Market inefficiencies associated with pricing oil stocks during shocks
Qiao, Kenan; Sun, Yuying; Wang, Shouyang

Published in:
Energy Economics

DOI:
10.1016/j.eneco.2019.04.016

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

Publication date:
2019

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Market inefficiencies associated with pricing oil stocks during shocks

Kenan Qiao\textsuperscript{a,b,c}, Yuying Sun\textsuperscript{a,b,d,*}, Shouyang Wang\textsuperscript{a,b,d}

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

\textbf{ARTICLE INFO}

\textbf{Article history:}
Received 16 April 2018
Received in revised form 29 March 2019
Accepted 14 April 2019
Available online 9 May 2019

\textbf{JEL classification:}
C10
C49

\textbf{Keywords:}
Crude oil shocks
Interval-valued factor pricing models
Market efficiency
Oil stocks
Quantile regression

\textbf{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 pricing when all these different types of oil price changes occur.

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:

\[ NOP_{It} = \max \left(0, \ln \frac{p_t}{\max (p_{t-1}, p_{t-2}, p_{t-3}, p_{t-4})} \right), \]

where \( p_t \) denotes the oil price for quarter \( t \). If the current oil price exceeds the maximal oil 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

\[ NOP_{Di} = \min \left(0, \ln \frac{p_t}{\min (p_{t-1}, p_{t-2}, p_{t-3}, p_{t-4})} \right). \]

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:

\[ r_t = \mu + \sum_{i=1}^{n} \rho_i r_{t-i} + \epsilon_t, \]
\[ \epsilon_t = \sigma_t \epsilon_t, \]
\[ \sigma_t^2 = \omega + \sum_{i=1}^{p} \theta_i \sigma_{t-i}^2 + \sum_{i=1}^{q} \gamma_i \epsilon_{t-i}^2, \]

where \( r_t = \ln \frac{p_t}{p_{t-1}} - 1 \) is the oil price growth rate. Then the normalized growth rate is simply defined as the innovation process \( \epsilon_t \). 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 $\tau$th and upper $(1-\tau)$th quantiles of $r_{t+1}$, conditional on $\mathcal{F}_t$, can be calculated as

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

(6)

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

(7)

where $\tau$ 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 $q_{t+1}^\tau$ or a positive shock 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 changes. Suppose $\mathcal{F}_t$ is an information set observed at time $t$. The lower and upper $0.05$th quantiles of $r_{t+1}$, respectively, are

$$q_{t+1}^{0.05} = f(x_{t+1}^1, x_{t+1}^2, \ldots, x_{t+1}^n, \beta^\tau),$$

(8)

$$q_{t+1}^{0.95} = f(x_{t+1}^1, x_{t+1}^2, \ldots, x_{t+1}^n, \beta^{1-\tau}),$$

where $f$ is a predetermined function, and $\beta^\tau$ is the vector of unknown parameters. With some regularity conditions, $\beta^\tau$ can be estimated according to the following equation:

$$\hat{\beta}^\tau = \arg \min_{\beta^\tau} \left[ \sum_{n < q_{t+1}^{0.05}} (r_{t+1} - q_{t+1}^{0.05}) + \sum_{r_{t+1} > q_{t+1}^{0.95}} (r_{t+1} - q_{t+1}^{0.95}) \right].$$

(9)

Thus, the lower and upper $\tau$th quantiles of $r_{t+1}$ can be calculated as $q_{t+1}^\tau = f(x_{t+1}^1, x_{t+1}^2, \ldots, \hat{\beta}^\tau)$ and $q_{t+1}^{1-\tau} = f(x_{t+1}^1, x_{t+1}^2, \ldots, \hat{\beta}^{1-\tau})$, 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 $r_{t+1}$. 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 $x_{t+1}^1, x_{t+1}^2, \ldots, x_{t+1}^n$. Third, without the assumption of normally distributed error terms, quantile regression performs well in fitting $r_{t+1}$’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 $r_{t+1}$.

