ISSN 2616-5872 Vol.2, Issue 1: 6-15, DOI: 10.25236/AJEE.2020.020102

Model of Octane Number Loss Based on Margin Analysis

Dacheng Song

Department of Library, Information and Archives, Shanghai University, Shanghai 200444

ABSTRACT. *This article focuses on the question of how to optimize the main operating steps of the desulfurization process. According to the correlation analysis, the residual analysis method based on the stepwise forward regression and the generalized least square method are used to establish the octane loss model, and finally the genetic algorithm is used to accelerate Calculate and reasonably analyze how the desulfurization process reduces the loss of octane number in the operation steps.*

KEYWORDS: *margin analysis method, maximum information coefficient, genetic algorithm*

1. Background

A good ecological environment is the fundamental foundation for the sustainable development of people and society, and resource conservation and environmental protection have always been my country's basic national policies. However, with the economic and social development of various countries and the increase in the number of cars, the consumption of gasoline is showing a rapid increase. The sulfur oxides and nitrogen oxides produced by gasoline combustion have become the most important causes of pollution in major cities. It also has a serious impact on the atmospheric environment. In order to make automobile exhaust emissions meet the standards, countries all over the world have formulated increasingly strict gasoline quality standards, one of the main purposes of which is to achieve clean gasoline fuel. The focus of gasoline cleaning is to reduce the sulfur and olefin content in gasoline while maintaining its octane number as much as possible.

Octane number (RON) is the most important indicator reflecting the combustion performance of gasoline. According to the difference of octane number, gasoline products can be divided into 89#, 92#, 95# and other commercial brands. my country's crude oil imports mainly rely on sulfur-containing and high-sulfur crude oil from the Middle East, but the crude oil in this region contains high heavy oil impurities and is difficult to directly utilize. In order to effectively utilize heavy oil

ISSN 2616-5872 Vol.2, Issue 1: 6-15, DOI: 10.25236/AJEE.2020.020102

resources, my country has vigorously developed heavy oil lightening technology with catalytic cracking as the core, which converts heavy oil into gasoline, diesel and low-carbon olefins. More than 70% of gasoline is produced by catalytic cracking, so refined gasoline More than 95% of sulfur and olefins come from catalytic cracking gasoline. However, my country's existing technology generally reduces the gasoline octane number in the process of desulfurization and olefin reduction of FCC gasoline. Generally speaking, every unit of octane reduction is equivalent to a loss of about 150 yuan/ton. Therefore, the octane number is not only related to the quality of the manufactured products, but also closely related to the economic benefits of the manufacturer, the control of gasoline intermediate products, and the blending of refined oil.

This paper establishes an octane loss model based on the margin analysis method, which takes the linear model as the main part of the "trend" and the margin as the fluctuation part of the "trend". First, the correlation matrix diagram is used to analyze the "involution" between inoperable variables and operable variables, and the generalized least square method is used to decrypt the "involution" of nonlinear variables and linear variables. Then, using the idea of linear forward regression, linear variables, monotonic non-linear variables and periodic non-linear variables that have an impact on the octane number loss are gradually added. The non-linear variables are all based on the margin obtained from the previous model. The residuals are fitted nonlinearly. Finally, the sensitivity analysis of the established octane number loss model shows that most of the errors of the established model fall within one standard deviation.

2. Problem analysis

This article first divides the 11 inoperable variables into three different categories: linear variables, monotonic nonlinear variables, and periodic non-monotonic variables, and then uses the partial least squares method to determine whether the inoperable variables and operable variables are inoperable. Whether there is a correlation between the variables themselves, finally, a residual approximation method is proposed, which uses a linear model to construct the general trend of the octane number (RON) loss prediction model, and gradually approximates the residual with monotonic nonlinear variables, and periodic non-monotonic variables approximate the remaining residuals, thereby establishing the octane number (RON) loss prediction model.

3. Problem solving

It can be seen from the second question that we selected 19 main operable sample variables, plus our own 11 inoperable variables, and selected a total of 30 main operable variables of the RON loss prediction model. After the main indicators are selected, the octane number (RON) loss prediction model is then established. The model construction steps are as follows:

ISSN 2616-5872 Vol.2, Issue 1: 6-15, DOI: 10.25236/AJEE.2020.020102

3.1 Use correlation coefficients to divide inoperable variables into different categories

(1) Use Pearson correlation coefficient to find the linear relationship between inoperable variables. In order to more clearly calculate the relationship between the main variables (operated variables, inoperable variables) of the octane number (RON) loss prediction model, we need to divide the inoperable variables into different categories.

