ISSN 1749-3889 (print), 1749-3897 (online)
International Journal of Nonlinear Science
Vol.29(2020) No.2/3,pp.115-120

Analysis of Factors Affecting Oil Price Short-term Fluctuations in Different Periods—Based on Principal Component Regression Model

Jie Zhou *
Energy Development and Environmental Protection Strategy Research Center, Jiangsu University, Zhenjiang, 212013 Jiangsu, China

(Received 3 April 2020, accepted 2 June 2020)

Abstract: As one of the commodities with the greatest impact on the world, Crude oil has an important economic status. However, oil prices often fluctuate frequently in a short period of time under the influence of various factors. Studying the impact of short-term factors on oil prices is of great significance to stability of crude oil market and prediction of crude oil prices. This paper sorts out the main factors that affect the short-term fluctuations of oil prices. After principal component analysis and multiple regression analysis, the relationship between various factors and oil prices is determined. After scientifically dividing the oil price fluctuations by the regression discontinuity, the relationship between different factors and crude oil price fluctuations in different periods is compared. The study found that the dollar index has a more significant effect on the severe fluctuations in oil prices. In recent years, we need to pay more attention to changes in the US dollar index and stocks.

Keywords: Oil price; Cubic Spline Interpolation; Principal Component Analysis; Regression Discontinuity

1 Introduction

Since the beginning of the 21st century, the rapid development of developing countries and the turbulence of the international political situation have exacerbated the uncertainty of the operation of the crude oil market, and the international crude oil price has fluctuated frequently under the influence of various risk factors[1].In addition to basic supply and demand factors, crude oil prices are largely affected by the alternating effects of many factors such as the US dollar index, geopolitics, speculation, and extreme weather. Especially in the short term, the continued rise in crude oil prices has brought more speculation to the financial markets, which has reduced the operating efficiency of the market while generating price bubbles, and what may even cause global inflation and hinder the industrialization of emerging economies.Miao et al.[2]roughly divided the factors that affect crude oil price fluctuations into six categories, tested the importance of various factors using various prediction models, and used the time-varying relationship between the influencing factors and oil prices to explain crude oil prices fluctuations. Zhang et al.[3] determined that there is a significant long-term equilibrium cointegration relationship between the US dollar exchange rate and international crude oil prices, and found that the depreciation of the US dollar is the key factor pushing up international crude oil prices. Kaufmann et al.[4] examined the causal relationship between crude oil prices in the spot and futures markets in North America, Europe, Africa and the Middle East, and found that the long-term relationship between crude oil spot prices and futures prices changed after September 2004, speculators realized that the possibility of rising oil prices over time is increasing, thereby exacerbating the shock of oil prices. Orbaneja et al.[5] focused on the impact of terrorism on the oil market. The discovery of bombings, major deaths, and incidents near oil facilities will cause crude oil prices to fluctuate significantly due to terrorist attacks in the Middle East. In general, the medium and long-term trend of international crude oil prices still conforms to the law that supply and demand determine prices, while the short-term rises and falls under the influence of financial fluctuations, emergencies and other factors[6].In recent years, the frequent occurrence of geographical conflicts, the instability of crude oil stocks, the increasing size of speculative funds and the continued turmoil in the exchange rate

