FORECASTING RETURN VOLATILITY OF CRUDE OIL FUTURE PRICES USING ARTIFICIAL NEURAL NETWORKS; BASED ON INTRA MARKETS VARIABLES AND FOCUS ON THE SPECULATION ACTIVITY

Authors
Hamed Shafiee Hasanabadi, Saqib khan, Rene V. Mayorga

Faculty of Engineering and Applied Science

University of Regina

May 15, 2014
Abstract

Considering the strong linkages between commodity and equity markets during the few past years, the motivation of the study in this Chapter is to forecast the crude oil future prices return volatilities of based on the information from the intra markets variables. According to the recent allegation that speculators participations affect the market trends and commodity prices, it is also necessary to analyze the speculation activity impact on the volatility prediction in this study. Speculation activity which is measured by workings “T” absorbed many attentions recently. The historical value of some explanatory variables other than the historical volatilities, which is normally used in the volatility forecasting models such as GARCH models, has been used to forecast the future volatilities of crude oil. The empirical data of light sweet crude oil future prices from the NEW YORK mercantile exchange market has been used to train and test the model. The results of the Radial Basis Function Network and the Feed Forward Back Propagation Network have been compared with each other and with the results from the GARCH model as well. The findings of this part of the study show that both Radial Basis Function Network and the Feed Forward Back Propagation Network are working better than the GARCH model in forecasting the crude oil future prices return volatilities and speculation plays an important role in forecasting volatility in absence of historical volatility. But when the historical volatilities are available it is observed that they have enough information of future and the information of the speculative activity is encompassed by the historical volatilities. The explicit contribution of this part of the study is the application of two types of the Artificial Neural Networks, including Feed Forward Back Propagation Network and Radial Basis Function Network in forecasting return volatilities of crude oil future prices using some intra markets variables especially with focus on the speculation activity.
1. Introduction

The impact of speculative activity in commodity markets has been a matter of controversy for quite some time. It has become even more contagious as the financialization of commodities has attracted even more interest to this matter, by both academics and practitioners. Especially the role of speculative activity in the run up of crude oil prices in 2008, followed by a sharp decline and again a reversal, has fuelled a renewed debate by market participants, regulators, and academics.

Speculative activity is not only blamed for affecting the price levels, but also volatility. However, a number of studies report mixed results in this regard. De Long et al. (1990) show that speculative activity can increase volatility. Tang and Xiong (2010) and Singleton (2011) find that speculative activity has a significant impact on both prices and volatility. Other studies find that speculative activity does not increase volatility. In fact Buyuksahin and Harris (2011) and Brunetti et al. (2010) suggest that speculative activity reduces volatility.

In this paper we adopt a different approach to study the relationship between speculative activity and volatility. We employ artificial neural networks (ANNs) to forecast crude oil returns volatility, using an information set of market variables for training. One of these variables is a measure of speculative activity. If the speculative activity did impact crude oil return volatility, we would expect that the information content of the speculative activity variable used in training the artificial neural networks, would improve the quality of forecast.

We use two different types of neural networks; feed forward back-propagation and radial basis neural networks. We find that using the selected market variables, including speculative activity, as information set to train the neural networks results in a forecasting performance that is better
than the GARCH model forecast. However, using the speculative activity variable does not improve the forecast by the neural networks that were trained by using the historical realized volatility only. This shows that the information content of the speculative activity variable is not significant in forecasting crude oil returns volatility, suggesting that the speculative activity does not significantly impact volatility.

The rest of the paper is organized as follows. Section I describes the data used. Section II discusses the methodology. Section III presents our results, and section IV provides a discussion of the results. The paper is concluded in section V.