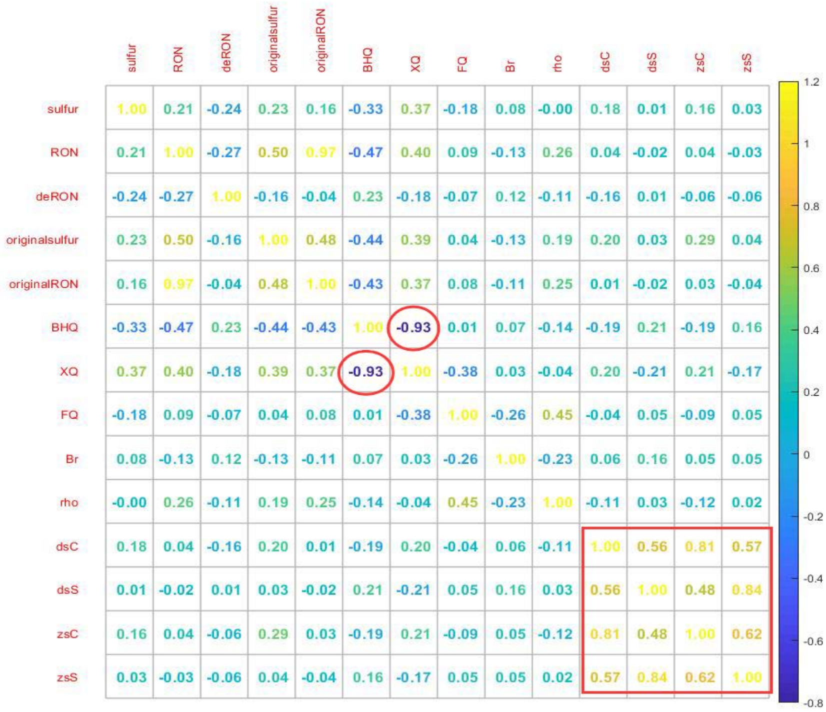

Figure. 1 The linear relationship between the dependent variable and the independent variable

First, the Pearson correlation coefficient is used to find the linear relationship between the inoperable variables, and then the variables that have a large contribution to the RON loss prediction model are found in turn. That is, the linear relationship between the dependent variable and the independent variable.

From Figure 1, we can know that the four variables in the lower right corner have extremely high correlations, but from their impact on octane loss (variable dsC has a much greater impact on octane loss than the other three variables), so it shows that the function combination of the remaining three variables is linearly related to

ISSN 2616-5872 Vol.2, Issue 1: 6-15, DOI: 10.25236/AJEE.2020.020102

the variable dcC in terms of the impact on the octane loss, which leads to obvious differences in the degree of influence between them. Below we first study the functional relationship between these four variables, through Eviews software, we can build the following model.

3.2 Determine the relationship between operating variables

After using the correlation coefficient to classify the inoperable variables, the partial least squares method is then used to determine whether there is a correlation between the inoperable variables and the operable variables, and between the inoperable variables themselves, that is, linear relationships, monotonic non- Linear relationship, and periodic non-monotonic relationship.

Partial least squares regression is a new type of multivariate statistical data analysis method. It mainly studies the regression modeling of multiple independent variables. Especially when the variables are highly linearly correlated, it is more effective to use partial least squares regression [4]. In addition, the partial least squares method is superior to three analysis methods including principal component analysis, canonical correlation analysis and multiple linear regression analysis. Both it and the principal component analysis method try to extract the maximum information that reflects the variation of the data, but the principal component analysis method only considers a matrix of independent variables, while the partial least square method also has a "response" matrix, so it has a predictive function.

3.3 Constructing the RON loss prediction model by the margin approximation method

In order to describe the model more accurately, we propose a margin approximation method to gradually establish an octane number (RON) loss prediction model. According to the correlation, first we use a linear model to construct the general trend of the octane number (RON) loss prediction model, secondly use monotonic nonlinear variables to gradually approximate the residuals, and finally use periodic non-monotonic variables to approximate the remaining residuals, thereby establishing Octane number (RON) loss prediction model.

(1) The linear model makes the general trend of the octane loss model

First, we use linear variables to fit the octane loss model. This part is used as the main trend part of the model. The simulation results and the significance verification of each parameter are given below.