To specify the quantile regression, $\tau$ is set to 5%. Technically, $\tau$ 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 $x_{t+1}$, such that

$$f(x_{t+1}^1, x_{t+1}^2, \ldots, x_{t+1}^n; \beta^\tau) = \alpha^\tau + \beta_{x_1}^\tau x_{t+1}^1 + \beta_{x_2}^\tau x_{t+1}^2 + \ldots + \beta_{x_n}^\tau x_{t+1}^n.$$ 

(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 $x_{t+1}$: the AR(n)-GARCH(p,q)-based conditional expectation ($\bar{r}_t$) and variance ($\sigma^2_t$) of $r_{t+1}$, or else the empirical moments of $r_{t+1}$ in an n-period rolling window $[t-n, t-1]$. Because the conditional expected values and volatility of $r_{t+1}$ 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 $r_{t+1}$ reveal more information about $r_{t+1}$’s conditional distribution, especially its third- and fourth-order moments. Thus, it is possible to calculate the sample mean, variance, skewness, and kurtosis of $r_{t+1}$ in the n-period rolling window, then include them in the quantile regression as well. The skewness and kurtosis are calculated as

$$\text{skew}_{r_t} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{r_{t+i} - \text{ave}_{t+i}}{\text{vol}_{t+i}} \right)^3,$$

$$\text{kurt}_{r_t} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{r_{t+i} - \text{ave}_{t+i}}{\text{vol}_{t+i}} \right)^4,$$

(11)

where $\text{ave}_{t+i}$ and $\text{vol}_{t+i}$ are sample mean and standard deviation of $r_{t+i}$ in rolling window $[t-i, t-1]$.

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 ($\sigma^2_t$) of $r_{t+1}$, as in Eqs. (3) to (5), relies on the sample mean (ave$_t$), 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:

$$q_{t+1}^{0.05} = \alpha^{0.05} + \beta_{\bar{r}_t}^{0.05} \bar{r}_t + \beta_{\sigma^2_t}^{0.05} \sigma^2_t + \beta_{\text{ave}_{t+1}}^{0.05} \text{ave}_{t+1} + \beta_{\text{vol}_{t+1}}^{0.05} \text{vol}_{t+1}$$

$$+ \beta_{\text{skew}_{t+1}}^{0.05} \text{skew}_{t+1} + \beta_{\text{kurt}_{t+1}}^{0.05} \text{kurt}_{t+1},$$

(12)

and

$$q_{t+1}^{0.95} = \alpha^{0.95} + \beta_{\bar{r}_t}^{0.95} \bar{r}_t + \beta_{\sigma^2_t}^{0.95} \sigma^2_t + \beta_{\text{ave}_{t+1}}^{0.95} \text{ave}_{t+1} + \beta_{\text{vol}_{t+1}}^{0.95} \text{vol}_{t+1}$$

$$+ \beta_{\text{skew}_{t+1}}^{0.95} \text{skew}_{t+1} + \beta_{\text{kurt}_{t+1}}^{0.95} \text{kurt}_{t+1}.$$ 

(13)

Next, the process calculates the lower and upper 0.05th quantiles, $q_{t+1}^{0.05}$ and $q_{t+1}^{0.95}$. Finally, if $r_{t+1} > q_{t+1}^{0.95}$, $r_{t+1}$ is identified as a negative oil price shock; if $r_{t+1} < q_{t+1}^{0.05}$, $r_{t+1}$ is identified as a positive oil price shock; and otherwise, $r_{t+1}$ 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 assigning a functional relationship between oil price shocks and pricing errors. A misspecified model might distort 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 $R_R$ as the return on the risk-free asset and $R_i$ as the return on a risky asset. Let $\{Y_t \mid t = [R_R, R_i]\}$ and $\{X_t \mid X_t^{(n)} \}$ are respectively defined by their left bounds (i.e., $R_R$ or $X_0$) and right bounds (i.e., $R_i$ or $X_n$), 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, and $X_t$ and $X_t^{(n)}$ are defined by their left bounds (i.e., $R_R$ or $X_0$) and right bounds (i.e., $R_i$ or $X_n$), 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, and $X_t$ and $X_t^{(n)}$ are defined by their left bounds (i.e., $R_R$ or $X_0$) and right bounds (i.e., $R_i$ or $X_n$), 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, and $X_t$ and $X_t^{(n)}$ are defined by their left bounds (i.e., $R_R$ or $X_0$) and right bounds (i.e., $R_i$ or $X_n$), respectively (see Yang et al. (2016), Sun et al. (2018)).

