Proceedings of the 2001 IEEE Congress on Evolutionary Computation Seoul, Korea • May 27-30, 2001

Compumetric Forecasting of Crude Oil Prices

M. A. Kaboudan
MS&IS, Penn State
8380 Mohr Lane, Fogelsville, PA 18051
mak7@psu.edu

Abstract- This paper contains short term monthly forecasts of crude oil prices using compumetric methods. Compumetric forecasting methods are ones that use computers to identify the underlying model that produces the forecast. Typically, forecasting models are designed or specified by humans rather than machines. Compumetric methods are applied to determine whether models they provide produce reliable forecasts. Forecasts produced by two compumetric methods – genetic programming and artificial neural networks – are compared and evaluated relative to a random walk type of prediction. The results suggest that genetic programming has advantage over random walk predictions while the neural network forecast proved inferior.

1 Introduction

Crude oil price (COP) is a globally important variable for which accurate forecasts are needed by decision-makers and planners in business as well as government. Price forecasts are used in projections of energy supply and demand. During this time in history when dependency on oil is practically universal and crude oil prices are near their peak one would expect COP to be the most statistically analyzed and forecasted variable. Yet a literature search finds other variables such as sunspot numbers, exchange rates, prices of precious metals, and even weather forecasting receiving much more attention especially from academic forecasters when compared with practitioners. Examples of academic forecasting of COP can be found in Sedriks (1998), MacDonald and Marsh (1993), Helkie (1991), and Okogu (1990). Sedriks approach is based on two key concepts. The first is that the underlying price trend is set by marginal producer's costs, which typically fall along an experience curve. The second concept is that deviations from the trend and cycling around the trend line are due to expectations about shortages, for which a representative function typically can be derived in terms of a supply-demand balance parameter. Serdriks' model provides projections over 15 years. MacDonald and Marsh examine the efficiency of oil price forecasts. Helkie conducts sensitivity analysis to determine the impact of oil market disruptions on COP. Okogu examines the nature of the relationship between official and spot prices during the latest weak market phase using quarterly data. Users of COP forecasts seem to be dependent on a handful of sources that provide scenarios of periodic COP outlooks. Perhaps ranking top among those sources are forecasts by DOE, DRI, WEFA, and GRI. The Energy Information Administration (EIA) of the U.S. Energy Department (DOE) developed the Short-Term Integrated Forecasting (STIFS) model to generate short-term (up to 8 quarters), monthly forecasts of U.S. supplies, demands, imports, stocks, and prices of various forms of energy. In their energy forecasting model (EIA, 2000) COP is exogenously determined and subjectively projected on the basis of factors that influence it such as new oil fields coming into production (as well as rates of decline in existing fields), regional economic growth rates and planned shifts in production from major producing nations. In short thier COP forecast is qualitative and is reached by analyzing the world petroleum demand and supply balance over forecast period. (EIA's study was designed to forecast prices of petroleum products and not COP.) DRI is Standard and Poor's Data Resources Inc., WEFA is Wharton Econometric Forecasting Associates, and GRI is the Gas Research Institute of the US. DRI and WEFA are two competing outfits with professional teams of forecasters who employ huge econometric models to produce COP forecasts. GRI produce their own in-house forecast using a combination of methods that are mainly statistically driven. Dependency on these traditional forecasters of COP dates back to the 1970s when oil prices first started to escalate more rapidly than ever before and when modeling and forecasting techniques were not sufficiently sophisticated to tackle the complexity of COP dynamics. If this is the case, it may be beneficial to investigate whether more recent techniques that rely less on humanly designed models and more on computational powers can provide better forecasts of COP. Such techniques are appropriately described here as compumetric forecasting methods. This paper will evaluate forecasts produced by two competing compumetric forecasting methods: genetic programming (GP) and artificial neural networks (ANN). Descriptions of these methods are available elsewhere. (See for example Koza, 1992, for an introduction to GP, Langdon, 2001, for all GP publications, and Principe et al., 2000 on ANN). This paper focuses more on the possible use of these techniques in producing one-step-ahead forecasts using monthly data. It also provides an objective comparison of such techniques.

More on these methods is found below in Section 2. The data used to produce forecasts is described in Section 3. Fits and forecasts and their comparison with a naïve random walk monthly predictions are in Section 4. The final Section contains concluding remarks.

