Price Discovery in China's Crude Oil Futures Markets: An Emerging Asian Benchmark?

(December 12, 2021)

Ziliang Yu, Assistant Professor of Finance
School of Finance, Nankai University
Email: yuziliang@nankai.edu.cn

Jian Yang*, J.P. Morgan Endowed Chair
J.P. Morgan Center for Commodities, Business School, University of Colorado Denver
Email: jian.yang@ucdenver.edu

Robert I. Webb, Paul Tudor Jones II Research Professor of Finance
McIntire School of Commerce, University of Virginia
Email: riw4j@virginia.edu

**Corresponding author*. Business School, PO Box 173364, University of Colorado Denver, Denver, CO 80217-3364. Email: jian.yang@ucdenver.edu. Tel: (303) 315-8423; Fax: (303) 315-8084

Acknowledgments: We acknowledge financial support from the National Natural Science Foundation of China (No. 72001119 & No. 72103105).

Abstract

We examine the price discovery performance of China's crude oil futures traded on the Shanghai International Energy Exchange (INE) for the spot prices of 19 types of deliverable and non-deliverable Asian crude oil. We find evidence for the INE crude oil futures price discovery function even at the early stage for almost all the deliverable crudes and some non-deliverable crudes. Both the INE crude oil futures price and spot price significantly contribute to the price discovery process, with the spot prices playing a somewhat more important role. While the price discovery performance was severely damaged around a period of the COVID-19 pandemic shock intensification in China with the temporary cancellation of the nighttime trading, it actually improved to some extent after China started the recovery from the shock. Further analysis reveals that both economic fundamentals (e.g., total warehouse inventory) and trading characteristics of the contract are significant determinants of price discovery performance. The findings imply that the INE crude oil futures have evolved into a useful and important information source in pricing Asian crudes and are emerging as an Asian benchmark.

Keywords: Crude oil futures; Time-varying price discovery; Asian benchmark; warehouse inventory; COVID-19

JEL Classification: G13, G14, C32

1. Introduction

Price discovery is a fundamental function of futures markets, as futures markets are supposedly react more promptly to new information because of centralized trading with higher liquidity and lower transaction costs (e.g., Working (1949); Black (1976)). The price discovery performance of commodity futures has been extensively investigated and empirically confirmed by the voluminous literature for agricultural futures (e.g., Adämmer and Bohl (2018); Yang, Li and Wang (2021)), and particularly relevant to this study, for crude oil futures (e.g., Schwarz and Szakmary (1994); Ng and Pirrong (1996); Switzer and El-Khoury (2007); Coppola (2008); Silverio and Szklo (2012); Shrestha (2014); Caporale, Ciferri and Girardi (2014)), among others.

The major crude oil futures contracts are extremely liquid and tied to a single crucial product that is widely consumed on a global basis (Armstrong, Cardella and Sabah (2021)). While crude oil futures markets are globally dominated by West Texas Intermediate (WTI) crude oil in the United States and Brent crude oil in the UK, a new contender emerged in March 2018 when the Shanghai International Energy Exchange (INE) launched its crude oil futures contract based on the type of crude oil widely used in Asia. This was also a response to the fact that Asia is the world's largest and fastest-growing oil consumer, and China has become the largest crude oil importer since 2017 (Yang and Zhou (2020)). Being China's first commodity futures market open to international investors (Yang and Zhou (2020); Fan et al. (2020)), the INE crude oil futures has quickly become the world's third-largest crude oil futures market with rapidly growing trading volume and open interest. Nevertheless, a major portion of the world's oil trade and long-term oil contracts, including those in Asia, are traditionally based on the WTI and Brent futures prices (Fattouh (2011)). Two questions naturally

arise. First, can INE's crude oil futures market perform the price discovery function reasonably well, particularly as almost all (seven during the sample or eight) underlying spot crudes are produced outside China (with only one exception)? Second, could the INE crude oil futures price emerge as a benchmark for all major types of crude used in Asia, even if they are not deliverable against the INE crude oil futures? This study tries to answer these intriguing questions, little explored in the price discovery literature.

This study examines the price discovery performance of China's crude oil futures traded on the INE for 19 types of Asian crude spot prices from its inception in March 2018 to the end of 2020. The study contributes to the price discovery literature in the following aspects. First, we comprehensively investigate the price discovery function of the INE's crude oil futures for both deliverable and major non-deliverable crude oils traded on the spot market in Asia. Most price discovery studies on commodity futures markets typically focus on a single popular deliverable spot price (Yang, Li and Wang (2021)), including all the aforementioned price discovery research in the crude oil futures markets. Despite the growing literature on price discovery of financial and commodity futures markets in China, the world's second-largest economy (e.g., Jin, Li, Wang and Yang (2018); Yang, Li and Wang (2021); Yang, Yang and Zhou (2012)), little research has been conducted on price discovery in China's crude oil futures market.[1] However, being claimed by some as "the single biggest change in capital markets," the success of this market potentially has great global implications and arguably may challenge the dominance of the US dollar in the international monetary system over

[1] One exception is Yang, Lv, Fang, and Shang (2020), who examine the price discovery of the INE crude oil futures, but only for two deliverable crudes rather than all seven (or currently eight) deliverable crudes. The spot prices in their study are Daqing, Shengli, and Oman. These few deliverable spot crudes are not main deliverable crudes, either (INE, 2021). Moreover, they do not examine potential time variation in the price discovery performance and its potential determinants (as we do in this study). Clearly, their analysis has a completely different focus from our study, as it only explores the existence of cointegration.

time (Reuters (2018a)), while good price discovery performance obviously is key to securing the success.

Second, this study also contributes to the literature on cross-border price discovery. The literature on cross-border price discovery mainly focuses on dual-listed stock prices (e.g., Cai, McGuinness and Zhang (2011); Li and Shi (2021); Eun and Sabherwal (2003); Frijns and Zwinkels (2018)), on-shore versus offshore foreign exchange markets (e.g., Chen and Xu (2021)), or similar futures traded on multiple exchanges (e.g., Ito and Lin (2001); Yang and Zhou (2020)). Cross-border price discovery of commodity futures for spot prices in other countries, to our knowledge, has not yet been thoroughly examined. Previous studies on price discovery on commodity futures, including crude oil futures, are typically based on the location of the futures markets and the underlying spot markets within the same country. Nevertheless, recent studies on agricultural commodity futures markets (e.g., Garcia, Irwin and Smith (2015); Yang, Li and Wang (2021)) reveal that poor delivery warehousing conditions can negatively affect the price discovery process, which is highly relevant to this study, as there is no international delivery warehousing for the INE crude oil futures during the sample period. Indeed, such a challenge in cross-border price discovery was cited as a major reason for not being optimistic about the potential success of the INE crude oil futures (Reuters (2018b)). Therefore, our examination of the 19 major Asian crudes, including all deliverable and major non-deliverable crudes, sheds fresh light on the cross-border price discovery of futures markets.

Finally, responding to the recent call for examining dynamic price discovery of Frijns and Zwinkels (2018) (albeit for the case of dually listed stocks), this study documents an intriguing time-varying price discovery pattern of the INE crude oil futures around the COVID-19 pandemic shock, and equally importantly, further

examines both economic and market-related determinants of the time-varying price discovery performance. The component share method of Gonzalo and Granger (1995) is combined with a modified version of recursive cointegration analysis of Hansen and Johansen (1999) to estimate the time-varying information contribution share of the futures market with allowance for the COVID-19 crisis. There is a gradual emergence of daily cash-futures cointegration over time before the COVID-19 pandemic shock, the damaging impact on the price discovery process due to the temporary suspension of INE's nighttime trading during several months of the COVID-19 intensification in China, and the resumed price discovery functioning after the pandemic shock was largely under control in China with resumed nighttime trading. Somewhat surprisingly and yet consistent with the dramatic increase in trading volume and open interests, the COVID-19 era generally improved the price discovery performance of the INE crude oil futures, while the cancellation of the nighttime trading negatively affected it. Hence, the study contributes to a rapidly growing literature examining the impacts of the COVID-19 pandemic on financial markets and firms (e.g., Bae, Ghoul, Gong and Guedhami (2021); Umar, Manel, Riaz and Gubareva (2021); Li, Liu, Mai and Zhang (forthcoming)), as its impacts on futures markets in general and price discovery performance in particular have been little explored.[2]

Furthermore, while a few studies (e.g., Silverio and Szklo (2012); Caporale, Ciferri and Girardi (2014)) examine time-varying price discovery performance of WTI or Brent crude oil futures, none of them have examined "the factors behind the time variation in the estimated time-varying price discovery measures" (Caporale, Ciferri and Girardi (2014), p.95).[3] Extending the literature (e.g., Alquist and Kilian (2010);

[2] As the pandemic is a truly exogenous shock of unprecedented magnitude, it provides a unique opportunity to test the resilience of the price discovery function in the INE's crude oil futures market.
[3] As reviewed by Caporale, Ciferri, and Girardi (2014), most of earlier studies on price discovery performance of crude oil futures employ conventional cointegration analysis and ignores potential

Armstrong, Cardella, and Sabah (2021)), we show, for the first time, that both economic fundamentals (e.g., China's domestic warehouse inventory) and market characteristics of the INE crude oil futures significantly affect the time-varying crude oil price discovery process.[4]

The remainder of this paper is organized as follows. Section 2 describes the study's data. Section 3 presents the empirical methodology used. Sections 4 and 5 interpret the empirical results for deliverable and non-deliverable crude oils, respectively. Section 6 examines the determinants of price discovery. Finally, section 7 concludes the paper.

2. Data

This study uses three types of data to examine the price discovery performance of the INE crude oil futures and its determinants from March 26, 2018, to December 31, 2020. First, we use continuous futures prices, measured by the daily closing prices of the most active futures contracts. The most active contracts are obtained by comparing the trading volumes of the first nearby futures contracts, the second nearby futures contracts, …, until the sixth nearby futures contracts. A unique feature of Chinese commodity futures markets is that the nearby futures contracts are typically not the most liquid and active contracts traded on the market (Fan and Zhang (2020)). Yang, Li, and Wang (2021) empirically confirm that using the prices of nearby futures contracts could lead to serious misleading inferences for the price discovery of China's agricultural futures markets. Similar to the pattern of the agricultural futures markets, many of the most active INE contracts are also not nearby contracts (see Appendix, Figure A-1).

time variation. Yang, Li, and Wang (2021) made a similar observation in the case of agricultural futures.

[4] The INE crude oil futures differ from Brent and WTI in that it has multiple (rather than a single) locations for delivery. Goldman Sachs (2018) considered such a warehouse delivery mechanism to be a key feature for the INE crude oil futures.

Second, we use 19 spot crude oil prices, consisting of the seven underlying deliverable spot crudes, one forthcoming deliverable spot crude starting in mid-2021, and 11 major Asian spot crudes. During the sample period, the seven deliverable crudes for the INE crude oil futures are UAE Dubai crude oil, Upper Zakum crude oil, Oman crude oil, Qatar Marine crude oil, Yemen's Masila crude oil, Iraq's Basrah light, and China's Shengli crude oil. Among these, Iraq's Basrah, UAE's Upper Zakum, and Oman crude are the three main crudes deliverable for settlement on the contract expiration dates, accounting for 71.5%, 16.2%, and 8.1% in 2020, respectively (see INE (2021)). On December 1, 2020, the INE announced that the UAE's Murban crude oil would be the eighth deliverable crude since June 1, 2021 (INE (2021)). Thus, for the convenience of discussion, we classify the Murban crude into the deliverable group when examining the price discovery, yet still classify it into the non-deliverable crude group when investigating the related determinants, as our sample ends in December 2020. In addition to domestic Daqing crude oil, the 11 non-deliverable major Asian crudes under consideration are from Saudi Arabia, Russia[5], Kuwait, Iran, Malaysia, and Indonesia, where these countries are among the top countries China imports oil from. We focus on the underlying deliverable crudes and discuss the other 11 major Asian crudes later.

