Price Discovery in Crude Oil Futures

John Elder*[,1], Hong Miao[1], Sanjay Ramchander[1]

Abstract

This study examines price discovery among the two most prominent price benchmarks in the market for crude oil, WTI sweet crude and Brent sweet crude. Using data on the most active futures contracts measured at the one-second frequency, we find that WTI maintains a dominant role in price discovery relative to Brent, with an estimated information share in excess of 80%, over a sample from 2007 through 2012. Our analysis is robust to different decompositions of the sample, over pit-trading sessions and non-pit trading sessions, segmentation of days associated with major economic news releases, and data measured to the millisecond. We find no evidence that the dominant role of WTI in price discovery is diminished by the price spread between Brent that emerged in 2008.

Key words:

Crude Oil Futures, WTI, Brent, Information Sharing, Inventory Level, Spread

JEL: G15, O13, Q43

*Corresponding Author: Tel.: +001 970 491 2952

Email addresses: `john.elder@colostate.edu` (John Elder), `hong.miao@colostate.edu` (Hong Miao), `sanjay.ramchander@colostate.edu` (Sanjay Ramchander)

[1]Department of Finance and Real Estate, Colorado State University, Fort Collins, Colorado, 80523-1272, United States

1. Introduction

Several studies have examined the time series properties and statistical relationships among various crude oil prices. For instance, Bachmeier and Griffin (2006) examine daily prices for five different crude oils - WTI, Brent, Alaska North Slope, Dubai Fateh, and the Indonesian Arun - and conclude that the world oil markets are tightly linked with each other. Similarly, Hammoudeh, Ewing and Thompson (2008) find evidence of cointegration in four oil benchmark prices (WTI, Brent, Dubai and Maya, see also Kleit, 2001; Bentzen, 2007). An obvious implication of this result is that supply and demand shocks that affect prices in one region quickly spillover to other regional markets.

The fact that crude oil markets are geographically fragmented, and yet remain susceptible to common global risk factors, poses somewhat of a challenge to market participants in determining precisely how price discovery is established. Price leadership of a benchmark is important to establish given its implications for reference pricing in the trade of physical and financial contracts. Furthermore, from a market microstructure perspective, the benchmark's contribution to price discovery provides insights into its ability to process information and attract informed traders in markets where they are traded.

There has been a great deal of interest in examining the dynamics between WTI and Brent prices. It has been argued that in economic terms the spread between WTI and Brent prices should reflect a quality differential, and is driven by underlying factors that are specific to each market. In equilibrium, the price of WTI should equal the price of Brent after adjusting for carrying cost and the quality discount (Alizadeh and Nomikos, 2004). Any mispricing in the relationship is likely to attract arbitrage opportunities in spot and derivative markets, thus forcing convergence. Historically Brent has traded at a slight discount to WTI,[2] although the relationship reversed in recent years with Brent trading

[2] Both Brent and WTI are classified as a 'light sweet' oil blend which means that they are easy to refine compared to heavier and sour oil blends. However, since Brent is relatively denser and has a higher sulfur content than WTI, based purely on its physical properties Brent is expected to trade at a discount to WTI.

at a substantial premium to WTI. The inversion in the price spread has been attributed to localized factors such as the dramatic increase in U.S. oil production combined with capacity constraints in the transportation and storage infrastructure of domestic crude oil (cf. Baumeister and Kilian, 2013). As a result of these changes, some studies cast doubt on the continued viability of WTI as an international benchmark (Bentzen, 2007), and argue that the ongoing decoupling of WTI from other U.S. and international crude grades is evidence that WTI is a 'broken benchmark' (Fattouh, 2007, 2011). Borenstein and Kellogg (2014) also question the leading role of WTI by showing that the relative price decrease of WTI does not pass through to wholesale gasoline and diesel prices.

It is important to note that the discussions surrounding the relative merits of WTI and Brent as price benchmarks are closely intertwined with the price discovery function in crude oil futures markets. This paper examines the price discovery relationship between two of the most widely referenced international oil price benchmarks – West Texas Intermediate (WTI) and Brent. Specifically, we apply the Hasbrouck (1995) information share (IS) model to estimate the degree of price discovery. This model is based on the econometrics of cointegrated vector autoregressions, assuming that cointegrated price series fluctuate around a common, unobserved "efficient" price. Hasbrouck defines the information share as the proportion of the variance in the common price process that is attributable to a particular price series. Additional details on the model and its applicability are provided in section 3. Our sample is January 2, 2007 to April 27, 2012, a period during which there has been a remarkable surge in U.S. oil production. Since both WTI and Brent have highly liquid futures markets, we use futures prices sampled at the one-second interval.

In a related study, Kao and Wan (2012) also apply the Hasbrouck IS model to daily prices of WTI and Brent futures over the 1991-2009 sample. These authors find that price discovery in WTI has been impaired due to production, transportation and inventory bottlenecks in the U.S., and conclude that since 2004 Brent has led the price discovery process. We extend their analysis in two important dimensions that have important implications for the empirical results.

First, we use high frequency data, at the one-second and millisecond frequency. The data in Kao and Wan (2012) are daily, and so do not capture intraday dynamics that are most relevant in price discovery. That is, intraday dynamics are important because oil futures markets are very liquid, fully reflecting new information within minutes (cf. Elder, Miao and Ramchander, 2013). Hasbrouck (1995) uses data at a one-second frequency, cautioning that "if the observation interval is so long that the sequencing cannot be determined... the initial change and the response will appear to be contemporaneous." Second, our use of high frequency data permits us to avoid the rolling estimation prodedure in Kao and Wan (2012), which uses windows of 1 to 5 years. Such a long window imposes excessive structure on the underlying dynamics, likely rendering the estimates of information share unreliable (cf. Hasbrouck, 1995).

Our primary empirical result is that we find evidence that WTI maintains a dominant role in price discovery relative to Brent, with an estimated information share in excess of 80%. Our analysis is robust to different decompositions of the sample, over pit trading sessions and non-pit trading sessions, segmentation of days associated with major economic news releases, and data measured to the millisecond. We find no evidence that the dominant role of WTI in price discovery is diminished by the price spread between Brent that emerged in 2008.

The remainder of the study is organized as follows. The data and methodology are presented in Sections 2 and 3, respectively. Section 4 discusses the empirical results. The final section concludes.

2. Data

The key data utilized is the intraday transaction futures prices for WTI and Brent light sweet crude oil for the period January 2, 2007 to April 27, 2012.[3] The data is obtained from TickData. The WTI futures (Ticker: CL) are traded simultaneously on the electronic (CME Globex and ClearPort) and open outcry markets. The electronic market is open Sunday to Friday, 6:00 pm - 5:15 pm and the open outcry market is open Monday to Friday, 9:00 am - 2:30 pm (all times U.S. Eastern Time). The Brent futures contracts (Ticker: B) are traded on the InterContinental Exchange (ICE) electronic platform, Sunday to Friday, 8:00 pm to 6:00 pm on the following day.[4] The contract unit for both WTI and Brent is 1,000 barrels and the prices are quoted in U.S. dollars. For majority of the sample, January 2, 2007 to June 30, 2011, transaction prices are available at 1-second intervals. Beginning July 1, 2011 trades are reported at 1/1000 of each second. We use this latter subsample to conduct robustness tests.

At any given point in time there are many outstanding futures contracts with different expirations and transaction prices. The WTI crude oil futures are listed nine years forward using the following listing schedule: consecutive months are listed for the current year and the next five years; in addition, the June and December contract months are listed beyond the sixth year. The Brent crude oil futures are listed in consecutive months up to 7 years forward, although most of the longer-dated contracts are thinly traded. The first nearest (front) contracts are typically the most liquid. Following standard procedures, we form a continuous series by splicing price observations from contracts with the most number of transactions.

[3] We start the sample in 2007, when transaction volume is also available. Volume is used to identify the most active contracts in constructing the futures price time series.

[4] Due to the difference between the period of British Summer Time (BST) and the daylight saving time (DST) in the U.S., the InterContinental Exchange makes temporary changes to the trading hours. BST begins at 01:00 GMT on the last Sunday of March and ends at 01:00 GMT on the last Sunday of October. DST begins on the second Sunday of March and ends on the first Sunday of November.

