Modelling and Measuring Price Discovery on the NYMEX and IPE Crude Oil Markets

by

Isabel Figuerola-Ferretti[A]

and

Jesús Gonzalo[B]

May 2008

The current surge of soaring crude oil prices has renewed interest on the information content of competing markets. In this paper we adopt the methodology developed by Figuerola and Gonzalo (2007) to model and measure price discovery on the NYMEX and IPE crude oil markets. We present an equilibrium model with finite elasticity of arbitrage services mainly arising due to transport cost and grade differences in the underlying asset. This leads on to price discovery measures that are solely determined by the relative liquidity in the two closely related markets. Applied to three month delivery prices, our empirical estimates demonstrate that NYMEX and IPE prices are cointegrated with a unit cointegrating vector, implying that there is a long run arbitrage equilibrium relation underlying both markets. Estimation of the price discovery measures indicates that there is a preponderant adjustment of the IPE towards NYMEX. This is consistent with figures for volumes traded and rationalizes the standard interpretation of NYMEX as the leading crude oil market.

JEL classification: G13, G14, C11, C22

Keywords: Commodity Futures, Crude Oil Market, Price Discovery, Permanent Transitory Decomposition.

[A] Business Department, Universidad Carlos III de Madrid, Spain
[B] Economics Department, Universidad Carlos III de Madrid, Spain

1. Introduction:

The subject of this paper is modelling price discovery in to closely related markets. Recent surge of bulling prices in the crude oil market has renewed interest in the information content of the two main setting markets: London International Petroleum Exchange (IPE) and New York Mercantile Exchange (NYMEX). Te former trades Brent Blend which for North Sea tanker-delivered oil. The later trades West Texas Intermediate or "light sweet crude" for delivery in the US eastern seaboard. Prices for the two contracts move together relatively closely, but transportation costs and grade differences imply the existence of a variable differential, and potential restrictions for arbitrage.[1] Two interesting questions arise: how does arbitrage ensure adjustment to the long-run path given location and grade differences? Which of the two markets is the market leader, or the most important contributor to price discovery? Both issues are important in evaluation of the risk profile of agents and institutions that engage in arbitrage in the two markets.

To answer these questions, we re-examine the general equilibrium model introduced by Figuerola-Gonzalo 2006 (FG thereafter) to measure price discovery. Price discovery is the process of uncovering an asset's full information or permanent value. The unobservable permanent component reflects the underlying fundamentals. It differs from the observable price which measures the permanent or efficient price with some transitory error. Whereas FG considered price discovery between spot and future prices traded in the same exchange, in this paper we quantify price discovery between crude oil contracts with same maturity but traded on different exchanges. Price discovery in US equity markets been considered by Hasbrouck (1995) and Harris (1997). See also Baillie et al (2002), Harris et al (2002) and Sapp (2002). Most of this studies use high frequency equities data on NYSE and regional exchanges to determine where price discovery occurs. The consensus is that price discovery concentrates in markets with highest market share (by trading volume). FG are able to justify this finding in a general equilibrium model that relates the degree of price discovery to the number of players in each market. In this paper we reconsider their framework to quantify price discovery in the NYMEX and IPE markets. Whereas in FG

[1] Issues of trading synchornization are important. Trading hours for NYMEX in the open outcry sesion are 9:45 to 15:10. There is also an electronic trading system starting at 16:00 on Monday through Thrusday and concluding at 8:00 the following day. On Sundays electronic trading starts at 19:00. All times are in EST. We consider NYMEX three month closing prices. The IPE starts trading at 10:02 GMT (05.02 EST) and closes at 20:13 GMT (15.13 EST). We consider IPE closing prices established at the end of each trading date.

finite elasticity of arbitrage services is explained by the existence of convenience yields, in this paper we focus on *arbitrage risk* arising from grade differences and transport cost. Other studies of the relationship between NYMEX and IPE prices include Brunetti and Gilbert (2000), Lin and Tamvakis (2004). The former looks at fractional cointegration of the two volatility processes. The later applies the ACD model using tick by tick data. Both studies support the view that NYMEX is the more important oil futures market. However non of these approaches quantify price discovery and theoretically relate the metric to relative volumes traded. This is the main contribution of the present paper.

Our findings may be summarized as follows

a) Both NYMEX and IPE prices are cointegrated with a unit cointegrating vector, confirming the existence of a long run equilibrium between the two markets.
b) As the model of FG predicts, price discovery takes place at NYMEX, the market that concentrates greater liquidity.