Third, some futures contracts' trading-related variables (e.g., trading volume) and other demand- and supply-related factors (e.g., China's oil imports) are used to investigate the determinants of price discovery. We discuss these variables later in this study.

All the data are collected from two sources. The INE futures contract prices,

[5] We treat Russia's crude oil through the Eastern Siberia–Pacific Ocean (ESPO) oil pipeline as an Asian crude, because of the importance of ESPO crude to Asian crude oil markets as well as to China's oil import.

transaction information, and the spot crude oil prices are collected from Bloomberg while the other variables examined in this study are collected from the Wind data service. As the INE futures prices are denominated in China's Renminbi while the spot crude prices are denominated in the U.S. dollar, we convert the INE futures prices into equivalent dollar values using the appropriate daily exchange rates collected from Bloomberg. Both the futures prices and the spot prices are transformed into logarithmic values.

Figure 1 plots the eight pairs between (the logarithmic) values of futures prices and the spot prices. The futures and spot prices show strong co-movements during the sample period except for January to May 2020, when China was under the COVID-19 pandemic shock. Table 1 reports the summary statistics for both the (log-transformed) futures prices and spot prices. The unit-root test results reveal that both the INE futures prices and spot crude price series have a unit root, which, along with the strong co-movements shown in Figure 1, imply both the potential existence of price discovery and that the price discovery process might be affected by the COVID-19 pandemic shock.

[*Insert Table 1 about here*]

[*Insert Figure 1 about here*]

3. Empirical Methodology

This study combines conventional and recursive cointegration analysis, generalized forecast error variance decomposition, and a time-varying component share method to investigate the price discovery performance of INE crude oil futures and related determinants. The improved recursive cointegration analysis can reveal how the price discovery process evolves with time, while the forecast error variance decomposition provides extra information on the economic significance and dynamic adjustment to

shocks.

As summarized in Yang, Bessler, and Leatham (2001), examining price discovery for commodity futures markets is equivalent to examining two relationships between futures and spot prices. First, the unbiasedness hypothesis that futures prices are an unbiased estimate of future spot prices. This hypothesis is equivalent to a cointegration relationship between futures prices and spot prices with a cointegrating vector of $(1, -1)$ (Brenner and Kroner (1995)). Second, the prediction hypothesis is that futures prices lead changes in the spot price for the existing cointegration relationship.

Let $P_t = \begin{pmatrix} P_{ft} \\ P_{st} \end{pmatrix}$, where P_{ft} is the logarithm value of the INE crude oil futures price, and P_{st} is the logarithm value of the underlying crude oil spot price. If both prices have a unit root (i.e., $I(1)$), they can be modeled in the following form:

$$\Delta P_t = \alpha_0 + \alpha\beta' P_{t-1} + \sum_{t=1}^{k-1} \Gamma_t \Delta P_{t-1} + \mu + \varepsilon_t \quad (t = 2, \cdots, T) \qquad (1)$$

where α_0 is the vector of the constant. The rank of $\Pi = \alpha\beta'$ determines the number of cointegration vectors of the cointegration relationships. The α and β vectors lie at the core of the cointegration analysis for price discovery.

As detailed in Yang et al. (2001), we can use two tests to examine the unbiasedness hypothesis. First, we examine the existence of a cointegration relationship using the trace tests (Johansen (1991)). Assuming the number of cointegration relations is r, then:

$$H_0(r): \Pi = \alpha\beta', \qquad (2)$$

where $\alpha\beta' = \begin{pmatrix} \alpha_f \\ \alpha_s \end{pmatrix} (\beta_f \quad \beta_s)$. A trace test is used to test the null hypothesis that there are (at most) r $(0 \leq r < p)$ cointegration vectors, where p is the number of variables in P_t and is two in this study. The trace test statistic is as follows:

$$\lambda_{trace} = -\mathrm{T} \sum_{\mathrm{t=r+1}}^{\mathrm{p}} \log(1 - \lambda_{\mathrm{i}}) \tag{3}$$

where T is the number of observations, and λ_{i}s are the eigenvalues. Testing the rank of Π in Equation (2) also needs to consider whether μ enters the VECM (Equation (1)) as a constant in the cointegrating vector or as a time trend in the original representation in levels. We follow the sequential hypothesis testing procedure used by Yang et al. (2001).

Second, the unbiasedness hypothesis can be tested regarding the cointegrating vector $\beta' = (\beta_f \quad \beta_s)$. The hypothesis $\beta' = (1 \quad -1)$, mathematically, can be summarized as follows:

$$H_1|H_0: R'\beta = 0 \tag{4}$$

where $R' = (1 \quad 1)$. An appropriate likelihood ratio test statistic can be applied to test whether such a hypothesis holds. The prediction hypothesis can be formulated as statistical tests with respect to the loading vector $\alpha = \binom{\alpha_f}{\alpha_s}$. There were three cases of interest in this study. (1) $\alpha_f = 0$ if futures prices are weakly exogenous; (2) $\alpha_s = 0$ if spot prices are weakly exogenous; and (3) $\alpha_f \neq 0$ and $\alpha_s \neq 0$ if there is a bidirectional long-run information flow between futures and spot prices, which can be further analyzed as follows:

$$H_2|H_1: \mathrm{B}'\alpha = 0 \tag{5}$$

where $\mathrm{B}' = (1 \quad 0)(\text{or } (0 \quad 1))$ if P_{ft} (or P_{st}) is a weakly exogenous series for Case 2 (or 1). According to Gonzalo and Granger (1995), in the first two cases above, a weakly exogenous series would be the primary source of information in the long run. Again, an LR test statistic can be applied. Nevertheless, in the more common case of $\alpha_f \neq 0$ and $\alpha_s \neq 0$, we follow Eun and Sabherwal (2003) and use a modified version

of the component share method of Gonzalo and Granger (1995) $\frac{|\alpha_s|}{|\alpha_f|+|\alpha_s|}$ to measure the contribution of futures prices to price discovery) and test the prediction hypothesis.

In this study, we focus on using the Gonzalo and Granger (1995)'s method for the following reasons. First, it greatly facilitates the comparison with the existing literature most relevant to this study. While Hasbrouck's (1995) information share is also widely used and often together with Gonzalo and Granger (1995)'s method by many other studies (e.g., Chen and Tsai (2017)), price discovery on commodity markets in general and crude oil futures in particular (e.g., Switzer and El-Khoury (2007); Coppola (2008); Figuerola-Ferretti and Gonzalo (2010); Caporale, Ciferri and Girardi (2014); Adämmer and Bohl (2018); Yang, Li and Wang (2021)) typically focus on Gonzalo and Granger (1995)'s method (or the closely related error correction terms). Figuerola-Ferretti and Gonzalo (2010) also demonstrate that the price discovery process in the framework of Garbade and Silber (1983) coincides exactly with the permanent component of the Gonzalo and Granger (1995). Second, compared with Gonzalo and Granger (1995)'s method, Hasbrouck's (1995) information share has an obvious technical challenge as it does not provide a unique measure but merely upper and lower bounds of a market's information share (Grammig and Peter (2013)). These bounds often can diverge considerably and thus prevent from detecting the market leading price discovery. As reviewed recently in Li and Shi (2021), while various improvements (e.g., Lien and Shrestha (2009); Grammig and Peter (2013)) have been proposed, they are also subsequently challenged, and thus whether the improvements are satisfactory is yet to be settled.[6] Third, the focus on Gonzalo and Granger (1995)'s method also naturally

[6] For example, Lien and Wang (2016) find that under the chosen data generation processes, Grammig and Peter (2013) perform poorly, while Lien and Shrestha (2009) provide at most marginal improvement over the method based upon the upper/lower bound midpoint of the Hasbrouck. The point of the unsettled nature can be extended to other proposed improvements along this line. Putniņš (2013) introduces a more robust measure named the Information Leadership Share (ILS)

fits well with the recursive (or rolling) cointegration analysis to produce a time-varying dynamic price discovery measure.

In addition to the previous standard cointegration test, we further apply recursive cointegration analysis (Hansen and Johansen (1999)) to investigate the possible time-varying cointegration relationship between futures prices and spot prices.[7] We modify the method with allowance for two additional dummy variables related to the COVID-19 and the temporary suspension of the nighttime trading. Such recursive cointegration analysis not only produces time-varying estimates of a cointegration relationship, but more importantly in this study, also time varying recursive estimates of α_f and α_s. Thus, we are able to construct time-varying component shares based on $\frac{|\alpha_s|}{|\alpha_f|+|\alpha_s|}$ at each point of time, which is key to the further analysis of the determinants in the price discovery process in this study. Consistent with the emphasis on the component share method, we use the so-called "R-representation" for recursive cointegration analysis to generate time-varying component shares. In this "R-representation", the short-run parameters Γ_i are kept fixed to their full sample values, and only the long-run parameters in the model (1) are re-estimated (Hansen and Johansen, 1999).

Following the literature (e.g., Hodgson, Masih and Masih (2003); Tong, Wang and Yang (2016)), we also use the generalized vector autoregression (VAR) forecast error

based on the combination of Gonzalo and Granger (1995) and Hasbrouck (1995). However, as pointed out by Li and Shi (2021), a significant issue of the ILS is still the lack of uniqueness, as it is based on Hasbrouck's (1995) information share. According to Li and Shi (2021), under certain circumstances, the ILS could lead to the worst estimation accuracy based on the simulated data, when compared with other methods.

[7] We use recursive rather rolling cointegration analysis, because Kim and Swanson (2014) find that though not always, recursive estimation often "dominates" rolling estimation in generating better fit models. This is because more data points and thus more information are exploited. Recursive cointegration analysis may be particularly well suited for examining how China's nascent crude oil futures markets may evolve in improving price discovery performance over time, and has a general advantage in addressing the potential mixed evidence on the evolving lead-lag relationship between futures and spot prices (Yang, Yang, and Zhou (2012); Yang, Li and Wang (2021)).

variance decomposition (Pesaran and Shin (1998)) to explore dynamic adjustments of the price discovery process. The generalized VAR forecast error variance decomposition approach can eliminate the possible dependence of the results on the ordering of variables (i.e., Cholesky factor orthogonalization). Specifically, for each VECM model of equation (1), the generalized VAR *H*-step-ahead error variance decomposition, $\theta_{ij}^{g}(H)$ for $\mathrm{H} = 1,2,\cdots$, is:

$$\theta_{ij}^{g}(H) = \frac{\sigma_{jj}^{-1}\sum_{h=0}^{H-1}\left(e'_{i}A_{h}\sum e_{j}\right)^{2}}{\sum_{h=0}^{H-1}\left(e'_{i}A_{h}\sum A'_{h}e_{j}\right)} \tag{6}$$

where Σ is the variance matrix of ε, σ_{jj} is the standard deviation of the error term for the *j*-th equation, and e_i is the selection vector, with the *i*-th element equal to one and all other elements equal to zero. As the sum of all the elements in each row of the variance decomposition table under the generalized VAR framework is not equal to one, following related literature (e.g., Yang, Hsiao, Li and Wang (2006)), we normalize each entry of the variance decomposition matrix by its row sum:

$$\tilde{\theta}_{ij}^{g}(H) = \frac{\theta_{ij}^{g}(H)}{\sum_{j=1}^{N}\theta_{ij}^{g}(H)} \tag{7}$$

4. The price discovery performance for all deliverable crude oils

This section examines the price discovery of the INE crude oil futures price for the seven underlying deliverable spot crude oil prices, plus UAE's Murban, which was not deliverable in our sample period, but was the first new grade to be added as a deliverable grade (starting from June 1, 2021) since the launch of the INE crude oil futures in March 2018. We first report the full-sample results using the conventional cointegration tests, generalized VAR forecast error variance decomposition, and then return to the recursive cointegration analysis to further explore how the price discovery process evolves with time.