Figure 1 plots the end-of-month WTI and Brent prices and the spread (left axis) for the full sample period. The two prices track each other closely between 2007 and 2010. Beginning 2011, the spread between the two price series widens considerably. The bottom two panels of Figure 1 also plot the monthly total volume and numbers of trades of the most active contracts for both WTI and Brent. The data indicate a slight increase in both the volume and number of trades for WTI relative to Brent.

Insert Figure 1 about here.

The summary statistics reported in Table 1 confirm these observations. Panel A of Table 1 reports the annual maximum, minimum and average prices of the most active WTI and Brent contracts. During 2007 and 2008, the mean difference between the WTI and Brent prices is positive and relatively small in magnitude (less than $2). The mean difference becomes slightly negative in 2009 and 2010, and then widens considerably in 2011 and 2012, to -$15. Throughout the sample period both WTI and Brent prices are volatile. Price were particularly volatile in 2008 when the maximum prices for both WTI and Brent exceeded $140. The minimum prices for WTI and Brent were $33.55 (in 2009) and $36.20 (in 2008). Panel B of Table 1 presents the daily average volume, number of trades and trade size. Trade size, which provides an indication of the type of market participant, is defined as the daily average volume divided by the total number of trades. The volume for WTI tended to increase through the sample, whereas the volume for Brent was relatively stable, expect for a large drop in 2009. A comparable drop in volume did not occur for WTI. From 2007 to 2011, the average daily volume of WTI relative to Brent increased from 1.58 times to 2.20 times, and until 2010, the trade size in WTI was larger than Brent. Beginning 2011 there is a reversal in the trade-size relationship between WTI and Brent, coinciding with the expanding negative spread.

Insert Table 1 about here.

We also use two sets of economic news announcements as proxies for information arrival. The first relates to the U.S. Employment Situation Report which is typically released at 8:30 am on the first Friday of each month. This report is widely followed by financial markets, and represents a broad measure of economic activity that includes data on the unemployment rate, labor force participation, the duration of unemployment as well as data from both the household and establishment surveys. Ex ante, we expect this report to contain a relatively high level of independent information about the state of the economy. The other proxy is the EIA (U.S. Energy Information Administration) weekly petroleum status report. The report provides information on weekly changes in petroleum inventories in the U.S., produced both locally and abroad. Market analysts and investors follow the inventory report to draw inferences on the supply and demand fundamentals in the oil market (Kaufmann, 2011). This report is generally released each Wednesday at 10:30 am. The exact release dates of the employment situation report and the weekly petroleum status report are obtained from Bloomberg. Our study period includes 342 news release dates - 64 employment reports and 278 EIA weekly petroleum status reports.

3. Price Discovery

Price discovery is the process by which security markets establish permanent changes in equilibrium transaction prices. The analysis is often based on the econometrics of cointegrated vector autoregressions under the assumption that intermarket arbitrage keeps asset prices (either the same asset or closely related assets) traded in different markets from drifting apart. If the prices are found to be integrated of order one, I(1), this implies that they are are non-stationary while price changes are covariance stationary - i.e., the price series share one or more common stochastic factors. If there is only one common factor this is referred to as the implicit efficient price. Hasbrouck (1995) defines the information share of a market as the proportion of the efficient price innovation variance that can be attributed to that market.

Hasbrouck's (1995) model can be applied to any finite number of price series, although this study involves only two price series (WTI and Brent). Assume that we observe a price vector $P_t = [P_{1,t}, P_{2,t}]'$, where P_1 and P_2 refer to the time series of the most active futures WTI and Brent contracts, respectively. In the original Hasbrouck (1995) paper, the two prices refer to observations on different markets of the same security. In Hasbrouck (2003), this condition is relaxed as long as the quantity $P_{1,t} - P_{2,t}$ does not diverge over time, or formally, the prices are cointegrated. In our study, the two prices involved, WTI and Brent, do diverge over time, but since the model is reestimated on a daily basis, this divergence is accommodated by the cointegrating vector. More importantly, the above cointegration tests indicate that the two price series are cointegrated on a daily basis. The cointegration of prices implies that they may be represented in a vector error correction model (VECM) of order K:[5]

$$\Delta P_t = \alpha\beta' P_{t-1} + \sum_{k=1}^{K} \mu_k \Delta P_{t-k} + u_t \tag{1}$$

where P_t is a vector of prices, α is the error correction vector that measures the speed of adjustment to the error correction term, $\beta = (1, -1)'$ is the cointegratiing vector, μ_k are matrices of autoregressive coefficients, and u_t are innovations with constant variance Ω. The VECM mode has two parts: the first part, $\alpha\beta' P_{t-1}$, represents the equilibrium dynamics between the price series, and the second part, $\sum_{k=1}^{K} \mu_k \Delta P_{t-k}$, depicts the short-term dynamics induced by market imperfections. This model can be represented as an integrated vector moving average process (see Watson (1994)):

$$P_t = \Psi(1) \sum_{v}^{t} u_v + \Psi^*(L) u_t, \tag{2}$$

[5] K is chosen to be 1200 in our calculation. This results in a large number of parameters to be estimated. To reduce the number of parameters (coefficients), we follow Hasbrouck (2002) by constraining a set of coefficients to be constant and constraining a set of coefficients to lie on a polynomial function of the lag. We should note that we are estimating a model over the course of one trading day with approximately 40,000 observations sampled at 1-second intervals for each day. This is repeated for each of 1,035 trading days. For this reason, we do not believe that structural change or breaks is an issue, since the parameters of the model are permitted to vary in an unrestricted fashion from one day to the next.

where $\Psi(1)$ is the sum of moving average coefficients $\Psi(1) = (1 + \Psi_1 + \Psi_2 + \cdots)$, and $\Psi^*(L)$ is a matrix polynomial in the lag operator (L). The first expression in equation (2) measures the long run impact of an innovation in prices, and, therefore, represents the common factor component among the price series. The second expression $\Psi^*(L)$ is transitory, and therefore measures the temporary influence on prices. Hasbrouck (1995) defines the information share of a price series as the proportion of the variance in the common price process that is attributable to that particular price series. Defining a row of $\Psi(1)$ as ψ, then the IS for the i^th time series is

$$IS_i = \frac{\psi_i^2 C_{ii}}{\psi \Omega \psi'}, \tag{3}$$

where C is the lower triangular Choleski factorization of Ω. The Choleski factorization orthogonalizes the variance in the common price process that attributed to each innovation, but since it is dependent on the arbitrary ordering of the price series in the VECM, the result is an estimate of IS_i that is not uniquely defined. Instead, upper and lower bounds of IS_i are calculated by applying the Cholesky factorization to all possible orderings. Baillie et al. (2002) find that the mean IS from across all orderings is a reasonable estimate of that price series's contribution to price discovery. Additional details on information share are provided in Hasbrouck (1995, 2003) and Baillie et al. (2002).

Finally, on a methodological note, it is useful to point out that the information share model is based on the common permanent component of all market prices. Price changes that are transitory are not viewed as aiding price discovery. This is also one reason why analyzing temporal dependence is not useful since it is not capable of distinguishing between transitory and permanent changes in prices, so that lead-lag relationships may capture only transitory fluctuations (such as bid-ask bounce) that are not information related.

4. Empirical Results

4.1. A First Look

Following Hasbrouck (1995, 2003), we estimate the IS model at the 1-second frequency level. In the event that there are multiple transactions during a 1-second interval, the last transaction price is used in the analysis. If there are no trades reported during a 1-second interval, the last price from the previous 1-second interval is taken. If trading is inactive this approach may create a sequence of constant prices. In order to minimize the effect of stale prices, we analyze trade activity patterns to identify time intervals when trading is reasonably active. Specifically, we calculate the average trading volume and number of trades per hour. The results are presented in Figures 2 and 3. In general the figures show that the volumes for both WTI and Brent are very high during the pit trading session, 9:00 am to 2:30 pm. In contrast, the hourly average volumes for both contracts are very low from 8:00 pm to 2:00 am, and the volumes dramatically shrink again after the pit market closes. To examine whether this pattern is different for the sub-sample when the spread between WTI and Brent is negative, we plot the average hourly trading volumes and number of trades for the period of 04/2010 to 04/2012. The plots show similar patterns with slightly higher trading volumes and much larger number of trades (perhaps in order to accommodate the decreasing trade sizes). Considering these issues, we estimate the IS model over the 3:00 am to 2:30 pm time interval.

