_t'} \right\rangle_K \right)^{-1} \sum_{t=1}^{T} \left\langle s_{Z_t}, s_{Y_t} \right\rangle_K, \qquad (24)$$

where $s_A(u)$ is the following support function:

$$s_A(u) = \begin{cases} \sup_{a \in A} \{u \cdot a | u \in s^0\} & \text{if } A_L \leq A_R, \\ \inf_{a \in A} \{u \cdot a | u \in s^0\} & \text{if } A_L \leq A_R, \end{cases} \qquad (25)$$

and it follows that $s_A(u) = A_R$ if $u = 1$, $s_A(u) = -A_L$ if $u = -1$. In empirical application, we follow the spirit of Han et al. (2016) and Sun et al. (2018) to use a two-stage minimum D_K-distance method to estimate parameters with a preliminary choice

of kernel K. Specifically, in first stage, we use a preliminary choice of kernel K such as $(a, b, c) = (5, 3, 4)$ to estimate an interval-valued factor model and calculate the estimated residuals $\{\hat{u}_t\}$. In second stage, we estimate an optimal kernel $\hat{K}_{opt}$ with $\hat{K}_{opt}(1,1) = T^{-1} \sum_{t=1}^{T} \hat{u}_{Lt}^2$, $\hat{K}_{opt}(1,-1) = \hat{K}_{opt}(-1,1) = T^{-1} \sum_{t=1}^{T} \sum_{t=1}^{T} \hat{u}_{Lt} \hat{u}_{Rt}$, and $\hat{K}_{opt}(-1,-1) = T^{-1} \sum_{t=1}^{T} \hat{u}_{Rt}^2$ and employ this optimal kernel to estimate parameters again. After several iterations, we can obtain a stable optimal kernel and then calculate parameter estimators. For detailed discussions, see Sun et al. (2018, 2019).

3. Empirical results

3.1. Data analysis

The empirical application investigates Oil & Gas industry stocks' market prices to examine market efficiency. The Oil & Gas industry is defined by firms categorized into the Petroleum and Natural Gas standard industrial code[1]. There are two reasons that drive us to focus on the Oil & Gas industry. First, the aim of the paper is to test the market efficiency under different types of oil shocks. Intuitively, after an oil price shock, the most affected sectors in the economy should be the consumers and producers of oil and gas. Hence, the response of oil producers, namely the Oil & Gas industry, is an ideal subject for our study. Second, the consumers of oil and gas are affected across a wide range of industry sectors, but its producers only exist in the Oil & Gas industry. We thus expect to find very strong co-movements of oil and gas producers after an oil price shock, because those firms are in same industry sector. Therefore, the

---

[1] See http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_49_ind_port.html for further details of Kenneth R. French data library's description of 49 industry's definition.

Table 4
Estimation of quantile regressions for oil price growth rates and oil shocks identification.

Panel A: estimation of quantile regressions for oil price growth rates		
	Lower 5%th quantile	Upper 5%th quantile
α	−0.026***	0.025***
	(−16.040)	(15.640)
$\beta_{\bar{r}}$	0.640	0.130
	(0.610)	(0.130)
β_{σ^2}	−20.364***	18.140***
	(−10.260)	(9.310)
β_{ave}	0.018	−0.571***
	(0.140)	(−4.470)
β_{vol^2}	0.861	5.812***
	(0.440)	(3.000)
β_{skew}	−0.001	0.000
	(−0.640)	(0.120)
β_{kurt}	0.001*	−0.001**
	(1.710)	(−2.570)
Panel B: oil price shocks identification		

	Mean	S.D.	Sum
Lower 5%th quantile	−0.036	0.016	
Upper 5%th quantile	0.037	0.019	
d_t			394
D_t			394

