CRUDE OIL PRICE FORECASTING WITH ANFIS

B. Zimberg

Administración Nacional de Combustibles, Alcohol y Portland ANCAP

Humboldt 3900, 11900, Montevideo, Uruguay

ABSTRACT

Crude oil pricing is commonly expressed as a formula referenced to Brent or WTI crude oil. The final price of these two qualities and the spread between WTI and Brent can drive the decision when the purchase of a crude oil cargo is evaluated. A crude oil price-forecasting model is presented. It is based on past data, inventory level and volatility index and it is derived with a neuro fuzzy inference system. The fuzzy model allows the visualization and analysis of the set of rules that govern the prediction. Results are compared with the prediction based on an econometric model.

Keywords: Forecasting, Crude oil price

1. INTRODUCTION

Generally, crude oil is priced in the period around the process of crude oil loading . This situation could take place two monthes after the evaluation of the crude purchasing. In some cases the decision can be correct or not depending on how different are these final prices with respect to the initial guess. In other cases the decision depends on the spread WTI – Brent, more predictable than the crude oil benchmarks.

It is important to estimate the value of these two price references because they define the final cost of the cargo. The WTI and Brent benchmarks are published daily by the Platts services [1] [2].

This work refers specifically to the following two benchmarks: first line WTI (West Texas Intermediate) crude oil spot price and Brent DTD (Brent Dated) crude oil spot price.

2. THE PROBLEM. THE MODEL

The problem under consideration is the prediction of the value of two benchmarks in the next period, based on information available for a previous period. Changes in the value of the benchmarks with time comes as the result of a set of events. The set of events, can be associated with input variables in a model. The model must be simplified because it is not possible to follow all the variables involved in the real problem. The model will take into account only the variables with major impact in the prediction.

The notation assumes the suffix +1 for a variable in the next period, a 0 for the present period and the suffix –1 for the previous period.

Based on daily data available for the period 1991-2003, a fifteeen days average was calculated for each benchmark This was the result of a compromise between the error of the approach and the number of parameters required. In the rest of the text, WTI and Brent refer to their averages.

The first variable to consider is the period (1-24). This variable can be correlated with a seasonal behaviour.

The Relative Inventory (IR), is defined as the difference between the inventory level of crude and products with respect to a base level in a certain period. The inventory level is the result of the balance between supply and demand:

$$\text{Inventory}(+1) = \text{Inventory}(0) - \text{Demand}(+1) + \text{Supply}(+1) \tag{1}$$

Inventory increases if supply exceeds demand. In occassions, crude oil production is reduced and the inventory decreases at fixed demand. Additionaly, industrial development is associated with an increase in the demand of petroleum, unless the efficiency of energy consumption increases.

Due to the evolution of supply and consumption over time, there is an expected Inventory level called Inventory* . Inventory* is assumed to follow a linear trend with time.

$$\text{IR} = \text{Inventory} - \text{Inventory*} \tag{2}$$

Reference [3] shows a correlation between WTI and IR for the OECD countries (Organization for Economic Cooperation and Development).

Commonly, crude oil price decreases when IR increases and when IR declines then prices increase. The IR variable is continuous by definition.. The EIA (Energy Information Administration) publishes monthly values of the inventory levels of crude oil and derivatives in the OECD and the estimations for the next quarter [4] [5].

In a first approach, monthly values of IR are calculated based on the evolution of Inventory* [3] and the Inventory levels. The final expresion of IR for each fifteen days period, is a discrete variable in the interval [-1,1]. This approach shows to be useful in order to represent the lack of information in a forecast (Figure 1).
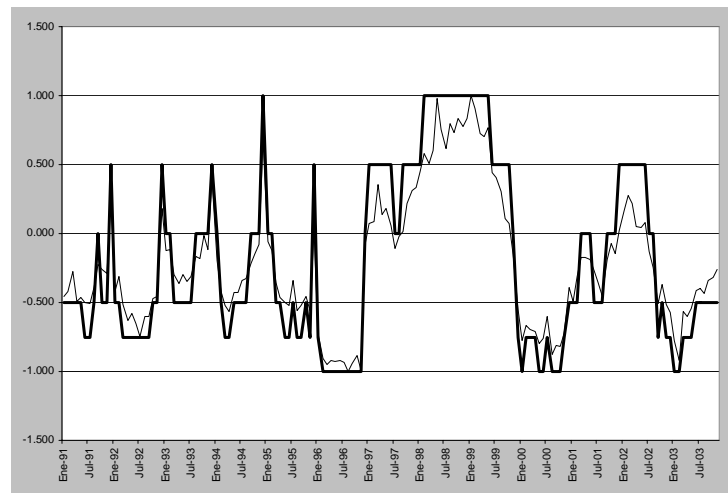