Insert Figures 2 and 3 about here.

Hasbrouck (2003) indicates that the VECM price discovery model is most appropriate within a trading session, so we estimate the model separately for each day. In order to establish the suitability of the information share model for our data, we first perform cointegration tests by utilizing the Stock and Watson (1988) test for common trends and the Johanson (1991) trace test. The cointegration tests are performed on the 1-second frequency data estimated over each trading day. The summary statistics of the results are presented in Table 2. The results indicate that on the vast majority of days, WTI and Brent sweet crude oil prices share a single common cointegrating vector. For instance, the Stock and Watson (1988) test indicates the presence of a single common trend within the two price time series for 1089, 1173, and 1221 days at the 1%, 5% and 10% levels of significance, respectively, out of 1345 trading days in the sample. Similar conclusions are achieved using the Johansen test. Together, these results indicate the assumption of single cointegrating vector over a one-day horizon is a reasonable characterization of the series.

Insert Table 2 about here.

In the Hasbrouck information sharing model, the Choleski factorization of the covariance matrix provides a means for orthogonalizing the residuals, and provides estimates the upper bound and the lower bound of the information shares. As pointed out by Ballie et al. (2002), the spread between the two bounds is positively related to the degree of correlation, and is zero if the residuals are uncorrelated. The correlation is driven by information flows between the markets and the frequency of the price data, with very high frequency data typically being less highly correlated. On the other hand, if the sampling frequency is low, the differences between the upper and lower bounds may be very large.[6] In the presence of modest correlation, the average of the upper and lower bounds provides a reasonable estimate of the information share, and therefore of each markets' role in the production of the efficient price.

Table 3 reports summary statistics on the aggregated estimates of the information share for each trading session, including the means, standard deviations, medians and standard

[6]For instance, Huang (2002), uses one-minute intervals to examine the price discovery between the electronic communications networks (ECNs) and various Nasdaq dealers. The lower and upper bounds of the Island (an ECN) for Yahoo, are 79.5% and 30.6%, respectively, for the month of January 1998. For the month of November 1999, the upper and lower bounds are 47.7% and 8.4%. Booth et al. (2002) study the price discovery between the Finnish upstairs and downstairs stock markets using trading intervals averaging approximately 30 minutes. The reported information average share upper and lower bounds for the downstairs market are 99.2% and 13.0%, respectively.

errors of the mean across days for both the upper bound, lower bound and the average of lower and upper bound. The model is first estimated for each trading session (3:00am to 2:30pm). During this time interval Brent trades solely on an electronic trading platform, while WTI trades on the electronic trading platform (Npit) and on Open Outcry (Pit) from 9:00am to 2:30pm. To control for effects associated with pit trading, we segment each trading day from 3:00am to 9:00am and 9:00am to 2:30pm and estimate the model separately over these intervals.

The results in Table 3 indicate that WTI has the dominant information share. For example, the first column reports the estimated information share for WTI when the model is estimated over both the pit and non-pit trading sessions for each trading day in the sample. This provides 1345 estimates of the information share for WTI (for a SMALL number of days, the model does not converge). The middle four rows report summary statistics on the mean information share for each day, in which the mean information share for each is the average of the estimated lower and upper bounds for each day. The mean information share for WTI, averaged over 1345 estimates of daily mean information share, is 81.6%. The median of the daily mean is 82.3%. The variation in the daily mean estimate of the information share is not large, with a standard deviation of 8.3%. This clearly indicates the dominant information share of WTI.

A more conservative estimate of the information share for WTI is based on the estimates of the daily lower bound. Summary statistics on the lower bound of WTI, reported in the bottom four rows of Table 3, show a mean lower bound of 69.3%. For comparison, the mean lower bound for the information share of Brent is 6.1%. Summary statistics on the upper bound of the information share, reported in the top four rows of Table 3, indicate a mean upper bound of 93.9% for WTI versus 30.7% for Brent.

The dominance of WTI in the information share of price discovery is not sensitive to the pit versus non-pit trading sessions, as similar results are obtained when the model is estimated separately over the pit trading session and the non-pit trading session. The mean of the daily average information share during the pit session is 80% for WTI and 20% for Brent, versus 78.7% for WTI and 21.3% for Brent in the non-pit session.

Overall, the results from Table 3 clearly suggest that WTI dominates Brent in the price discovery process for oil, with an information share of about 80%, versus Brent, with an information share of about 20%. We believe that our results differ from Kao and Wan (2012), who find a lower for information share for WTI beginning in 2004, for two primary reasons. First, we use high frequency data in order to capture intraday dynamics, as opposed to the daily data in Kao and Wan (2012). Second, we reestimate our model each day, whereas Kao and Wan (2012) employ a rolling regression with a window length up to five years.

Insert Table 3 about here.

4.2. Evolution of Information Shares

We have established that over the five year period, 2007-2012, WTI has tended to lead Brent in price discovery, with an information share in excess of 80%. Kao and Wan (2012) report some variation in the information share of WTI, finding that, with a 750-day moving window, the IS of WTI has decreased over time, falling below 50% starting in the second half of 2004. We calculate and plot the IS for WTI and Brent over our sample period (January 2007 to April 2012) in Figure 4. The top panel of Figure 4 presents the monthly averages of the upper bounds, lower bounds and the averages of the bounds for both WTI and Brent. The average IS of WTI tends to vary between 65% to about 90%, with the upper and lower bounds relatively close in the first half of the sample and wider in the second half, particularly in 2009 and 2010. The lower bound of the IS approaches a low of about 50% for only two brief periods during the sample. In contrast, the mean IS for Brent is always less than 40%, with an upper bound that around 50% only twice during the sample. In particular, there is only one month, January 2010, for which the average upper bound of Brent (51.2%) exceeds the average of the lower bound of WTI (48.8%).

This month also has the lowest average IS for WTI, of 67.4%. Interestingly, WTI does not lose its IS advantage over Brent, even during periods that it traded at a discount to Brent. This finding contrasts with Kao and Wan (2012).

The bottom panel in Figure 4 plots the daily time series average of the upper and lower bound of the information shares for both WTI and Brent. It reinforces our finding that WTI dominates Brent in the price discovery process. There is a slight drop in the IS of WTI during the period of 07/2009 to 07/2010, but there are only five days when the information share of WTI drops below that of Brent. These are 12/12/2008 (37.2% for WTI vs. 62.8% for Brent), 4/9/2009 (40.9% vs. 59.1%), 5/12/2009 (39.1% vs. 60.9%), 5/11/2009 (31.4% vs. 68.6%), 1/21/2010 (33.9% vs. 66.1%) and 1/22/2010 (33.9% vs. 66.1%).

In summary, there appears to be strong support to the price leadership of WTI over Brent and this result is robust to the sign of the spread between the two benchmark prices. This does raise the question of whether WTI's dominant role may be driven, at least partly, by its higher trading volume.[7] However, the empirical evidence that WTI maintains a dominant role in price discovery relative to Brent does not appear to be due to changes in both the volume and number of trades for WTI relative to Brent. Combining Figures 1 and 4 provides a clear answer to this question. From Figure 1, we observe that from 2007 to 2012, the volume and number of trades of the most active WTI contracts increase more than Brent. In particular, the difference in number of trades of the WTI and Brent becomes consistently wider over time. On the other hand, Figure 4 shows that the information share of WTI relative to Brent does not follow a similar pattern. Instead we find that there is considerable variation in the monthly means of information during this time period. Putting both pieces of evidence together suggests that there is no consistent relationship between relative increase of volume and number of trades and the information share of WTI.

[7] We thank the reviewer for raising this question.

Another possible explanation is that the EIA/DOE releases frequent and detailed analysis of oil markets which may drive most price discovery, causing WTI to lead Brent. Our paper considers such an indirect test within the information share framework. This is formally evaluated in the next subsection.