Note: Panel A reports the quantile regression for oil price growth rates r_t

$$q_t^{0.05} = \alpha^{0.05} + \beta_{\bar{r}}^{0.05}\bar{r}_t + \beta_{\sigma^2}^{0.05}\sigma_t^2 + \beta_{ave}^{0.05}ave_t + \beta_{vol^2}^{0.05}vol_t^2 + \beta_{skew}^{0.05}skew_t + \beta_{kurt}^{0.05}kurt_t,$$
$$q_t^{0.95} = \alpha^{0.95} + \beta_{\bar{r}}^{0.95}\bar{r}_t + \beta_{\sigma^2}^{0.95}\sigma_t^2 + \beta_{ave}^{0.95}ave_t + \beta_{vol^2}^{0.95}vol_t^2 + \beta_{skew}^{0.95}skew_t + \beta_{kurt}^{0.95}kurt_t,$$

where $\bar{r}_t$ and σ_t^2 denote the AR(1)-GARCH(1,1) based conditional expectation and variance of r_t; ave_t, vol_t^2, $skew_t$ and $kurt_t$ denote the 21-day rolling window based conditional expectation, variance, skewness and kurtosis, respectively. Panel B reports the descriptive statistics of estimated quantiles of r_t and oil price shocks identification dummy variables d_t and D_t. If $r_t < q_t^{0.05}$, it is identified as a negative shock and set $d_t = 1$; if $r_t > q_t^{0.95}$, it is identified as a positive shock and set $D_t = 1$; otherwise, d_t and D_t are both zeros. The growth rates of oil prices are calculated as $r_t = \frac{p_t - p_{t-1}}{p_{t-1}}$, where p_t is the WTI oil price. *, ** and *** denote 10%, 5% and 1% significance respectively. The t-statistics of coefficients' estimators are presented in the brackets.

Oil & Gas industry is a better choice for our study compared to oil and gas consumers.

This empirical analysis uses daily data from various sources. Daily crude oil prices reflect the WTI spot prices, downloaded from FRED Economic Data[2] Daily risk-free rates are proxied by one-month T-bill rates, collected from Kenneth R. French's website. The daily aggregated price index of all oil stocks listed on NYSE/NASDAQ stock exchanges and various pricing factors (for the Fama-French three- and five-factor models) also are available from Kenneth R. French's website. [3] The sample runs from January 2nd, 1986 to October 31st, 2017. The description of the main variables appears in Table 2, revealing that the average excess return on the oil stock index is around 0.037 %, slightly larger than the risk premium of the market portfolio (0.034 %). The standard deviation of the oil stock index's excess returns is about 1.472%, also larger than the standard deviation of the market portfolio's excess returns (1.109%). The Sharpe ratios of the oil stock index and market portfolio are 2.519% and 3.022%. Thus, the market portfolio appears more mean-variance efficient than the oil stock index, consistent with a traditional CAPM model. In addition, the growth rate of oil prices is around 0.039 % on average, very close to the average excess returns on the oil stock index. Yet the standard deviation of 2.509% is much greater than the standard deviation of oil stock index returns. Thus, crude oil prices are more volatile than oil stock prices.

3.2. Identification for oil shocks

As noted, this study calculates the regressors of the quantile regression: the AR(1)-GARCH(1,1)-based conditional expectation and volatility of r_t, and the sample moments of r_t in a rolling window. Panel A of Table 3 contains the AR(1)-GARCH(1,1) estimation for r_t, demonstrating that the estimator of ρ is negatively significant; that is, oil prices' growth rates exhibit a mean-reversion pattern. The coefficients θ and γ are both positively significant, implying a volatility clustering effect in oil price growth rates. Panel B of Table 3 also provides the descriptive statistics of the conditional moments of r_t, including AR(1)-GARCH(1,1)-based conditional expectation and variance, and the 21-day rolling window-based expectation, variance, skewness, and kurtosis. The mean AR(1)-GARCH(1,1)-based conditional expectation is close to the mean of the rolling-window-based conditional expectation, but the latter is more volatile, with a larger standard deviation. Furthermore, the AR(1)-GARCH(1,1)-based variance and rolling-window-based variance have similar means and standard deviations. Rolling-window-based skewness is negative on average, which implies an asymmetric distribution of r_t. Finally, the mean of the rolling-window-based kurtosis is greater than 3, reflecting the fat tail of r_t's distribution.

Table 4 contains the results of the quantile regressions for oil price growth rates and oil shock identification. Panel A offers the estimates of the quantile regressions from Eqs. (12)–(13). The signs and significance of coefficients are capricious, seemingly caused by the multicollinearity of the regressors, which is outside the scope of this study. The focus instead is on the forecast values of the lower and

---

Table 5
D_K-minimum distance estimation of interval-valued factor pricing models.