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

$$
Y_t = \alpha_0 + \alpha_1 d_{t-1} + \alpha_2 D_{t-1} + \sum_{k=1}^{n} \beta_k X_t^{(n)} + \epsilon_t,
$$

where $Y_t$ is the interval-valued return constructed by the risk-free rate and asset return; $X_t^{(n)}$ is the $k$th 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; $Z_t = \{(1, 1), l_0, I_{t-1}, D_{t-1}, X_t, \cdots, X_t^{(n)}\}$; $l_0 = [-1/2, 1/2]$ is a constant, unit interval; and $\theta = (\alpha_0, \alpha_1, \alpha_2, \beta_1, \beta_2, \cdots, \beta_n)$. Eq. (15) is based on the assumption that $u_t = [u_{id}, u_{dt}]$ is an interval martingale difference sequence with respect to the information set $I_{t-1} = \{I_{t-1}, I_{t-1}, D_{t-1}, X_{t-1}, \cdots, X_t^{(n)}\}$, $\mathbb{E}$ is defined to reflect the Hadamard product for matrices based on the support function $s_A$ (see Section 2.4). In turn, $E_{t-1}(\cdot)$ is set at a unit interval

$$
E_{t-1}(0_{1 \times (n+3)}) = 0, E_{t-1}(0_{1 \times (n+3)}) = 0. \text{ 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 the two assets. Thus, the interval-based model can derive a traditional point-valued factor pricing model as follows:}

$$
R_t = R_R = \alpha + \alpha_1 d_{t-1} + \alpha_2 D_{t-1} + \sum_{k=1}^{n} \beta_k X_t^{(n)} + \epsilon_t,
$$

where $X_t^{(n)} = X_{t-1} - X_{t-1}, d_{t-1} = 1$ 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 $I_{t-1} = \left[-\frac{1}{2}, \frac{1}{2} \right]$, and $ID_t = \left[-\frac{1}{2}, \frac{1}{2} \right]$, $Id_t = \left[-\frac{1}{2}, \frac{1}{2} \right]$, as a special case of traditional factor models under state-based pricing errors (Ferson and Korajczyk, 1995).

Furthermore, $\alpha$ 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 $\alpha$ 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 $\alpha < 0$, the asset is overpriced.

The analysis of pricing error under different types of oil shocks relies on Eq. (14). If $\alpha$ is the pricing error under negative oil price shocks, and $\alpha^0$ is the pricing error under positive oil price shocks. Using $\alpha = 0$ as a benchmark, Table 1 summarizes the economic implications of $\alpha$, $\alpha_1$, and $\alpha_2$. It is worth noting that a significant $\alpha$ or $\alpha^0$ 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
where a factor model, and interval-valued Fama-French five-factor model. The pricing errors (Eq. (14)) also can produce some special cases, including interval data, which contain more information than ranged data. Thus, information set \( \mathcal{I}_{t-1} \), i.e., \( E(u_1|\mathcal{I}_{t-1}) = [0,0] \) 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:

\[
R_t - R_f = \alpha + \beta (R_{mt} - R_f) + \alpha^d d_{t-1} + \alpha^D D_{t-1} + \epsilon_t, \quad t = 1, \ldots, T, \tag{18}
\]

where \( E(\epsilon_t|\mathcal{I}_{t-1}) = E(u_1 - u_2|\mathcal{I}_{t-1}) = 0 \). The CAPM beta is measured by the coefficient \( \beta \) 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:

\[
Y_t = \alpha_0 + \alpha_1 + \alpha_2 d_{t-1} + \alpha^d d_{t-1} + \alpha^D D_{t-1} + \epsilon_t, \quad t = 1, \ldots, T, \tag{19}
\]

where

\[
Y_t = \begin{bmatrix} \alpha_0, \alpha_1, \alpha_2, d_{t-1} \end{bmatrix} = \begin{bmatrix} R_t - R_f, R_{mt} - R_f, \delta_{t-1} \end{bmatrix},
\]

\[
X_t = \begin{bmatrix} R_{HMLt}, R_{ARMWt}, R_{CMA_t}, R_{RMWt} \end{bmatrix}, \quad X_t^2 = \begin{bmatrix} (B/L_t + B/M_t + B/H_t), \frac{1}{2}(S/L_t + S/M_t + S/H_t) \end{bmatrix}.
\]