Insert Figure 4 about here.

4.3. News and Information Share

In this subsection, we examine whether the information share tends to vary with relevant economic news. The motivation for this analysis is that economic news tends to drive major price changes. If WTI dominates price discovery relative to Brent, then we should expect WTI to dominate during periods of major economic news releases that tend to move oil prices. If the price discovery of WTI is stable, then we would expect the IS share to not vary with news announcements. To investigate these issues, we consider two sets of scheduled economic news releases - the employment situation report and EIA weekly petroleum status report. Both of these news announcements have been found to drive oil prices (cf., Elder, Miao and Ramchander, 2013).

We conduct this analysis by separating the sample into days without news releases from the EIA on inventories, and those with news releases. The results on IS calculated on this days are reported in Table 4. The results in Table 4 indicate that there is little variation in IS between the two sample groups. For instance, the average information share for WTI for days without and with EIA inventory reports, in columns 1 and 7, is 81.7% and 81.1%, respectively, This is also little variation in information share in pit versus non-pit trading hours, although IS of WTI is slightly higher than for during pit hours than non-pit hours.

Table 5 reports the comparable results or days with and without the releases of the employment situation report. The reason for selecting the employment situation report is that its importance as a source of information for oil prices has been documented in earlier studies. For instance, Andersen and Bollerslev (1998) refer to the Employment Situation

Report as the "king" of all announcements because of the significant sensitivity of most asset prices to its public release. Elder, Miao and Ramchander (2013) find that the Change in Nonfarm Payrolls, a major component of the Employment Situation Report, is the only one of ten major macroeconomic factor which has significant impact on intraday jumps in crude oil prices. The results reported in Table 5, reaffirm our earlier results. The average information share of WTI remains at about 80%, independent of the news release or other trading session (pit or non-pit).

Overall, our results in Tables 4 and 5 show that the IS of WTI is dominant and stable across major news releases that impact crude oil prices.

Insert Tables 4 to 5 about here.

4.4. The WTI-Brent Price Spread

The role of price discovery in crude oil markets is intertwined with the recent debate on the relative merits of Brent and WTI as price benchmarks for crude oil. It has been argued that as the production of shale oil in the US has increased dramatically, a significant price difference between WTI and Brent has developed. This price spread has been attributed to export restrictions on WTI as well as bottlenecks in the domestic transportation infrastructure, and has affected the relevance of WTI as a benchmark measure of the level of world oil prices (see, for example, Baumeister and Kilian, 2013). This raises several issues. One issue is whether the empirical model is appropriate in the presence of a consistently widening price spread. The price spread is captured in the Hasbrouck information share model by the cointegrating vector. Since, the model is reestimated each day, the point estimates for the cointegrating vector may evolve to accommodate the longer term trends in the spread. Another issue is whether the price spread affects the role of WTI in price discovery. We consider this hypothesis by examining the relationship between the IS of WTI and an indirect measure of the price spread – inventory levels in Cushing, Oklahoma – as well as the WTI-Brent price spread.

Figure 5 plots the monthly average IS and inventory levels at Cushing. Visually, there does not appear to a strong relationship, although there is a small increase in IS while inventory levels drop from May 2007 to October 2007. However, there is no discernible relationship between the rapid increase of inventories beginning the second-half of 2008 through 2011 and the IS of WTI. We also conduct a more formal investigation, testing for Granger-causality from the price spread to the IS, using appropriate transformations to render the series stationary. We are unable to reject the null hypothesis of no causal relationship between Cushing inventory levels and the information share of WTI.[8]

Insert Figure 5 about here.

Next, we examine the relationship between the information share of WTI and the direction of the price spread, using a procedure similar to Kao and Wan (2012). That is, we segment the sample into days when the price spread is positive and negative. A positive (negative) spread is defined as one in which the average daily trade prices of WTI's most active futures contracts is greater (smaller) than the average daily trade price of Brent. These result are reported in Table 6. There appears to be little evidence that the information share of WTI is contingent on the relative spread between WTI and Brent prices. Rather, we find that the IS between the days with positive spreads and negatives spreads is virtually identical, with the mean of the average of the upper and lower bound of WTI's IS is 81.9% for the 772 days with positive spreads, and 81.1% for the 573 days with negative spreads.

Insert Table 6 about here.

4.5. Sampling Frequency

The above results suggest that WTI has dominate role in price discovery, when prices measured at 1-second intervals. Baillie et al. (2002) analytically show that upper and lower

[8] We test for stationarity using Augmented Dickey-Fuller tests, which suggests that both series require first-differencing. The results are available from the authors upon request.

bounds in the IS can differ substantially when the correlation between contemporaneous price innovations are high. They suggest that higher frequency data, which tends to have lower contemporaneous correlation, should result in IS shares that are estimated more precisely, since the model is less dependent on the Cholesky factorization to orthogonalize the price innovations. A downside of higher frequency data is that results may be highly dependent on very accurate time-stamping and may potentially be contaminated by other microstructure effects. Another downside of the higher frequency data in our application is the short time span, as described below.

This analysis uses higher frequency data, but is restricted to a smaller sample (207 trading days) between July 2, 2011 to April 27, 2012, during which futures transaction prices are available on a tick-by-tick basis and time-stamped to the millisecond. All available trades from the most active contracts are used to estimate the model. If for a given point in time, say 9:30:011, there is a trade for WTI and no corresponding trade for Brent, then we use the WTI trade and the last trade for Brent. This results in a sample size of nearly 15 million observations for the 3:00 am to 2:30 pm trading interval. Separate models are estimated for three different trading intervals - 3:00 am to 2:30 pm (sample size is 15 million observations), 3:00 am to 9:00 am (non-pit trading hours), and 9:00 am to 2:30 am (pit trading hours) intervals. The summary statistics on the information share are reported in Table 7.

The results in Table 7 indicate that the higher frequency enables the information share to be estimated with greater precision. For example, the average IS for WTI is 98.1% with an average upper bound of 98.3% and lower bound of 98.0%; compared to bounds of 93.9% and 69.3% for the overall sample using 1-second frequency (cf. Table 3). This shrinking of the bounds is consistent with effects described by Baillie et al. (2002) for the Hasbrouck model with higher frequency data. The lower correlation of the higher frequency price data permits the model to better attribute price discovery to a single series. The results also show that IS during the pit trading session is slightly greater than during the non-pit trading sessions for WTI, with the average IS between 98.9% and 94.1% for pit and non-pit, respectively. The corresponding values for Brent are only 1.1% and 5.9%.

Insert Table 7 about here.

5. Conclusions

The recent inversion of the spread between WTI and Brent crude oil prices have led market participants to question the continued viability of WTI as an international benchmark. This debate is closely tied to the price discovery function of the two benchmarks. This paper uses Hasbrouck's (1995) information share model to investigate the mechanics of price discovery, defined in terms of each market's relative contribution to the variance of the innovations to a common factor. Using the most active futures contracts, we find that WTI maintains a dominant role in price discovery relative to Brent, with an estimated information share of approximately 80%, over a sample from 2007 through 2012. Our analysis is robust to different decompositions of the sample, include pit-trading sessions versus non-pit trading sessions and segmentation of days associated with major economic news releases. We also aggregate the information by month, revealing that the information share of WTI has tended to vary between 65% to 90%, and almost always dominates Brent.

Finally, we examine whether the well-known price spread between WTI and Brent, along with inventories levels at Cushing, influences the price discovery role of WTI. We find no evidence for such an effect. Finally, our evidence that WTI exhibits a dominant role in price discovery is robust to higher frequency data, time-stamped at the millisecond.

Overall our results strongly support the leading role of WTI in incorporating new information into oil prices. Based on this evidence WTI still plays an important role as a benchmark for world crude oil prices.[9]

[9] At the time the paper's final revision, the spread between WTI and Brent about -$3.00, which is much narrower than 2011.

References

[1] Alizadeh, A.H., Nomikos, N.K., 2004. Cost of carry, causality and arbitrage between oil futures and tanker freight markets. Transportation Research Part E: Logistics and Transportation Review 40, 297-316.