	Column A	Column B	Column C
	(Interval-CAPM)	(Interval-FF3 model)	(Interval-FF5 model)
$\alpha_0(\%)$	0.000	0.000	0.000
	(0.640)	(0.456)	(0.418)
$\alpha\,(\%)$	0.010	0.010	0.010
	(1.199)	(0.951)	(0.906)
$\alpha^d\,(\%)$	−0.200***	−0.210***	−0.200***
	(8.497)	(9.225)	(9.000)
$\alpha^D\,(\%)$	0.070	0.070	0.070
	(1.255)	(1.504)	(1.214)
β^1	0.900***	0.885***	0.890***
	(1.556×10^3)	(1.406×10^3)	(1.489×10^3)
β^2		−0.231***	−0.284***
		(91.528)	(109.815)
β^3		0.256***	0.140***
		(111.687)	(10.037)
β^4			−0.098***
			(8.196)
β^5			0.266***
			(23.375)

Note: This table shows the two-stage minimum D_K-distance estimation of interval-valued factor pricing models. The sample period is from January 2nd, 1986 to October 31st, 2017. These three interval-valued factor models (i.e., interval-valued CAPM, interval-valued 3-factor, interval-valued 5-factor) are obtained as follows:

$$Y_t = \alpha_0 + \alpha I_0 + \beta_1 X_t + \alpha^d Id_{t-1} + \alpha^D ID_{t-1} + u_t,$$

$$Y_t = \alpha_0 + \alpha I_0 + \beta_1 X_t^1 + \beta_2 X_t^2 + \beta_3 X_t^3 + \alpha^d Id_{t-1} + \alpha^D ID_{t-1} + u_t,$$

and

$$Y_t = \alpha_0 + \alpha I_0 + \beta_1 X_t^1 + \beta_2 X_t^2 + \beta_3 X_t^3 + \beta_4 X_t^4 + \beta_5 X_t^5 + \alpha^d Id_{t-1} + \alpha^D ID_{t-1} + u_t.$$

Asterisks *, ** and *** denote 10%, 5% and 1% significance respectively. The Wald-statistics of coefficients' estimators are presented in the brackets.

upper 5%th quantiles of oil price growth rates, instead of the economic implications of these coefficients. Panel B reports the descriptive statistics of the forecast values of the lower and upper 5%th quantiles; the average lower (upper) 5%th quantile is around 0.036 (0.037). As noted, if $r_t < q_t^{0.05}$, the interval-valued dummy variable Id_t is set at a unit interval $[-1/2, 1/2]$; if $r_t > q_t^{0.05}$, the interval-valued dummy variable ID_t is set at a unit interval $[-1/2, 1/2]$; and otherwise, they are both $[0, 0]$. This sample reveals 788 oil price changes that can be identified as oil prices shocks (394 positive and 394 negative).

3.3. Results for market efficiency

This section reports the main empirical results according to the interval-valued factor pricing model. Robust results based on classic factor models are presented subsequently.

The minimum D_K-distance estimators of the interval-valued factor pricing models are in Table 5, and they offer some notable findings. First, the hypothesis $\alpha^d = 0$ is rejected at 1% level in all three models. The magnitude of α^d is about −0.002, which implies an annualized loss of approximately $−0.002 \times 240 \approx −48$ %. The efficient market hypothesis α, α^D and α^d are all zero if the pricing model is correctly specified. Of the vast number of tests of three-factor pricing models, most reveal their fairly good empirical performance. The annualized loss of 48% is too large here. This finding thus violates the prediction of the efficient market hypothesis. A negative α^d, reflecting the pricing error related to negative oil price shocks, indicates that oil stocks tend to be overpriced in response to negative oil price shocks. Two potential explanations might address this anomalous finding: investors' underreaction or an anchoring effect.

Specifically, substantial decreases in crude oil prices often occur together with extremely bad news about the supply of or demand for crude oil. The intrinsic value of oil industrial shares thus should slump simultaneously. But if the market does not efficiently absorb the news, or investors underreact to it, oil stocks remain overpriced. Furthermore, investors often prefer to maintain their initial forecasts, leading to an anchoring effect, which could be related to "anchoring". This common human tendency reflects an overreliance on specific piece of information for making decisions. If more new information arrived, investors may update their revisions smoothly instead of over-adjusting; see Tversky and Kahneman (1974) and Capistrán and López-Moctezuma (2014).