The paper is organized as follows. Section 2 reconsiders the model of FG to describe the equilibrium model with finite elasticity of supply of arbitrage services incorporating the existence of arbitrage risk. It shows that under this specification price discovery will depend on the number of players in each market. Section 3 presents empirical estimates of the model developed in section 2. It tests for cointegration and the presence of unit cointegrating vectors, estimates the participation of the NYMEX and IPE prices in the price discovery process and tests the hypothesis of one of the markets being the sole contributor to price discovery. Section 4 concludes. Graphs are collected in the appendix.

2. Theoretical Framework: a model for price discovery in closely related markets

The goal of this section is to characterize NYMEX and IPE price relationships re-examining the framework developed by FG.

2.1. Equilibrium Prices with finitely elastic supply of arbitrage services

Let F_{at} be the natural logarithm of the future market price in market a of a commodity in period t and let F_{bt} be the natural log of the contemporaneous price of future contract for that commodity in market *b*. To facilitate interpretation markets *a* and *b* may be considered to be the NYMEX and IPE markets respectively. In order to find the non-arbitrage equilibrium condition the following set of standard assumptions apply:

- (a.1) No taxes or transaction costs
- (a.2) No limitations on borrowing
- (a.3) No arbitrage risk other than grade and transport cost risk
- (a.4) No limitations on short sale of the commodity in both markets
- (a.5) Arbtirage risk is determined by $r_t = b + I(0)$ where b is the mean of r_t and I(0) is an stationary process with mean zero and finite positive variance.[2]
- (a.6) F_{at} and F_{bt} are I(1) processes

If $r_{t,}$ is arbitrage risk, by the above assumptions a1-a4, non-arbitrage equilibrium conditions imply

$$F_{at} = F_{bt} + b + I(0)\,, \qquad (1)$$

From (a.5) and (a.6), equation (1) implies that F_{at} and F_{bt} are cointegrated with the standard cointegrating relation (1, -1).

.

To describe the interaction between future prices in markets *a* and *b* we must first specify the behavior of agents in the marketplace. There are N_a participants market *a* and N_b participants in market *b*. Let $E_{i,t}$ be the endowment of the i^{th} participant inmediately prior to period t and R_{it} the reservation price at which the that participant is willing to hold the endowment $E_{i,t}$., then the demand schedule of the i[th] participant in market *a* in period t is

$$E_{i,t} - h\left(F_{at} - R_{i,t}\right), \quad h > 0, \quad i = 1,....., N_s \qquad (4)$$

where ? is the elasticity of demand, assumed to be the same for all participants.

The aggregate cash market demand schedule of arbitrageurs in period t is

$$H\left((F_{bt} + b\ ) - F_{at}\right), H > 0 \qquad (5)$$

[2] Note that this assumption is consistent with the interest rate being deterministic (see FG).

Where H is the elasticity of market *a* demand by arbitrageurs. It is finite when the arbitrage transactions of buying in market *a* and selling in market *b* or vice versa are not risk less. Market *a* will clear at the value of F_{at} that solves

$$\sum_{i=1}^{N_a} E_{i,t} = \sum_{i=1}^{N_a} \{E_{i,t} - h(F_{at} - R_{i,t})\} + H\left((F_{bt} + b) - F_{at}\right), \tag{6}$$

Market *b* will clear at the value of F_{at} and F_{bt} such that

$$\sum_{i=1}^{N_F} E_{i,t} = \sum_{i=1}^{N_F} \{E_{i,t} - h(F_{bt} - R_{i,t})\} - H\left((F_{bt} + b) - F_{at}\right), \tag{7}$$

Equations (9) and (10) can be solved for F_{at} and F_{bt} as a function of the mean reservation price participants in market $a\left(R_t^a = N^{-1}{}_a \sum_{i=1}^{Na} R_{i,t}\right)$ and the mean reservation price for participants in market $b\left(R_t^b = N^{-1}{}_b \sum_{j=1}^{N_b} R_{j,t}\right)$

And solving for market conditions we get

$$F_{at} = \frac{(hN_a + H)N_a R_t^a + HN_b R_t^F + HN_b b}{(H + hN_s)N_F + HN_S}, \tag{8}$$

$$F_{bt} = \frac{HN_a R_t^a + (H + AN_a)N_b R_t^b - HN_a b}{(H + hN_a)N_b + HN_b}, \tag{9}$$

Following FG, dynamic price relationships at the mean reservation price are derived as,

$$\begin{aligned} R^a{}_t &= F_{a,t-1} + v_t + w^a{}_t, i = 1,...., N_a \\ R^b{}_t &= F_{b\,t-1} + v_t + w^b{}_t, j = 1,...., N_b \end{aligned} \tag{10}$$

where , $\quad w_t^a = \frac{\sum_{i=1}^{Na} w_{i,t}^a}{N_S}, \; w_t^b = \frac{\sum_{i=1}^{N_b} w_{i,t}^b}{N_F}$