[2] Andersen, T.G., Bollerslev, T., 1998. DM-dollar intraday volatility: activity pattern, macroeconomic announcements, and longer run dependences. Journal of Finance 53, 219–265.

[3] Bachmeier, L.J., Griffin, J.M., 2006. Testing for market integration: Crude oil, coal, and natural gas. Energy Journal 27, 55-72.

[4] Baillie, R.T., Booth, G.G., Tse, Y., Zabotina, T., 2002. Price discovery and common factor models. Journal of Financial Markets 5, 309–321.

[5] Baumeister, C., Kilian, L. 2013. Are Product Spreads Useful for Forecasting? An Empirical Evaluation of the Verleger Hypothesis? University of Michigan, working paper.

[6] Bentzen, J.B., 2007. Does OPEC influence crude oil prices? : Testing for co-movements and causality between regional crude oil prices. Applied Economics 39, 1375-1385.

[7] Booth, G.G., Lin, J., Martikaines, T., Tse, Y., 2002. Trading and pricing in upstairs and downstairs stock markets. Review of Financial Studies 15, 1111-1135.

[8] Borenstein, S., Kellogg, R., 2014. The incidence of an oil glut: Who benefits from cheap crude oil in the midwest? Energy Journal 35, 15-33.

[9] Elder, J., Miao, H., Ramchander, S., 2013. Jumps in oil prices: the role of economic news. Energy Journal 34, 217-237.

[10] Fattouh, B., 2007. WTI benchmark temporarily breaks down: Is it really a big deal? Oxford Energy Comment, Oxford Institute for Energy Studies, April.

[11] Fattouh, B., 2010. The dynamics of crude oil price differentials. Energy Economics 32, 334-42.

[12] Fattouh, B., 2011. An Anatomy of the Crude Oil Pricing System." Working Paper WPM40, Oxford Institute for Energy Studies, January.

[13] Hammoudeh, S., Thompson, M., Ewing, B., 2008. Threshold cointegration analysis of crude oil benchmarks. Energy Journal 29, 79-95.

[14] Hasbrouck, J., 1995. One security, many markets: Determining the contributions to price discovery. Journal of Finance 50, 1175-1199.

[15] Hasbrouck, J., 2003. Intraday price formation in U.S. equity index markets. Journal of Finance 58, 2375-2399.

[16] Huang, R.D., 2002. The quality of ECN and Nasdaq market maker quotes. Journal of Finance 57, 1285-1319.

[17] Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrica 59, 1551–1580

[18] Kao, C., Wan, J., 2012. Price discount, inventories and the distortion of WTI benchmark. Energy Economics 34, 117-124.

[19] Kleit, A.N., 2001. Are regional oil markets growing closer together? An arbitrage cost approach. The Energy Journal 22, 1-15.

[20] Stock, J.S., Watson, M.W., 1988. Variable trends in economic time series. Journal of Economic Perspectives 2, 147–174.

Table 1: Summary Statistics of Prices, Volume and Trades

This table reports the summary statistics of prices, volume and trades each year in the sample.

Panel A: Max, Min and Average Prices Over Trades

Year	Max			Mean			Min		
	WTI	Brent	Diff	WTI	Brent	Diff	WTI	Brent	Diff
2007	99.29	96.65	2.64	74.83	73.29	1.54	50.28	50.75	-0.47
2008	147.27	147.50	-0.23	102.56	101.66	0.90	35.13	36.20	-1.07
2009	81.99	80.26	1.73	62.45	64.04	-1.59	33.55	39.35	-5.80
2010	92.06	94.74	-2.68	79.11	79.91	-0.80	67.15	67.88	-0.73
2011	114.83	127.02	-12.19	94.78	110.67	-15.88	74.95	92.37	-17.42
2012	110.55	128.40	-17.85	103.09	118.65	-15.56	95.44	108.35	-12.91

Panel B: Daily Average Volume, Number of Trades and Trade Size

Year	Volume (Thousands)			Trades (Thoursands)			Trade Size		
	WTI	Brent	Ratio	WTI	Brent	Ratio	WTI	Brent	Ratio
2007	151.39	95.61	1.58	51.18	38.22	1.34	2.96	2.50	1.18
2008	185.49	96.72	1.92	92.69	55.17	1.68	2.00	1.75	1.14
2009	187.58	64.56	2.91	91.43	43.66	2.09	2.05	1.48	1.39
2010	228.91	81.16	2.82	136.58	56.39	2.42	1.68	1.44	1.16
2011	232.92	105.82	2.20	178.16	70.57	2.52	1.31	1.50	0.87
2012	195.69	100.59	1.95	154.83	69.13	2.24	1.26	1.46	0.87

Table 2: Summary Statistics of Cointegration Test Results

This table reports results from alternative cointegration tests using the Stock-Watson (Panel A) and the Johansen (Panel B) methods. The cointegration tests are conducted on each day of the sample. For the Stock-Watson test, the null hypothesis in our study is that there are 2 common trends against the alternative that there is 1 trend. The null hypothesis for Johansen's Trace test is that the number of cointegrating vectors zero. The last three columns in the table indicate the number of days during which the null hypothesis is rejected at the 10%, 5% and 1% levels of statistical significance.

Statistics	Mean	Std	Min	Max	Sig at 1%	Sig at 5%	Sig at 10%
Panel A: Stock-Watson test for common trends							
χ^2	-148.25	154.96	-1040.50	-2.56	1089	1173	1221
Panel B: Johansen Trace test							
Trace	104.42	107.59	5.01	778.53	1112	1143	1051

Table 3: Information Share: Pit vs. Non-pit Trading Sessions

This table reports the summary statistics on the estimated information share. Statistics are based on a vector error correction model of prices for WTI and Brent nearest futures contract prices estimated at 1-second resolution. All prices are the last-sale prices.The model is estimated for each of the 1345 trading days (3:00 to 14:30) in the sample (January 2, 2007 through April 27, 2012). The model is then estimated separetely for non-pit (3:00-9:00) and pit (9:00-14:30) sessions. The table reports summary statistics on the daily estimates of the information share. The values in "( )" are the number of days in the sample when the model converges.

	Statistics	All (1345)		Pit (1343)		NPit (1343)	
		WTI	Brent	WTI	Brent	WTI	Brent
Upper Bound	Mean	0.939	0.307	0.936	0.336	0.891	0.316
	Median	0.951	0.303	0.956	0.330	0.912	0.307
	SEM	0.002	0.003	0.002	0.004	0.002	0.003
	Std. dev.	0.063	0.121	0.086	0.152	0.088	0.122
Avg of Bounds	Mean	0.816	0.184	0.800	0.200	0.787	0.213
	Median	0.823	0.177	0.809	0.191	0.800	0.200
	SEM	0.002	0.002	0.003	0.003	0.003	0.003
	Std. dev.	0.083	0.083	0.109	0.109	0.096	0.096
Lower Bound	Mean	0.693	0.061	0.664	0.064	0.684	0.109
	Median	0.697	0.049	0.670	0.044	0.693	0.088
	SEM	0.003	0.002	0.004	0.002	0.003	0.002
	Std. dev.	0.121	0.063	0.152	0.086	0.122	0.088

Table 4: Information Share: Days with/without EIA Inventory Announcements

This table reports the summary statistics on the daily estimates of information sharel. Statistics are based on a vector error correction model of prices for WTI and Brent nearest futures contract prices estimated at 1-second resolution. All the prices are the last-sale prices.The model is then estimated separately for each trading-day (3:00 to 14:30) in the sample (January 2, 2007 through April 27, 2012). The model is then estimated separetely for non-pit (3:00-9:00) and pit (9:00-14:30) sessions. The summary statistics are grouped into days with EIA Annoucements and days without EIA inventory announcements. The values in "( )" are the number of days in the sample when the model converges.

