The effects of market sentiment on crude oil futures markets

Nathan Somayaji

University of Richmond

Follow this and additional works at: https://scholarship.richmond.edu/honors-theses

Part of the Economics Commons

Recommended Citation
https://scholarship.richmond.edu/honors-theses/1055
Abstract

This paper uses a behavioral finance approach to examine the effect of psychological factors on pricing in futures markets. Specifically, I assess the impact of the contemporaneous market sentiment on price discovery in crude oil futures markets. A considerable amount of previous research has shown that futures prices in crude oil markets lead spot prices, as futures act as a mechanism for determining spot prices. My analysis addresses whether the lead-lag relationship between futures and spot varies with differing market sentiment. I hypothesize that futures pricing will lead in times of increased uncertainty due to lower transactions costs and greater business decision making flexibility relative to futures. Based on the NBER recession classifications, I categorize historical time periods in terms of the two categories of Optimism and Pessimism, using Granger Causality tests to determine the price discovery in crude oil markets. The findings suggest that futures lead spot prices during times of greater uncertainty and over the long-term.
Introduction

One of the most fundamental assumptions of economics states that when individuals and businesses are faced with decisions concerning scarcity, these economic agents act rationally. What if, however, this assumption does not hold? An increasing amount of evidence conveys that under certain conditions, economic agents do not behave according to expectations based on rational economic models. Researchers have begun to incorporate psychological factors into economic theories in order to account for these discrepancies. In their Prospect Theory, Kahneman and Tversky (1979) theorized that individuals treat losses asymmetrically to gains, with losses having a greater absolute impact on utility than gains, all else equal. Even though it may be the rational decision to participate in an economic transaction, this individual may act irrationally and avoid a beneficial net gain if the negative effect of the loss on utility is greater than the positive effect of the gain. This theory has been consistent with empirical studies conducted over the last thirty years, such as those by Hanging and Xun Yu (2008) and Seror (2008), and continues to be a strong alternative to the von Neumann and Morgenstern theory (1944) which does not factor in irrational behavior due to psychological factors.

The successes of behavioral models such as Prospect Theory show that purely rational models often do not paint the full picture of economic decision making. Often, we forget that it is humans that participate in these economic transactions, and that humans do not always act rationally. In crude oil markets, it is ultimately humans, influenced by both private and public information, that drive price discovery. Since humans drive price discovery, emotions and other psychological factors may filter through into our assessment of value. Specifically I look to determine if increased uncertainty of future scarcity and of the future economic landscape impact the price-discovery relationship between crude oil futures and spot prices.
Much of the literature has shown that futures dominate spot prices in price discovery in crude oil markets. I hypothesize that futures pricing will dominate the lead-lag relationship during times of increased uncertainty due to several factors. Futures contracts have lower transaction costs than that of spots and allow for more near-term flexibility as delivery is not immediate. If there is a shock that changes the underlying economic landscape such that the level of uncertainty is increased, businesses will look to lock in future rates for their future needs and thus this shock will lead to price changes in the futures market versus the spot, with arbitrageurs quickly eliminating any disparities.

Using Granger Causality tests under the Error Correction Model form, I test the futures-spot lead-lag relationship using the National Bureau of Economic Research recession dates as a proxy for periods of high levels of uncertainty. I believe that recessions serve as a good proxy for gauging times of increased market uncertainty because the negative implications on economic factors such as output and unemployment that are readily observable by participants in markets and the concurrent inability to gauge the trough. I find that in each of the past three recessions futures prices have led spot prices.

Literature Review

As exchange-traded derivative products gain prominence in financial markets, financial instruments such as futures, arrangements that facilitate the future exchange of a good or service under the terms of a standardized contract, have become more prevalent. Used for various purposes such as speculation and hedging, futures play an especially significant role in commodities, as oil, wheat and gold as well as many other major commodities are traded in futures markets.
The efficiency of futures markets as a mechanism for price discovery has been under a tremendous amount of scrutiny, as researchers continually attempt to discern the effect of the forward looking price, represented by the futures price, on the current spot price. Participants in commodity futures markets bet on the direction commodities will head, leading researchers to ask the following question: How accurately do forward price outlooks predict movements in spot commodity prices? Do futures prices, in effect, lead spot prices and how significant is the causal relationship if such a relationship does indeed exist?