Substituting expressions in (10) into (8-9) yields

$$\begin{pmatrix} F_{at} \\ F_{bt} \end{pmatrix} = \frac{Hb}{d}\begin{pmatrix} N_a \\ -N_b \end{pmatrix} + (M)\begin{pmatrix} F_{at-1} \\ F_{bt-1} \end{pmatrix} + \begin{pmatrix} u_t^a \\ u_t^b \end{pmatrix}, \qquad (11)$$

where

$$\begin{pmatrix} u_t^a \\ u_t^b \end{pmatrix} = M\begin{pmatrix} v_t + w_t^a \\ v_t + w_t^b \end{pmatrix},$$

$$M = \frac{1}{d}\begin{pmatrix} N_a(H + hN_b) & HN_b \\ HN_a & (H + AN_a)N_b \end{pmatrix},$$

and

$$d = (H + AN_a)N_b + HN_a,$$

Expression (11) can be rewritten as a Vector Error Correction Model (VECM) by subtracting $(F_{a,t-1}, F_{b,t-1})'$ from both sides,

$$\begin{pmatrix} \Delta F_{at} \\ \Delta F_{bt} \end{pmatrix} = \frac{Hb_3}{d}\begin{pmatrix} N_b \\ -N_a \end{pmatrix} + (M - I)\begin{pmatrix} S_{t-1} \\ F_{t-1} \end{pmatrix} + \begin{pmatrix} u_t^a \\ u_t^b \end{pmatrix}, \qquad (12)$$

with

$$M - I = \frac{1}{d}\begin{pmatrix} -HN_b & HN_b \\ HN_a & -HN_a \end{pmatrix},$$

Rearranging terms

$$\begin{bmatrix} \Delta F_{at} \\ \Delta F_{bt} \end{bmatrix} = \frac{H}{d}\begin{bmatrix} -N_b \\ N_a \end{bmatrix}\begin{bmatrix} 1 & -1 & -b \end{bmatrix}\begin{bmatrix} F_{a,t-1} \\ F_{b,t-1} \\ 1 \end{bmatrix} + \begin{bmatrix} u_t^a \\ u_t^b \end{bmatrix}, \qquad (13)$$

Applying the PT decomposition (described in the next section) to the VECM, the permanent component will be the linear combination *(F_{at}, F_{bt})* formed by the orthogonal vector (properly scaled) of the adjustment matrix ($-N_b$, N_a). In other words the permanent component is

$$\frac{N_a}{N_a + N_b} F_{at} + \frac{N_b}{N_a + N_b} F_{bt} \qquad (14)$$

This is the price discovery metric proposed by FG. Note that this metric does not depend on the value of the elasticities h and H (>0) or on the equilibrium relationship between F_{at} and F_{bt}. These elasticities do not affect the long-run equilibrium relationship, only the adjustment process and the error structure. For modelling purposes is important to notice that the long run equilibrium is determined by expression (1) , and it is the rest of the VECM (adjustment processes and error structure) that is affected by the different market assumptions on elasticities, participants, etc.

4. Empirical Price Discovery in non-ferrous metal markets

The data include daily observations of three month crude oil NYMEX prices and three month crude oil IPE prices. Prices are available from June 1988 to December 2008. The data source is Ecowin. Figure 1 in the appendix depict both data series. It suggests that NYMEX and IPE crude oil series are I(1) and that they are closely related.

The empirical analysis is based on the VECM (13) of the theoretical section. Lags of the vector $\begin{pmatrix} \Delta F_{a\,t} \\ \Delta F_{b\,t} \end{pmatrix}$ are added until the error term is a vector white noise.

The econometric details of the estimation and inference of (13) can be found in Johansen (1996), and Juselius (2006). The procedure to estimate $a_{\perp}$, and test the hypothesis on it are in Gonzalo and Granger (1995).

Results are presented in tables 1-4, following a sequential number of steps that correspond to those that we propose for the empirical analysis and measuring of price discovery.

A. Univariate unit root test:

Non of the Log-prices reject the null of a unit root. The results are available upon request.

B. Determination of the rank of cointegration

Before testing the rank of cointegration in the VECM specified in (13) two decisions are to be taken: (i) To choose the number of lags of $\begin{pmatrix} \Delta F_{a\,t} \\ \Delta F_{b\,t} \end{pmatrix}$ necessary to obtain white noise errors and, (ii) To decide how to model the deterministic elements in the VECM. For the former we use an information criteria (the AIC), and for the later we restrict the constant term to be inside the cointegrating relationship as the economic model in (13) suggests. The results for the trace test are presented in Table 1. Critical values are taken from Juselius (2006). NYMEX and IPE markets (F_{at} and F_{bt}) are clearly cointegrated as the model in section 2 predicts.