	Statistics	Days without DOE Inventory Ann.						Days with DOE Inventory Ann.					
		All (1068)		Pit (1065)		NPit (1066)		All (277)		Pit (278)		NPit (277)	
	Statistics	WTI	Brent	WTI	Brent	WTI	Brent	WTI	Brent	WTI	Brent	WTI	Brent
Upper Bound	Mean	0.938	0.304	0.936	0.331	0.890	0.318	0.942	0.320	0.939	0.355	0.894	0.310
	Median	0.951	0.301	0.956	0.324	0.912	0.307	0.952	0.320	0.957	0.354	0.911	0.306
	SEM	0.002	0.004	0.003	0.005	0.003	0.004	0.003	0.008	0.004	0.009	0.004	0.007
	Std. dev.	0.066	0.120	0.091	0.151	0.092	0.125	0.047	0.126	0.065	0.155	0.074	0.112
Avg of Bounds	Mean	0.817	0.183	0.802	0.198	0.786	0.214	0.811	0.189	0.792	0.208	0.792	0.208
	Median	0.825	0.175	0.813	0.187	0.800	0.200	0.816	0.184	0.801	0.199	0.799	0.201
	SEM	0.003	0.003	0.003	0.003	0.003	0.003	0.005	0.005	0.006	0.006	0.005	0.005
	Std. dev.	0.083	0.083	0.111	0.111	0.099	0.099	0.080	0.080	0.102	0.102	0.084	0.084
Lower Bound	Mean	0.696	0.062	0.669	0.064	0.682	0.110	0.680	0.058	0.645	0.061	0.690	0.106
	Median	0.699	0.049	0.676	0.044	0.693	0.088	0.680	0.048	0.646	0.043	0.694	0.089
	SEM	0.004	0.002	0.005	0.003	0.004	0.003	0.008	0.003	0.009	0.004	0.007	0.004
	Std. dev.	0.120	0.066	0.151	0.091	0.125	0.092	0.126	0.047	0.155	0.065	0.112	0.074

Table 5: Information Share: Days with/without Employment Situation Report

This table reports the summary statistics on the estimated information share. Statistics are based on a vector error correction model of prices for WTI and Brent nearest futures contract prices estimated at 1-second resolution. All the prices are the last-sale prices.The model is estimated for each trading-day (3:00 to 14:30) in the sample (January 1, 2007 through April 27, 2012). The model is then estimated separetely for non-pit (3:00-9:00) and pit (9:00-14:30) sessions. The table reports summary statistics for these daily estimates by grouping the days into days with and without the releases of monthly Employment Situation reports. The values in "( )" are the number of days in each of the samples when the model converges.

	Statistics	Days without Empl. Sit. Report						Days with Empl. Sit. Report					
		All (1284)		Pit (1282)		NPit (1282)		All (61)		Pit (61)		NPit (61)	
	Statistics	WTI	Brent	WTI	Brent	WTI	Brent	WTI	Brent	WTI	Brent	WTI	Brent
Upper Bound	Mean	0.939	0.306	0.936	0.335	0.890	0.314	0.946	0.328	0.945	0.345	0.899	0.359
	Median	0.951	0.302	0.956	0.330	0.912	0.306	0.952	0.315	0.952	0.342	0.929	0.329
	SEM	0.002	0.003	0.002	0.004	0.002	0.003	0.004	0.017	0.005	0.017	0.011	0.020
	Std. dev.	0.064	0.121	0.088	0.153	0.088	0.120	0.031	0.136	0.040	0.134	0.085	0.159
Avg of Bounds	Mean	0.816	0.184	0.800	0.200	0.788	0.212	0.809	0.191	0.800	0.200	0.770	0.230
	Median	0.823	0.177	0.810	0.190	0.800	0.200	0.814	0.186	0.800	0.200	0.778	0.222
	SEM	0.002	0.002	0.003	0.003	0.003	0.003	0.010	0.010	0.010	0.010	0.014	0.014
	Std. dev.	0.083	0.083	0.110	0.110	0.095	0.095	0.079	0.079	0.081	0.081	0.111	0.111
Lower Bound	Mean	0.694	0.061	0.665	0.064	0.686	0.110	0.672	0.054	0.655	0.055	0.641	0.101
	Median	0.698	0.049	0.670	0.044	0.694	0.088	0.685	0.048	0.658	0.048	0.671	0.071
	SEM	0.003	0.002	0.004	0.002	0.003	0.002	0.017	0.004	0.017	0.005	0.020	0.011
	Std. dev.	0.121	0.064	0.153	0.088	0.120	0.088	0.136	0.031	0.134	0.040	0.159	0.085

Table 6: Information Share: Positive Spread Days vs. Negative Spread Days

This table reports the summary statistics on the estimated information share. Statistics are based on a vector error correction model of prices for WTI and Brent first nearest futures contract prices estimated at 1-second resolution. All the prices are the last-sale prices.The model is estimated for each trading-day (3:00 to 14:30) in the sample (January 2, 2007 through April 27, 2012). The model is then estimated separetely for non-pit (3:00-9:00) and pit (9:00-14:30) sessions. The table reports summary statistics for these daily estimates by grouping the days into positive spread days (when the average 1-second last-trade prices of WTI are higher than Brent) and negative spread days. The values in “( )” are the number of days in the sample when the model converges.

	Statistics	Days with Positive Spreads						Days with Negative Spreads					
		All (772)		Pit (770)		NPit (771)		All (573)		Pit (573)		NPit (572)	
	Statistics	WTI	Brent	WTI	Brent	WTI	Brent	WTI	Brent	WTI	Brent	WTI	Brent
Upper Bound	Mean	0.945	0.306	0.941	0.336	0.901	0.314	0.932	0.309	0.929	0.336	0.876	0.320
	Median	0.953	0.306	0.959	0.333	0.921	0.301	0.947	0.292	0.951	0.325	0.897	0.317
	SEM	0.001	0.004	0.003	0.005	0.003	0.004	0.003	0.006	0.004	0.007	0.004	0.006
	Std. dev.	0.041	0.098	0.069	0.129	0.076	0.112	0.083	0.147	0.105	0.178	0.101	0.135
Avg of Bounds	Mean	0.819	0.181	0.803	0.197	0.794	0.206	0.811	0.189	0.797	0.203	0.778	0.222
	Median	0.822	0.178	0.809	0.191	0.806	0.194	0.825	0.175	0.810	0.190	0.788	0.212
	SEM	0.002	0.002	0.003	0.003	0.003	0.003	0.004	0.004	0.005	0.005	0.004	0.004
	Std. dev.	0.062	0.062	0.091	0.091	0.087	0.087	0.104	0.104	0.130	0.130	0.107	0.107
Lower Bound	Mean	0.694	0.055	0.664	0.059	0.686	0.099	0.691	0.068	0.664	0.071	0.680	0.124
	Median	0.694	0.047	0.667	0.041	0.699	0.079	0.708	0.053	0.675	0.049	0.683	0.103
	SEM	0.004	0.001	0.005	0.003	0.004	0.003	0.006	0.003	0.007	0.004	0.006	0.004
	Std. dev.	0.098	0.041	0.129	0.069	0.112	0.076	0.147	0.083	0.178	0.105	0.135	0.101

Table 7: Information Share: Pit vs. Non-pit Trading Sessions- tick frequency

This table reports the summary statistics of the information share model. Statistics are based on a vector error correction model of prices for WTI and Brent first nearest futures contract prices estimated at the tick level resolution. All the prices are the last-sale prices.The model is estimated for each trading-day (3:00 to 14:30) in the sample (July 1, 2011 through April 27, 2012). The model is then estimated separetely for non-pit (3:00-9:00) and pit (9:00-14:30) sessions. The table reports summary statistics for these daily estimates. The values in "( )" are the number of days in the sample when the model converges.