Garbade and Silber (1983) studied several commodities including wheat, corn and orange juice concentrate. They found that in these markets, 75% of new information is first incorporated into futures prices and that price changes in these future markets leads to changes in spot prices. While Silvapulle and Moosa (1999) state a bi-directional relationship between crude futures and spot prices, they ultimately concur with Garbade and Silber (1983), indicating that the effect of futures prices has a stronger effect on price discovery. Silvapulle and Moosa (1999) state that the logic fueling the empirical evidence, which shows a causal relationship between futures and spot prices, is predicated on the lower transaction costs associated with futures than with spot purchases. This is due to several reasons including: the holder of a future does not receive physical delivery until the end of the contract, there are less initial outlays required to purchase a future, and purchasing a future takes less time than a spot. Subsequently, arbitrageurs and speculators who do not wish to hold the physical commodity and prefer greater liquidity purchase futures over spots, as well as hedgers with limited physical capacity (Silvapulle and Moosa, 1999).
A follow up study completed by Moosa (2002) shows that in the crude futures market, futures prices influence sixty percent of price discovery based on his model. Moosa’s findings are strengthened by a broader study of petroleum market futures conducted by Schwarz and Szakmary (1994), who found in their analysis that price discovery for sweet crude oil, heating oil #2, and unleaded gasoline, is driven by their respective future markets. Through their work, Schwarz and Szakmary refute a claim by Quan (1992) that crude oil futures do not significantly contribute to price discovery in crude oil markets. They convey that using better data and a better time-series than those used in Quan’s model allow one to demonstrate future prices leading spot prices. In terms of convergence of the spot price and futures price of crude oil, Schwarz and Szakmary (1994) and Moosa (2002) find that 31 and 34 percent, respectively, of the price differential disappears within the first day of price divergence. This level of price convergence is considered to be a fairly quick, and without any further shocks should be completed after a week (Moosa, 2002).

Cost of Carry Model

Schwarz and Szakmary (1994) introduce the following theoretical model. Under perfect market assumptions, futures and spot prices are said to be in partial equilibrium where the following condition is met,

\[ F_t = S_t(1+r)^{T-t} \]  

(1)

According to this equation, the futures price at time \( t \) is a function of the spot price, \( S_t \), and yield \( r \). This yield is the interest received as compensation to the seller of the barrel of oil for the deferred payment from the buyer and is dependent on the remaining time until maturity of the
contract, $T-t$. This premium diminishes as delivery approaches, where the spot price should converge to the ending month futures contract price with $T-t$ approaching zero and $(1+r)^{T-t}$ approaching one (Schwarz and Szakmary, 1994).

If perfect market conditions do not hold and pricing discriminations occur such that $F_t \leq S_t(1+r)^{T-t}$ or $F_t \geq S_t(1+r)^{T-t}$, then assuming the cointegration of these markets, this deviation should be readily corrected by arbitrage activity (Moose, 2002). This situation often creates a lead-lag relationship between two integrated markets as one or both of the prices must converge towards the other to restore equilibrium. If a consistent leader exists, then previous price movements in the leading time series will influence the current price of the lagging series (Gujarati, 2005). Stated another way, the lagged values of the leading time series influence the price of the lagging series some $k$ periods later where $k$ is the length of the lag. To determine if a lead-lag relationship exists between futures and spot prices, a Granger Causal VECM model is implemented.