Figure 1. The variable IR, continuous and discrete.

The Volatility Index, another variable in the model, measures the contribution of the instability of the market. Prices can raise or decrease quickly from one period to the next. Real events or the speculation associated with the probability of its occurrence, can have a strong impact on the price of the crude oil benchmark. Such effects can be expressed as a variable called R.

R(+1) = -1. If Crude(+1)-Crude(0) < -2 U$S/bbl.   (3)

R(+1) = +1. If Crude(+1)-Crude(0) > +2 U$S/bbl.

R(+1) = 0. Otherwise.

Figure 2 shows, as small circles, the points where R is not zero (Brent data series).
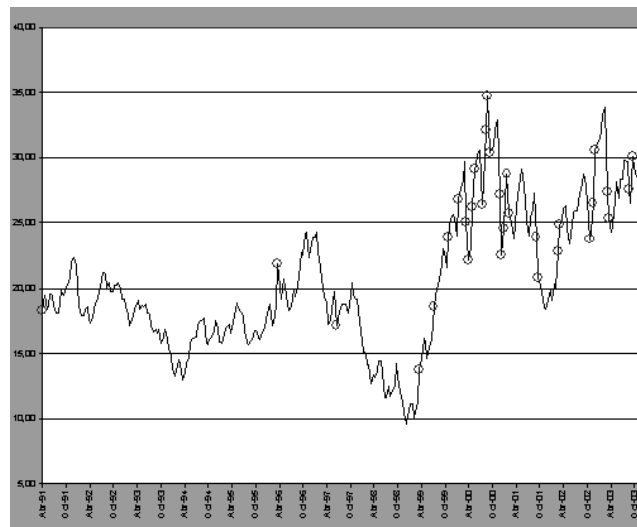

Figure 2. The variable R.

With reference to prices in different periods, for the Brent or WTI series, the correlation between the price evaluated in (+1) and the price evaluated in (0), (-1) or (-2) was greater than 0.85.

Based on this results, the proposal of input variables for the prediction of Brent(+1) is the following:

- Period(0), Period(+1).
- Brent(-2), Brent(-1), Brent(0).
- IR(0), IR(+1).
- R(+1).

In order to estimate the price in the next period and due to the complexity and non linearity of the problem, a neuro-fuzzy approach was considered.

The neuro-fuzzy approach proved to be better than the neuro models in terms of lower mean square error (MSE) of the training process and lower error distribution. The Feedforward network was tried first, followed by an Elman neural network, but the ANFIS (Adaptive Neuro-Fuzzy Inference System) approach was finally selected [6].

3. IMPLEMENTATION AND RESULTS

The problem was implemented in Matlab [7] . The Cross Validation technique showed that the number of points in the data series was enough for the training process. Before the training process, all data was normalized. Cross validation was applied in order to improve the forecast of the network The data set was divided at random into two subsets, one for training and another for testing, in a relation of 2 to 1. Subtractive Clustering provided a reduction in the number of rules [8].

Regarding the detection of outliers in the Brent and WTI series, three tests were applied, each of them based on an interval of acceptance: a range of three standard deviations   from the mean or similar expressions. In addition, the Grub test (extreme studentized deviate) did not detect outliers in the series [9].

Different combinations of variables were tested on a trial an error basis and the MSE results were compared. The variables Period(0), Period(+1), Brent(-2) did not contribute substantially to the reduction of the training error and were rejected. IR(+1) seemed to be marginally more appropiate than IR(0). R(+1) contributed significantly.