*Corresponding author: jiechou128@163.com

IJNS.2020.06.15

of the US dollar have all caused crude oil prices to gradually deviate from the long-term movement trend and change the supply and demand relationship of oil in the short term. This makes it necessary for us to understand the impact of these short-term influencing factors on the short-term fluctuations of crude oil prices and the changing laws between them. For the research on the dynamic changes of crude oil prices and their influencing factors, most of them focus on econometric methods, such as vector autoregressive model, cointegration test and general ARCH model. For example, Bollerslev[7] proposed a multivariate time series model with conditional changes and covariance changes with time, but the conditions are unchanged, and is used to capture the transmission of fluctuations between oil prices and foreign exchange rates. In fact, most economic time series have experienced structural changes under the influence of economic structure and social development, and empirical analysis of structural changes in parameters in the sample interval is very important. When making economic predictions, we generally use the data within the most recent stable interval as a reference to achieve the desired prediction effect. Generally speaking, different influencing factors affect the fluctuation of crude oil prices to different extents, and the influences also have obvious differences in different fluctuation times. Since most researchers consider the crude oil price fluctuation and its influencing factors to be more comprehensive and extensive, and they pay less attention to the short-term fluctuation of crude oil price or focus on studying the relationship between certain influencing factors and oil price fluctuations in a short period of time. In addition, when studying crude oil price fluctuations, the time span of the sample data selected is large, and even the division of time is mostly based on priori information, which often has errors in judgment. Considering the above problems, this paper sorts out the main factors that affect the short-term fluctuations of crude oil prices. The principal component analysis method is used to summarize various factors into several principal components, and then multiple regression analysis is conducted with crude oil prices. At the same time, the breakpoint regression method is applied to make a more scientific interval division of crude oil fluctuations, and the above analysis and comparison are repeated for the relationship between oil price fluctuations and their factors in different periods.

2 Method

Principal Component Analysis (PCA)[8]performs dimensionality reduction on the high-dimensional variable space under the principle of minimizing data information loss, that is, a comprehensive indicator composed of a few linear combinations of the research index system. In general, in order to eliminate the influence of dimensions between indicators, the original data set needs to be standardized.Suppose there are p indicators, denoted as $x_1, x_2, \cdots, x_p$. The regrouped comprehensive indicators are denoted as $z_1, z_2, \cdots, z_m (m \leq p)$. The new comprehensive indicator $(m = p)$ can be represented by a linear combination of the original indicators.

$$\begin{cases} z_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1p}x_p \\ z_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2p}x_p \\ \vdots \\ z_p = a_{p1}x_1 + a_{p2}x_2 + \cdots + a_{pp}x_p \end{cases} \tag{1}$$

The correlation coefficient matrix between the variables is shown by Eq(2), where $r_{ij}(i, j = 1, 2, \cdots, p)$ is the correlation coefficient between i and j , and $R = (r_{ij})_{p\times p}$ is a real symmetric matrix.

$$r_{ij} = \frac{\sum_{k=1}^{n} (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^{n} (x_{ki} - \bar{x}_i)^2 \sum_{k=1}^{n} (x_{kj} - \bar{x}_j)^2}} \tag{2}$$

Eigenvalues λ_i and eigenvectors ξ_i are obtained by solving the characteristic equation $|\lambda I - R| = 0$ .The eigenvalues λ_i are arranged as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. The contribution rate reflects the percentage of the total information provided by the principal component. c_i is the contribution rate of the i-th principal component, and the cumulative contribution rate of the first m principal components.

$$c_i = \frac{\lambda_i}{\sum_{k=1}^{p} \lambda_k} \quad (i = 1, 2, \cdots, p) \tag{3}$$

The load of the principal component a_i reflects the degree of influence of the principal component on the variable.

$$a_{ij} = \rho(z_i, x_j) = \sqrt{\lambda_i}\xi_{ij} \quad (i, j = 1, 2, \cdots, p) \tag{4}$$

The larger the load value, the more the variable explains the principal component and the greater the contribution rate. Then the scores of m principal components can be obtained.

$$Z = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1m} \\ z_{21} & z_{22} & \cdots & z_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{nm} \end{bmatrix} \tag{5}$$

With the help of the multiple linear regression model, the change of m independent variables (principal components) is used to explain the change of dependent variable Y_t , and the quantitative relationship between these variables is determined by estimating the related random variables. We construct an m-ary linear regression model(Eq(6)).

$$Y_t = \beta_0 + \beta_1 Z_{1t} + \cdots \beta_m Z_{mt} + \mu_t \tag{6}$$

Where $\beta_s (s = 0, 1, \cdots, m)$ is the partial regression coefficient, and when the other respective variables are kept constant, the influence of the change of a specified independent variable Z_{st} on the change of Y_t . In addition, μ_t is the random error term and Q in Eq(7) is the sum of squared errors. The Ordinary Least Square (OLS) is used to estimate the regression parameter so that the sum of squared deviations between the observed and estimated values of the dependent variable is minimized.

$$\min \left\{ Q\left(\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_p\right) = \sum_{s=1}^{n} \left(Y_t - \hat{Y}_t\right)^2 = \sum_{t=1}^{n} u_t^2 \right\} \tag{7}$$

In addition, we use Regression Discontinuity (RD)[9] to check that there are certain breakpoints in the time series graphic changes, and to ensure that the range of the cabinet is reasonably divided so that there is a significant jump in the sample at the breakpoint. Then the regression estimation of the sample data on both sides of the breakpoint is used to divide the time series scientifically and rationally.