is the return on the portfolio of big market value firms, \( B/(L_t + S/M_t + S/H_t) \) is the return on the portfolio of small market value firms, \( S/(L_t + B/M_t + B/H_t) \) is the return on the portfolio of low book-to-market ratio firms, and \( S/(H_t + B/H_t) \) 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:

\[
R_t - R_f = \alpha + \alpha^d d_{t-1} + \alpha^D D_{t-1} + \beta_1 (R_{RMW} - R_f) + \beta_2 (R_{CMA} - R_f) + \beta_3 (R_{HML} - R_f) + \epsilon_t, \quad t = 1, \ldots, T, \tag{20}
\]

where \( \beta_1 \), \( \beta_2 \), and \( \beta_3 \) 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 \( \beta_{hi,t} = 1, 2, 3 \) for each factor. Small firms usually have relatively large factor loadings \( \beta_2 \), and high book-to-market ratio firms tend to have relatively large \( \beta_3 \).

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:

\[
Y_t = \alpha_0 + \alpha_1 + \alpha^d d_{t-1} + \alpha^D D_{t-1} + \epsilon_t, \quad t = 1, \ldots, T, \tag{21}
\]

where

\[
X_t^2 = \begin{bmatrix} \frac{1}{2}(S/W+B/W), \frac{1}{2}(S/R+B/R), \frac{1}{2}(S/A+B/A), \frac{1}{2}(S/C+B/C), \frac{1}{2}(S/W+B/W) \end{bmatrix},
\]

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:

\[
R_t - R_f = \alpha + \beta_1 (R_{RMW} - R_f) + \beta_2 (R_{CMA} - R_f) + \beta_3 (R_{HML} - R_f) + \beta_4 (R_{ARMW} - R_f) + \beta_5 (R_{RMW} - R_f) + \epsilon_t, \quad t = 1, \ldots, T, \tag{22}
\]

where \( \beta_1 \) is the difference between the returns on diversified portfolios of stocks with robust and weak profitability, and \( \beta_4 \) 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 \( \beta_4 \), but conservative firms have relatively large \( \beta_5 \).

### Table 1

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

<table>
<thead>
<tr>
<th>Oil price change</th>
<th>Negative oil shock</th>
<th>Moderate change</th>
<th>Positive oil shock</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha^2 &lt; 0 )</td>
<td>Over-priced</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha^2 &gt; 0 )</td>
<td>Under-priced</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha &lt; 0 )</td>
<td>Over-priced</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha &gt; 0 )</td>
<td>Under-priced</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha^2 &lt; 0 )</td>
<td>Over-priced</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha^2 &gt; 0 )</td>
<td>Under-priced</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: This table shows the economic implications of the pricing errors under different types of oil price changes. The pricing errors \( \alpha^2 \) and \( \alpha \) are derived by

\[
R_t = \alpha + \alpha^d d_{t-1} + \alpha^D D_{t-1} + \epsilon_t, \quad t = 1, \ldots, T, \tag{20}
\]

where \( SMBt \) and \( HMLt \) are the differences between the returns on 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 \( \beta_{hi,t} = 1, 2, 3 \) for each factor. Small firms usually have relatively large factor loadings \( \beta_2 \), and high book-to-market ratio firms tend to have relatively large \( \beta_3 \).

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:

\[
Y_t = \alpha_0 + \alpha + \alpha^d d_{t-1} + \alpha^D D_{t-1} + \epsilon_t, \quad t = 1, \ldots, T, \tag{21}
\]

where

\[
X_t^2 = \begin{bmatrix} (S/W+B/W), (S/R+B/R), (S/A+B/A), (S/C+B/C), (S/W+B/W) \end{bmatrix},
\]

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

\[
R_t - R_f = \alpha + \beta_1 (R_{RMW} - R_f) + \beta_2 (R_{CMA} - R_f) + \beta_3 (R_{HML} - R_f) + \beta_4 (R_{RMW} - R_f) + \beta_5 (R_{CMA} - R_f) + \epsilon_t, \quad t = 1, \ldots, T, \tag{22}
\]

