International Conference on Applied Energy 2019
Aug 12-15, 2019, Västerås, Sweden
Paper ID: 446

OIL PRICE FORECASTING BASED ON CAPUTO'S ARFIMA MODEL

Xue-Hui Chen[1], Yuan Meng[1]

1 School of Mathematics and Physics, University of Science and Technology Beijing, Beijing 100083, China

ABSTRACT

After a major event affecting the world economy, oil prices tend to fluctuate due to the event in the next few months or even years. It can be seen that oil prices may have long-term correlation. In the processing of time series, the traditional ARMA model cannot accurately describe long memory, which leads to the deviation of parameter estimation during the modeling process. In order to better describe the long memory in the time series, this paper establishes the ARFIMA model to perform fractional difference on the series, and obtains the series satisfying the zero-mean ARMA process, then estimates the parameters. Further research shows that the Caputo fractional difference process is a specialized Grunwald-Letnikov (G-L) fractional differential process. Therefore, this paper introduces the Caputo fractional L1 formula into the time series model, and constructs a new fractional difference method to deal with Brent futures price return rate and perform ARFIMA modeling. This method works better in predicting than the traditional ARMA model and the G-L differential ARFIMA model. It can provide more effective assessments in economic markets such as oil price risk measurement and control, helping investors to better avoid market risks and obtain greater returns.

Keywords: Oil price; Long memory; Caputo fractional difference; ARFIMA model

NONMENCLATURE

Abbreviations	
APEN	Applied Energy
Symbols	
n	Year

1. INTRODUCTION

The fluctuation of oil prices has an important impact on world economic growth and social stability. Therefore, oil price forecasting has always been a research hotspot and focus of oil economy. However, there are various factors that cause the volatility of oil price, making it difficult to predict. The “effective market hypothesis” does not explain the price volatility brought by unexpected events in the oil market. Moreover, the oil market is mostly nonlinear. An important feature of nonlinear systems is long memory, so traditional linear models are not effective in explaining the nature of the market.

The characteristic of long memory is a slow decay of its autocorrelation function at a hyperbolic rate, especially in financial market. Where fractional derivatives and integrals can be used to describe random phenomena with long memory or self-dependence, the concept of fractal is proposed. Hurst et al. adopts the Rescaled range (R/S) analysis method to derive the long-memory relationship of time series [1]. Lo improves the R/S analysis method based on this, which made the improved method have stronger long memory test ability [2]. In terms of the time domain, many scholars have done a lot of research on the long memory of financial markets, and explored ways to improve the accuracy of fixed orders [3, 4]. In terms of the frequency domain, a semi-parametric estimation method proposed by Geweke & Porter-Hudak uses spectral density to solve long memory parameters, which has the advantage that the indirect estimation is faster than the maximum likelihood estimation in the time domain [5].

The autoregressive fractional integral moving average model (ARFIMA) proposed by Granger and Joyeux as the basic model of fractional dynamics, which overcomes the shortfall that the traditional measurement model cannot express the long memory, and makes up for the defect of the over-differentiation of ARIMA model. Granger conducted a more in-depth study in this aspect[6]. And Baillei reviews the characteristics of long memory and ARFIMA model [7]. In the ARFIMA modeling process, discretized fractional derivatives use the G-L formula to differentiate the original series. The two types of fractional differential processes often used in the fractional order domain are the G-L fractional derivative and the Caputo fractional derivative. The former can only achieve first-order precision, while the latter can achieve second-order precision [8,9]. Based on the previous scholars' discussion of ARFIMA and the knowledge of fractional calculus, this paper uses Caputo's L1 formula to replace the G-L fractional differential formula in the previous ARFIMA model, so as to achieve better difference effect and improve the prediction effect. (The G-L differential ARFIMA model is recorded as ARFIMA_G, and the Caputo differential ARFIMA model is recorded as ARFIMA_C).

2. ARFIMA_C MODEL

2. ARFIMA_C model is a fractional difference autoregressive moving average (ARMA) process, which proposes the concept of long memory dependence, considers the power law correlation structure, and is widely used in the dynamic behavior modeling of various complex systems.