3 Date

3.1 Data selection

We use the Brent crude oil spot price as the dependent variable. Since the beginning of the 21st century, oil prices have been fluctuating frequently. Especially in the 13 years from 2005 to 2017, Brent International crude oil price has experienced a change of more than $100.Coupled with China's increasing dependence on foreign crude oil, it has surpassed the United States as the world's largest oil importer. We have selected some major factors affecting the short-term fluctuations of international crude oil from multiple perspectives such as supply, demand, finance, speculation and geopolitics. Taking financial factors as an example, the US dollar index is the settlement currency of the international crude oil market, and its changes will directly affect the actual price of crude oil and its derivatives. SP500 and MSCI, as representatives of the US and world stock markets, respectively, can reflect the stability of the financial market to a certain extent, and are also a barometer of a country's economy. The volatility of the stock market will exacerbate the concerns of producers and investors, pushing up futures and causing inflation, which will further affect the crude oil market. Therefore, the time span of the sample data we selected is 2005/01/04-2017/12/29. Table 1 lists various data indicators, brief descriptions and data sources.

3.2 Data processing

Since Energy Information Administration only publishes weekly data of U.S. Ending Stocks of Crude Oil, in order to maintain the consistency of data frequency, this paper uses cubic spline interpolation method [10] to interpolate crude oil stocks weekly data into daily data. The interpolation method not only retains the advantages of piecewise low−order interpolation polynomials, but also improves the smoothness of the interpolation function and has better stability. Therefore,it is reasonable to apply the interpolated stocks data to the following research. In addition, considering the lack of partial time data in the sample data, these missing values only account for a very small part of the sample size. Therefore we chose to remove the missing objects, and also screen other factors for daily data to obtain alignment data. This not only avoids the trend smoothing caused by the differential compensation data, but also facilitates subsequent research.

Table 1: Time series data set.

Indicator	Brief description	Frequency	Source
OP(Y)	Brent crude oil spot price	Daily	Energy Information Administration
DXY(x_1)	US Dollar Index	Daily	Wind database
Stocks(x_2)	U.S. Ending Stocks of Crude Oil	Weekly	Energy Information Administration
SP500(x_3)	S&P 500 Index	Daily	Wind database
MSCI(x_4)	Morgan Stanley Capital International World Index	Daily	Wind database
Futures(x_5)	Cushing, OK Crude Oil Future Contract 1	Daily	Energy Information Administration
Terrorism(x_6)	Terrorist attack in the Middle East and North Africa	Daily	Global Terrorism Database

4 Empirical analysis

4.1 Full time analysis

We use Matlab programming according to the above method to perform principal component analysis on the six sets of influencing factor data of DXY, Stocks, SP500, MSCI, Futures and Terrorism. The eigenvalues obtained in this order are as follows:

$$\lambda_1 = 3.6847, \lambda_2 = 1.3785, \lambda_3 = 0.5147, \lambda_4 = 0.2950, \lambda_5 = 0.1086, \lambda_6 = 0.0185$$

The eigenvectors corresponding to different eigenvalues are as follows:

$$\begin{gathered}\xi_1 = (0.4399, 0.4493, 0.4857, 0.4056, -0.2846, 0.0185)^T \\ \xi_2 = (-0.3919, -0.1152, 0.2471, 0.4070, 0.6778, 0.3898)^T \\ \xi_3 = (-0.0164, 0.2220, -0.2803, -0.5411, 0.0136, 0.7609)^T \\ \xi_4 = (-0.2099, 0.8457, -0.0607, -0.1083, 0.3135, -0.3563)^T \\ \xi_5 = (0.7520, -0.1077, 0.0346, -0.2225, 0.6004, -0.1085)^T \\ \xi_6 = (-0.2077, -0.0932, 0.7902, -0.5620, -0.0258, -0.0853)^T\end{gathered}$$