### Table 2

Data description of main variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th># of obs.</th>
<th>Mean (%)</th>
<th>S.D. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth rate of oil price</td>
<td>8001</td>
<td>0.039</td>
<td>2.509</td>
</tr>
<tr>
<td>T-bill rate (( r_t ))</td>
<td>8026</td>
<td>0.013</td>
<td>0.010</td>
</tr>
<tr>
<td>Excess return on oil stock index (( % ))</td>
<td>8026</td>
<td>0.037</td>
<td>1.472</td>
</tr>
<tr>
<td>Excess return on market portfolio (( % ))</td>
<td>8026</td>
<td>0.034</td>
<td>1.109</td>
</tr>
<tr>
<td>Size factor (( SMkt ))</td>
<td>8026</td>
<td>0.002</td>
<td>0.586</td>
</tr>
<tr>
<td>Value factor (( HML ))</td>
<td>8026</td>
<td>0.011</td>
<td>0.573</td>
</tr>
<tr>
<td>Operating profitability factor (( RMW ))</td>
<td>8026</td>
<td>0.017</td>
<td>0.439</td>
</tr>
<tr>
<td>Investment factor (( CMA ))</td>
<td>8026</td>
<td>0.012</td>
<td>0.401</td>
</tr>
</tbody>
</table>

Note: This table reports the data description of main variables. Daily growth rate of oil price (\( r_t \)) is calculated as \( r_t = \frac{1}{2} \ln \left( \frac{P_{t+1}}{P_t} \right) \), where \( P_t \) is the WTI oil price which is downloaded from FRED Economic Data; daily T-bill rate (\( r_t \)). Excess return on oil stock index (\( \% \)). Excess return on market portfolio (\( \% \)). Size factor (\( SMkt \)). Value factor (\( HML \)). Operating profitability factor (\( RMW \)). Investment factor (\( CMA \)) are all collected from Kenneth R. French’s website. The excess return on Oil Stock Index (\( \% \)) is calculated as \( R_t - r_f \), where \( R_t \) is the net return on oil stock index. The sample is from January 2nd, 1996 to October 31st, 2017.
2.4. Estimation

Interval information supports estimates of the coefficients $\theta$. The challenge arises in the specification of the objective function to measure the sum of 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 distance between set-valued intervals $Y_t$ and its fitted value $Z_t' \theta$. Specifically, the $D_K$ metric is:

$$D_K^2(Y_t, Z_t') = \int_{(u,v) \in \mathcal{K}} \left[ s_Y(u), s_{Z'}(v) \right] \left[ s_Y(v), s_{Z'}(u) \right] dk(u,v),$$

$$= \left[ s_{Y_t - Z'_t \theta} \right]^2_{\mathcal{K}} = \left[ \| Y_t - Z'_t \theta \|_{\mathcal{K}} \right],$$

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 $(\cdot, \cdot, \cdot)$ 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:

$$\hat{\theta} = \left( \sum_{t=1}^T s_{Y_t}, s_{Z_t} \right)^{-1} \sum_{t=1}^T \left[ s_Y, s_{Z'_t} \right],$$

where $s_\theta(u)$ is the following support function:

$$s_\theta(u) = \begin{cases} \sup_{a \in A} \{ u - a | u \in s^0 \} & \text{if } A_L \leq A_R, \\ \inf_{a \in A} \{ u - a | u \in s^0 \} & \text{if } A_R \leq A_L, \end{cases}$$

and it follows that $s_\theta(u) = A_R$ if $u = 1$, $s_\theta(u) = -A_L$ 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 $K_{opt}$ with $K_{opt}(1,1) = T^{-1} \sum_{t=1}^T u_t^2$, $K_{opt}(-1,1) = T^{-1} \sum_{t=1}^T u_t u_{t-1}$, and $K_{opt}(-1,-1) = T^{-1} \sum_{t=1}^T u_t^2$ 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

Table 3

<table>
<thead>
<tr>
<th>Coef.</th>
<th>Conditional expectation</th>
<th>Conditional variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu(%)$</td>
<td>0.046**</td>
<td>$\rho$</td>
</tr>
<tr>
<td>(2.410)</td>
<td>(8.320)</td>
<td>$\omega$</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.891***</td>
<td>$\gamma$</td>
</tr>
<tr>
<td>(191.120)</td>
<td>(24.440)</td>
<td></td>
</tr>
</tbody>
</table>