ISSN 2616-5872 Vol.2, Issue 1: 6-15, DOI: 10.25236/AJEE.2020.020102

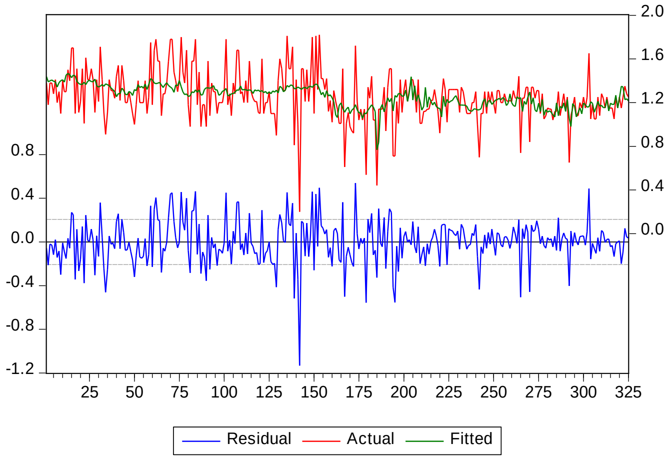

Figure. 2 Residual plot of linear mixed model

At this time, the blue line is our margin. From Figure 2, we can know that although the linear model is not very significant, the residuals are basically within one standard deviation, and the linear model can indeed represent the trend. The main part. Next, the nonlinear monotonic class is fitted. Since the monotonic class has many nonlinear terms and irregularities, we must consider which models need to use a new margin every time we do it.

(2) Monotonic nonlinear variables gradually approach the residual, see Figure 3 and Figure 4.

Dependent Variable: LOG(Y1^2)
Method: Least Squares
Date: 09/20/20 Time: 17:07
Sample: 1 325
Included observations: 325

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.548110	21.27020	0.448896	0.6538
LOG(DNL149^2)	3.272722	4.430973	0.738601	0.4607
LOG(DNL158^2)	0.080380	0.023910	3.361770	0.0009
LOG(DNL276^2)	0.010268	0.031766	0.323223	0.7467
LOG(DNL324^2)	0.031963	0.170057	0.187953	0.8510
LOG(DNL331^2)	0.003600	1.809430	0.001989	0.9984

R-squared	0.051058	Mean dependent var	-4.890818
Adjusted R-squared	0.036184	S.D. dependent var	2.470284
S.E. of regression	2.425179	Akaike info criterion	4.627977
Sum squared resid	1876.197	Schwarz criterion	4.697832
Log likelihood	-746.0463	Hannan-Quinn criter.	4.655856
F-statistic	3.432758	Durbin-Watson stat	1.708606
Prob(F-statistic)	0.004903		

Figure. 3 Logarithmic monotonic nonlinear model (1)

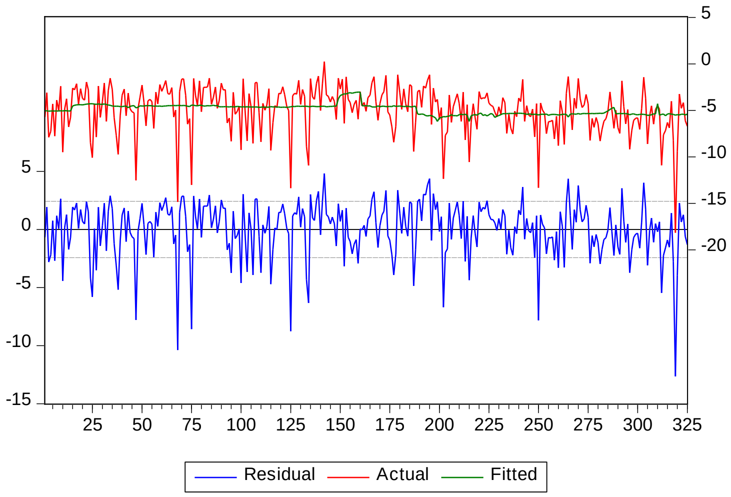

Figure. 4 Residual plot of logarithmic monotonic nonlinear model

The Figure 5 still estimates the monotonic nonlinear term.

Dependent Variable: LOG(Y1^2)
Method: Least Squares
Date: 09/20/20 Time: 17:32
Sample: 1 325
Included observations: 325

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-6.054388	0.397602	-15.22727	0.0000
LOG(DNL170^2)	0.055419	0.017808	3.111985	0.0020

R-squared	0.029110	Mean dependent var	-4.890818
Adjusted R-squared	0.026104	S.D. dependent var	2.470284
S.E. of regression	2.437828	Akaike info criterion	4.626227
Sum squared resid	1919.591	Schwarz criterion	4.649512
Log likelihood	-749.7619	Hannan-Quinn criter.	4.635520
F-statistic	9.684452	Durbin-Watson stat	1.668154
Prob(F-statistic)	0.002024		

Figure. 5 Logarithmic monotonic nonlinear model (2)

(3) The remaining residuals are approximated by periodic variables. For periodic nonlinear terms, we mainly construct trigonometric functions to simulate the equations, see Figure 6 and Figure 7.