2. Data

We use weekly crude oil futures price data from New York Mercantile Exchange (NYMEX) (US dollars per barrel) from January 4, 1995, to May 26, 2004. The data is divided into two periods for training and testing the ANNs. The subsample from January 1995 to February 2002 is used for training the ANNs and from March 20, 2002 to May 26, 2004, is used to evaluate the out-of-sample volatility forecasts. As is standard practice, the closest to maturity future price is considered as spot crude oil price. The data are obtained from the Energy Information Administration of the US Department of Energy. The crude oil future prices are plotted in Fig.1.
In order to calculate the volatility, first of all the weekly continuous return series is generated as follows:

$$r_t = \ln \left( \frac{P_t}{P_{t-1}} \right) \quad \text{(Eq.1)}$$

for $t = 1, 2, \ldots, T$

where, $r_t$ is the logarithmic return for crude oil from time $t-1$ to $t$ and $P_t$ is the price at time $t$.

The weekly continuous returns of crude oil are plotted in Fig.2.
The standard deviation of returns is used as the realized volatility (Daly, 2008). The rolling realized volatility return of crude oil future prices is calculated as the standard deviation of the crude oil weekly returns series for the last 6 month, as follows:

\[
V = \sqrt{\frac{\sum_{t=1}^{T} (r_t - \bar{r}_t)^2}{T-1}} \quad \text{(Eq. 2)}
\]

Where \(\bar{r}_t\) is the average of the previous six month crude oil weekly logarithmic return series.

Fig. 3 plots the weekly volatility series:

**Figure 3.** Crude oil volatility
Crude oil Futures prices weekly returns volatility (6 months volatility)

The crude oil implied volatility data was obtained from Bloomberg. S&P index, S&P value weighted return, S&P equal weighted return data was obtained from CRSP. The commitment of trader’s data was obtained from Commodity Futures Trading Commission (CFTC).

### 3. Methodology

First we forecast crude oil price volatility using the traditional GARCH model that is widely used by market participants and academics. Next we forecast the volatility using ANNs and a set of information variables. We also calculate forecasting errors from the two methodologies and compare them.
4. **GARCH model**

The Generalized Auto Regressive Conditionally Heteroscedastic (GARCH) model that was developed by Bollerslev (1986), has since become one of the most widely used techniques for modeling volatility. Hasanabadi et. Al. (2012) provide a discussion of the model and the estimation technique that we used.

5. **Artificial Neural Networks**

As a second approach to forecasting volatility, we use the artificial neural networks. We use two different types of neural networks, Feed Forward Back-Propagation and Radial Basis neural networks. A detailed discussion about these types of neural networks, and the forecasting process is provided in Hasanabadi et. Al. (2012).

6. **Developing forecasting models**

In order to use ANNs to forecast volatility, we first need to choose the input variables based on the information content. It is important to select an optimal number of input variables, as too many variables will clutter the system, and too few will reduce its forecasting ability. Based on the prior knowledge and understanding of the impact of different variables on the volatility, we select the input variables. The variables which we used for our forecasting model are as following:

**S&P 500 index:** S&P 500 is one of the most widely followed equity indices. It is considered to be a leading indicator for the American economy. As this index is an indicator of the US stock market, we include it to account for the increasing level of market integration pointed out in the
literature quoted earlier. In addition to the level of the index we also include the S&P equal weighted return and value weighed returns.

**Implied Volatility:** Implied volatility is a forward looking estimate of the volatility of underlying asset. Given the price of a traded option, it is possible to determine the volatility forecast over the lifetime of the option implied by the option’s price. We obtain the implied volatility data from Bloomberg.

**T-value:** The growth of speculative activity is calculated by an index called the Working’s (1960) “T”. T-value is a measure of the level of involvement of speculators in the futures market. We calculated the “T” based on the data for every Tuesday from 1995 to 2004 using the following formula, equation 3. (Bahattin Büyüksahin et al., 2011)

\[
T_{i,t} = \begin{cases} 
1 + \frac{SS}{HL_{i,t} + HS_{i,t}} & \text{if } HS_{i,t} \geq HL_{i,t} \\
1 + \frac{SL}{HL_{i,t} + HS_{i,t}} & \text{if } HL_{i,t} \geq HS_{i,t} 
\end{cases}
\]

(Eq.3)

SS stands for Speculators Short, SL for Speculators Long, HS for Hedgers Short and HL for Hedgers long. We obtained the position of traders’ data from the CFTC. The CFTC’s Division of Market Oversight collects information for each trader whose position exceeds a certain level, which is determined according to the specifications of each market. The CFTC also collects information from each large trader about the purpose of its positions in different U.S. futures markets (hedging or speculation). The CFTC data is gathered on a daily basis (Bahattin Büyüksahin et al., 2011). In our weekly analysis, we focus on the Tuesday reports.