The model can be expressed as a function F1:

$$Brent(+1) = F1(Brent(-1), Brent(0), IR(+1), R(+1)) \tag{4}$$

Results from the problem ANFIS- Brent showed a MSE of 0.60 or in an equivalent way, a root mean square error value (RMSE) of 0.8 U$S/bbl. Subtractive Clustering provided the system ANFIS-Brent with only 6 rules for the 4 input variables (instead of 1296) and 6 membership functions. Figure 3 presents the distribution of the error after training; 68% of the values of Brent(+1) fall in the range [-0.5, 0.5]. In other words, 68% of the estimations of the model differ from its actual value by a maximum of 0.5 U$S/bbl.

The ANFIS-WTI model is similar:

$$WTI(+1) = F2(WTI(-1), WTI(0), IR(+1), R(+1)) \tag{5}$$

RMSE= 0.8 U$S/bbl.

Figure 4 depicts the output of the model outlined in reference [3] . Figure 5 shows the results of model ANFIS-WTI for the same period. ANFIS-WTI improves the results.
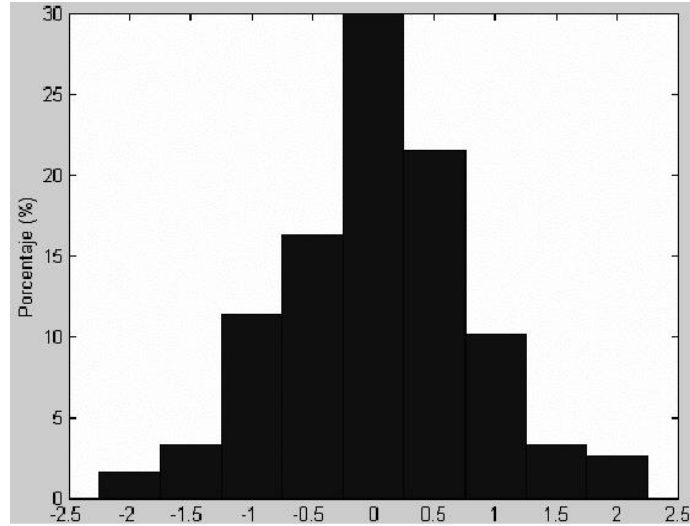

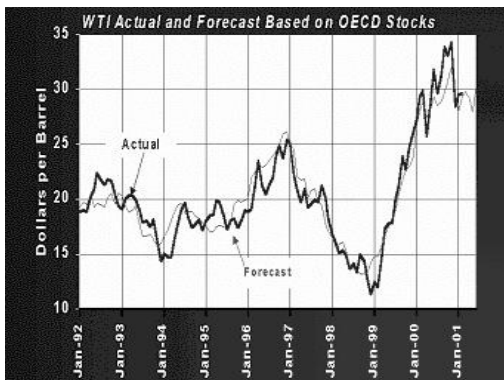**Figure 3. ANFIS-Brent. Error distribution.**
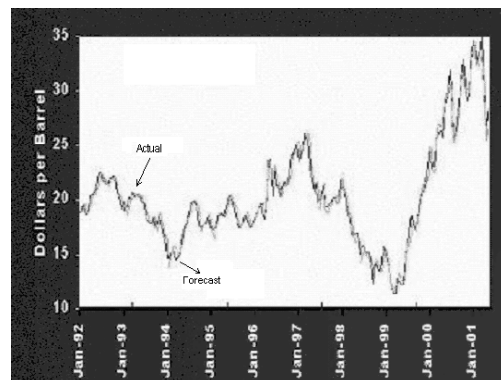

Figure 4. ANFIS-WTI. Output [3]

Figure 5. ANFIS-WTI. output.

The model for the spread WTI-Brent is the following:

Diff(+1) = F3(Diff(-1), Diff(0), IR(+1), Brent(+1))   (6)

In this case RMSE= 0.55 U$S/bbl and 83% of the values of Diff(+1) are in the range [-0.5, 0.5]. The error distribution is narrower than the distribution depicted in Figure 3.