**Empirical Model**

Granger (1988) shows that causal relationships between two time series $X$ and $Y$ can be tested by regressing the lagged values of $X$ and $Y$ on $X$, and $X$ and $Y$ on $Y$ at time $t$. If the lagged values of $X$ are found to be significant in explaining $Y$ at time $t$, then $X$ Granger causes $Y$. The same applies for time series $Y$. In order to apply Granger Causal tests for crude oil markets, it must be determined whether or not futures and spot prices exhibit stationarity. Since Granger (1988) states that the proper test for two cointegrated series is the Error Correction Model, I also test for cointegration between futures and spot prices. Since I want to look at how changes in the
futures time series lead to changes in the spot time series, I show that the first differences, $\Delta F_t$ and $\Delta S_t$ are stationary in order to apply the Granger ECM.

**A) Unit Root Test**

A unit root is said to exist if the series exhibits a random walk process and is therefore nonstationary. The following equation is an example of a pure random walk process.

$$ Y_t = p Y_{t-1} + u_t \quad (2) $$

This equation describes a system where the value of the variable, $Y_t$, depends on its lagged variable, $Y_{t-1}$, plus a completely unpredictable stochastic error term, $u_t$. If $p = 1$, then the system is said to have a unit root and the series is nonstationary as the variance of $Y_t$ grows as $t$ increases (Gujarati, 2005). Using the Dickey-Fuller method, we can test whether or not $p$ is equal to one, and thus whether or not the series is stationary, by taking the first difference of $Y_t$. Equation 3 shows that manipulating Eq. (2) by subtracting $Y_{t-1}$ from both sides yields

$$ Y_t - Y_{t-1} = p Y_{t-1} - Y_{t-1} + u_t \quad (3) $$

$$ \Delta Y_t = (p - 1) Y_{t-1} + u_t $$

$$ = g Y_{t-1} + u_t $$

An OLS regression is then be used to test $g$ to see if it is significantly different than zero. The null hypothesis is that $g$ is not significantly different than zero and the alternative is that it is significantly different than zero (Gujarati, 2005). If a single unit root exists, then the process is nonstationary and the first difference of this series is stationary because $\Delta Y_t = u_t$ and $u_t$ is
~N(0, \sigma^2). The example above is the case where there is no intercept, but the Dickey-Fuller test can be applied to random walks with drift as well as to systems where trend stationarity exists (Gujarati, 2005). In my paper I use DF tests to test both futures and spot prices for pure random walks, random walks with drift, and random walks with drift and trend. I further test the first differences for unit roots to ensure that no more than one unit root exists for both series. The critical values of the test are based on the Tau distribution and the cutoff point was determined to be the 5% level of significance in order to lessen the probability of type I error.

B) Cointegration

Two series are said to be cointegrated if a linear combination exists of the two variables such that this linear combination is stationary. If two series are found to be cointegrated, they are thought to share long-term trends and an equilibrium relationship which mitigates the stochastic differences between the two over time. Engle and Granger(1987) provide a simple test for cointegration in which the values of X regressed on Y are expressed in terms of the error term. This is demonstrated in the following equation:

\[ Y_t = \varphi_1 + \varphi_2 X_t + u_t \]  \hfill (4)

\[ u_t = \varphi_1 + \varphi_2 X_t - Y_t \]

Dickey-Fuller tests are then used to determine if the residual term is stationary. If the residual term is found to be stationary, then the two series are found to be cointegrated, with the cointegrating vector being the residual term (Gujarati, 2005). In this case, the Dickey-Fuller critical values must be adjusted because the error term tested is based on estimated values. The values, however, still follow the Tau distribution and the adjusted critical values are provided by
Engle and Granger (1987) with the cutoff remaining at the 5% level of significance order to lessen the probability of type I error.