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

$$\text{skew}_t = \frac{1}{n} \sum_{i=1}^n \left( \frac{r_t - \bar{r}_t}{\text{vol}_t} \right)^3, \text{kur}_t = \frac{1}{n} \sum_{i=1}^n \left( \frac{r_t - \bar{r}_t}{\text{vol}_t} \right)^4,$$

where $\bar{r}_t$ and $\text{vol}_t$ 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.
Estimation of quantile regressions for oil price growth rates and oil shocks identification.

Panel A: estimation of quantile regressions for oil price growth rates

<table>
<thead>
<tr>
<th></th>
<th>Lower 5%th quantile</th>
<th>Upper 5%th quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>-0.026***</td>
<td>0.025***</td>
</tr>
<tr>
<td>( \beta_0 )</td>
<td>0.640</td>
<td>0.130</td>
</tr>
<tr>
<td>( \beta_{10} )</td>
<td>-20.364***</td>
<td>18.140***</td>
</tr>
<tr>
<td>( \beta_{11} )</td>
<td>0.018</td>
<td>-0.571***</td>
</tr>
<tr>
<td>( \beta_{vol} )</td>
<td>0.861</td>
<td>5.812***</td>
</tr>
<tr>
<td>( \beta_{skew} )</td>
<td>0.001</td>
<td>-0.001</td>
</tr>
<tr>
<td>( \beta_{kurt} )</td>
<td>0.001</td>
<td>-0.001</td>
</tr>
</tbody>
</table>

Note: Panel A reports the quantile regression for oil price growth rates. The t-statistics of coefficients’ significance are **10%, ***5% and ****1% significance respectively. The t-statistics of coefficients’ skewness, kurtosis, respectively.

Panel B: oil price shocks identification

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>S.D.</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower 5%th quantile</td>
<td>0.037</td>
<td>0.016</td>
<td>394</td>
</tr>
<tr>
<td>Upper 5%th quantile</td>
<td>0.019</td>
<td>394</td>
<td></td>
</tr>
</tbody>
</table>

Note: Panel A reports the quantile regression for oil price growth rates. The t-statistics of coefficients’ significance are **10%, ***5% and ****1% significance respectively. The t-statistics of coefficients’ skewness, kurtosis, respectively.

Where \( r_t \) and \( q^0.05 \) denote the AR(1)-GARCH(1,1) based conditional expectation and variance of \( r_t \); \( ave_t \), \( vol_t \), \( skew_t \), 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 \( r_t < q^0.05 \), it is identified as a negative shock and set \( d_t = 1 \); if \( r_t > q^0.05 \), 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 \( r_t = \frac{p_{t+1} - p_t}{p_t} \), 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. 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 threeand five-factor models) also are available from Kenneth R. French’s website. 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 index stock is around 0.037 %, slightly larger than the risk premium of the market portfolio (0.034 %). The standard deviation of the oil index stock’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 2.51% and 3.02%. 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 \( \beta \) is negatively significant; that is, oil prices’ growth rates exhibit a mean-reversion pattern. The coefficients \( \theta \) and \( \gamma \) 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 upper 5% quantiles.
The annualized loss of approximately 48% is too large here. This finding thus violates the prediction of the efficient market hypothesis. A negative $\alpha$ reflects the pricing error related to negative oil price shocks, indicating 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).

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 $\alpha = 0$ is rejected at 1% level in all three models. The magnitude of $\alpha$ is about −0.002, which implies an annualized loss of approximately $-0.002 \times 240 \approx -4.8\%$. The efficient market hypothesis $\alpha = 0$ and $\alpha = 0$ 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 $\alpha$, 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 $\alpha$ and $\alpha = 0$ 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 $\beta_3$ 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 $\beta_3$ 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 $\beta_3$ is constantly positive and significant, consistent with oil companies’ higher book-to-market ratios and financial leverages, relative to most industries.