Second, the coefficients α and α^D are both insignificant at the 10% level in all three models. Virtually no significant price errors arise with regard to moderate oil price changes or positive oil price shocks. This finding complies with the efficient market hypothesis. It also is worth noting that 95% of the observations of oil price changes are moderate or positive, so in the vast majority of cases, it is hard to reject this efficient market hypothesis, consistent with previous studies (see Fama and French (1997) and Arshanapalli et al. (1998)).

Third, the significance and signs of the betas are consistent with expectations. The estimator of β_1 is approximately 0.89 in all three models, indicating that oil stocks' factor loading on the market portfolio is slightly smaller than 1. The estimated value of β_2 is significantly negative at the 1% level in the three- and five-factor models. This finding is to be expected, considering the relatively large market values of oil industry firms. The estimator of β_3 is constantly positive and significant, consistent with oil companies' higher book-to-market ratios and financial leverages, relative to most industries.

Table 6
Robustness checks of pricing efficiency under different oil price changes.

	Column A	Column B	Column C	Column D
	(CAPM)	(FF3 model)	(FF5 model)	(Time-varying FF3 model)
α (%)	0.015	0.010	−0.009	0.000
	(1.200)	(0.810)	(−3.780)	(0.695)
α^d (%)	−0.208***	−0.213***	−0.197***	−0.185***
	(−3.710)	(−3.910)	(−3.780)	(−4.625)
α^D (%)	0.068	0.076	0.068	0.060
	(1.230)	(1.390)	(1.310)	(1.391)
β^{mkt}	0.903***	0.920***	1.059***	0.952
	(82.800)	(86.160)	(92.060)	
β^{siz}		−1.117***	0.009	−0.106
		(−5.800)	(0.460)	
β^{val}		0.437***	0.194***	0.548
		(21.170)	(8.420)	
β^{pro}			0.463***	
			(15.810)	
β^{inv}			0.684***	
			(19.460)	

Note: This table presents the tests for pricing efficiency of oil shocks based on various point-valued pricing models. The sample period is from January 2nd, 1986 to October 31st, 2017. Column A shows the estimation of

$$R^e_{t+1} = \alpha + \alpha^d d_t + \alpha^D D_t + \sum_{k=1}^{n} \beta^k F^k_{t+1} + \epsilon_{t+1},$$

$$E(\epsilon_{t+1}) = E(\epsilon_{t+1} d_t) = E(\epsilon_{t+1} D_t) = E\left(\epsilon_{t+1} F^1_{t+1}\right) = \ldots = E(\epsilon_{t+1} F^n_{t+1}) = 0,$$

based on Fama-French three-factor model; Column B shows the results based on traditional CAPM; Panel C shows the results based on Fama-French five-factor model. The models are estimated by OLS estimation. The t-statistics of coefficients' estimators are presented in the brackets. Panel D shows the result based on conditional Fama-French three-factor model,

$$R^e_{t+1} = \alpha + \alpha^d d_t + \alpha^D D_t + \beta^{mkt}_t Mktrf_{t+1} + \beta^{siz}_t SMB_{t+1} + \beta^{val}_t HML_{t+1} + \epsilon_{t+1},$$

$$E_t(\epsilon_{t+1} \otimes (1, Mktrf_{t+1}, SMB_{t+1}, HML_{t+1})) = 0_{1\times4},$$

$$\beta^{mkt}_t = a_1 + b_1 \beta^{mkt}_{t-1} + e'_{1t}; \beta^{siz}_t = a_2 + b_2 \beta^{siz}_{t-1} + e'_{2t}; \beta^{val}_t = a_3 + b_3 \beta^{val}_{t-1} + e'_{3t}.$$

The equations are estimated by Kalman filtering method. The loadings on three factors are reported as the time-series average smoothed state estimates. The z-statistics of coefficients' estimators are presented in the brackets. *, ** and *** denote 10%, 5% and 1% significance respectively.