C) Granger Causality with ECM Representation

Conducting a Granger test between two time series allows us to examine whether or not the lagged variables of each series is significant in explaining the other. As previously mentioned, both underlying series are assumed to be stationary under the application of the model (Granger, 1988). The general representation of this theorem is as follows:

\[ X_t = \alpha + \sum_{i=1}^{n} \phi_i X_{t-i} + \sum_{j=1}^{n} B_j Y_{t-j} + \epsilon_{1t} \]  

\[ Y_t = \alpha + \sum_{i=1}^{n} \lambda_i X_{t-i} + \sum_{j=1}^{n} \delta_j Y_{t-j} + \epsilon_{2t} \]  

Equation 5.1 states that the value of \( X_t \) is a function its own lagged values and the lagged values of \( Y \) with \( \epsilon_{1t} \) representing the variation in \( X \) unexplained by model. Eq (5.2) is similar and states the representation for \( Y_t \). Thus if the coefficients for the lagged variables of \( Y, B_j \), are significant in explaining the variation in \( X \), \( Y \) is said to Granger Cause \( X \). If the coefficients of the lagged variables of \( X, \lambda_i \), are found to be significant in explaining the variation in \( Y \), then \( X \) Granger Causes \( Y \). This significance of the inclusion of variables is determined by using an F-test for Joint Significance. The null hypothesis is that, and that the sum of the coefficients of the lagged terms of \( Y \) are not significantly different from zero when regressed on \( X \). If the F-value is found to be significant, then including the lagged variables of the competing series is significant to the model.
When two series are cointegrated, Granger (1988) has shown that the basic form should be adapted because there exists some vector that is a linear combination of the two series. The lagged value of this linear combination must be integrated into the framework of the model in order to account for the correcting effects of this vector in response to previously experienced short-term disequilibrium. Implementing this gives the following equation.

\[ X_t = \alpha + \psi u_{t-1} + \sum_{i=1}^{n} \phi_i X_{t-i} + \sum_{j=1}^{n} B_j Y_{t-j} + \varepsilon_{1t} \]  
\[ Y_t = \alpha + \psi u_{t-1} + \sum_{i=1}^{n} \lambda_i X_{t-i} + \sum_{j=1}^{n} \delta_j Y_{t-j} + \varepsilon_{2t} \]  

If it can be shown that \( F_t \) and \( S_t \) are cointegrated and that \( \Delta F_t \) and \( \Delta S_t \) are stationary, then this model can be extended to assess price leadership in crude oil markets by the following equation.

\[ \Delta F_t = \alpha + \psi u_{t-1} + \sum_{i=1}^{n} \phi_i \Delta F_{t-i} + \sum_{j=1}^{n} B_j \Delta S_{t-j} + \varepsilon_{1t} \]  
\[ \Delta S_t = \alpha + \psi u_{t-1} + \sum_{i=1}^{n} \lambda_i \Delta F_{t-i} + \sum_{j=1}^{n} \delta_j \Delta S_{t-j} + \varepsilon_{2t} \]  

The number of lagged terms was determined empirically to be two and was done in an attempt to minimize the autocorrelation in the error terms as well as maximize adjusted \( R^2 \). If values of \( B_j \) are found to be jointly significant and \( \lambda_i \) jointly insignificant, then spot prices are thought to be price leaders, and if values of \( \lambda_i \) are found to be jointly significant while values of \( B_j \) are found to be insignificant, than futures lead spot prices. The possibility also exists that both values of \( B_j \) and \( \lambda_i \) are found to be jointly significant, in which case there bi-directional price leadership (Gujarati, 2005). The hypothesis contends that values of \( \lambda_i \) will be jointly significant and values of \( B_j \) jointly insignificant during times of increased uncertainty, as gauged by recessions. Given
that the recession dating process is of a somewhat arbitrary nature, one month was added to the beginning and to the end of each recession to help account for slight misdating by the NBER. The NBER dates provided six distinct periods, three recessionary and three expansionary, that allowed for several points of comparison.

Data

The data for WTI futures and spot prices was compiled from the Energy Information Administration, a government association that monitors commodities prices, and dates back to January of 1986. The specific contract used in this study is Contract 1, which expires at the end of each month. These contracts are traded on the futures market until their expiration at the end of the month, at which point the next months contract serves as its replacement. The data for the spot prices represents the daily closing price of the immediately deliverable contract. Holidays and other non-trading days were omitted.