ISSN 2616-5872 Vol.2, Issue 1: 6-15, DOI: 10.25236/AJEE.2020.020102

Dependent Variable: Y1
Method: Least Squares
Date: 09/20/20 Time: 17:42
Sample: 1 325
Included observations: 325

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.003594	0.012135	-0.296169	0.7673
SIN(CNL120)	0.023682	0.016437	1.440840	0.1506
COS(CNL250)	0.030652	0.017775	1.724428	0.0856
COS(FQ)/SIN(XQ)	0.000809	0.000781	1.036716	0.3007
COS(ORIGINALRON)	0.017595	0.018510	0.950579	0.3425

R-squared	0.025614	Mean dependent var	6.62E-16
Adjusted R-squared	0.013434	S.D. dependent var	0.201590
S.E. of regression	0.200232	Akaike info criterion	-0.363418
Sum squared resid	12.82967	Schwarz criterion	-0.305206
Log likelihood	64.05548	Hannan-Quinn criter.	-0.340186
F-statistic	2.102961	Durbin-Watson stat	1.993856
Prob(F-statistic)	0.080242		

Figure. 6 Nonlinear statistics of periodic functions

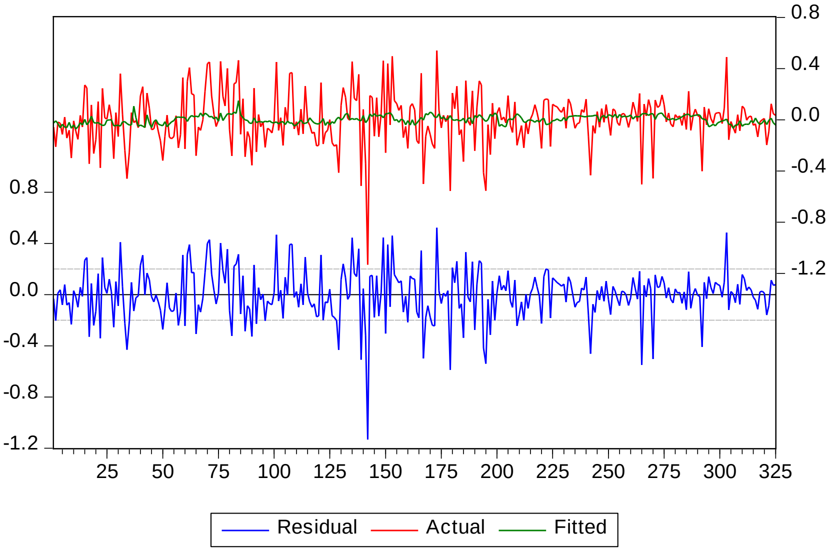

Figure. 7 Nonlinear residuals diagram of periodic function

ISSN 2616-5872 Vol.2, Issue 1: 6-15, DOI: 10.25236/AJEE.2020.020102

Note that we are only exploring the function format above, without setting the coefficients, so we need to put all the function items together and do linear regression to get the corresponding weight coefficients, see Figure 8 and Figure 9.

Dependent Variable: DERON
Method: Least Squares
Date: 09/20/20 Time: 20:15
Sample: 1 325
Included observations: 325

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.827801	4.569863	0.618793	0.5365
ORIGINALSULFUR	-0.000283	0.000246	-1.150409	0.2509
BHQ	0.003183	0.003943	0.807211	0.4202
BR	-0.000516	0.001658	-0.311391	0.7557
RHO	-0.002115	0.003022	-0.699806	0.4846
DSC	-0.020737	0.012764	-1.624635	0.1053
DSS^2	0.000462	0.000724	0.638261	0.5238
ZSC	0.040151	0.017647	2.275184	0.0236
ZSS^(-1/2)	0.441184	0.261206	1.689030	0.0923
L37	5.00E-05	1.35E-05	3.705900	0.0003
L41	-0.000431	0.000199	-2.170188	0.0308
L177	0.021060	0.010671	1.973626	0.0493
L195	6.63E-06	3.70E-06	1.793473	0.0739
L207	0.000761	0.000736	1.033022	0.3024
L325	-0.008580	0.010437	-0.822120	0.4117
ABS(DNL170)	1.88E-09	2.81E-08	0.066795	0.9468
DNL149*DNL158*DNL276*DNL324*DNL...	2.34E-08	5.02E-08	0.467052	0.6408
DNL165^(-3)	2.34E-08	5.02E-08	0.465852	0.6417
DNL257^(-3)	473408.8	366111.8	1.293072	0.1970
DNL5^3	3.75E-08	1.30E-08	2.872189	0.0044
DNL71^3	1.16E-06	1.55E-06	0.748517	0.4547
DNL84^3	1.46E-13	5.68E-13	0.256777	0.7975
SIN(CNL120)	0.034789	0.018715	1.858903	0.0640
COS(CNL250)	0.046012	0.023610	1.948806	0.0523
COS(FQ)/SIN(XQ)	0.000842	0.000831	1.013428	0.3117
COS(ORIGINALRON)	0.002100	0.025964	0.080874	0.9356