2 Compumetric Forecasting

When forecasting any variable for many analysts concern is mainly with producing the most accurate forecasts first while identifying factors responsible for variations in the variable of interest if possible second. Ideally one would like to accomplish both. However, realistically the ideal situation is rarely possible and the opportunity cost of obtaining accurate forecasts is usually identifying the explaining factors. Alternatively, measuring the effect of factors impacting the dynamics of a variable may result in less accurate forecasts. If one is able to identify models that explain variations but forecast poorly, it is easy to argue that such models are of little value and for all practical purposes may include incorrect specifications of any hypothesized relationships. That said it is easy to see why it is more important to obtain a model difficult to interpret and explain that forecasts well than the reverse. Support for this argument is enhanced by the fact that our human skills and abilities to comprehend processes remain rather limited.

The two competing compumetric techniques used in this study offer forecasts produced by models that may be difficult to explain or sometimes justify. But the techniques are quickly gaining popularity because they are capable of producing forecasts that compete aggressively with conventional statistical methods. GP produces strange and lengthy specifications of models that have been shown to forecast rather accurately. An explanation or interpretation of models GP evolves is for the most impossible. They may include variables that can logically and intuitively be considered as reasonable and justifyable in causing variations in the dependent variable. However, the final functional form of the evolved model is too complex to explain. ANN produces accurate forecasts without providing information about the underlying model. One has some control in selecting a neural network structure, but the final specification is hidden in a black box. Whether forecasts by one method or the other are superior is the concern here. The forecasts are judged relative to each other and relative to the naïve random walk forecast.

Oil prices follow cyclical patterns over time. They tend to escalate for an extended period, reverse direction then perhaps escalate again. Periodicity is not constant and variations within an escalating or a decreasing period are typical. In this preliminary investigation, a lag structure of up to nine months is assumed. Three forecasts are produced assuming that time series type models are capable of producing the forecasts. In time series models, no assumptions are made about any explanatory variables and no distinction is made between periods when prices are rising and others when prices are falling. They are based on the assumption that past history of the dependent variable contains sufficient information to predict the future from it. More sophisticated models that include other than lagged dependent variables or modeling periods of increasing and decreasing prices separately are left for future research.

3 The Data

All data for this study were obtained from two sources: the US Department of Energy and Erogmagic.com: Economic Time Series. The variable to forecast is crude oil FOB price in cents/barrel. The series used start January 1993 and taken at monthly closing price. Data prior to 1993 contains outliers due to the effect of the Gulf War on COP. During the Gulf War, prices rose sharply for a few months then dipped to their levels prior to the War shortly after it ended. Attempting to include data for that period produced totally unreliable models and forecasts. An attempt was also made to use data prior to the War and after the War thus excluding the War effect. The resulting models and forecasts were also unreliable.

It was also necessary to decide on whether to use nominal or real (deflated) prices. The decision was made against adjusting for inflation. Nominal prices are used for two reasons. First, crude oil prices fell between 1985 and the Gulf War when they suddenly spiked then fell again. Second, deflating oil prices when using monthly data is of little value since decisions are short term in nature and inflation has little to no impact. Further, using constant prices aggrevates forecast inaccuracies due to errors in forecasting the applicable price index. Because of these problems there may be some advantages in forecasting nominal instead of constant prices.

The data to fit ends December 1998 (or a total of 60 months are used to evolve GP models and train ANN models). Data of 1993 are used to account for lags (up to 12 months for these time series models) while data for 1999 are used to compare *ex post* forecasts (or forecasts beyond data used to train or fit models with). The dependent variable in the time series models is: Y_t or FOB Crude Oil Price of US imports measured in cents/barrel. The independent variables were determined through trial and error. Initially, 18 variables were considered to evolve a GP model to forecast with. They included different lags (from 1-3 periods) of what seemed to be reasonable explanatory variables. More specifically, these variables were monthly world crude production, OECD consumption, world crude stocks, monthly change in known US stocks, and lagged FOB crude oil price of US imports. All variables except for lagged prices were not helpful in evolving an acceptable equation. The decision was then made to include only lagged price variables. Different lag-lengths were tested to find that best equation to forecast with. The final independent variables that produced the best possible forecasts were: Y_{t-1}, Y_{t-2}, …, Y_{t-12}. The same experiment was conducted to determine if ANN methodology will utilize explanatory variables other than lagged prices. The outcome led to the same conclusion where only lagged prices were helpful in producing meaningful fits and forecasts. For consistency in comparison, identical data was employed to evolve the GP models and obtain the ANN forecasts.