Fourth, oil stocks tend to be overpriced in response to negative oil price shocks, while oil stocks are efficiently priced in response to positive oil price shocks. Positive oil price movements tend to have detrimental effect on aggregate equity market. Previous studies (Jones and Kaul, 1996; Sadorsky, 1999; Park and Ratti, 2008) reported that oil price increases have significantly negative impact on various stock markets from different countries and regions. Traditional mean-variance analysis implies the positive relationship between expected return of market portfolio and investors' utility (or Sharpe Ratio). Therefore, according to the theory of behavioral finance, investors will be more sensitive to positive oil price movements because oil price increases signal bad news to the investors in financial market. On the contrary, investors tend to underreact to the negative oil price movements.

Furthermore, another possible explanation is an asymmetric transmission from crude oil prices to stock prices. It is widely acknowledged that oil price shocks are mainly driven by global aggregate demand shocks and precautionary demand shocks (Kilian, 2009b). An increase in precautionary demand for oil would cause an increase in oil prices, which is considered as a positive oil shock. As Kilian (2009a) pointed out that this can lead to persistently negative U.S. stock returns. Empirically, investors are much more sensitive to negative news in financial markets. Thus, the oil stock prices respond quickly to positive demand shocks for oil and therefore this results in efficient market movements, while the oil stock prices respond slowly to negative demand shocks for oil and this results in inefficient market movements.

4. Robustness checks

Various robustness checks check some alternative model specifications: ordinary least square (OLS) regression for the traditional CAPM and Fama-French three- and five-factor models, based on commonly used point-valued excess returns and factors; a state-space model to estimate a time-varying beta Fama-French three-factor model; and classic point-based parameters.

First, the OLS estimations for traditional (i.e., point-valued) CAPM and the Fama-French three- and five-factor models appear in Columns A, B, and C, respectively, of Table 6. The significance of the pricing errors α, α^d and α^D is consistent with the findings from the interval-valued models: Both α and α^D are constantly insignificant, and α^d is negatively significant at a 1% level. This result affirms the previous conclusions that the market is efficient for moderate oil price changes and positive oil price shocks, but it tends to overprice oil stocks after negative oil price shocks. The sign and magnitude of the factor loadings also appear capricious in the point-valued models. Specifically, the point-valued CAPM and Fama-French three-factor model both indicate loadings on the market factor of less than 1. But the point-valued Fama-French five-factor model indicates a factor loading that is greater than 1. This contradictory finding suggests the need to identify the investment style of the oil industrial portfolio, as aggressive or passive. Unlike point-valued factor pricing models, the interval-valued models consistently imply a passive investment style, according to the factor loadings below 1. In addition, the point-valued Fama-French five-factor model produces a

Table 7
Estimation efficiency of parameters in interval-valued and point-valued factors models.

CAPM		α_0	α	β_1	α^d	α^D			
Interval	Bias	0.000	0.000	0.000	0.000	0.000			
	MSE	0.001	0.001	0.097	0.000	0.000			
	SD	0.000	0.000	0.010	0.000	0.000			
Point	Bias		0.000	0.002	0.000	0.000			
	MSE		0.001	0.123	0.006	0.005			
	SD		0.000	0.012	0.001	0.001			
3-Factor-model		α_0	α	β_1	β_2	β_3	α^d	α^D	
Interval	Bias	0.000	0.000	0.001	0.000	−0.000	0.000	0.000	
	MSE	0.001	0.001	0.113	0.029	0.027	0.001	0.001	
	SD	0.000	0.000	0.011	0.003	0.003	0.000	0.000	
Point	Bias		0.000	0.001	0.001	0.002	0.000	0.000	
	MSE		0.001	0.117	0.187	0.236	0.005	0.006	
	SD		0.000	0.012	0.019	0.023	0.001	0.001	
5-Factor-model		α_0	α	β_1	β_2	β_3	β_4	β_5	α^d $\quad\alpha^D$
Interval	Bias	0.000	0.000	−0.001	−0.001	0.001	−0.000	0.000	0.000 $\quad$ 0.000
	MSE	0.001	0.001	0.099	0.041	0.069	0.053	0.085	0.001 $\quad$ 0.001
	SD	0.000	0.000	0.010	0.004	0.007	0.005	0.009	0.000 $\quad$ 0.000
Point	Bias		0.000	0.001	−0.002	0.003	0.000	0.001	0.000 $\quad$ 0.000
	MSE		0.001	0.105	0.177	0.240	0.251	0.376	0.005 $\quad$ 0.005
	SD		0.000	0.010	0.018	0.024	0.025	0.038	0.001 $\quad$ 0.001