R-squared	0.256760	Mean dependent var	1.254769
Adjusted R-squared	0.194616	S.D. dependent var	0.225890
S.E. of regression	0.202721	Akaike info criterion	-0.277356
Sum squared resid	12.28762	Schwarz criterion	0.025350
Log likelihood	71.07036	Hannan-Quinn criter.	-0.156546
F-statistic	4.131712	Durbin-Watson stat	2.005925
Prob(F-statistic)	0.000000		

Figure. 8 The octane loss prediction model

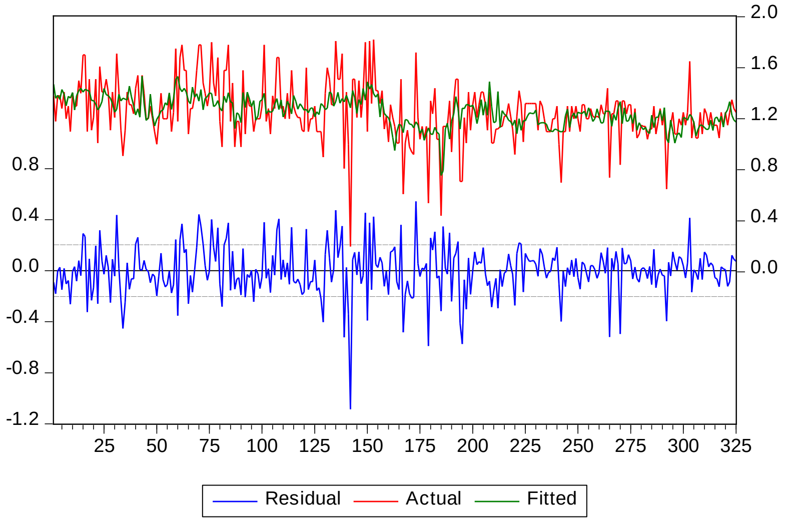

Figure. 9 Octane loss prediction model residual diagram

4. Model evaluation

The biggest advantage of the octane number (RON) loss prediction model is that it takes into account the correlation between variables, that is, the variables are not only linearly correlated, but may also be monotonic non-linear correlations, periodic non-linear correlations, etc., The constructed model has a good fitting relationship. The model can be extended to solve the problem of multiple variables that are related and lagging, and it can be applied to prediction problems in actual problems. By optimizing the main operating variables to achieve the purpose of reducing the octane number, reducing enterprise production costs, improving the utilization rate of heavy oil, and reducing pollution source emissions, it has huge economic and social value.

Reference

[1] Han Zhongzhi, Wan Jianhua, Liu Kangwei. Gasoline octane number prediction and model optimization based on near-infrared spectroscopy [J]. Analysis Laboratory, 2015, 34 (11): 1268-1271.
[2] Jiang Feng, Zhu Jialing, Hu Kaiyong, Xin Guanglei, Zhao Yangyang. Application study of Pearson correlation coefficient to evaluate the characteristics of ORC system evaporator [J]. Acta Solar Energy, 2019, 40 (10): 2732-2738.
[3] Reshef, Reshef Y A, Finucane H K, et al. Detecting Novel Associations in Large Data Sets [J]. Science, 2011, 334 (6062): 1518-1524.

ISSN 2616-5872 Vol.2, Issue 1: 6-15, DOI: 10.25236/AJEE.2020.020102

[4] J J E Slotine, W Li.Applied nonlinear control [M]. EnglewoodCliffs, USA: Prentice-Hall, 1991.
[5] Wang Tiefang. Computer genetics: Grid trust model based on family genes [M]. Beijing: Intellectual Property Press, 155, 2016.
[6] Shi Xianjun, Liao Jian, Ma Changli, et al. Generalized Least Squares Parameter Identification and Simulation Based on MATLAB [J]. Computer and Digital Engineering, 2009 (08): 179-181.