Empirical Results

A) Unit Root Tests

The unit root tests were conducted in order to determine if stationarity exists for futures and spot prices as well as their first differences. This test was administered against the three primary cases of random walk: no drift, drift, and drift with trend. The computed Tau statistics for the parameters and their level of significance are included in the following table.
Table 1: Critical Tau values for Unit Root DF tests

<table>
<thead>
<tr>
<th></th>
<th>No Drift</th>
<th>Drift</th>
<th>Drift and Trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ft</td>
<td>-0.71</td>
<td>-0.16</td>
<td>-2.59</td>
</tr>
<tr>
<td>St</td>
<td>-0.74</td>
<td>-1.66</td>
<td>-2.63</td>
</tr>
<tr>
<td>ΔFt</td>
<td>-80.41**</td>
<td>-80.42**</td>
<td>-80.42**</td>
</tr>
<tr>
<td>ΔSt</td>
<td>-80.32**</td>
<td>-80.33**</td>
<td>-80.33**</td>
</tr>
</tbody>
</table>

**Represents significance at the 1% level

As this tests for whether or not $g$ is significantly different than zero, the Tau statistics reveal that the null hypothesis cannot be rejected for futures and spot prices, but can be rejected for the first differences of these prices. In the case of futures and spot prices, the null hypothesis is accepted and the two series are both said to contain a unit root. In the case of the first differences, the alternative hypothesis is accepted and the first differences of both the futures and the spot prices are said to not contain a unit root. The unit root tests, therefore, indicate that the futures and spot prices are nonstationary, but that the first differences are stationary. This demonstrated stationarity allows us to implement the first differences into the Granger Causal ECM.

B) Cointegration

The test for cointegration was done by determining if residual values of spot prices regressed on futures prices are stationary. If these residuals are found to be stationary, then the these residuals estimate a cointegrating vector which represents the long-term equilibrium
The following table displays the critical Tau values calculated from the Engle-Granger test.

Table 2: Critical Tau values for Cointegration EG test

<table>
<thead>
<tr>
<th></th>
<th>No Drift</th>
<th>Drift</th>
<th>Drift and Trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ut</td>
<td>-43.49**</td>
<td>-43.49**</td>
<td>-43.53**</td>
</tr>
</tbody>
</table>

**Represents significance at the 1% level

The tests indicate that the two series, futures and spots, are cointegrated, as the residuals from spot prices regressed on futures prices is found to be stationary. In each of these tests, the null hypothesis that $g$ is not significantly different than zero is rejected in favor of the alternative. Because these two series show empirical signs of cointegration and are theoretically thought to be cointegrated, then given the stationarity of the first differences and the cointegrated nature of futures and spot prices, Granger Causality can now be demonstrated in the Error Correction Model framework.

C) Granger Causality with ECM

Granger (1988) shows that the proper way to conduct a Granger Causality test when two series are cointegrated is through the Error Correction Model framework. This framework was used to determine if changes in spot prices lead to changes in futures prices, if changes in futures prices lead to changes in spot prices, if the process is bi-directional, or if there is no Granger Causal relationship. This study looks to examine whether futures prices lead spot prices during
times of increased uncertainty, with these time periods categorized by utilizing the NBER recession dates. For the first and second expansionary periods, autocorrelation of the residuals was found to exist using the Breusch-Godfrey method. Since expansionary periods were not of main importance, however, the lag-length of two was used in these periods to maintain consistency.

The following table displays the categorized periods of expansion and recession and the F-values of the Granger Causality tests. The Granger Causal column indicates the price leader.