4.1 Cointegration tests

Table 2 reports the cointegration tests for the eight pairs between INE futures price and underlying deliverable spot crude oil prices (including Murban) during the full-sample period. For each test, the optimal lag length is chosen by the Bayesian information criteria (BIC). At the conventional 5% significance level, the results of both the trace tests and the maximum eigenvalue tests reveal one cointegration relationship ($r = 1$) for each pair. Despite the concern that the domestic crude oil spot prices might not yet be fully market-oriented as the crude oil industry in China is probably still controlled by the government and is dominated by a few large state-owned enterprises, the INE futures is still cointegrated with China's Shengli crude, the only deliverable domestic crude oil on the INE.

Seven out of the eight underlying spot crudes are produced by, and imported from, Middle Eastern countries. Note that the Brent futures price is generally used as the benchmark price for crude oil trade in the Middle East (Fattouh, 2011). The existence of cointegration relationships for all eight pairs, with a possible exception of Yemen[8], reveals that China's INE futures price probably performs, to some extent, price discovery for the seven imported Middle Eastern crude oils. The price discovery performance of INE is consistent with China's role as the world's largest crude oil importer, which helps establish the linkage between the INE futures price and spot prices of oil-exporting countries in the Middle East.

[*Insert Table 2 about here*]

As the cointegration relationships have been found to be statistically significant at the 5% significance level, we use the LR tests to further examine the unbiasedness

[8] The test statistics for r = 1 could be rejected at the 10% significance level for Yemen in the case of no linear trend. Further recursive analysis in Figure 3 also demonstrates that the concern might be well grounded. The spot price series for Yemen was not available in the more recent period of the sample.

hypothesis and the prediction hypothesis of price discovery. Panel A of Table 3 reports the results of the unbiasedness hypothesis with the restriction that $\beta' = (1 \quad -1)$ and the test results with other restrictions on the β vector. The other restrictions include whether the COVID-19 dummy, allows for the impact of the COVID-19 pandemic shock, or whether the trend in the cointegration space can be excluded. The unbiasedness hypothesis of price discovery is soundly rejected even at the 1% significance level with or without considering a linear trend in the cointegration space, except for Yemen's Masila, which is rejected at the 5% significance level. The hypothesis that the linear trend can be excluded cannot be rejected at the conventional 5% significance level, except for the Shengli crude. Thus, we can conclude that there is generally no linear trend in the cointegration relationship. The hypothesis that the COVID-19 dummy can be excluded from the cointegration space is also rejected at the conventional 5% significance level for each pair in the cointegration space without a linear trend. This hypothesis is accepted for all cases in cointegration with a linear trend. Thus, we use cointegration without a linear trend as our benchmark model, in which the unbiasedness hypothesis and the hypothesis of no impact from the COVID-19 pandemic shock are soundly rejected.

Panel B of Table 3 reports the testing results of the prediction hypothesis that impose restrictions on the α vector. The imposed restriction is $B'\alpha = 0$, where $\alpha' = (\alpha_f \quad \alpha_s)$. $\alpha_f = 0$ indicates that the futures price leads the spot price; $\alpha_s = 0$ indicates that the spot price leads the futures price; if neither α_f nor α_s is zero, it implies that the information transmission between futures price and spot price is bidirectional. The results from Panel B reveal that the hypothesis of either $\alpha_f = 0$ or $\alpha_s = 0$ is significantly rejected at the 5% significance level for each pair with or without considering the liner tread in the cointegration space. Therefore, the INE

futures price does not dominate in the futures-spot price information transmission; the transmission is bidirectional.

In summary, we can obtain preliminary evidence that although there is a cointegration relationship between China's INE futures price and underlying spot crude oil prices (including UAE's Murban), the information flow is bidirectional; the futures price does not dominate in the price discovery process with a leading information role, it also reacts to the information from the spot prices.

[*Insert Table 3 about here*]

4.2 Forecast error variance decomposition

This section analyzes the eight error correction models (VECMs) using the generalized VAR forecast error variance decomposition to further explore the bilateral information flow between the INE crude oil futures price and underlying spot crude prices.

Table 4 reports the generalized VAR forecast error variance decomposition at 360 steps, approximately equivalent to trading days of one year and a half, revealing two noticeable results of how bilateral information is transmitted from one market to the other at the short- and long-term horizons. First, echoing the prediction hypothesis results (Panel B of Table 3), the impact of futures price on the spot price is approximately equivalent to that of the impact of spot price on futures price at the short horizon. For instance, within five steps (i.e., a week), the movement of spot price can be explained by the futures price from 29 % (Murban) to 39% (Yemen), while the spot price can explain the movement of the futures price from 30% (Iraq) to 38% (Yemen). This again is consistent with the earlier preliminary evidence that the INE does not play a dominant leading information role.

Second, the explanatory power of spot prices for futures prices is much higher than that of futures prices for spot prices in the long-term horizon. On average, at 360 steps,

approximately 60% (Shengli) to 78% (Dubai) futures price changes can be explained by the spot price, while only about 20% (Dubai) to 39% (Yemen) spot price changes can be explained by futures prices. The explanatory power of futures prices for spot prices in the long-term horizon is quite comparable with that of the short-term horizon. The explanatory power of spot prices for futures price triples from the short-term to long-term horizon. The result further reveals that although the INE futures price contributes noticeably to the price discovery process, it does not dominate in the price discovery process.

[*Insert Table 4 about here*]

4.3 Recursive cointegration analysis and time-varying component shares

This section applies an improved recursive analysis to further explore how INE performs price discovery over time. As the unbiasedness hypothesis has been rejected previously, our recursive cointegration analysis focuses on exploring the cointegration relationships and the prediction hypothesis. The analysis is based on the "R-representation" where only the long-run parameters in the model are re-estimated. We use the first 85 observations (i.e., from March 26 to July 31, 2018) for the initial estimation period for the recursive analysis. The use of the first 85 observations as the initial sample period because China's INE emerged as the world's third-largest futures market within this period (Yang and Zhou (2020)). We also want to explore how the price discovery process evolved before the first INE futures contract matured in September 2018. We then expanded the sample period with one additional data point for each recursive test.

Figure 2 plots the recursive trace tests for the eight underlying spot crudes, with the test statistics normalized by the critical value at the 5% significance level. Thus, a statistic value larger than one implies rejecting the null hypothesis at a conventional 5%

significance level. Given the well-known power of Johansen's cointegration test, the cointegration results reveal that the INE crude oil futures price was cointegrated with the Dubai, Upper Zakum, Oman, Qatar, Yemen, and Murban spot crude prices since around March 2019, about one year after its inception, to the outbreak of COVID-19 pandemic shock in January 2020. Such cointegration relationships were severely interrupted by the COVID-19 pandemic shock but continued when China recovered from the shock and re-opened the nighttime trading of INE. More interestingly, we can find that the cointegration stays relatively more stable for all the spot crudes after the COVID-19 pandemic shock, as the test statistics become less volatile than that before the shock. This pattern holds even for Iraq's Basrah and China's Shengli, which show relatively weak (Iraq) or no (Shengli) cointegration relationship before the pandemic shock but exhibit certain cointegration relationships after the shock. Extending our previous full-sample static cointegration test results, the recursive cointegration analysis reveals that the COVID-19 shock significantly affects the price discovery performance of INE's crude oil futures. The more stable cointegration relationship in the last half of 2020 might be explained by the worldwide outbreak of the COVID-19 pandemic when China nevertheless largely put it under control and recovered from the shock. Put differently, as the world economy was severely affected by the COVID-19 pandemic shock while only China recovered, China's strong oil consumption demand might help INE crude oil futures to perform the price discovery function again.

[*Insert Figure 2 about here*]

Figure 3 depicts the estimated adjusting coefficients (i.e., values in α); Figure 4 depicts the time-varying values of INE's crude oil futures contribution to the price discovery process calculated from the component share method (i.e., $\frac{|\alpha_s|}{|\alpha_f|+|\alpha_s|}$). Again, confirming the previous full-sample results, we find that both futures prices (left

column of Figure 3) and spot prices (right column of Figure 3) adjust to the price deviation from the long-run relationship between spot and futures prices. The extent of adjustment for each pair, as shown in Figure 4, is similar. The INE crude oil futures price contributes approximately 50% to the price discovery process of Dubai, Upper Zakum, Qatar, and Murban spot prices before and after the COVID-19 pandemic shock, while it is below 50% before the shock but around 50% after the shock for the Omani and Iraqi spot prices. In general, the contribution to price discovery apparently increased to some extent after the pandemic shock and remained relatively more stable (see Figure 4).

[*Insert Figure 3 about here*]

[*Insert Figure 4 about here*]

5. The price discovery performance for 11 non-deliverable Asian crude oils

Thus far, the analysis shows that the INE futures price contributes significantly, though not dominantly, to the price discovery for the underlying deliverable spot crude prices, including UAE's Murban crude. As INE was cointegrated with the two global crude oil futures markets, WTI and Brent, and became the world's third-largest crude oil futures market concerning trading volume within only three months since its inception (Yang and Zhou (2020)). The existence of a price discovery performance of INE for the underlying spot crudes from the Middle Eastern area may lead to another question: does INE also perform a price discovery function for other Asian crude oil spot prices? To answer this question, we re-conduct the examinations for the 11 major Asian crude oil spot prices, with the crudes chosen from the Asian countries that ranked among China's top 30 oil-importing countries, except for the domestic Daqing crude and Russia's ESPO. Figure A-2 in the appendix plots the 11 pairs between the (logarithmic) value of the futures price and the spot prices. Table A-1 in the appendix reports the description,

summary statistics, and unit-root tests for these 11 major Asian crude oil spot prices, revealing that all the spot price series have a unit root (i.e., $I(1)$).

Similar to the previous examination of deliverable spot crudes, we first conduct full-sample cointegration trace tests and then apply the LR test to examine the unbiasedness and prediction hypotheses. Table A-2 in the appendix reports the cointegration trace tests, and four cointegration pairs are found out of the 11 pairs under investigation: Saudi Arabia's medium crude to Asia, Russia's ESPO crude, Kuwait crude, and Iran's Foroozan crude (see Table A-2). The cointegration results for other spot prices are too good to be true and unreliable, as it is theoretically impossible for two time series with a unit root to have a cointegration rank of two. Further recursive cointegration analysis (available on request) confirms this finding. We find a cointegration relationship between Saudi Arabia's Medium to Asia, Russia's ESPO, Kuwait crude, and Iran's Foroozan crude before and after the COVID-19 pandemic shock. The cointegration relationship for each of these four pairs is quite mixed during the shock period. However, it has also become relatively stable since June 2020, when China recovered from the pandemic shock. In summary, INE performs a certain price discovery function for four out of the 11 other non-deliverable major Asian crudes similar to deliverable crudes.

6. Determinants of the dynamic price discovery process

The examination above shows that the INE crude oil futures perform the price discovery function for spot prices across all deliverable crudes over time. In what follows, we further investigate the factors that affect the price discovery process.

6.1 The COVID-19 and basic trading characteristics

Following Yang, Li, and Wang (2021), we used two dependent variables for the analysis. The first dependent variable is the recursive trace test results in Figures 2, with assigned

value 1 to denote a cointegration relationship ($I_{Trace>5\%}$) and 0 for others, for each time point at the 5% significance level. The second dependent variable is the contribution of the INE futures price to price discovery, calculated as $\frac{|\alpha_s|}{|\alpha_f|+|\alpha_s|}$ where α_f and α_s are the adjustment coefficients estimated from the recursive cointegration analysis. A higher value of this ratio reflects a greater contribution of price discovery from the INE crude oil futures price (Eun and Sabherwal (2003); Gonzalo and Granger (1995)).