Firstly, the time series is tested for long memory. Considering the advantages and disadvantages of the parameters and semi-parametric methods, this paper intends to use the semi-parametric estimation method to determine the ordering of the parameters from the aspect of spectral density. The semi-parametric estimation method is adopted in this paper, and the ordering parameters are determined from the aspect of spectral density. The GPH method is a semi-parametric estimation of fractional single product in the frequency domain, based on the logarithmic period diagram and the Fourier transform. The spectral regression function of the constructed time series is $\ln(I_x(\lambda_j)) = \ln f_\mu(0) - d\ln(4\sin^2(\lambda_j/2)) + \varepsilon_j$ and the long memory parameter is estimated by least squares method.

Next, a fractional difference process on the parameter d is performed to eliminate the long memory feature of the series. In typical fractional calculus, the Caputo fractional differential with higher precision and simplified computational process is combined with the ARFIMA model.

Give the definition of the Caputo fractional derivative:

$$ {}_a^C D_t^\mu f(t) = \frac{1}{\Gamma(n-\mu)} \int_a^t \frac{f^{(n)}(\xi)}{(t_k - j)^{\mu-n+1}} d\xi \tag{1} $$

The range of the long memory parameter in the ARFIMA model is $d \in (-0.5, 0.5)$, and the corresponding Hurst index range is $H \in (0,1)$. Therefore, the value of n in equation (1) is taken as 1 , $\mu \in (0,1)$, and divide the interval [a, t] into n equal parts. Then the discrete form of the Caputo fractional L1 differential discrete equation (2) is as follows:

$$ {}_a^C D_t^\mu f(t_k) = \frac{(\Delta t)^{-\mu}}{\Gamma(2-\mu)} \sum_{j=0}^{k-1} [(j+1)^{1-\mu} - j^{1-\mu}][f(t_{k-j}) - f(t_{k-j-1})] + R_k \tag{2} $$

Let $b_j^{1-\mu} = (j+1)^{1-\mu} - j^{1-\mu}$, then the equation (2) can be changed to the following equation (3) :

$$ \begin{aligned} (1-L)^\mu x_k &\approx \frac{1}{\Gamma(2-\mu)} \left(x_k - b_{k-1} y_0 + \sum_{j=1}^{k-1} [b_j - b_{j-1}] x_{k-j} \right) \\ &= \frac{1}{\Gamma(2-\mu)} \left(1 - \sum_{j=1}^{k-1} [b_j - b_{j-1}] L^j \right) x_k \end{aligned} \tag{3} $$

Where x_k is the original rate of return series, L is the lag operator, μ is a fractional parameter. The new series $\{W_t\}$ can be obtained.

Finally, according to the ARMA modeling process, the model parameters p, q can be obtained. These parameters are both the ARMA model parameter of series $\{W_t\}$, and the ARFIMA_C model parameter of original time series $\{x_t\}$.

3. EMPIRICAL ANALYSIS

Based on the above-mentioned ARFIMA_C theory derivation, the empirical process of this chapter is mainly divided into two major modules, which are data sources and oil price return prediction.

3.1 Data Sources

This article obtains the daily data of Brent futures price from January 4, 2006 to September 28, 2008 in Wind. Using its logarithmic return as sample data for modeling studies, the total sample size is 3,248.

Perform a basic statistical analysis of the series to obtain descriptive statistics on the representative global sample. From the results of statistical analysis: The skewness is 0.08996 and the kurtosis is 6.15283, which reflects that the time series is a skewed spike (normal distribution kurtosis is 3). And the Jarque–Bera statistic shows the null hypothesis of refusing to follow the standard normal distribution at a confidence level of 5%. Therefore, it can be considered that the return series does not obey the normal distribution, and shows the characteristics of peak and thick tail.

Unit root test is first performed to verify the stationarity of the series. The Augmented Dickey Fuller (ADF) test, the Philips–Perron (PP) test, and the Kwiatkowski, Phillips, Schmidt and Shin (KPSS) values are compared with the critical values to show that the series is stable.