Note: This table shows the bootstrap results for estimation efficiency of parameters in interval-valued factor models and point-valued factor models respectively. Bias, MSE and SD of each parameter are computed based on 1000 bootstrap replications.

positive (though insignificant) loading for the size factor. This finding contradicts the reality, in which most oil industry firms have relatively large market values. Unlike these point-valued models, the interval-valued models always confirm a significantly negative loading on the size factor. Therefore, the interval-valued factor pricing models provide better empirical performance than traditional point-valued models.

Second, Column D of Table 6 contains the results of a Kalman filtering estimation of a time-varying beta Fama-French three-factor model,

$$R_{t+1}^e = \alpha + \alpha^d d_t + \alpha^D D_t + \beta_t^{mkt} Mktrf_{t+1} + \beta_t^{siz} SMB_{t+1} + \beta_t^{val} HML_{t+1} + \epsilon_{t+1}, \quad (26)$$

$$E_t(\epsilon_{t+1} \otimes (1, Mktrf_{t+1}, SMB_{t+1}, HML_{t+1})) = 0_{1\times 4}, \quad (27)$$

$$\beta_t^{mkt} = a_1 + b_1 \beta_{t-1}^{mkt} + e'_{1t}; \beta_t^{siz} = a_2 + b_2 \beta_{t-1}^{siz} + e'_{2t}; \beta_t^{val} = a_3 + b_3 \beta_{t-1}^{val} + e'_{3t}, \quad (28)$$

which assumes the factor loadings are mean-reverting. Previous studies indicate that unconditional OLS estimates of a factor pricing model might produce biased pricing error α if the true model is conditional (Jensen, 1968; Dybvig and Ross, 1985; Lettau and Ludvigson, 2001; Lewellen and Nagel, 2003). This robustness check helps rule out the possibility that the findings are driven by time-varying beta, not α. According to the Kalman filtering estimation, the coefficient α^d is significantly negative; α and α^D are both insignificant. That is, oil stocks are overpriced in response to negative oil price shocks but fairly priced for moderate oil price changes and positive oil price shocks. These results are consistent with the main findings.

Third, with a bootstrap method, it is possible to compare the estimation efficiency of the parameters of the interval-valued factor models with those of classic point-valued factor models. The point-valued innovations $\{\epsilon_t\}_{t=1}^T$ for Eqs. (18), (20) and (22), and the interval-valued innovations $\{u_t\}_{t=1}^T$ from Eqs. (17), (19) and (21) are as described in Section 2. In line with Chen and Hong (2012), this

study generated 1000 bootstrap samples and obtains 1000 bootstrap estimates for each parameter, to compute the key criteria (bias, mean square error (MSE), standard deviation (SD)) for the parameter estimators. For each bootstrap sample, OLS provides the estimate of the model parameters. The estimators also reveal the same set of model parameters for the interval CAPM, three-factor interval model, and five-factor interval model, using interval sample data and the minimum D_K-distance estimation procedure. Table 7 reports the bias, MSE, and SD values of the estimators. A comparison of interval-based and classic point-based estimators, according to the minimum D_K-distance estimation method using interval information for three models, yields more efficient estimates than the OLS estimators based on point-valued difference data. This finding highlights the need to gather the industry information contained in interval data, even if the goal is to model the difference between asset returns and risk-free interest rates.

5. Conclusion

This article proposes a two-stage procedure to examine stock market efficiency in pricing oil stocks in response to different types of crude oil price changes. The novel first step relies on quantile regression to identify oil shocks and their directions. In the second stage, interval-valued factor pricing models evaluate market efficiency, which supports a novel, minimum distance estimation. Improving on traditional point-valued data, the interval-valued observations contain more information and produce more efficient parameter estimates.