4 Fits and Forecasts

Three fits and forecasts are compared here: the first one is produced by best GP evolved model, the second is produced by the best ANN model trained and tested, and the third is a random walk forecast that acts as a point of reference. To evolve the GP models, the number of generations used was 220 while population sizes were 1,200. The operators included were +, -, *, (protected) /, sin, cos, and (protected) sqrt. Other parameters used were as follows: maximum expression = 80, mutation rate = 60 %, cross-self rate = 20 %, tournament size = 8, kill tournament = 3, and maximum age of an individual = 2000. The equation below is the best obtained in 100 runs. Ephemeral constants between 127 and –127 were utilized.

Here is the best resulting GP equation found after it was converted to infix notation and modified for easy readership:

$$Y_t = X_{t-1} + A1 + X_{t-9} + (X_{t-8} / X_{t-1}) + A2 + \cos(X_{t-5})/\sqrt{X_{t-1}} + A3$$

where

$$A1 = \cos[\, (X_{t-11}-X_{t-1}-\cos(X_{t-5})- (X_{t-1}/(X_{t-1}-X_{t-7}) + \cos(X_{t-12}-X_{t-5}) / X_{t-12}/(30*X_{t-8}) + X_{t-8}$$

$$A2 = \cos(X_{t-4}) / \sqrt{} (X_{t-10}-3*X_{t-2}+2*X_{t-7}+X_{t-9})$$

$$A3 = \sin(X_{t-4}) / \sqrt{} [\cos(X_{t-6}) -\{X_{t-9}/\sin(\sqrt{} (X_{t-10}-X_{t-4}+X_{t-8}))\}]$$

The final ANN model selected to produce a forecast is a multilayer perceptron (MLP) type that is a layered feedforward network typically trained with static backpropagation. Its main advantage is that it is easy to use, and that it can approximate many input maps. (More than a dozen other run specifications were tested before selecting this configuration that produced the best forecast.) Only one hidden layer is used. The parameters for a layer of processing elements (PEs) are as follows:

> Input and output layers PE transfer function is SigmoidAxon, learning rule = momentum, step size = 0.1, momentum = 0.9, with a maximum number of epochs for supervised learning = 10,000 and number of epochs for learning = 15,000.

A comparison of the forecasting performance of the two methods with the naïve random walk forecast shows that GP produces the best forecasting model. Fitted values from the evolved GP equation and the trained ANN compared with the random walk (RW) fit produced the following mean square errors (MSE):

MSE from the GP = 0.24
MSE from the ANN = 1.29
MSE from random walk fit = 0.91

Plots of the resulting fits are presented in Figure 1. ANN's performance was rather disappointing. The fit in Figure 1 is the best out of fourteen different ones attempted using different run specifications with different transfer functions as well as different learning and training epochs. In all runs including the one reported here, ANN's performance was especially worst toward the end of the training period. As evident from Figure 1, ANN's plot of fitted prices in 1998 were higher than the actual, GP fit, or their random walk values.

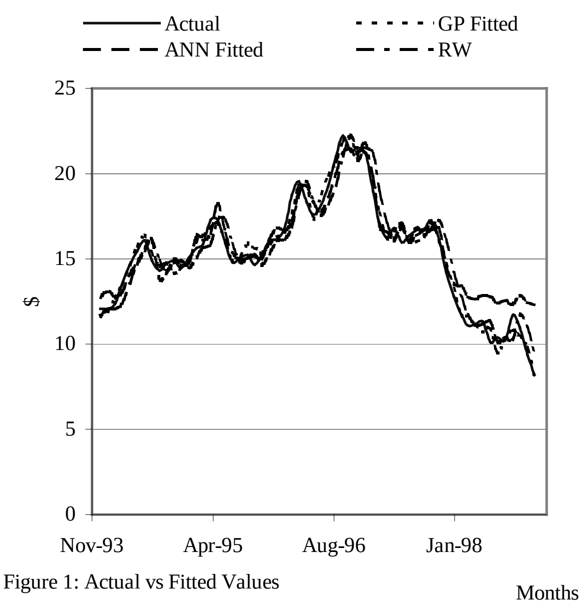

Figure 1: Actual vs Fitted Values

The resulting forecasts are in Figures 2 and 3. Figure 2 compares GP and RW forecasts while Figure 3 compares GP and ANN forecasts. A complete comparison of the forecasts for 1999 using the three methods is in Table 1. In the Table Y represents actual COP, GP is the forecast the genetic program produces, ANN is the forecast by the neural network model, and RW is the random walk prediction. ANN's forecast badly overestimated the first few periods. Both GP and the RW forecast slightly under estimated the first two periods.