Second, a long memory test is performed. The d value can be calculated by observing the autocorrelation function and partial autocorrelation function plots or by semi-parametric estimation to determine whether it has long memory or not. In this paper, the GPH estimation method is used for long memory test and the series is verified by RH method, which makes it more convincing.

Table 1. Series long memory parameter estimation

Long memory Test	m	d_{GPH}	d_{RH}
$[T^{0.45}]$	38	0.03897	0.01198
$[T^{0.50}]$	57	0.12545	0.12319
$[T^{0.55}]$	86	0.12611	0.13369

It can be seen from Table 1 that the values of the parameters d under different bandwidths are all within $(0,0.5)$, and are significantly different from 0. This indicates that the Brent price return series has long memory, but the long-memory performance is different. According to the reference, the d value of GPH estimation when bandwidth $m=[T^{0.50}]$ is selected as the fractional parameter of ARFIMA$_C$ modeling.

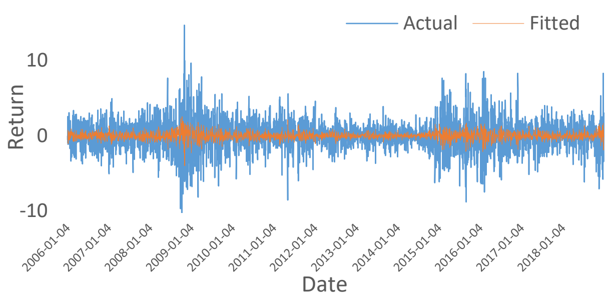

Fig 1 ARFIMA$_C$ model sample test fitting effect

At this point, the differenced series $\{W_t\}$ can be obtained, which is a new time series for eliminating part of the long memory. According to the ARMA modeling process, the in-sample test results of Brent's price return are obtained, as shown in Fig 1. It can be seen from the figure that there are 2-3 large fluctuations in the return series, and the ARFIMA$_C$ model can reflect this phenomenon in different degrees.

3.2 Comparative results

Using Brent futures price daily data, the ARFIMA$_C$ model of its return has been realized. In order to illustrate the quality of ARFIMA$_C$ modeling, this paper uses the same data interval to analyze ARMA. Intercepting the data with obvious fluctuations from 2013 to 2018, the fitting degree of ARMA model and ARFIMA$_C$ model is compared in Fig 2. It is obvious that the ARMA model cannot express long memory, and the model based on Caputo fractional difference can better reflect the phenomenon of fluctuating aggregation.

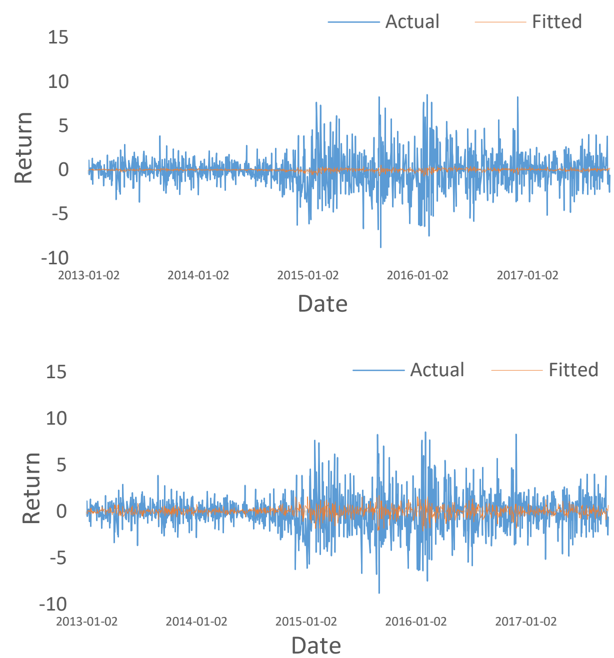

Fig 2 Comparison of ARMA model and ARFIMA$_C$ model sample test

Through the comparison of the two models, it can be seen intuitively that the ARFIMA$_C$ model has better fit effects than ARMA model in the in-sample test. Next, in order to evaluate the predicted performance of the proposed ARFIMA$_C$ model, we use the original series to predict the price return rate for the next three months (the total number of days on the trading day is 64 days). And the prediction goodness of different models is evaluated from three error analysis indicators: root mean square error (RMSE), mean absolute error (MAE) and error variance (EV) [10]. Table 2 shows the comparison of the values of ARMA, ARFIMA$_G$ and

$ARFIMA_C$ models under different measurement errors. It can be seen that the $ARFIMA_C$ has the smallest value of the error indicators in all aspects, and the improvement in EV is more obvious. Therefore, compared with the first two models, the proposed $ARFIMA_C$ model has the best effect on Brent oil price return forecasting.