This newly proposed method in turn reveals some interesting insights. First, the findings challenge the efficient market hypothesis. Oil stocks tend to be overpriced after negative oil price shocks. Second, the pricing error rarely differs from zero after moderate oil price changes or positive oil price shocks, so the efficient market hypothesis is frequently supported. Third, oil stocks' factor loadings on the market portfolio are slightly smaller than 1, implying that oil stocks represent a conservative investment. Fourth, the factor loadings on size are significantly negative, due to the relatively large market values of oil industry firms. Finally, the factor loadings on the book-to-market ratio are significantly positive, consistent with the

notion that oil companies tend to have higher book-to-market ratios and financial leverage than firms in other industries.

By taking a fine-grained approach to investigate market efficiency in pricing oil stocks, this study reveals that oil stocks are overpriced when negative oil price shocks occur. Oil companies produce crude oil; in turn, a natural next question involves the market's performance in pricing stocks issued by their consumers, such as steel works, machinery industries, or the aircraft industry. This interesting issue will be a focus of continued research.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eneco.2019.04.016.

References

Aloui, C., Jammazi, R., 2009. The effects of crude oil shocks on stock market shifts behaviour: a regime switching approach. Energy Econ. 31 (5), 789–799.

Arshanapalli, B.G., Coggin, T.D., Doukas, J., 1998. Multifactor asset pricing analysis of international value investment strategies. J. Portf. Manag. 24 (2), 10–23.

Bernanke, B.S., Gertler, M., Watson, M.W., 2004. Reply: oil shocks and aggregate macroeconomic behavior: the role of monetary policy. J. Money Credit Bank. 36 (2), 287–291.

Bollerslev, T., Chou, R., Kroner, K., 1992. Arch modelling in finance: a review of the theory and empirical evidence. J. Econ. 52 (1-2), 5–59.

Boyer, M.M., Filion, D., 2007. Common and fundamental factors in stock returns of Canadian oil and gas companies. Energy Econ. 29 (3), 428–453.

Capistrán, C., López-Moctezuma, G., 2014. Forecast revisions of Mexican inflation and GDP growth. Int. J. Forecast. 30 (2), 177–191.

Ceylan, N.B., Berument, N.D.M.H., 2010. The impact of oil price shocks on the economic growth of selected MENA countries. Energy J. 31 (1), 149–176.

Cong, R.G., Wei, Y.M., Jiao, J.L., Fan, Y., 2008. Relationships between oil price shocks and stock market: an empirical analysis from China. Energy Policy 36 (9), 3544–3553.

Cunado, J., Gracia, F.P.D., 2003. Do oil price shocks matter? Evidence for some European countries. Energy Econ. 25 (2), 137–154.

Dybvig, P.H., Ross, S.A., 1985. Differential information and performance measurement using a security market line. J. Financ. 40 (2), 383–399.

Engemann, K.M., Owyang, M.T., Wall, H.J., 2014. Where is an oil shock? J. Reg. Sci. 54 (2), 169–185.

Fama, E.F., 1970. Efficient capital markets: a review of theory and empirical work. J. Financ. 25 (2), 383–417.

Fama, E.F., 1991. Efficient capital markets: II. J. Financ. 46 (5), 1575–1617.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. J. Financ. Econ. 33 (1), 3–56.

Fama, E.F., French, K.R., 1997. Industry costs of equity. J. Financ. Econ. 43 (2), 153–193.

Fama, E.F., French, K.R., 2015. Incremental variables and the investment opportunity set. J. Financ. Econ. 117 (3), 470–488.

Ferson, W.E., Korajczyk, R.A., 1995. Do arbitrage pricing models explain the predictability of stock returns? J. Bus. 68 (3), 309–349.

Hamilton, J.D., 1996. This is what happened to the oil price-macroeconomy relationship. J. Monet. Econ. 38 (2), 215–220.

Hamilton, J.D., 2003. What is an oil shock? J. Econ. 113 (2), 363–398.

Hammoudeh, S., Dibooglu, S., Aleisa, E., 2004. Relationships among U.S. oil prices and oil industry equity indices. Int. Rev. Econ. Financ. 13 (4), 427–453.

Han, A., Hong, Y., Wang, S., Yun, X., 2016. A vector autoregressive moving average model for interval-valued time series data. Adv. Econ. 417–460.

He, Y., Wang, S., Lai, K.K., 2010. Global economic activity and crude oil prices: a cointegration analysis. Energy Econ. 32 (4), 868–876.