<table>
<thead>
<tr>
<th>Table 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_k$-minimum distance estimation of interval-valued factor pricing models.</td>
</tr>
<tr>
<td>Column A</td>
</tr>
<tr>
<td>(Interval-CAPM)</td>
</tr>
<tr>
<td>$\alpha_0(%)$</td>
</tr>
<tr>
<td>(0.640)</td>
</tr>
<tr>
<td>$\alpha(%)$</td>
</tr>
<tr>
<td>(1.199)</td>
</tr>
<tr>
<td>$\alpha^*(%)$</td>
</tr>
<tr>
<td>(8.497)</td>
</tr>
<tr>
<td>$\alpha^0(%)$</td>
</tr>
<tr>
<td>(1.255)</td>
</tr>
<tr>
<td>$\beta_1$</td>
</tr>
<tr>
<td>$(1.556 \times 10^3)$</td>
</tr>
<tr>
<td>$\beta_2$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$\beta_3$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$\beta_4$</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

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:

\[
Y_t = \alpha_0 + \alpha_1 \beta_1 Y_t + \alpha_2 \beta_2 Y_t + \alpha_3 \beta_3 Y_t + \alpha_4 \beta_4 Y_t + \alpha_5 \beta_5 Y_t + \alpha_6 \beta_6 Y_t + u_t,
\]

\[
Y_t = \alpha_0 + \alpha_1 \beta_1 Y_t + \alpha_2 \beta_2 Y_t + \alpha_3 \beta_3 Y_t + \alpha_4 \beta_4 Y_t + \alpha_5 \beta_5 Y_t + \alpha_6 \beta_6 Y_t + u_t,
\]

and

\[
Y_t = \alpha_0 + \alpha_1 \beta_1 Y_t + \alpha_2 \beta_2 Y_t + \alpha_3 \beta_3 Y_t + \alpha_4 \beta_4 Y_t + \alpha_5 \beta_5 Y_t + \alpha_6 \beta_6 Y_t + u_t.
\]

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

upper 5% 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% quantiles; the average lower (upper) 5% quantile is around 0.036 (0.037). As noted, if $r_t < q_{0.05}$, the interval-valued dummy variable $I_{d,t}$ is set at a unit interval $[−1,2,1/2]$; if $r_t > q_{0.05}$, the interval-valued dummy variable $I_{d,t}$ is set at a unit interval $[−1/2,1,2/1]$; and otherwise, they are both $[0,0]$. This sample reveals 788 oil price changes that can be identified as oil price shocks (394 positive and 394 negative).
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 to aggregate equity 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, 2009a). 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 slowly 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-factor model both indicate loadings on the market factor of less than 1. In addition, the point-valued Fama-French five-factor model produces 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: Both CAPM and the Fama-French three- and five-factor models appear in 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
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,

$$ R_t^{4,1} = \alpha + \alpha^d d_t + \alpha^p p_t + \beta_{t}^mkt \times t + \beta_{t}^{SMB} SMB_{t+1} + \beta_{t}^{HML} HML_{t+1} + \varepsilon_{t+1}. $$

$$ E(\varepsilon_{t+1} \otimes \{1, Mkt_{t+1}, SMB_{t+1}, HML_{t+1}\}) = 0_{4 \times 4}. $$

$$ \beta_t^{mkr} = a_1 + b_1/\beta_{t-1}^{mkr} + e_{t,1}^{mkr} = a_2 + b_2/\beta_{t-1}^{SMB} + e_{t,2}^{SMB}, \beta_t^{val} = a_3 + b_3/\beta_{t-1}^{val} + e_{t,3}^{val}. $$

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 $\alpha$ 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 $\alpha$. According to the Kalman filtering estimation, the coefficient $\alpha^d$ is significantly negative; $\alpha$ and $\alpha^p$ 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 $\{\varepsilon_t\}_{t=1}^T$ for Eqs. (18), (20) and (22), and the interval-valued innovations $\{d_t\}_{t=1}^T$ 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

We thank the editor, two referees, Lammertjan Dam, Yongmiao Hong, Arjan Trinks and some participants at the Third International Symposium on Interval Data Modelling: Theory and Applications at Beijing for their comments and suggestions. This work was partially supported by National Natural Science Foundation of China (Nos. 71703156, 71701199, 71871213) and funds No. 201601 provided by Fujian Provincial Key Laboratory of Statistics (Xiamen University).

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