The independent variables are the natural logarithm value of the daily number of contracts traded, $\log(Volume)$; the logarithm value of daily open interest, $\log(Open\ Interest)$; the ratio that measures the speculative intensity, $\frac{Volume}{Open\ Interest}$; daily volatility, $(\log(\frac{Price_t}{Price_{t-1}}))^2$, where $Price_t$ is the daily closed futures price; $\log(\frac{Price_{open,t}}{Price_{settle,t-1}})$, which measures the information flow from the end of the last trading day to the beginning of the current trading day; COVID, a dummy variable which indicates the time since January 23, 2020 ($\mathrm{COVID} = 1$; 0 otherwise) capturing the impact of the COVID-19 pandemic shock; and $\mathrm{NigCancel}$, a dummy variable which indicates the period from February 3 to May 6, 2020 ($\mathrm{NigCancel} = 1$; 0 otherwise) capturing the impact of the cancelation of nighttime trading.

Table 5 reports the results based on the deliverable crudes. As the Murban crude oil was still not a deliverable crude during the sample period, we exclude it from the deliverable group here. For the logistic regression with $I_{Trace>5\%}$ as the dependent variable, we can find that almost all the coefficients of the independent variables under consideration are estimated significantly in the simple bivariable regressions except for $\log(Open\ Interest)$. In the following multiple regression with all these detected significant determinants, only the coefficient of $\log(\frac{Price_{open,t}}{Price_{settle,t-1}})$ loses its

significance while the others show a similar result as that in the simple regressions. The results thus show that the COVID-19 pandemic shock improved the price discovery process of INE futures price, while the cancelation of nighttime trading, along with some futures-market-related factors, negatively affect the price discovery process. For the simple regression analysis with the information contribution (i.e., $\frac{|\alpha_s|}{|\alpha_f|+|\alpha_s|}$) as the dependent variable, $\log(Open\ Interest)$, $\frac{Volume}{Open\ Interest}$, $\log(\frac{Price_{open,t}}{Price_{settle,t-1}})$, COVID, and $\mathrm{NigCancel}$ are the detected significant factors. But only $\frac{Volume}{Open\ Interest}$, COVID, and $\mathrm{NigCancel}$ remain significant in the following multiple regression analysis. But the estimated coefficient of $\frac{Volume}{Open\ Interest}$ changes its sign. This result suggests COVID and $\mathrm{NigCancel}$ are the two significant determinants for information contribution.

In summary, the multiple regression results show that the COVID-19 pandemic shock and other futures-market-related trading characteristics significantly affect the INE's futures price discovery performance. The impact of COVID-19 pandemic shock is also consistent with Figure 4, where the information contribution share of the futures prices increased for some markets.

[*Insert Table 5 about here*]

6.2 Quality, import, and delivery inventories

We continue to explore the other three types of determinants of INE's price discovery. The first is the quality of spot crudes, measured by the American Petroleum Institute (API) gravity. API gravity is an inverse measure of petroleum liquid density relative to water, and a higher value implies a better degree of the crudes. Second, China's oil demand was measured by its oil imports. We consider three variables for imports. Specifically, China's total oil imports, $\log(\mathrm{Import}_{\mathrm{Total,t}})$; China's oil imports from a

specific country, $\log(\text{Import}_{\text{Specific,t}})$[9]; and the portion of China's oil imports from a specific country, $\frac{\text{Import}_{\text{Specific,t}}}{\text{Import}_{\text{Total,t}}}$. Third, we also consider three other inventory variables: the total inventory in the warehouse, $\log(\text{Inventory}_{\text{Total,t}})$; the inventory of a specifically deliverable crude in the warehouse, $\log(\text{Inventory}_{\text{Specific,t}})$; and the portion of a specifically deliverable crude in the total inventory, $\frac{\text{Inventory}_{\text{Specific,t}}}{\text{Inventory}_{\text{Total,t}}}$.

Table 6 reports the results. For the logistic regression with $\text{I}_{\text{Trace}>5\%}$ as the dependent variable, API gravity, China's oil imports from a specific country, and the total inventory in the warehouse are among the most significantly positive determinants. But only the total inventory in the warehouse remain significant in the multiple regression. For the simple regression analysis with the information contribution (i.e., $\frac{|\alpha_s|}{|\alpha_f|+|\alpha_s|}$) as the dependent variable, we find API gravity and the total inventory in the warehouse are the two significant determinants. But in the following multiple regression, again, only the total inventory in the warehouse remain significant. Therefore, the domestic warehouse inventory is the only significant determinant among the factors under consideration here.

The significant influence of inventory confirms the importance of storage and delivery for price discovery on commodity markets in general (Garcia, Irwin and Smith (2015); Wright and Williams (1982); Yang, Li and Wang (2021)) and on crude oil futures in particular (Alquist and Kilian, 2010; Armstrong, Cardella, and Sabah, 2021). In particular, Alquist and Kilian (2010) find that WTI crude oil futures prices tend to be less accurate in forecasting spot prices in the future and thus there is the variability of

[9] We add 1 to the value of oil import from a specific country to address that the logarithm value might not be calculated if the import value is zero. We address the following inventory-related variables in the same way.

the futures price about the spot price, due to the marginal convenience yield of oil inventories. Armstrong, Cardella, and Sabah (2021) show that the rapid and significant response of the WTI crude oil futures prices to the weekly news of changes in the inventory, as part of the price discovery process.

[*Insert Table 6 about here*]

7. Conclusions

This study examines the price discovery performance of China's INE crude oil futures for 19 deliverable and non-deliverable Asian crude prices from 2018 to 2020. We find a cointegration relationship between the INE crude oil futures price and spot prices for all deliverable crudes (with the possible exception of Yemen) and some non-deliverable crudes. However, the unbiasedness hypothesis of price discovery does not hold for any of these spot price series. The results of the prediction hypothesis show that during the sample period, both the INE crude oil futures price and spot prices significantly contribute to the price discovery process, with the spot prices playing a somewhat more important role. The price discovery process was severely affected by the COVID-19 pandemic shock in early 2020, but the price discovery performance was even somewhat improved after China's recovery from the shock, demonstrating the resilience of China's crude oil futures market in the price discovery performance. Further investigation reveals that inventory of deliverable crudes and other factors significantly determine the price discovery performance of INE futures prices.

Overall, as the INE futures price performs a certain price discovery function for the deliverable and some non-deliverable crudes, the INE futures price has evolved into a significant informational source in Asia, while arguably it is on the path to emerge as an Asian benchmark. It might be worth noting that rarely is a futures contract dominant in price discovery immediately after its introduction. The relative contribution of the

INE crude oil futures market is possible to grow over time, as market participants get more used to using the futures market for the price discovery process of different types of crude oil. Future research may revisit the issue when INE crude oil futures contracts are further developed when INE introduces new requirements or new features to crude oil futures. Other aspects of the INE crude oil futures market, particularly those in the context of market microstructure, would be worthwhile to pursue and shed more light on its price discovery process.

References

Adämmer, Philipp, and Martin T Bohl, 2018, Price discovery dynamics in european agricultural markets, *Journal of Futures Markets* 38, 549-562.

Alquist, Ron, and Lutz Kilian, 2010, What do we learn from the price of crude oil futures? *Journal of Applied Econometrics* 25, 539-573.

Armstrong, Will J., Laura Cardella, and Nasim Sabah, 2021, Information shocks, disagreement, and drift, *Journal of Financial Economics* 140, 916-940.

Bae, Kee-Hong, Sadok El Ghoul, Zhaoran Gong, and Omrane Guedhami, 2021, CSR in times of crisis: Evidence from the COVID-19 pandemic, *Journal of Corporate Finance* 67, article 101876.

Black, Fischer, 1976, The pricing of commodity contracts, *Journal of Financial Economics* 3, 167-179.

Brenner, Robin J, and Kenneth F Kroner, 1995, Arbitrage, cointegration, and testing the unbiasedness hypothesis in financial markets, *Journal of Financial and Quantitative Analysis* 30, 23-42.

Cai, Charlie X, Paul B McGuinness, and Qi Zhang, 2011, The pricing dynamics of cross-listed securities: The case of chinese A-and H-shares, *Journal of Banking & Finance* 35, 2123-2136.

Caporale, Guglielmo Maria, Davide Ciferri and Alessandro Girardi, 2014, Time-varying spot and futures oil price dynamics, *Scottish Journal of Political Economy* 61, 78– 97.

Chen, Yu-Lun, Tsai, Wei-Che, 2017. Determinants of price discovery in the VIX futures market, *Journal of Empirical Finance* 43, pp. 59-73.

Chen, Yu-Lun and Ke Xu, 2021. The impact of RMB's SDR inclusion on price discovery in onshore-offshore markets, *Journal of Banking & Finance* 127, 106124.

Coppola, Andrea, 2008, Forecasting oil price movements: Exploiting the information in the futures market, *Journal of Futures Markets* 28, 34-56.

Eun, Cheol S, and Sanjiv Sabherwal, 2003, Cross-border listings and price discovery: Evidence from us-listed canadian stocks, *Journal of Finance* 58, 549-575.

Fan, John Hua, Adrian Fernandez-Perez, Ivan Indriawan, Neda Todorova, 2020, Internationalization of futures markets: Lessons from China, *Pacific-Basin Finance Journal* 63, 101429.

Fan, John Hua, and Tingxi Zhang, 2020, The untold story of commodity futures in china, *Journal of Futures Markets* 40, 671-706.

Fattouh, Bassam, 2011, *An anatomy of the crude oil pricing system* (Oxford institute for energy studies).

Figuerola-Ferretti, Isabel, and Jesús Gonzalo, 2010, Modelling and measuring price discovery in commodity markets, *Journal of Econometrics* 158, 95-107.

Frijns, Bart, and Remco C.J. Zwinkels, 2018, Time-varying arbitrage and dynamic price discovery, *Journal of Economic Dynamics and Control* 91, 485-502.

Garbade, Kenneth D., Silber, William L., 1983, Price movements and price discovery in futures and cash markets, *Review of Economics and Statistics* 65, 289-297.

Grammig, Joachim, Peter, Franziska J., 2013, Telltale tails: a new approach to estimating unique market information shares, *Journal of Financial and Quantitative Analysis* 48 (2), 459-488.

Garcia, Philip, Scott H Irwin, and Aaron Smith, 2015, Futures market failure?, *American Journal of Agricultural Economics* 97, 40-64.