The six eigenvalues corresponding to the principal component information contribution rate are as follows:

$$c_1 = 61.41\%, c_2 = 22.97\%, c_3 = 8.58\%, c_4 = 4.92\%, c_5 = 1.81\%, c_6 = 0.31\%$$

Where λ_1 and λ_2 respectively correspond to the principal component cumulative information contribution rate greater than 80%. Taking F_1 as the first principal component and F_2 as the second principal component, these two principal components basically retain the information of $x_1, x_2, \cdots, x_6$, so that the original six indicators are converted into two new indicators, which plays a role in dimensionality reduction. The linear combination of the F_1 and F_2 are as follows:

$$\begin{gathered}F_1 = 0.4399x_1 + 0.4493x_2 + 0.4857x_3 + 0.4056x_4 - 0.2846x_5 + 0.3509x_6 \\ F_2 = -0.3919x_7 - 0.1152x_2 + 0.2371x_3 + 0.4070x_4 + 0.6778x_5 + 0.3898x_6\end{gathered}$$

Taking F_1 and F_2 as the independent variables and Brent crude oil spot price as the dependent variable, a multiple linear regression analysis can be carried out to obtain the linear regression equation.

$$Y = 70.336 - 0.505F_1 + 0.946F_2$$

After the Student's test, the probability values of the two partial regression coefficients are both 0.000. According to the given significance level of 0.1, they have statistical significance. However, according to the tolerance and VIF value, it can be found that the problem of collinearity between independent variables is more obvious. The value of the coefficient of determination R^2 is 0.884 , the value of the complex linear correlation coefficient R is 0.94. It can be seen that the absolute values of the coefficient of determination and the linear correlation coefficient are close to 1, indicating that the fitting effect is good(Figure 1) and the degree of correlation is also quite high. Finally, the fitting equation of the Brent international crude oil spot price is obtained after the substitution calculation.

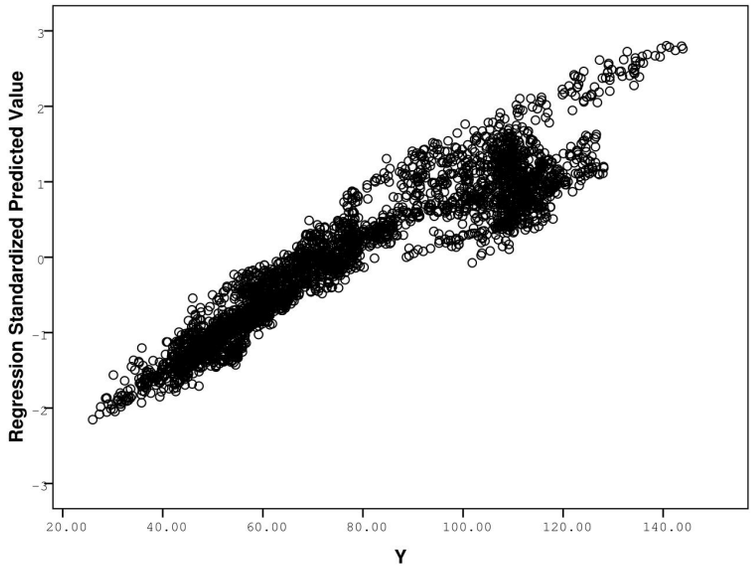

Figure 1: Scatterplot of dependent variable and regression normalized predicted value.

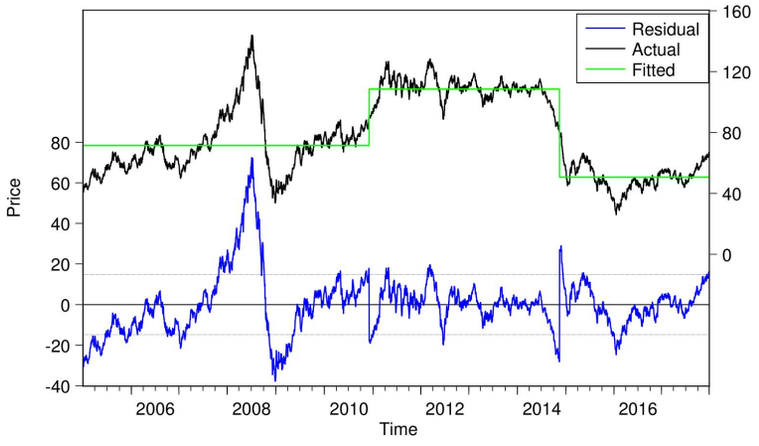

Figure 2: Results of division of crude oil price fluctuation period.