One of the advantages of ANFIS is that it provides a set of rules that can be helpful in the understanding of the problem under study. It follows an analysis of the rules from the model ANFIS-Brent. Figure 6 shows the 4 variables of the model, each of them with 6 membership functions. The variables are normalized.
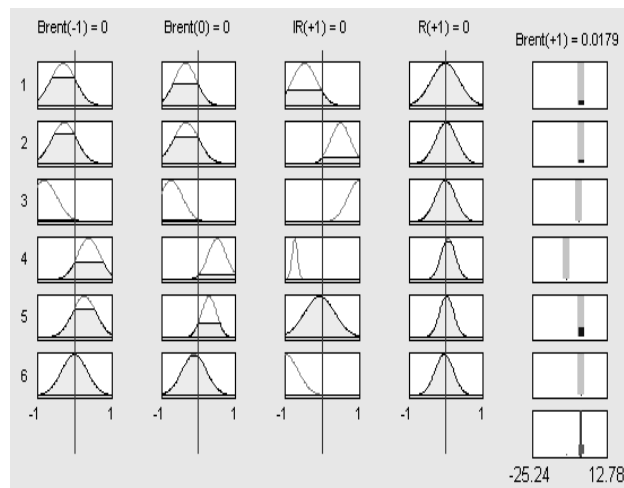

Figure 6. ANFIS-Brent. Fuzzy rules.

For rules 1 to 3 ,variables Brent(-1) and Brent(0) change in a low to medium range. For rules 4 to 6 ,variables Brent(-1) and Brent(0) vary in a medium to high range.

Some of the membership functions of IR(+1) are very narrow; this means that IR(+1) or the inventory level affect the evolution of Brent specially when inventory levels are to low or to high.

R(+1) is centered around zero. As a result, R(+1) contributes in most of the cases to the final value of Brent(+1).

In terms of fuzzy reasoning, the first three rules trigger low levels of Brent(+1) whereas high levels of Brent(+1) are inferred from the last three rules. Medium levels of Brent(+1) result from a combination of the rules. For instance, if Brent(-1) and Brent(0) are low then Brent(+1) is low (R(+1) = 0, IR(0) = 0).

But these rules also show that if Brent(-1) and Brent(0) are very low and IR increases then Brent(+1) increases.

4. CONCLUSIONS

A crude oil price forecasting model has been presented for crudes Brent and WTI. It is based on a neuro-fuzzy inference system. A model for the spread WTI-Brent has been obtained based on a similar approach. Four input variables were considered: R(+1), Brent(0), IR(+1) and Brent(-1). The model takes into account the effect of the inventory level, past and present values of the price benchmarks and the volatility of the market.

Variables R and IR are discontinuous in order to represent the lack of information during the forecast. The model ANFIS-WTI has shown a better approach than the model showed in reference [3], at least for the period 1992-2000.

The model ANFIS-Diff shows the lower error distribution and corresponds with the expected evolution of the time series.

These models can be usefull for sensitivity analysis.

In a next step, the model will be extended to consider the period 2004-2007.

BIBLIOGRAPHY

(1) *Platts Crude Oil Marketwire*. 1996 to 2003. *Platts Oil Price Handbook & Oilmanac*.**68 to 72**

(2) M.Ye, J. Zyren and J. Shore, Forecasting Crude Oil Spot Price Using OECD, Petroleum Inventory Levels. *Int'l Advances in Econ.Res*. **8(4)**, 2002.

(3) Petroleum Stocks in OECD Countries. Monthly Energy Review *http://www.eia.doe.gov*

(4) International Petroleum Supply and Demand. Short-Term Energy Outlook. *http://www.eia.doe.gov*

(5) B. Zimberg and N.Chiaro. Indicadores de Precios de Crudo. Post Graduate Course on Neuro-Fuzzy Systems. *UDELAR. Final Report.* 2003

(6) Matlab.version 6.1.0.450. *The Mathworks, Inc*, 2001

(7) J. Jang, C. Tsai Sun and E. Mizutani. Neuro – fuzzy and Soft Computing, a computational approach to learning and machine intelligence. *Prentice Hall*, 1997

(8) V.Barnett and T.Lewis .Outliers in Statistical Data. *Wiley*, 1994.