Goldman Sachs, 2018, Finally, an exchange traded price for Chinese crude oil. March

28, 2018.
Gonzalo, Jesus, and Clive Granger, 1995, Estimation of common long-memory components in cointegrated systems, *Journal of Business & Economic Statistics* 13, 27-35.
Hansen, Henrik, and Søren Johansen, 1999, Some tests for parameter constancy in cointegrated var-models, *The Econometrics Journal* 2, 306-333.
Hasbrouck, Joel, 1995, One security, many markets: Determining the contribution to price discovery, *Journal of Finance*, 50, 1175-1199.
Hodgson, A., Masih, A., and Masih, R., 2003, Price discovery between informationally linked markets during different trading phases, *Journal of Financial Research*, 26, 77–96.
Kim, H. H., Swanson, N. R., 2014, Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence. *Journal of Econometrics* 178, 352-367.
INE, 2021, *Shanghai crude oil futures market development report*, (Shanghai International Energy Exchange).
Ito, Takatoshi, and Wen-Ling Lin, 2001, Race to the center: Competition for the nikkei 225 futures trade, *Journal of Empirical Finance* 8, 219-242.
Li, K., X. Liu, F. Mai and T. Zhang, forthcoming. The role of corporate culture in bad times: evidence from the COVID-19 pandemic, *Journal of Financial and Quantitative Analysis*.
Li, Hong and Yanlin Shi, 2021, A new unique information share measure with applications on cross-listed Chinese banks, *Journal of Banking & Finance* 128, 106141.
Lien, Donald, and Keshab Shrestha, 2009, A New Information Share Measure, *Journal of Futures Markets* 29, 377–395.
Lien, Donlad and Zijun Wang, 2016, Estimation of market information shares: acomparison, *Journal of Futures Markets* 36, 1108–1124 .
Jin, Muzhao, Youwei Li, Jianxin Wang, and Yung Chiang Yang, 2018, Price discovery in the Chinese gold market, *Journal of Futures markets* 38, 1262-1281.
Johansen, Soren, 1991, Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models, *Econometrica* 59, 1551-80.
Ng, V. K., Pirrong, S.C., 1996, Price dynamics in refined petroleum spot and futures markets. *Journal of Empirical Finance* 2, 359–388.
Pesaran, H. Hashem, and Yongcheol Shin, 1998, Generalized impulse response analysis in linear multivariate models, *Economics Letters* 58, 17-29.
Putniņš, Tālis J. 2013. What do price discovery metrics really measure? *Journal of Empirical Finance* 23, 68-83.
Reuters, 2018a, China oil futures launch may threaten primacy of U.S. dollar: UBS.
Reuters, 2018b, China's flawed futures contract pushes oil trade to record high in 2018.
Schwarz, T. V., and Szakmary, A. C., 1994, Price discovery in petroleum markets: Arbitrage, cointegration, and the time interval of analysis, *Journal of Futures Markets* 14, 147–167.
Silve´rio, R., and A. Szklo, 2012, The Effect of the Financial Sector on the Evolution of Oil Prices:Analysis of the Contribution of the Futures Market to the Price Discovery Process in the WTI Spot Market, *Energy Economics* 34, 1799–1808.
Shrestha, Keshab, 2014, Price discovery in energy markets, *Energy Economics* 45, 229-233.
Switzer, Lorne N, and Mario El-Khoury, 2007, Extreme volatility, speculative efficiency, and the hedging effectiveness of the oil futures markets, *Journal of*

Futures Markets 27, 61-84.
Tong, Jiadong, Zijun Wang, and Jian Yang, 2016, Information flow between forward and spot markets: Evidence from the chinese renminbi, *Journal of Futures Markets* 36, 695-718.
Umar, Zaghum, Youssef Manel, Yasir Riaz, Mariya Gubareva, 2021. Return and volatility transmission between emerging markets and US debt throughout the pandemic crisis, *Pacific-Basin Finance Journal* 67, 101563.
Working, Holbrook, 1949, The theory of price of storage, *The American Economic Review* 39, 1254-1262.
Wright, Brian D, and Jeffrey C Williams, 1982, The economic role of commodity storage, *The Economic Journal* 92, 596-614.
Yang, Jian, David A. Bessler, and David J. Leatham, 2001, Asset storability and price discovery in commodity futures markets: A new look, *Journal of Futures Markets* 21, 279-300.
Yang, Chen, Fei Lv, Libing Fang, and Xingxing Shang, 2020, The pricing efficiency of crude oil futures in the shanghai international exchange, *Finance Research Letters* 36, 101329.
Yang, Jian, Cheng Hsiao, Qi Li, and Zijun Wang, 2006, The emerging market crisis and stock market linkages: Further evidence, *Journal of Applied Econometrics* 21, 727-744.
Yang, Jian, Zheng Li, and Tao Wang, 2021, Price discovery in chinese agricultural futures markets: A comprehensive look, *Journal of Futures Markets* 41, 536-555.
Yang, Jian, Zihui Yang, and Yinggang Zhou, 2012, Intraday price discovery and volatility transmission in stock index and stock index futures markets: Evidence from china, *Journal of Futures Markets* 32, 99-121.
Yang, Jian, and Yinggang Zhou, 2020, Return and volatility transmission between china's and international crude oil futures markets: A first look, *Journal of Futures Markets* 40, 860-884.

Table 1: Variable description, summary statistics, and unit-root test

Series	Definition, log-transformed of	Obs	Mean	Standard Error	Minimum	Maximum	Skewness	Kurtosis (excess)	API	level		Difference	
										ADF	PP	ADF	PP
$FuturePrice$	Most active future prices	656	4.055	0.255	3.374	4.457	-0.734	-0.822	-	-1.273	-1.389	-8.468***	-23.677***
$Dubai$	UAE Dubai crude oil spot price	656	4.018	0.296	2.699	4.42	-1.380	1.854	31	-1.872	-1.726	-6.087***	-25.170***
$UpperZakun$	UAE Upper Zakum crude oil spot price	656	4.018	0.329	2.47	4.412	-1.631	2.700	32.9	-2.321	-1.839	-4.181***	-24.954***
$Oman$	Oman crude oil spot price	656	4.029	0.29	2.837	4.429	-1.378	1.807	34	-1.978	-1.741	-5.776***	-27.563***
$Qatar$	Qatar Marine crude oil spot price	656	4.018	0.321	2.606	4.409	-1.588	2.488	35.8	-2.283	-1.767	-4.581***	-26.428***
$Yemen$	Yemen Masila crude oil spot price	509	4.131	0.29	2.845	4.458	-2.394	5.410	31.4	-3.262**	-1.164	-3.291**	-21.431***
$Iraq$	Iraq Basrah Light crude oil spot price	656	4.026	0.32	2.583	4.426	-1.665	2.839	30.5	-2.328	-1.767	-4.524***	-26.724***
$Shengli$	China Shengli crude oil spot price	656	4.003	0.24	2.979	4.347	-1.302	1.760	24.2	-1.584	-1.508	-9.699***	-24.934***
$Murban$	UAE Murban crude oil spot price	656	4.036	0.337	2.495	4.437	-1.556	2.358	40.2	-2.247	-2.055	-4.186***	-24.973***

Notes: This table reports the basic information about the variables (series). All the series are denominated in U.S. dollars and log-transformed. Obs. denotes observations. *** and ** imply rejecting the null hypothesis of with a unit root at 1% and 5% significant level, respectively.

Table 2: Cointegration Tests

Panel A: Between INE futures price and spot price of **Dubai** crude oil

Without linear trend			With linear trend	
λ_{max}	λ_{trace}	H_0	λ_{max}	λ_{trace}
0.043***	36.320***	$r = 0$	0.049***	47.505***
0.011	7.745	$r = 1$	0.023	14.863

Panel B: Between INE futures price and spot price of **Upper Zakum** crude oil

Without linear trend			With linear trend	
λ_{max}	λ_{trace}	H_0	λ_{max}	λ_{trace}
0.038**	30.919**	$r = 0$	0.045***	44.888***
0.008	5.568	$r = 1$	0.022	14.614

Panel C: Between INE futures price and spot price of **Oman** crude oil

Without linear trend			With linear trend	
λ_{max}	λ_{trace}	H_0	λ_{max}	λ_{trace}
0.041***	34.407***	$r = 0$	0.046***	45.012***
0.011	7.316	$r = 1$	0.021	13.954

Panel D: Between INE futures price and spot price of **Qatar** crude oil

Without linear trend			With linear trend	
λ_{max}	λ_{trace}	H_0	λ_{max}	λ_{trace}
0.038**	30.554**	$r = 0$	0.045***	44.921***
0.008	5.132	$r = 1$	0.022	14.635

Panel E: Between INE futures price and spot price of **Yemen** crude oil

Without linear trend			With linear trend	
λ_{max}	λ_{trace}	H_0	λ_{max}	λ_{trace}
0.040**	33.264**	$r = 0$	0.040**	35..374**
0.025	12.552	$r = 1$	0.029	14.905

Panel F: Between INE futures price and spot price of **Iraq** crude oil

Without linear trend			With linear trend	
λ_{max}	λ_{trace}	H_0	λ_{max}	λ_{trace}
0.037**	28.961**	$r = 0$	0.044***	43.480***
0.007	4.380	$r = 1$	0.021	13.922

Panel G: Between INE futures price and spot price of **Shengli** crude oil

Without linear trend			With linear trend	
λ_{max}	λ_{trace}	H_0	λ_{max}	λ_{trace}
0.033**	28.687**	$r = 0$	0.049***	49.936***
0.010	6.520	$r = 1$	0.025	16.829

Panel H: Between INE futures price and spot price of **Murban** crude oil

Without linear trend			With linear trend	
λ_{max}	λ_{trace}	H_0	λ_{max}	λ_{trace}
0.039**	31.345**	$r = 0$	0.045***	45.823***
0.009	5.605	$r = 1$	0.024	15.585

Notes: λ_{max} and λ_{trace} are the statistics of maximum eigenevalue test and trace test, respectively. r is the number of cointegration vectores. The impact of COVID-19 pandemic shock is controlled in all the tests. **, and *** denote rejecting the null hypothesis at the 5%, and 1% significance level, respectively.

Table 3: Restriction tests for the cointegration relationships

Panel A: Restriction tests on β-vectors

	Without linear trend		With linear trend		
	COVID-19	$\beta' = (1, -1)$	COVID-19	Trend	$\beta' = (1, -1)$
VECM1: Dubai	14.400***	14.864***	0.002	2.139	12.692***
VECM2: Upper Zakum	13.530***	14.330***	0.203	2.761	9.864***
VECM3: Oman	13.162***	13.435***	0.128	2.232	11.875***
VECM4: Qatar	13.892***	14.682***	0.195	2.732	9.769***
VECM5: Yemen	3.901**	5.198**	0.388	1.833	5.451**
VECM6: Iraq	13.693***	14.357***	0.327	2.703	9.385***
VECM7: Shengli	10.857***	10.840***	0.056	7.778**	12.074***
VECM8: Murban	13.477***	14.686***	0.168	2.549	9.140***

Panel B: Restriction tests on α-vectors

	Without linear trend, Exclude		With linear trend, Exclude	
	Futures Prices	Spot Prices	Futures Prices	Spot Prices
VECM1: Dubai	28.279***	17.876***	32.274***	23.378***
VECM2: Upper Zakum	25.115***	13.063***	30.078***	20.600***
VECM3: Oman	26.732***	15.119***	30.489***	21.534***
VECM4: Qatar	25.202***	12.480***	30.115***	20.331***
VECM5: Yemen	20.657***	15.272***	20.019***	15.782***
VECM6: Iraq	24.476***	11.038***	29.411***	19.468***
VECM7: Shengli	19.715***	11.313***	26.989***	23.350***
VECM8: Murban	25.611***	12.578***	30.122***	20.802***

Notes: Panel A reports the results of unbiasedness hypothesis with the restriction that $\beta' = (1 \quad -1)$ along with other testing results of restrictions on the β vector. Panel B reports the prediction hypothesis's testing results that impose restrictions on the α vector. *, ***, and *** denote rejecting the null hypothesis at 5% and 1% significant level, respectively.