Table 1: Trace Cointegration rank test

	NYMEX-IPE
Trace test	
r =1 vs r=2 *(95% c.v=9.14)*	6.03
r = 0 vs r=2 *(95% c.v=20.16)*	172.05

C. Estimation of the VECM:

The VECM specified in (13) is a reduced rank model and its estimation is done following Johansen (1995). Estimation results are reported in table 2. Two points are worth mentioning: (i) the cointegrating relationship has a slope close to one, suggesting that there is a unit cointegrating vector between NYMEX and IPE crude oil prices. This is formally tested within the next step (ii) IPE prices react significatively to the equilibrium error and NYMEX doesn't, suggesting that NYMEX is the market leader.

Table 2: Estimation of the VECM specified in (13)

NYMEX (Fa) IPE (Fb)

$$\begin{bmatrix} \Delta F_{a\,t} \\ \Delta F_{b\,t} \end{bmatrix} = \begin{bmatrix} 0.01 \\ \mathit{(0.38)} \\ -0.04 \\ \mathit{(2.63)} \end{bmatrix} \left[\hat{z}_{t-1}\right] + k \; lags\, of \begin{bmatrix} \Delta S_{t-1} \\ \Delta F_{t-1} \end{bmatrix} + \begin{bmatrix} \hat{u}_t^S \\ \hat{u}_t^F \end{bmatrix}$$

with $\hat{z}_t = S_t - 0.98F_t + 0.17$ and $\boldsymbol{k}$(AIC)=20

Note: p values are given in parenthesis.

D. Hypothesis testing on beta

Results reported in table 3 show that, the standard cointegrating vector (1, -1) is accepted at the 10% significance level. This suggests that arbitrage activities ensure deviations from the long run path of two prices are adjusted and equilibrium prices are restored.

Table 3: Hypothesis testing on the cointegrating vector

	NYMEX-IPE
Cointegrating vector	
$(\beta_1, -\beta_2, -\beta_3)$	
β_1	1.000
β_2	-0.98
SE (β_2)	0.002
β_3 (*constant term*)	-0.17
SE (β_3)	(*0.015)*
Hypothesis testing	
H_0:β_2=1 vs H_1:β_2>1 *(p-value)*	(0.060)

E. Estimation of $\alpha_\perp$ and hypothesis testing on $\alpha_\perp'$=(0,1)

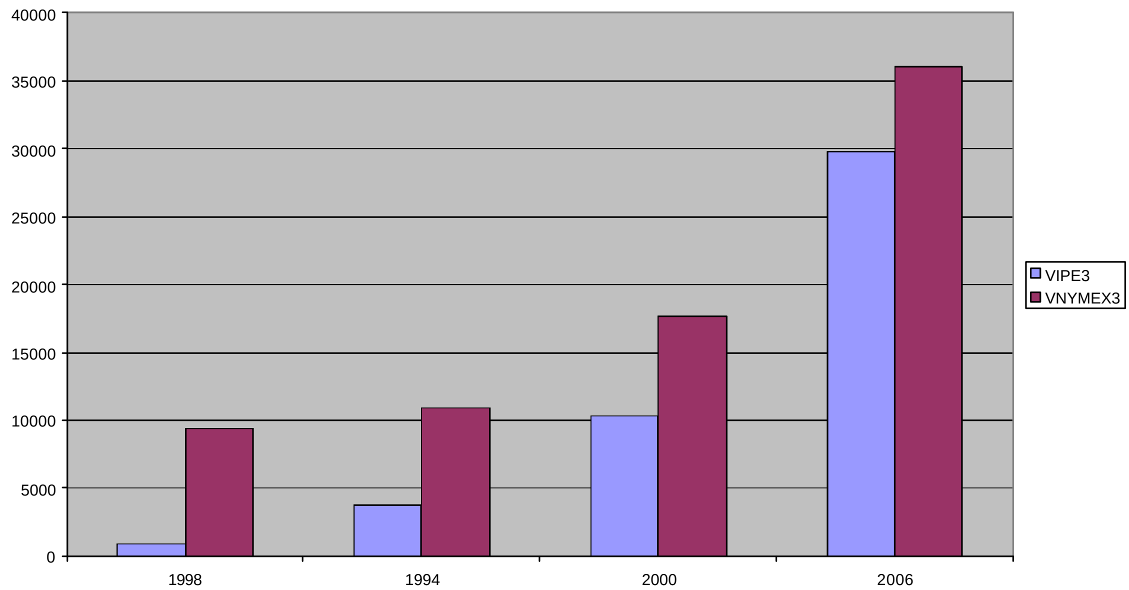
Table 4 shows the contribution of IPE and NYMEX prices to discovery function. The estimated parameter of $\alpha_\perp$ suggests that the NYMEX price is the main contributor to price discovery. In fact the hypothesis that the NYMEX price is the sole contributor to price discovery $\alpha_\perp'$=(1,0) fails to be rejected and the hypothesis that the IPE price is the sole contributor to price discovery $\alpha_\perp'$= (0,1) is rejected.