Henriques, I., Sadorsky, P., 2008. Oil prices and the stock prices of alternative energy companies. Energy Econ. 30 (3), 998–1010.

Jensen, M.C., 1968. The performance of mutual funds in the period 1945–1964. J. Financ. 23 (2), 389–416.

Jones, C.M., Kaul, G., 1996. Oil and the stock markets. J. Financ. 51 (2), 463–491.

Kaucher, E., 1980. Interval Analysis in the Extended Interval Space IR. Fundamentals of Numerical Computation (Computer-Oriented Numerical Analysis), , pp. 33–49.

Kilian, L.and Park, C., 2009a. The impact of oil price shocks on the US stock market. Int. Econ. Rev. 50, 1267–1287.

Kilian, L., 2009b. Not all oil price shocks are alike: disentangling demand and supply shocks in the crude oil market. Am. Econ. Rev. 99 (3), 1053–1069.

Lee, K., Ni, S., 2002. On the dynamic effect of oil price shocks: a study using industry level data. J. Monet. Econ. 49 (4), 823–852.

Lee, K., Ni, S., Ratti, R.A., 1995. Oil shocks and the macroeconomy: the role of price variability. Energy J. 16 (4), 39–56.

Lettau, M., Ludvigson, S.C., 2001. Measuring and modelling variation in the risk-return trade-off. In: Constantinides, G., Harris, M., Stulz, R. (Eds.), Handbook of the Economics of Finance. Elsevier, pp. 617–690.

Lewellen, J., Nagel, S., 2003. The conditional CAPM does not explain asset-pricing anomalies. J. Financ. Econ. 82 (2), 289–314.

Li, Z., Sun, J., Wang, S., 2013. An information diffusion-based model of oil futures price. Energy Econ. 36, 518–525.

Lu, F., Hong, Y., Wang, S., Lai, K.K., Liu, J., 2014. Time-varying granger causality tests for applications in global crude oil markets. Energy Econ. 42, 289–298.

Nandha, M., Faff, R., 2008. Does oil move equity prices? A global view. Energy Econ. 30 (3), 986–997.

Nather, W., 1997. Linear statistical inference for random fuzzy data. Statistics 29 (3), 221–240.

Nather, W., 2000. On random fuzzy variables of second order and their application to linear statistical inference with fuzzy data. Metrika 51 (3), 201–221.

Park, J., Ratti, R.A., 2008. Oil price shocks and stock markets in the U.S. and 13 European countries. Energy Econ. 30 (5), 2587–2608.

Sadorsky, P., 1999. Oil price shocks and stock market activity. Energy Econ. 21 (5), 449–469.

Sadorsky, P., 2001. Risk factors in stock returns of Canadian oil and gas companies. Energy Econ. 23 (1), 17–28.

Sun, Y., Han, A., Hong, Y., Wang, S., 2018. Threshold autoregressive models for interval-valued time series data. J. Econ. 206, 414–446.

Sun, Y., Zhang, X., Hong, Y., Wang, S., 2019. Asymmetric pass-through of oil prices to gasoline prices with interval time series modelling. Energy Econ. 78, 165–173.

Timmermann, A., Granger, C.W.J., 2004. Efficient market hypothesis and forecasting. Int. J. Forecast. 20 (1), 15–27.

Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. Science 185 (4157), 1124–1131.

Yang, W., Han, A., Hong, Y., Wang, S., 2016. Analysis of crisis impact on crude oil prices: a new approach with interval time series modelling. Quant. Finan. 16 (12), 1917–1928.

Yu, L., Wang, S., Lai, K.K., 2008. Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. Energy Econ. 30 (5), 2623–2635.

Zhang, X., Lai, K.K., Wang, S., 2008. A new approach for crude oil price analysis based on empirical mode decomposition. Energy Econ. 30 (3), 905–918.

Zhang, X., Yu, L., Wang, S., Lai, K.K., 2009. Estimating the impact of extreme events on crude oil price: an EMD-based event analysis method. Energy Econ. 31 (5), 768–778.

Zhao, L., Zhang, X., Wang, S., Xu, S., 2016. The effects of oil price shocks on output and inflation in China. Energy Econ. 53, 101–110.