Table 4: Generalized forecast error variance decomposition, percentage (%)

	VECM1		VECM2		VECM3		VECM4		VECM5		VECM6		VECM7		VECM8	
	Forecast error variance decomposition of **Futures** prices															
	Dubai		**Upper Zakum**		**Oman**		**Qatar**		**Yemen**		**Iraq**		**Shengli**		**Murban**	
Step	Futures	Spot	Futures	Spot	Futures	Spot	Futures	Spot	Futures	Spot	Futures	Spot	Futures	Spot	Futures	Spot
1	74.32	25.68	77.65	22.35	78.03	21.97	78.13	21.87	69.74	30.26	78.23	21.77	69.02	30.98	77.85	22.15
2	67.24	32.76	71.11	28.89	70.13	29.88	71.33	28.67	64.75	35.25	71.96	28.04	67.61	32.39	71.42	28.59
3	65.73	34.27	69.81	30.19	68.55	31.45	70.09	29.91	63.54	36.46	71.05	28.95	67.13	32.87	70.13	29.87
5	63.29	36.71	68.42	31.58	65.95	34.05	68.73	31.27	61.97	38.04	69.87	30.13	66.15	33.86	68.76	31.25
10	57.503	42.50	64.61	35.39	60.03	39.97	65.01	34.99	58.56	41.44	66.54	33.46	63.76	36.24	64.92	35.09
15	52.736	47.26	61.30	38.71	55.23	44.77	61.83	38.18	55.88	44.12	63.62	36.38	61.60	38.40	61.59	38.41
20	48.819	51.18	58.38	41.62	51.33	48.67	59.04	40.96	53.70	46.31	61.04	38.96	59.68	40.32	58.67	41.34
30	42.933	57.07	53.59	46.41	45.55	54.45	54.47	45.53	50.45	49.56	56.73	43.27	56.43	43.57	53.87	46.13
60	33.688	66.31	44.69	55.31	36.73	63.27	46.05	53.96	45.48	54.52	48.53	51.47	50.15	49.85	45.05	54.95
90	29.556	70.44	40.04	59.96	32.90	67.10	41.66	58.34	43.39	56.61	44.11	55.89	46.80	53.20	40.48	59.52
180	24.769	75.23	34.129	65.87	28.54	71.46	36.075	63.93	41.133	58.87	38.34	61.66	42.625	57.38	34.678	65.32
360	22.085	77.92	30.558	69.44	26.13	73.87	32.697	67.30	39.948	60.05	34.78	65.22	40.240	59.76	31.186	68.81
	Forecast error variance decomposition of **Spot** prices															
	Dubai		**Upper Zakum**		**Oman**		**Qatar**		**Yemen**		**Iraq**		**Shengli**		**Murban**	
Step	Futures	Spot	Futures	Spot	Futures	Spot	Futures	Spot	Futures	Spot	Futures	Spot	Futures	Spot	Futures	Spot
1	25.68	74.32	22.35	77.65	21.97	78.03	21.87	78.13	30.26	69.74	21.77	78.23	30.98	69.02	22.15	77.85
2	29.11	70.89	25.34	74.66	27.71	72.29	25.83	74.18	33.58	66.42	26.43	73.57	33.77	66.23	25.10	74.90
3	31.75	68.25	28.14	71.86	31.29	68.71	28.91	71.09	37.28	62.72	29.93	0.07	36.80	63.20	27.85	72.15
5	32.25	67.75	29.58	70.42	32.24	67.77	30.35	69.65	39.06	60.94	31.52	68.48	38.59	61.41	29.34	70.66
10	30.873	69.13	30.24	69.76	31.83	68.18	31.37	68.63	40.14	59.86	32.80	67.20	39.60	60.40	30.12	69.88
15	29.120	70.88	30.00	70.00	30.62	69.38	31.32	68.68	40.20	59.81	32.85	67.15	39.69	60.31	29.99	70.01
20	27.655	72.35	29.65	70.36	29.56	70.44	31.10	68.90	40.08	59.92	32.70	67.30	39.60	60.40	29.72	70.28
30	25.557	74.44	29.00	71.00	28.03	71.97	30.65	69.35	39.81	60.19	32.32	67.68	39.32	60.68	29.21	70.80
60	22.664	77.34	27.91	72.09	25.95	74.05	29.86	70.14	39.33	60.67	31.60	68.40	38.69	61.32	28.33	71.67
90	21.550	78.45	27.44	72.56	25.17	74.83	29.52	70.48	39.13	60.87	31.29	68.71	38.36	61.64	27.96	72.04
180	20.386	79.61	26.958	73.04	24.37	75.63	29.172	70.83	38.929	61.07	30.95	69.05	38.003	62.00	27.579	72.42
360	19.787	80.21	26.710	73.29	23.96	76.04	28.994	71.01	38.827	61.17	30.79	69.22	37.818	62.18	27.383	72.62

Notes: This table reports the variance decomposition using the generalized VAR approach (Pesaran and Shin (1998)).

Table 5: Market related factors affecting the price discovery performance for deliverable crudes

	$I_{Trace>5\%}$								Contribution of futures price to price discovery ($\lvert\alpha_{c,t}\rvert/(\lvert\alpha_{c,t}\rvert+\lvert\alpha_{f,t}\rvert)$)							
$\log(Volume)$	**-0.116*** **							**-0.065***	0.000							
	(-7.85)							(-1.71)	(0.85)							
$\log(Open\ Interest)$		0.006								**0.005*** **						0.020
		(0.17)								(3.58)						(1.12)
$\frac{Volume}{Open\ Interest}$			**-0.132*** **					**-0.070*** **			**-0.001*** **					**0.001*** **
			(-15.29)					(-4.69)			(-2.75)					(4.19)
$(\log(\frac{Price_t}{Price_{t-1}}))^2$				**-88.734*** **				**-93.121*** **				3.934				
				(-3.90)				(-3.48)				(1.61)				
$\log(\frac{Price_{open,t}}{Price_{settle,t-1}})$					**0.099*** **			0.015					**0.006*** **			-0.026
					(3.22)			(0.19)					(5.46)			(-1.54)
COVID						**0.981*** **		**1.350*** **						**0.019*** **		**0.045*** **
						(14.23)		(10.62)						(11.64)		(5.29)
$NigCancel$							**-0.538*** **	**-1.559*** **							**-0.010*** **	**-0.031*** **
							(-5.04)	(-12.10)							(-3.48)	(-6.04)
constant	1.046***	-0.276	0.321***	-0.159***	-0.630***	-0.575***	-0.155***	0.456**	0.475***	0.428***	0.482***	0.478***	0.454***	0.471***	0.481***	0.363***
	(6.32)	(-0.77)	(6.77)	(-4.50)	(-4.72)	(-13.71)	(-4.51)	(2.30)	(84.53)	(28.84)	(364.97)	(350.08)	(89.62)	(356.91)	(525.52)	(3.15)
Observations	3857	3857	3857	3857	3857	3857	3857	3857	1723	1723	1723	1723	1723	1723	1723	1723
Pseudo/adj R^2	0.013	0.000	0.046	0.003	0.002	0.039	0.005	0.100	0.000	0.012	0.003	0.015	0.026	0.071	0.005	0.113

Notes: *t* statistics are in parentheses. *, ***, and *** denote the significant level of 10%, 5%, and 1%, respectively.

Table 6: Impact of quality, demand (import), and warehouse inventory on the price discovery performance

	$I_{Trace>5\%}$								Contribution of futures price to price discovery ($\lvert\alpha_{c,t}\rvert/(\lvert\alpha_{c,t}\rvert+\lvert\alpha_{f,t}\rvert)$)							
API	**0.161*** **							0.124	**0.004** **							-0.001
	(3.24)							(1.16)	(2.20)							(-0.21)
$\log(\text{Import}_{\text{Total,t}})$		0.648								-0.030						
		(1.00)								(-1.32)						
$\log(\text{Import}_{\text{Specific,t}})$			**0.076***					0.016			-0.005					
			(1.65)					(0.24)			(-1.47)					
$\frac{\text{Import}_{\text{Specific,t}}}{\text{Import}_{\text{Total,t}}}$				1.316								-0.270				
				(0.33)								(-1.24)				
$\log(\text{Inventory}_{\text{Total,t}})$					**0.318** **			**0.319** **					**0.008** **			**0.008** **
					(2.45)			(2.41)					(2.27)			(2.07)
$\log(\text{Inventory}_{\text{Specific,t}})$						0.038								0.001		
						(1.50)								(1.25)		
$\frac{\text{Inventory}_{\text{Specific,t}}}{\text{Inventory}_{\text{Total,t}}}$							0.000								-0.000	
							(0.36)								(-0.34)	
Constant	-5.359***	-6.451	-1.032*	-0.174	-4.702**	-0.156	0.120	-8.983**	0.371***	0.786***	0.553***	0.505***	0.373***	0.483***	0.492***	0.404**
	(-3.37)	(-1.02)	(-1.77)	(-0.75)	(-2.39)	(-0.59)	(0.63)	(-1.97)	(6.50)	(3.63)	(13.07)	(56.95)	(6.44)	(101.27)	(61.71)	(2.10)
Observations	203	174	174	174	145	145	145	145	88	82	82	82	78	78	78	78
Pseudo/adj R^2	0.046	0.004	0.013	0.000	0.031	0.011	0.001	0.039	0.018	0.001	0.051	0.012	0.014	-0.003	-0.012	0.002

Notes: *t* statistics are in parentheses. *, ***, and *** denote the significant level of 10%, 5%, and 1%, respectively.

Figure 1: INE futures price and the spot prices of deliverable crudes, 03/26/2018-12/31/2020