Table 4: Proportion of NYMEX and IPE prices in the Price Discovery Function

Estimation	**NYMEX-IPE**
$\alpha_{a\perp}$	0.68
$\alpha_{b\perp}$	0.32
Hypothesis testing *(p-values)*	
H_0: $\alpha_\perp'$=(0,1)	(*0.00*)
H_0: $\alpha_\perp'$=(1,0)	(*0.14*)

As the theoretical model suggests, price discovery metrics are directly related to the number of players which may be approximated by the volumes traded in each market. This is confirmed by Figure 1, which shows that NYMEX volumes for the three month delivery contract have been higher than those for the IPE three month over the whole sample period.

Figure1: Volumes traded in the three month delivery crude oil IPE and NYMEX contracts. Sample 1988-2008



4. Conclusions and Extensions

The process of price discovery is crucial for all participants in the currently bulling crude oil market where prices have recently increased by almost $25 within a five day period. The present paper re-examines the work of FG to measure price discovery in two major crude oil futures markets, which trade similarly defined futures contracts on closely related but non-identical underlying assets. The underlying assumption is arbitrage between these markets is restricted by *arbitrage risk* arising from grade and location differences. The model leads to price discovery metrics that are determined by the relative number of participants in each market. Applied to three month futures data we find that i) NYMEX and IPE prices are cointegrated with a unit cointegrating vector ii) NYMEX is the main contributor to price discovery in the crude oil markets, consistent with figures relative volumes traded in each market. This quantifies the standard interpretation of NYMEX as the leading crude oil future market.

Given that we model closely related markets to which unit cointegrating vectors may be imposed, we can construct a non linear or asymmetric Error Correction Model and obtain within such framework a “non-linear price discovery mechanisim”. This will allow us to determine where the price discovery in being produced when NYMEX>IPE and when IPE>NYMEX.

References:

Baillie R., G. Goffrey, Y. Tse, and T. Zabobina (2002). "Price discovery and common factor models." *Journal of Financial Markets*, 5, 309-325.

Brunetti C. Gilbert C.L (2000). "Bivariate FIGARCH and fractional cointegration." *Journal of Empirical Finance*, 7, 509-530

Figuerola-Ferretti and Gonzalo (2007). " Modelling and Measuring Price discovery in Commodity Markets." *074510 Working paper series*. Business Department, Universidad Carlos III de Madrid.

Garbade, K. D. & Silver W. L. (1983). "Price movements and price discovery in futures and cash markets." *Review of Economics and Statistics*, 65, 289-297.

Gonzalo, J. Granger C. W. J (1995). "Estimation of common long-memory components in cointegrated systems." *Journal of Business and Economic Statistics*,13, 27-36

Johansen (1996). "The cointegrated VAR Model, Methodology and Applications". *Oxford University Press*: Oxford.

Johansen, S, and Juselius, K (1990). "Maximum Likelihood Estimation and Inference Cointegration- With applications to the Dema nd For Money." *Oxford Bulletin of Ecoomics and Statistics*, 52, 169-210.

Juselius (2006)." The cointegrated VAR Model, Methodology and Applications". *Oxford University Press*. Oxford.

Lin S. Tamvakis M. (2004). "Effect of NYMEX trading on IPE Brent Crude futures markets: a duration analysis". *Energy Policy*. 32, 77-82

Harris F.H., T.H. McInish, and R.A. Wood (1997)."Common long-memory components of intraday stock prices: A measure of price discovery." *Wake Forest University Working Paper*.

Harris F. H., T.H. McInish, and R.A. Wood (2002). "Security price adjustment across exchanges: an investigation of common factor components for Dow stocks." *Journal of Financial Markets*, 5, 277-308.

Hasbrouck, J. (1995). “One security, many markets: Determining the contributions to price discovery.” *Journal of Finance*, 50, 1175-1199.

Graphical Appendix

Figure 1: NYMEX and IPE crude oil three month delivery prices