**Table 3: F-values for GC tests**

<table>
<thead>
<tr>
<th>Time Horizon</th>
<th>Expansion/Recession</th>
<th>Granger Relationship</th>
<th>ΔFt</th>
<th>ΔSt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 1986 - May 1990</td>
<td>Expansion</td>
<td>Bi-Directional</td>
<td>6.67**</td>
<td>8.66**</td>
</tr>
<tr>
<td>June 1990 - April 1991</td>
<td>Recession</td>
<td>Futures</td>
<td>4.29*</td>
<td>0.10</td>
</tr>
<tr>
<td>May 1991 - Jan 2001</td>
<td>Expansion</td>
<td>Spot</td>
<td>2.13</td>
<td>11.38**</td>
</tr>
<tr>
<td>Feb 2001 - Dec 2001</td>
<td>Recession</td>
<td>Futures</td>
<td>3.24*</td>
<td>2.65</td>
</tr>
<tr>
<td>Jan 2002 - Oct 2007</td>
<td>Expansion</td>
<td>Neither</td>
<td>1.77</td>
<td>1.62</td>
</tr>
<tr>
<td>Nov 2007 - Mar 2009</td>
<td>Recession</td>
<td>Futures</td>
<td>14.48**</td>
<td>2.41</td>
</tr>
</tbody>
</table>

**Represents Significance at the 1% level  *Represents Significance at the 5% level**

The results confirm the hypothesis that futures prices lead spot prices during times of increased uncertainty, as categorized by the recession dates. The futures Granger F-value was significant at the 5% level for the June 1990 – May 1991 (F = 4.29) and February 2001 – December 2001 (F = 3.24) recessionary periods. The most recent recessionary period of November 2007- March 2009 (F = 14.48), which represents the most current data, showed the strongest signs of futures prices leading spot prices and was significant at the 1% level.
According to the results of the Granger tests, past changes in futures price lead to changes in spot prices. If we are to believe that in times of increased uncertainty businesses look to lock in future rates for their future needs and thus participate in the futures market, then if spot and futures markets are cointegrated any price discrepancies created from increased activity in futures markets will be readily eliminated. The spot market plausibly plays a role in this mitigation.

No lead-lag relationship was hypothesized for expansionary periods. An important aspect of the hypothesis regarding recessionary periods is the sense of urgency in decision making. If economic growth is more prevalent than contraction, it is assumed that the macro sense of urgency is not present. Although no Granger Causality was hypothesized for expansionary periods, it should be noted that each expansionary period transitioned from a non-futures dominant relationship into a futures dominated relationship. This contributes to the validation of the hypothesis, that futures prices lead spot prices during times of greater uncertainty.

**Conclusion**

The empirical results validate the hypothesis that futures prices lead spot prices during times of greater economic uncertainty. While the tests used for this analysis are relatively sophisticated, this study could be improved if more robust methods such as the Johansson and Augmented Dickey-Fuller tests were used to examine cointegration and stationarity, respectively. There are also discrepancies in determining the lag length for Granger tests, as the number of included lags can have an extremely large impact on the significance of the inclusion of the variable. For the Granger tests used in this study, the lag length was chosen to be two by an Adjusted $R^2$ criterion while minimizing autocorrelation in the error terms. This was also
consistent with Schwarz and Szakmary (1994), who found that price discrepancies between futures and spot prices are resolved quickly within a matter of a couple days.

It is also possible that recession dates do not purely capture changes sentiment concerning uncertainty. Extensions of this study might look to improve this proxy by integrating a continuous variable such as the put-call ratio to account for sentiment. Recession dates do have the advantage, however, of creating discrete time periods, whereas categorizing time frames based on a continuous variable may prove challenging due to large variations in intraday data. Given that the three recessions from 1986-2009 used in the analysis all showed futures price leadership indicates that recessions might truly capture sentiment provided the hypothesis is true.

These findings, that futures prices lead spot prices during recessions and times of increased uncertainty, can be leveraged in several ways. If a run-up on crude oil occurs during a recession, for instance, public officials might better understand which market to investigate and implement policies. In terms of forecasting, understanding which market leads and which lags under various conditions might improve the quality of the forecasting.

If the results of this study accurately reflect the underlying circumstances, then a significant amount of insight has been achieved about how perceptions and mentality, both of which can be subjective and irrational, influence the behavior and price discovery of markets. This demonstrates the importance of the human element in market interactions.
Bibliography