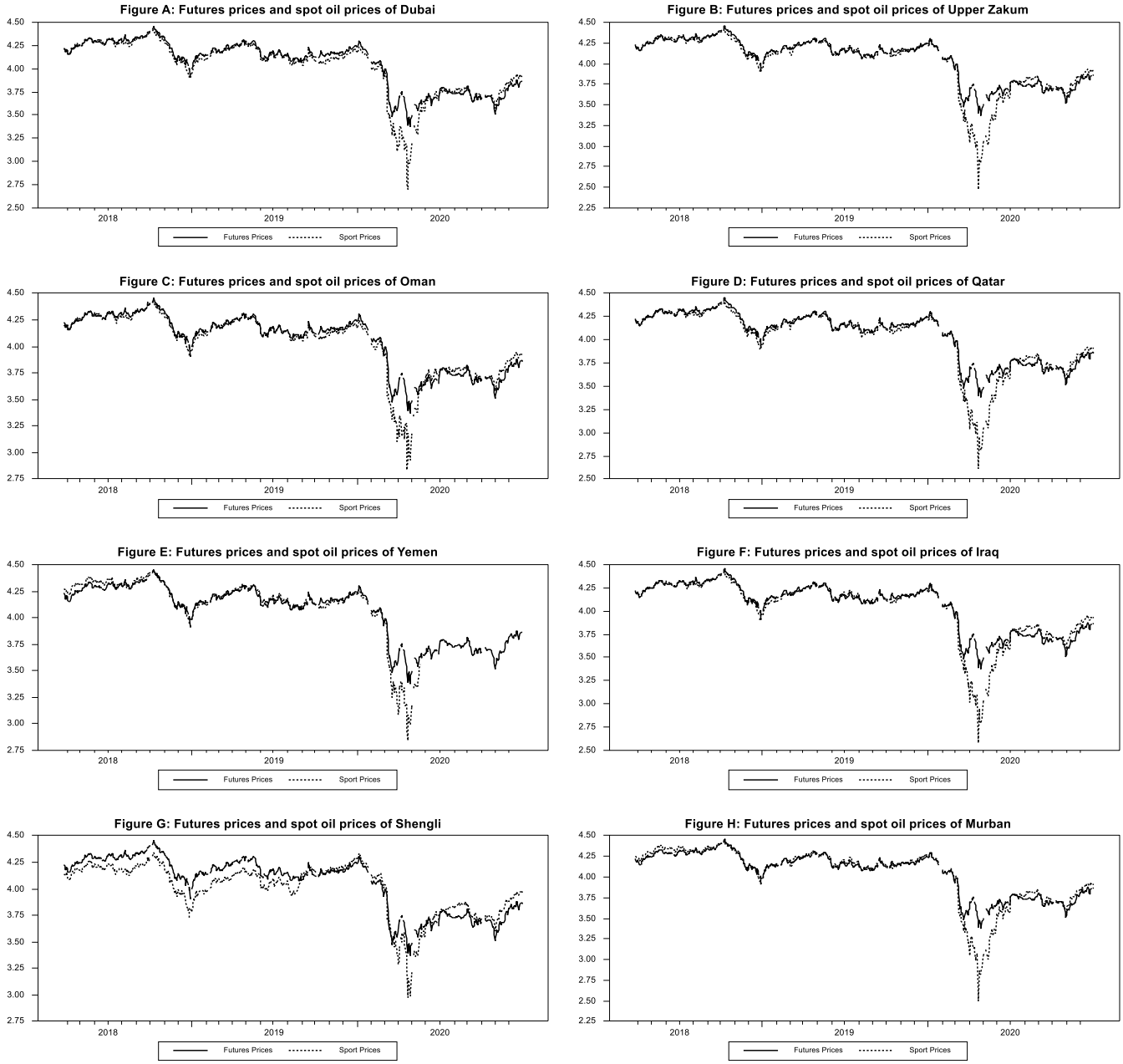
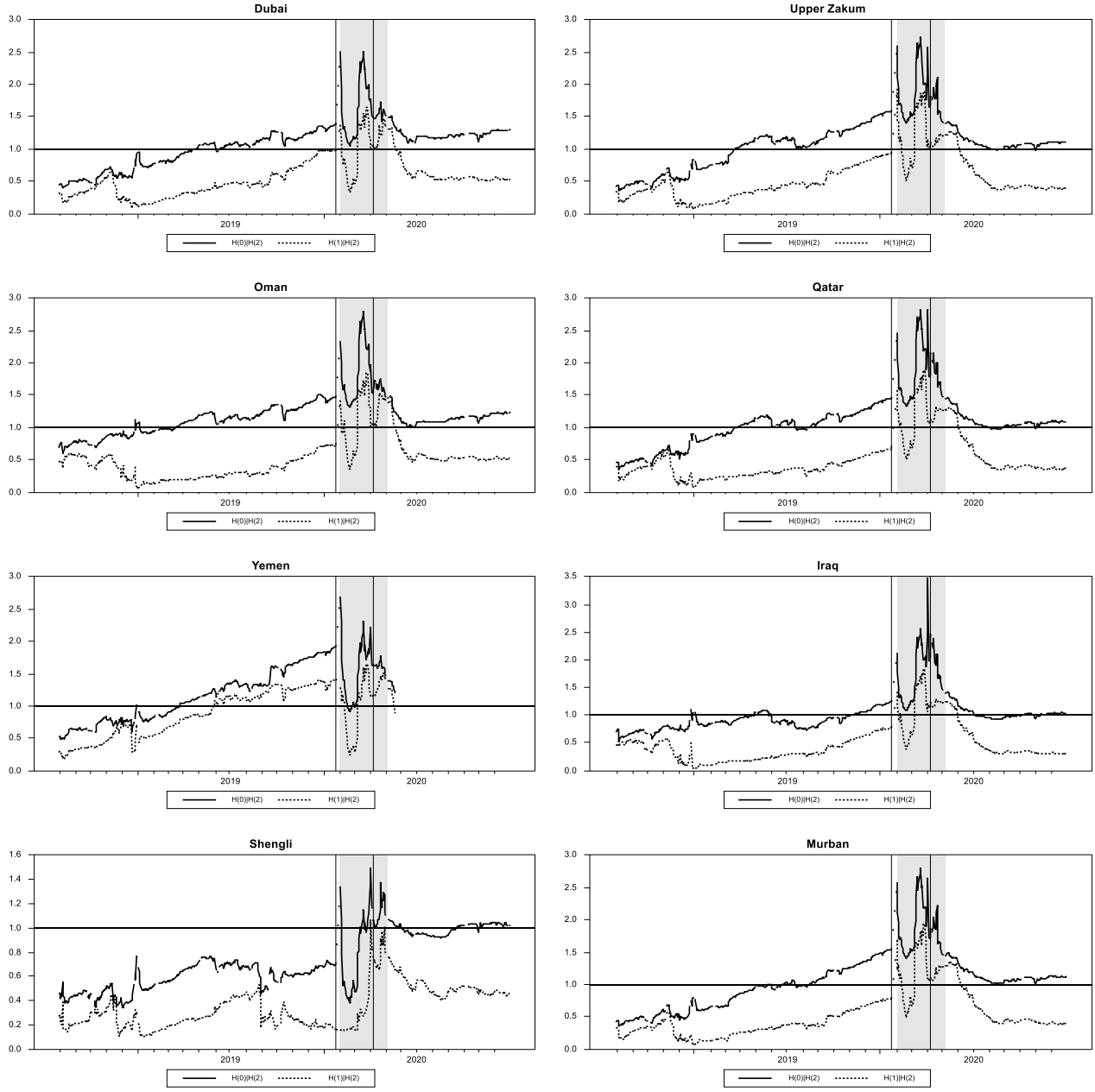
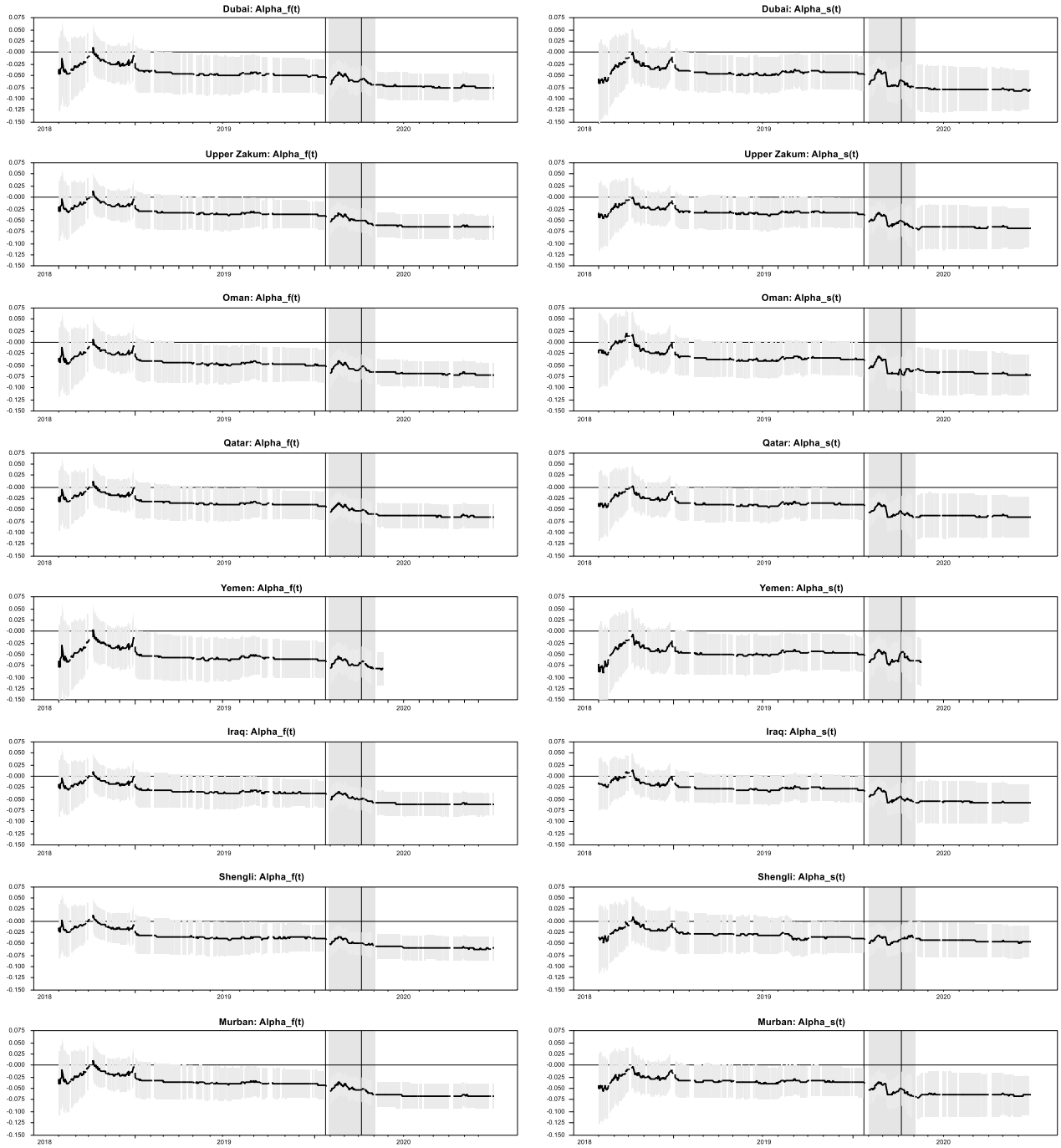


Figure 2: Recursive cointegration trace tests



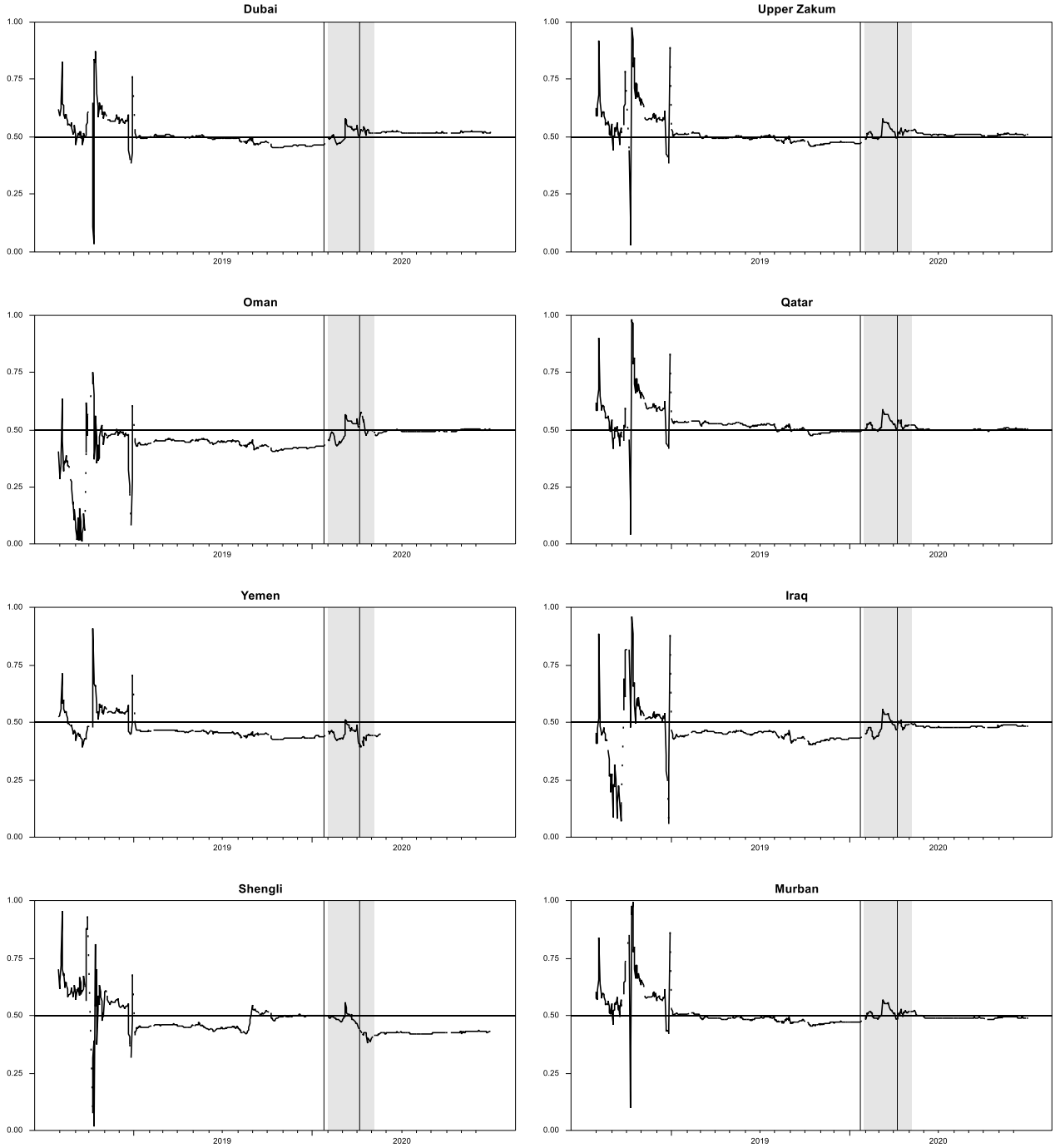
Notes: All the statistics are normalized by the critical value at 5% significant level. The two vertical lines highlight the lockdown (January 23, 2020) and re-open (April 8, 2020) day of the Wuhai city, China, which was the center of COVID-19 pandemic shock in 2020. The grey area highlights the period of China's government suspended the nighttime trading of INE futures from February 3 to May 6, 2020.