	Statistics	All (207)		Pit (207)		Non-Pit (207)	
		WTI	Brent	WTI	Brent	WTI	Brent
Upper Bound	Mean	0.983	0.020	0.989	0.012	0.944	0.062
	Median	0.985	0.018	0.990	0.012	0.956	0.050
	SEM	0.002	0.002	0.000	0.001	0.005	0.005
	Std. dev.	0.022	0.023	0.007	0.008	0.065	0.067
Avg of Bounds	Mean	0.981	0.019	0.989	0.011	0.941	0.059
	Median	0.983	0.017	0.989	0.011	0.953	0.047
	SEM	0.002	0.002	0.001	0.001	0.005	0.005
	Std. dev.	0.023	0.023	0.007	0.007	0.066	0.066
Lower Bound	Mean	0.980	0.017	0.988	0.011	0.938	0.056
	Median	0.982	0.015	0.988	0.010	0.950	0.044
	SEM	0.002	0.002	0.001	0.000	0.005	0.005
	Std. dev.	0.023	0.022	0.008	0.007	0.067	0.065

Figure 1: Monthly Prices and Spreads

This figure shows end of month prices of WTI and Brent and the spreads between them for the overall period 01/2007 - 04/2012

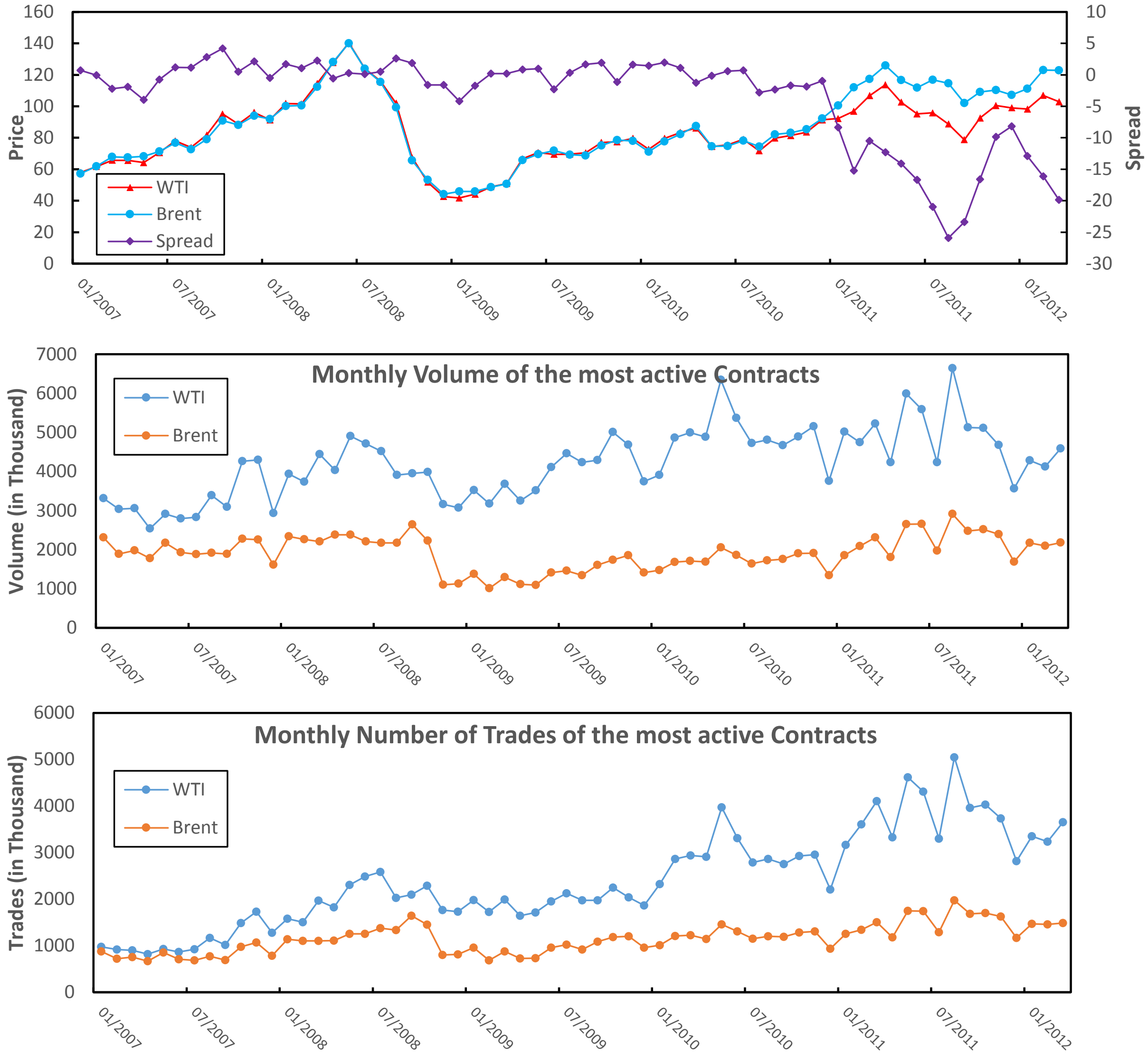
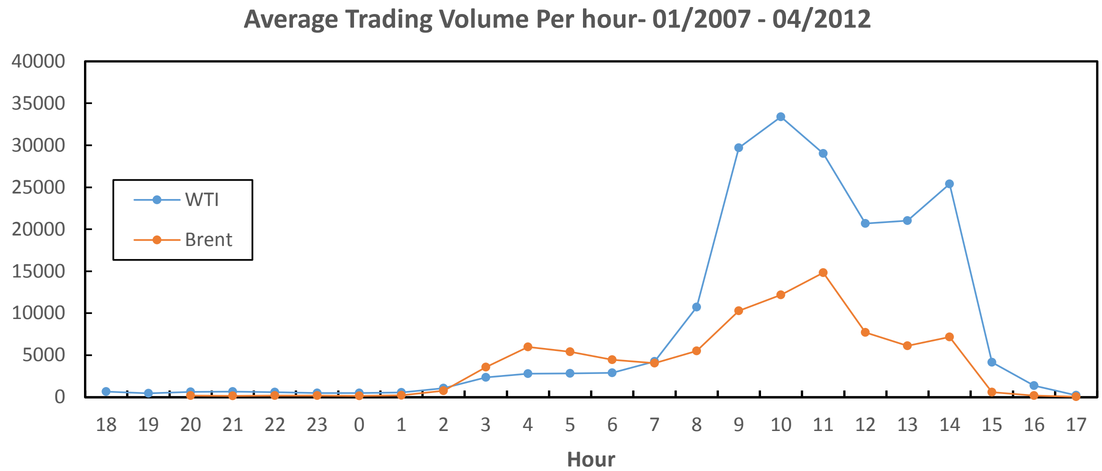


Figure 2: Hourly Average Trading Volume

This figure shows average hourly trading volume for both WTI and Brent over the overall period of 01/2007 - 04/2012 and the subperiod 04/2010 to 04/2012 when the price spreads were negative and wide.



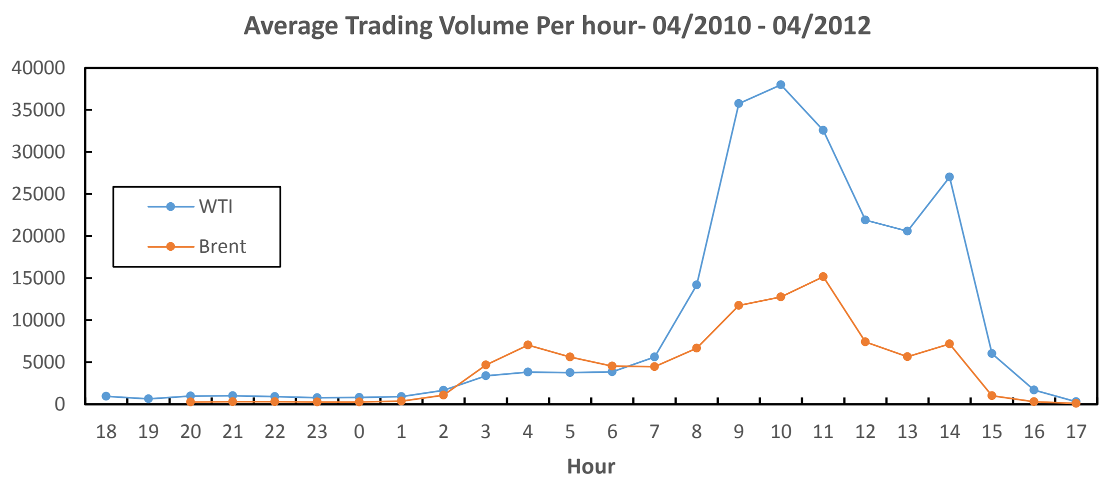
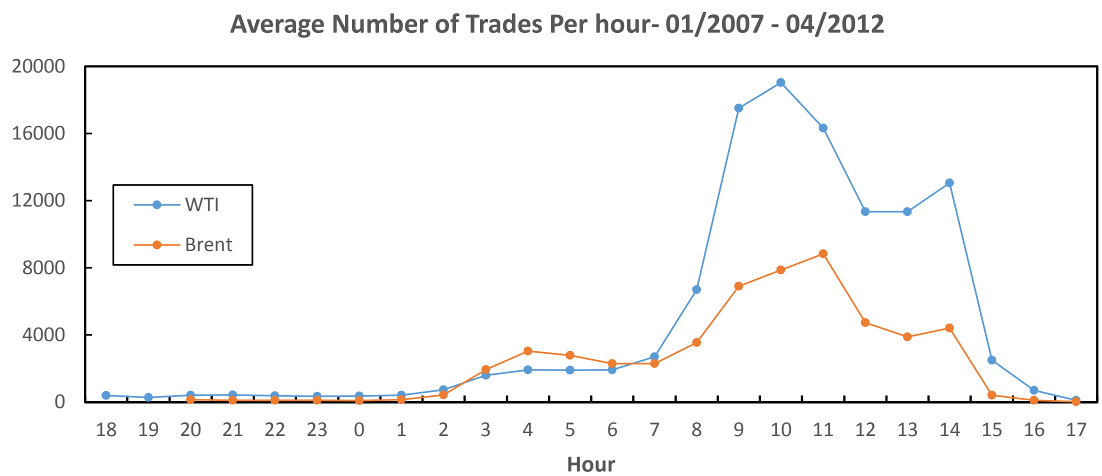