Table 2. Model prediction error analysis.

Models	ARMA	$ARFIMA_G$	$ARFIMA_C$
RMSE	2.60776	2.57814	2.56108
MAE	1.93695	1.92231	1.89252
EV	6.32938	6.25649	6.15957

4. CONCLUSION

In this paper, the long memory is verified by the GPH method of semi-parametric estimation and the fractional parameter *d* is given. Then, the ARFIMA model is established and the price return value of the next three months (64 days) is predicted. The results show that the Brent futures price return has long memory, which means there is a strong correlation between the long-distance observations, and the previous price return has a positive correlation with the gradual decline of the late price return.

In addition, when dealing with the original time series, this paper uses the Caputo-type fractional difference process. Furthermore, the ARFIMA model with fractional difference is compared with the traditional ARMA model, and the two classical fractional (Caputo’s L-1 algorithm and G-L fractional differential) difference process are compared. Solved the key problem of fractional difference modeling, get the following conclusion:

(1). When calculating the fractional lag of time series, the Caputo fractional order selected in this paper can better describe the long memory characteristics of time series than G-L fractional order, and the calculation process also avoids the problem of gamma function tending to zero in G-L type.

(2). Due to the existence of long memory, oil prices were affected by the economic crisis of 2008 and 2014, and brought about obvious fluctuations and aggregations in the following years. These phenomena can be reflected in the ARFIMA model, but not in the ARMA model. It shows that if the long memory of time series is neglected and the traditional ARMA model is directly used to model, the system deviation will be caused and the model will be invalid.

(3). When forecasting the price return rate in the coming months, the prediction error of ARFIMA model is the smallest, which is consistent with the views of some domestic and foreign scholars . Moreover, the prediction effect of the $ARFIMA_C$ model proposed in this paper is better than the $ARFIMA_G$ model.

In summary, the combination of the new fractional difference method and ARFIMA model can improve the prediction accuracy, which help investors and companies to judge the trend of oil prices, avoid market risks and bring economic benefits.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the financial support from the National Natural Science Foundation of China (nos. 71871020, 71403014, 71521002).

REFERENCE

[1] Hurst HE. Long-term storage capacity of reservoirs. Trans. Am. Soc. Civ. Eng. 1951;116:770–779.

[2] Lo AW. Long-term memory in stock market prices. Econometrical .1991;59(5):1279–1313.

[3] Cheong CW, Nor AHSM, Isa Z. Asymmetry and long-memory volatility: Some empirical evidence using GARCH. Physica A. 2007;373:651–664.

[4] Li M, Cattani C, Chen SY. Viewing sea level by a one-dimensional random function with long memory. Math. Probl. Eng. 2011.

[5] Geweke J, Porter-Hudak S. The estimation and application of long memory time series models. J. Time Series Anal. 1983;4(4):221–238.

[6] Granger CWJ. Long memory relationships and the aggregation of dynamic models. J. Econometrics. 1980:227–238.

[7] Baillie RT. Long memory processes and fractional integration in econometrics. J. Econometrics. 1996;73:5–59.

[8] Sun Z, Wu X. A fully discrete difference scheme for a diffusion-wave system. Appl Numer Math.2006;56(2): 193-209.

[9] Gao G-H, Sun Z-Z, Zhang Y-N. A finite difference scheme for the fractional sub-diffusion equations on an unbounded domain using artificial boundary conditions. J Comput Phys.2012;231(7): 2865-2879.

[10] Zhao L-T, Zeng G-R, He L-Y. Forecasting Short-Term Oil Price with a Generalised Pattern Matching Model Based on Empirical Genetic Algorithm. Comput Econ, 2018.