We used this variable as an input to our neural network to incorporate the widely held view of market participants that speculative activity affects commodity price volatility.
The weekly data of following variables was considered as the candidate inputs to the model.

1. S&P value weighted return
2. S&P equal weighted return
3. S&P index
4. Weekly future price of crude oil
5. net T-value index on crude oil future contracts
6. Implied volatility of crude oil price
7. 6 month rolling volatility of logarithmic daily returns of S&P index
8. Average of last 6 months returns crude oil future price
9. 1 month Lagged T-value
10. Volatility of T value

And the model output is 6 month rolling volatility of logarithmic weekly returns of nearest to maturity crude oil price.

We used the statistical correlation between the target and the potential input variables as a measure for choosing among the candidate input variables. In order to choose the optimal group of variables from the pool of these 10 variables, we calculate the correlations of these variables with both nearest to maturity crude oil returns and the 6 month rolling realized volatility of logarithmic weekly returns of nearest to maturity crude oil price. Also we applied a regression between each of these variables and the crude oil return volatility. A variable is selected if one of the following conditions is satisfied:

- The summation of absolute correlation between that variable and each of the return volatility and future prices goes over 10 percent (McCormick, 1992).
- The regression coefficient is significant.
Since the 1 month lagged T-value, outperformed the contemporaneous T-value in regression analysis, we replaced it with T and dropped T-value.

As shown in Table I, 8 out of our 9 variables satisfy the above mentioned conditions and are considered as network inputs.

S&P equal weighted return did not fulfill the criteria and was dropped off from the input variables. The following table shows the value of the correlation between input variable and the realized volatility as well as the crude oil future price and the regression between input variable and the realized volatilities.

| Variables                                      | Corr. with future price | Corr. with return volatility | $\sum | | Regression |
|-----------------------------------------------|-------------------------|-----------------------------|------|----------------|
| S&P value weighted return                     | -9.15%                  | -3.02%                      | 12.17% | √              |
| S&P equal weighted return                     | -6.39%                  | 0.01%                       | 6.40% | −              |
| S&P index                                     | 21.75%                  | 54.13%                      | 75.89% | √              |
| 6 month rolling volatility of logarithmic daily returns of S&P index | 14.19% | 36.25% | 50.44% | √ |
| 1 month Lagged T-value                        | 51.65%                  | 9.41%                       | 61.05% | √              |
| Volatility of T value                         | 23.11%                  | 9.25%                       | 32.36% | √              |
| Weekly crude oil future price                 | 100.00%                 | -0.53%                      | 100.53% | −             |
| Average of last 6 months returns crude oil future price | 42.33% | -17.55% | 59.88% | √ |
| Implied volatility of crude oil price         | 0.73%                   | 46.01%                      | 46.74% | √              |

After choosing these 8 variables, we applied regression model on these 8 candidate variables again to make sure that all of them have a significant effect on future return volatility, and we
dropped the following 3 variables because they didn’t show an important effect in regressing the return volatilities:

- S&P value weighted return
- 6 month rolling volatility of logarithmic daily returns of S&P index
- 1 month Lagged T-value

The final input variables for neural networks were

1. S&P index
2. Volatility of T value
3. Weekly crude oil future price
4. Average of last 6 months returns crude oil future price
5. Implied volatility of crude oil price

Using this final set of information variables we forecast the volatility using the two types of ANNs. We also forecast volatility using the GARCH model as in Hasanabadi et. Al. 2012.