Figure 3: Hourly Average Number of Transactions

This figure shows average number of hourly transactions for both WTI and Brent over the overall period of 01/2007 - 04/2012 and the subperiod 04/2010 to 04/2012 when the price spreads were negative and wide.



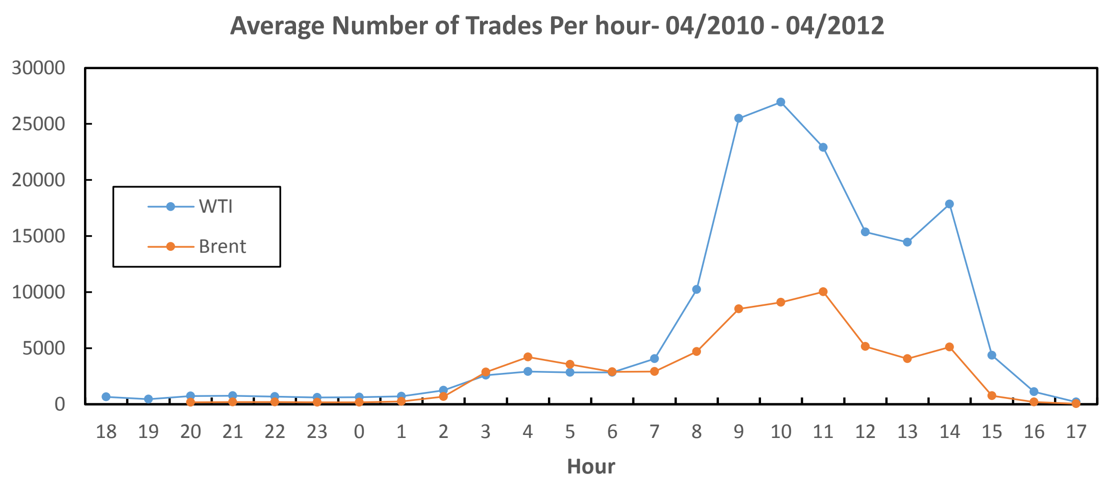


Figure 4: Information Share Across Time

The first plot shows the monthly average of the upper bound, lower bound, and the averages of the upper and lower bound of information shares for both WTI and Brent during the period of 01/2007 to 4/2012. The second plot shows the time series of the averages of the upper and lower bounds of information shares of both WTI and Brent for the same period. The reds lines are the 50% line for information share.

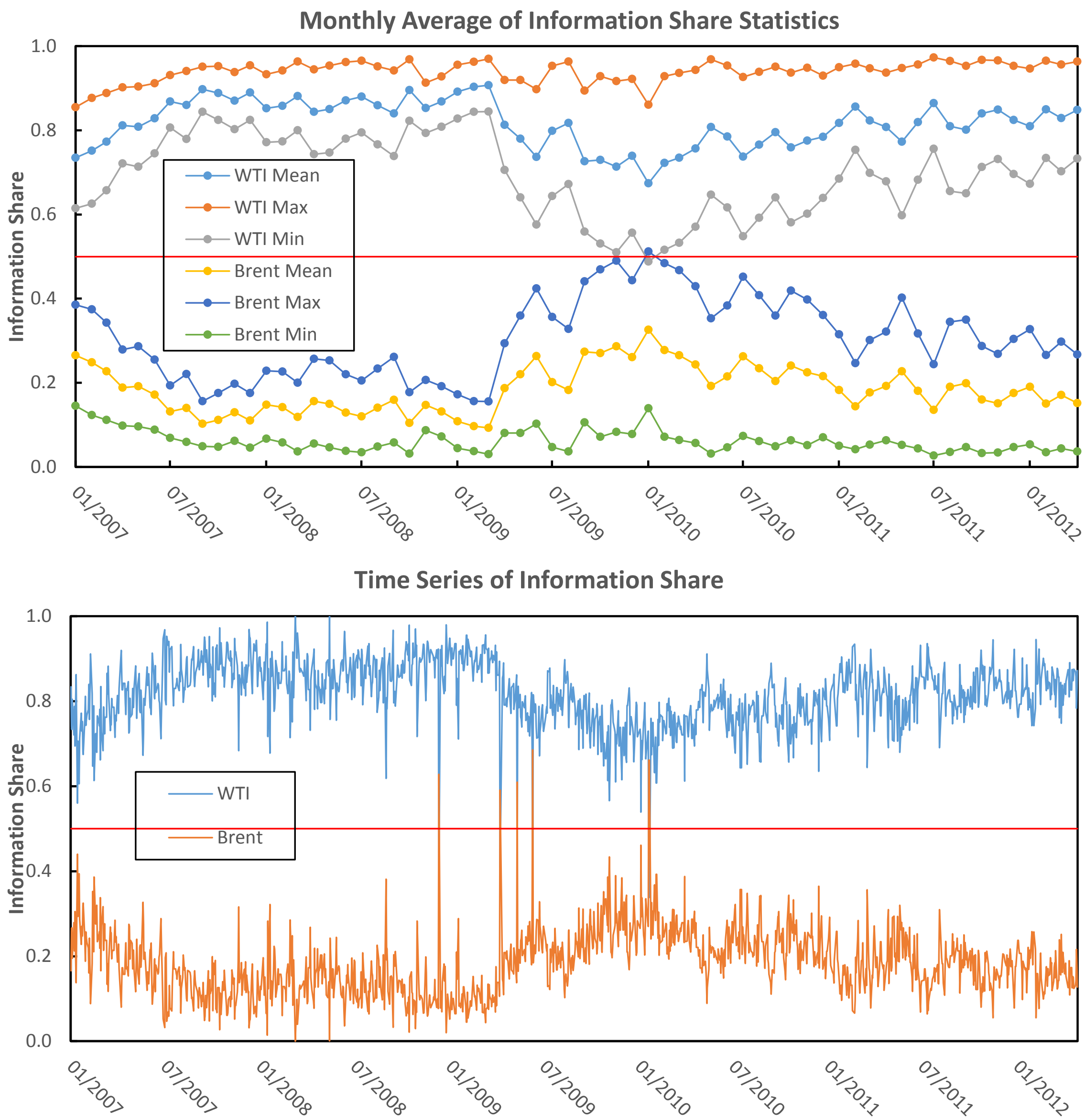


Figure 5: Cushing Inventory and WTI Information Share

This figure shows monthly average of WTI information share mean and the monthly Cushing crude oil inventory level for the period of 1/2007 to 4/2012. The red line is the 50% line for information share.