Figure 3: Recursive estimates of the adjustment coefficients and 95% confidential intervals



Notes: The two vertical lines highlight the lockdown (January 23, 2020) and re-open (April 8, 2020) day of the Wuhai city, China, which was the center of COVID-19 pandemic shock in 2020. The grey area highlights the period of China's government suspended the nighttime trading of INE futures from February 3 to May 6, 2020. "f(t)" and "s(t)" represent the futures and spot prices, respectively.

Figure 4: Recursive information contribution share of futures prices



Notes: The two vertical lines highlight the lockdown (January 23, 2020) and re-open (April 8, 2020) day of the Wuhai city, China, which was the center of COVID-19 pandemic shock in 2020. The grey area highlights the period of China's government suspended the nighttime trading of INE futures from February 3 to May 6, 2020. The information contribution share is based on the component share method of Eun and Sabherwal (2003) and Gonzalo and Granger (1995).

Appendix:

Table A-1: Description of the 11 Asian crude spot oil (logarithm) prices

crudes	Obs.	Mean	Standard Error	Min	Max	API	level ADF	level PP	difference ADF	difference PP
Medium to Asia, Saudi Arab	657	58.26	15.16	12.39	83.58	30.2	-2.35	-1.76	-4.52***	-26.74***
ESPO, Russia	657	61.21	15.55	10.37	89.33	34.8	-1.96	-1.98	-6.41***	-21.68***
Kuwait	657	57.91	15.16	11.79	83.18	30.2	-2.37	-1.78	-4.51***	-26.73***
Light, Iran	595	57.61	14.15	16.64	81.53	33.1	-1.61	-1.41	-5.12***	-20.12***
Forozan, Iran	657	57.55	14.89	12.44	82.88	29.7	-2.36	-1.79	-4.51***	-26.77***
Tapis, Malaysia	657	61.54	15.65	20.43	86.89	45.2	-2.44	-1.66	-19.13***	-25.00***
Labuan, Malaysia	657	62.81	15.48	21.83	86.32	32	-1.39	-1.39	-25.16***	-25.20***
Minas, Indonesia	657	57.78	15.48	16.64	86.52	35.3	-2.09	-2.03	-24.10***	-24.14***
Duri, Indonesia	657	59.44	12.27	27.71	81.44	20.8	-2.18	-2.22	-11.30***	-27.46***
Cinta, Indonesia	657	56.88	16.34	14.93	89.09	31.1	-2.07	-2.01	-15.52***	-24.10***
Daqing, China	657	56.29	14.67	15.79	81.35	32.2	-2.17	-2.11	-14.04***	-24.09***

Notes: This table reports the basic information about 11 spot Asian crude oil prices. All the series are denominated in U.S. dollars and log-transformed. Obs. denotes observations. *** and ** imply rejecting the null hypothesis of with a unit root at 1% and 5% significant level, respectively.

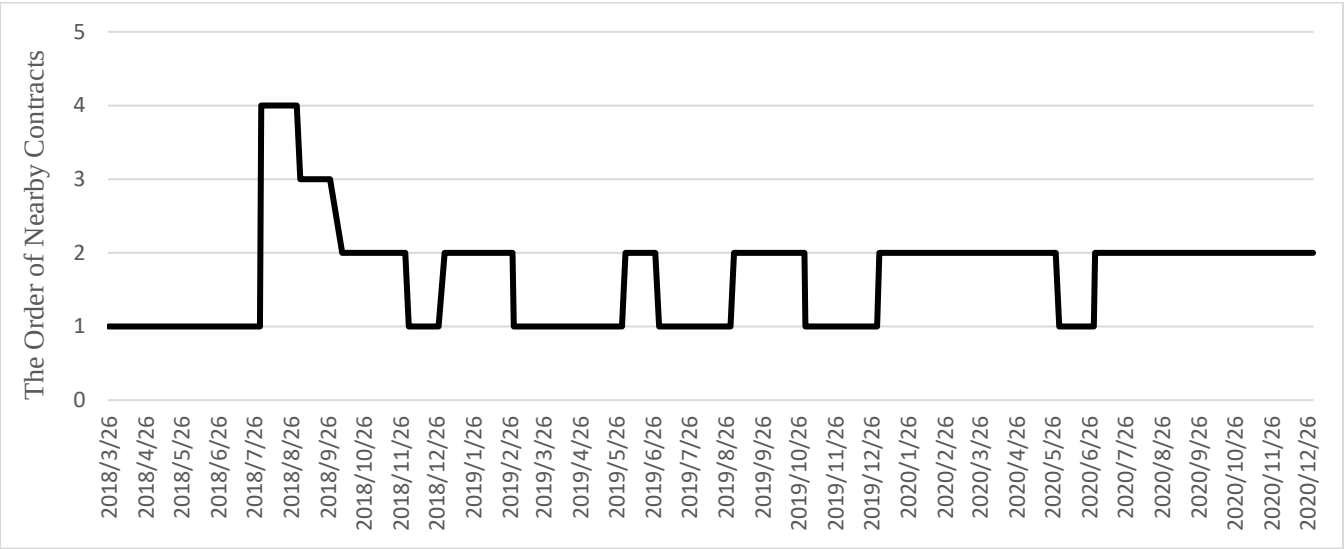
Appendix:

Table A-2: Cointegration trace test, 11 other Asian crudes

λ_{trace} Without linear trend	λ_{trace} With linear trend	$H_0 = r$	λ_{trace} Without linear trend	λ_{trace} With linear trend
Medium to Asia, Saudi Arab			**ESPO, Russia**	
29.001**	44.054***	**0**	35.399***	48.069***
4.585	14.64	**1**	11.025	18.413
Kuwait			**Light, Iran**	
28.894**	43.507***	**0**	39.468***	60.208***
4.534	14.129	**1**	14.438**	22.663**
Forozan, Iran			**Tapis, Malaysia**	
29.481**	43.034***	**0**	35.724***	53.663***
4.452	13.364	**1**	12.811*	24.611***
Labuan, Malaysia			**Minas, Indonesia**	
33.34***	50.643***	**0**	47.453***	73.903***
11.507*	23.624***	**1**	15.987**	25.521***
Duri, Indonesia			**Cinta, Indonesia**	
42.26***	62.802***	**0**	46.951***	72.153***
15.868**	21.51**	**1**	15.627**	25.542***
Daqing, China				
45.439***	71.135***	**0**		
17.187**	25.575***	**1**		

Notes: λ_{trace} is the statistics of trace test. r is the number of cointegration vectors. The impact of COVID-19 pandemic shock is controlled in all the tests. **, and *** denote rejecting the null hypothesis at 5%, and 1% significant level, respectively.

Figure A-1: INE's most active contracts in each month, 2018/3/26~2020/12/31



Appendix:

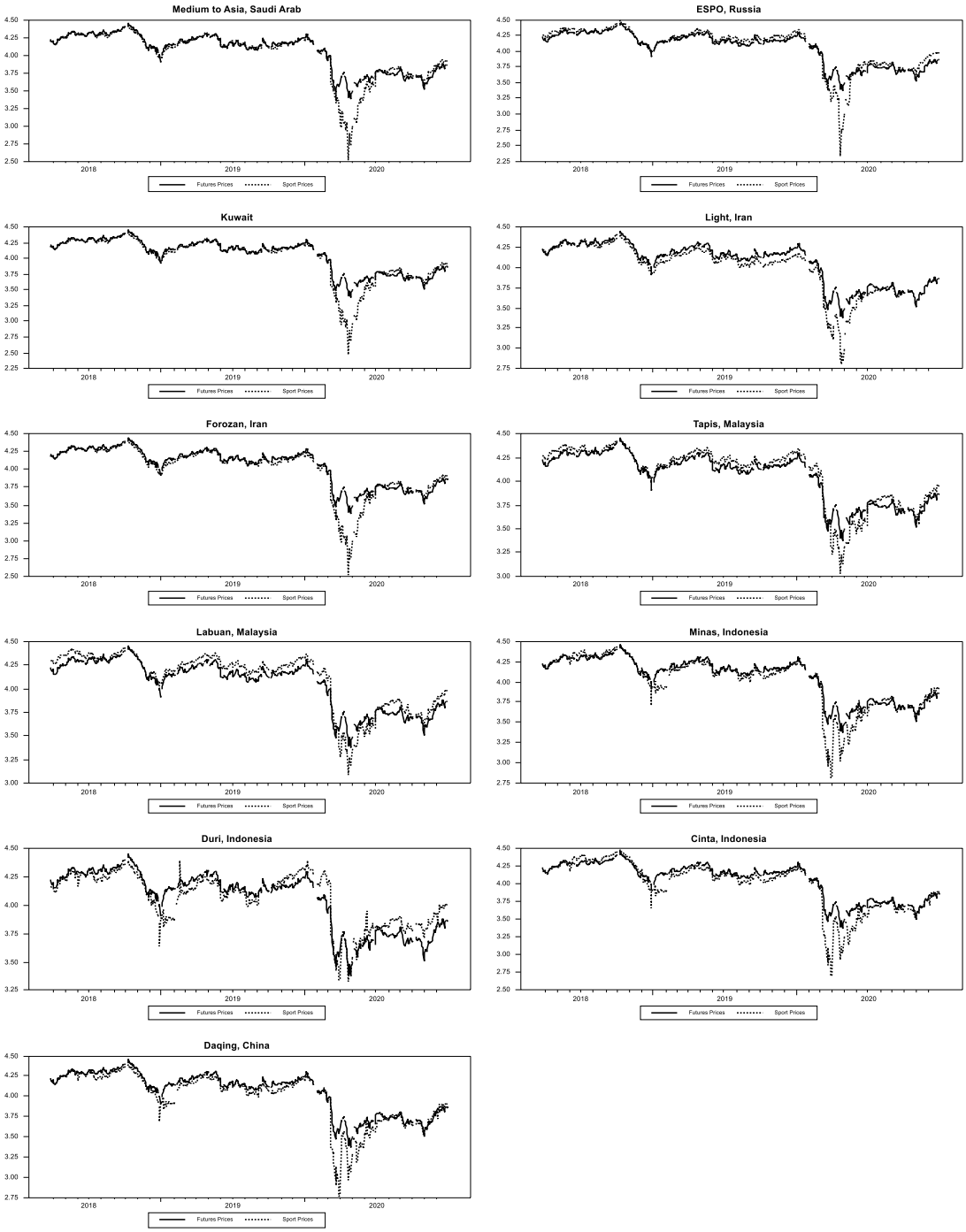



Figure A-2: INE futures price and 11 other Asian crude oil spot prices