One-month-ahead forecasts for a period of one year (12 months of 1999) produced the following results:

MSE from the GP model forecast= 1.85
MSE from the ANN forecast = 3.54
MSE from the random walk forecast = 2.29

The difference between the MSE of GP and RW forecasts is rather large thus eliminating worries about GP depicting RW behavior. To determine whether such difference is significant the differences between individaul points of the two forecasts were computed. The differences between each of the twelve points forecasted had a mean = 0.66 with a standard error = 0.24. The null hypothesis that the mean of the differences was equal to zero was rejected at the .02 level of significance with a t-statistic = 2.75 (= 0.66/0.24) (For further explanation on the t test for the mean difference used when comparing dependent or related samples, see Brenson and Levine, 1999). The t-statistic is computed as follows:

$$t = \text{D-bar} / (S_d/\sqrt{n})$$

where

$$\text{D-bar} = \Sigma D_t/n$$

$$S_d = \sqrt{[(\Sigma(D_t - \text{D-bar})^2/(n-1)]}$$

and where t = 1, …, n = 12 months, and D = (GP forecast in time period t – RW forecast in the same time period).

Table 1: **Actual versus the three forecasts**

		Forecasts		
	Y	**GP**	**ANN**	**RW**
Jan-99	9.17	7.31	12.20	8.18
Feb-99	9.34	9.43	12.38	9.17
Mar-99	11.83	9.93	12.26	9.34
Apr-99	14.14	11.61	13.46	11.83
May-99	14.43	14.30	14.66	14.14
Jun-99	15.13	16.49	14.63	14.43
Jul-99	17.30	16.86	15.74	15.13
Aug-99	19.10	18.46	17.12	17.30
Sep-99	21.04	19.72	17.92	19.10
Oct-99	20.89	21.94	20.56	21.04
Nov-99	22.46	20.95	20.22	20.89
Dec-99	22.91	23.92	21.53	22.46

The forecasting ability advantage GP has can be further supported by comparing the number of periods in which each of the forecasts produces an absolute error > 1.65 (or 10% of the mean price in 1999). The number of months out of the tweleve forecasted in which GP forecast exceeded that 10% is three compared with five periods for the ANN and RW forecasts.