$$Y = 70.336 - 0.593x_1 - 0.336x_2 - 0.021x_3 + 0.180x_4 + 0.785x_5 + 0.192x_6$$

4.2 Multi-period analysis

We use Regression Discontinuity test to divide the fluctuations of Brent international crude oil prices from 2005/01/04 to 2017/12/29. As shown in Figure 2, two breakpoints were tested, they divided oil price fluctuations into three different fluctuation periods, and different oil prices fluctuated up and down at different fluctuation levels.

We repeated the above principal component analysis and multiple linear regression on the divided three periods of data, to obtain the regression coefficient between crude oil prices and various influencing factors in each period (Table 2).

5 Conclusion

By analyzing the above research results, the following conclusions are obtained:

Table 2: Estimation results of regression coefficients in each period.

	Full time	2005/01/04-2010/12/02	2010/12/03-2014/11/12	2014/11/13-2017/12/29
DXY	-0.59300	-0.80700	0.50400	-0.55700
Stocks	-0.33600	0.16800	0.09800	-0.57500
SP500	-0.02100	-0.03000	0.00600	-0.03000
MSCI	0.18000	0.14200	-0.10600	0.13900
Futures	0.78500	0.70800	-0.79800	0.37200
Terrorism	0.19200	-0.00003	0.00400	0.00200

1. Among the various factors that affect the short-term fluctuations of oil prices, the influence of each factor on the short-term fluctuations of oil prices is different. The US dollar index, stocks and crude oil futures have an important impact on the short-term fluctuations of oil prices;

2. The dollar index has always had an important impact on oil price fluctuations. In the period of stable oil price fluctuations, its role is relatively stable; while in the period of severe oil price fluctuations, the dollar index has a more significant reverse effect on oil price fluctuations;

3. In recent years, the impact of stocks on oil price fluctuations has kept pace with the US dollar index; in contrast, the impact of oil futures has decreased significantly. When forecasting oil prices in the short term, we need to pay more attention to changes in the US dollar index and stocks.

References

[1] Hamilton, James D. Oil and the macroeconomy since World War II. *Journal of political economy* 91(2)(1983)228–248.

[2] Miao, Hong and Ramchander, Sanjay and Wang, Tianyang and Yang, Dongxiao. Influential factors in crude oil price forecasting. *Energy Economics* 68(2017)77–88.

[3] Zhang, Yue-Jun and Fan, Ying and Tsai, Hsien-Tang and Wei, Yi-Ming. Spillover effect of US dollar exchange rate on oil prices. *Journal of Policy Modeling* 30(6)(2008)973-991.

[4] Kaufmann, Robert K and Ullman, Ben. Oil prices, speculation, and fundamentals: Interpreting causal relations among spot and futures prices. *Energy Economics*31(4)(2009)(550–558).

[5] Orbaneja, José R Valdivia and Iyer, Subramanian R and Simkins, Betty J. Terrorism and oil markets: A cross-sectional evaluation. *Finance Research Letters* 24(2018)42–48.

[6] Martínez, Josué M Polanco and Abadie, Luis M and Fernández-Macho, Javier. A multi-resolution and multi-variate analysis of the dynamic relationships between crude oil and petroleum-product prices. *Applied Energy* 228(2018)1550–1560.

[7] Bollerslev, Tim. Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *The review of economics and statistics* (1990)498–505.

[8] Abdi, Hervé and Williams, Lynne J. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics* 2(4)(2010)433–459.

[9] Trochim, William MK. Research design for program evaluation: The regression-discontinuity approach. *SAGE Publications, Inc* 6(1984).

[10] Sarfraz, Muhammad. Interpolatory rational cubic spline with biased, point and interval tension. *Computers & graphics* 16(4)(1992)427–430.