7. Evaluating the results

We use two loss functions to evaluate the results and compare the models. These loss functions include; the mean absolute error (MAE) using equation 4 and the mean square error (MSE) using equation 5.

\[
\text{MAE} = \frac{\sum_{i=1}^{N}|V_{\text{forecast}} - V_{\text{realized}}|}{N} \quad (\text{Eq.} 4)
\]
MSE = \sum_{i=1}^{N} \left( V_{\text{forecast}} - V_{\text{realized}} \right)^2 (\text{Eq.5})

The forecasting performance is better when the value is smaller.

8. Empirical results

The feed forward back-propagation neural network (FFBPNN) results are shown in table II. The best network architecture which produces these results has a three layer, 5-3-1 architecture. The first row of the table shows the training process results and the second row shows the testing process results.

<table>
<thead>
<tr>
<th>Network Type</th>
<th>'MSE'</th>
<th>'MAE'</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFBPNN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td>1.32E-05</td>
<td>7.64E-03</td>
</tr>
<tr>
<td>Testing</td>
<td>1.03E-04</td>
<td>9.21E-03</td>
</tr>
</tbody>
</table>

The radial basis function neural network (RBFNN) results are shown in table III. The best network architecture which produces these results has a three layer, 5-218-1 architecture.

<table>
<thead>
<tr>
<th>Network Type</th>
<th>'MSE'</th>
<th>'MAE'</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBFNN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td>3.47E-05</td>
<td>4.57E-03</td>
</tr>
<tr>
<td>Testing</td>
<td>1.01E-04</td>
<td>9.01E-03</td>
</tr>
</tbody>
</table>

Table IV presents the results of applying GARCH model on the testing set:
Table IV: GARCH model results
Forecasting return volatilities using GARCH model.

<table>
<thead>
<tr>
<th>Model</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE</td>
</tr>
<tr>
<td>GARCH</td>
<td>1.07E-04</td>
</tr>
</tbody>
</table>

9. Discussion

Table V compares the results from the two types of neural networks with the GARCH model. We can see that both neural networks including feed forward back-propagation and radial basis neural network show better performance than the traditional GARCH model in forecasting the volatility of crude oil future prices.

Table V. Final results comparison:
Comparison analysis for neural networks and GARCH model’s results

<table>
<thead>
<tr>
<th>Model type</th>
<th>'MSE'</th>
<th>'MAE'</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFBPNN</td>
<td>1.03E-04</td>
<td>9.21E-03</td>
</tr>
<tr>
<td>RBFNN</td>
<td>1.01E-04</td>
<td>9.01E-03</td>
</tr>
<tr>
<td>GARCH</td>
<td>1.07E-04</td>
<td>9.62E-03</td>
</tr>
</tbody>
</table>

Comparing the FFBPNN and RBFNN results with each other we can see a better performance from the radial basis function neural network compared to the feed forward back-propagation model.

We then use the ANNs to forecast crude oil price volatility by only using historical volatility as the information variable that is input to the ANNs. The results presented in table VI, show that forecasting ability of the ANNs is better when using just historical volatility alone as information variable. Hence the measure of speculative activity does not improve the quality of forecast. This suggests that speculative activity does not significantly impact crude oil price volatility.
Table VI. ANNs results comparison for lagged volatilities:
Comparison analysis for neural networks when they have just lagged volatilities as input

<table>
<thead>
<tr>
<th>Model type</th>
<th>'MSE'</th>
<th>'MAE'</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFBPNN</td>
<td>5.04E-05</td>
<td>3.35E-03</td>
</tr>
<tr>
<td>RBFNN</td>
<td>4.98E-05</td>
<td>3.21E-03</td>
</tr>
</tbody>
</table>

10. Conclusion

In this paper we study the relationship between speculative activity and volatility. We employ artificial neural networks (ANNs) to forecast crude oil returns volatility, using an information set of market variables for training. One of these variables is a measure of speculative activity. If the speculative activity did impact crude oil return volatility, we would expect that the information content of the speculative activity variable used in training the artificial neural networks, would improve the quality of forecast. However, the results show that the speculative activity variable did not improve the quality of the forecast. We therefore do not find evidence that speculative activity impacts crude oil price volatility.
References