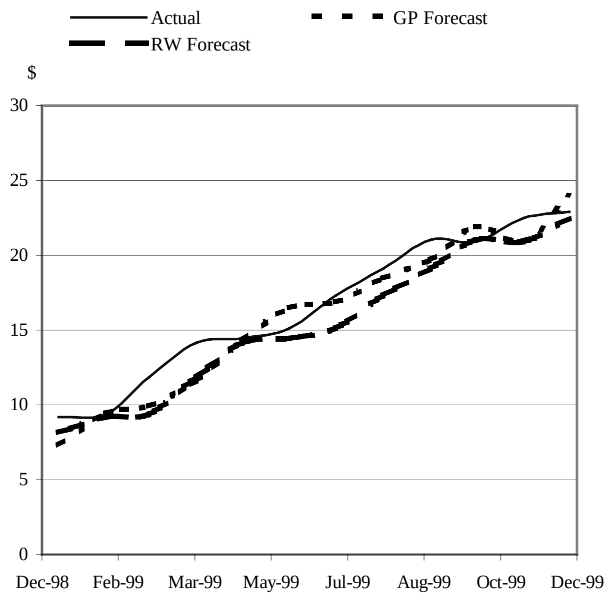

Figure 2: **GP vs RW Forecasts**

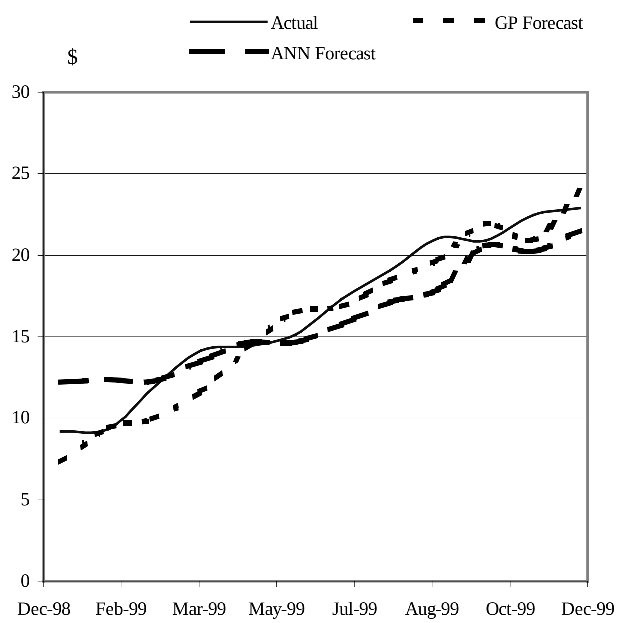

Figure 3: **GP vs ANN Forecasts**

5 Conclusion

This paper presented a preliminary first attempt to forecast crude oil prices using two compumetric techniques GP and ANN. The exercise was limited to producing one-month-ahead price forecasts. Attempts to produce forecasts for more than one-month-ahead suggested the need for further more serious analyses. Such analyses were not possible within the time available to complete this paper. Investigation of more complex models continues and results will be available in future work. Attempting to evolve such models that may produce more useful forecasts proved to be intricate and demanding further understanding of the dynamics of the world oil markets quickly. The dynamics of crude oil prices are simply too complex to forecast without a much more investment of time and resources.

The results presented in this paper show that GP is in fact capable of producing rather impressive one-month-ahead forecasts. This work invites more investigations in many directions. First, it invites investigating the possibility of producing more complex models that incorporate variables other than just lagged prices. Second, it invites investigating the changes in model structures over time. One can only hypothesize here that for a different time period, the dynamics of the market will be different. A model to forecast prices in 2000 for example is expected to be significantly different than one evolved to forecast 1999. This is so because while attempting to evolve the GP model reported here, the data used initially started in 1988. However, the number of periods used was reduced several times (each time by one year) to reach the model reported. The effect of the Gulf War on months prior to 1993 resulted in poor models. The evolved models improved only after deleting those years from the analysis. The effect of the War was a sudden increase in oil prices for a few months followed by a sudden drop in them when the War ended. Excluding the affected months from the analysis is not a reasonable solution since the effects of the exogenous disruption that occurred cannot be isolated while including monthly data of prior and later years. The decision was made to exclude data for all years prior to 1993 in this study. The situation is different for the year 2000 when prices are reaching new highs. Market dynamics are responsible for such increases and they are not the result of a war or even the threat of one. Therefore, a different model must be evolved to capture the effect of such change in dynamics.

Further, ideally when the dependent variable (price of oil here) investigated reveals cyclical behavior, it is prudent to divide the data into two groups: one representing periods when prices are expected to increase and the other when prices are expected to decrease. Expectations are determined statistically using historical correlations between different lags that help deciding which future periods will be experiencing price incerases and which experiencing price decreases. One then fits two models instead of one. The idea is based on the notion that forces causing pricses to increase are different from those causing prices to decrease. Examples of such forces may include changes in exchange rates, changes in supply, changes in stocks, prevailing world economic conditions, OPEC agreements, cheating by OPEC members, among others. Naturally some forces may be common in periods of increasing and of decreasing prices. Implementing this idea is only possible when the number of data points available to model is sufficiently large (or more than 80 according to prior experimentation). This was not possible to implement in this study because data prior to 1993 was dicarded as explained earlier. Implementing the two-model technique will be possible in a future study.

Bibliography

Sedriks, W., 1998. Projecting Crude-Oil Prices, *Chemtch*, 28, 47-53.

MacDonald, R., and Marsh, I., 1993. On the efficiency of oil price forecasts, *Applied Financial Economics*, 3, 293-292.

Helkie, W., 1991. The Impact of an Oil Market Disruption on the Price of Oil: A Sensitivity Analysis, *The Energy Journal*, 12, 105-117.

Okogu, B., 1990. Testing for Symmetry in OPEC Official Price Reaction to the Spot Price in Tight and Weak Market Conditions, *Applied Economics*, 22, 605-617.

EIA, 2000. Short-Term Energy Outlook Energy Prices Model Description, http://www.eia.doe.gov/emeu/steo/pub/document/textpr.html#WP57IUS.

Koza, J., 1992. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, Cambridge, MA: The MIT Press.

Langdon, B., 2001. http://liinwww.ira.uka.de/bibliography/Ai/genetic.programming.html.

Erogmagic.com: Economic Time Series, 2000. http://www.economagic.com.

Principe, J., Euliano, N., and Lefebvre, W., 2000. Neural and Adaptive Systems: ***Fundamentals Through Simulation***, John Wiley & Sons, New York.

US Department of Energy, 2000. http://www.eia.doe.gov/emeu/mer/petro.html.

Brenson, M. and Levine, D., 1999. *Basic Business Statistics: Concepts and Applications*, Seventh Edition. Pentice Hall, Upper Saddle River, New Jersey.