Quantitative Analysis of Influencing Factors of Steel Price Index based on Multiple Regression

Anna Li
School of Finance, Anhui University of Finance and Economics, Bengbu, 233030, China

Abstract: The Sino US trade war has a great influence on the price of domestic steel, but in recent years, China's steel industry has been in a very disadvantageous position in the international steel market. Steel price index, as the main measure index of steel industry, is affected by the relationship between steel supply and demand, macro-economy and some other aspects. In this paper, through the establishment of measurement model, the influencing factors of steel price index are analyzed, and the main influencing factors are found out. Finally, according to the obtained model, relevant application scope and suggestions are put forward.

Keywords: Steel price index; Multiple linear regression; Measurement test; Generalized difference method

1. Introduction

In recent years, the price fluctuation of iron and steel is large. As an important industry, the price fluctuation of iron and steel has an important impact on China's economic development. Therefore, it is necessary to carry out empirical analysis on the influencing factors of iron and steel price. The iron and steel industry has been oversupplied for a long time, resulting in overcapacity, low pricing and the lowest profit level in the industry. In addition, at the beginning of the Sino US trade war on March 8, 2018, the United States imposed a 20% tariff on Chinese steel. China's steel industry has been in a very unfavorable position. Steel price index, as the main measure index of steel industry, is affected by the relationship between steel supply and demand, macro-economy and some other aspects. In this paper, through the establishment of a quantitative model, we can quantitatively explore the impact of supply and demand, macroeconomic, labor costs and the output value of related industries on the steel price index.

2. Journals Reviewed

At present, there are few researches on the influencing factors of iron and steel price in the academic circle. There are two main viewpoints: one is that the iron and steel price is mainly affected by the supply and demand relationship of iron and steel; the other is that the iron and steel price is mainly affected by the national macro-economy and macro policies [1].

Su Ying and Zhao Shuting [2] think that steel price is mainly affected by macro-economy. They have studied the quantitative analysis of steel price on macro-economy, steel raw material cost, money supply (M2), downstream industry development of steel and steel supply and demand. They have not sorted the influencing factors of steel price. Xu Yaping and Wang Tian [3] think that in the long run, the impact of macroeconomic development fluctuation on steel price fluctuation is greater than that of supply and demand impact on steel price. Su Ying and Zhao Shuting all use unit root test and cointegration test, but Xu Yaping and Wang Tianduo use variance decomposition to get the contribution degree of each influencing factor to the fluctuation of steel price. Li Yujun [4] believes that in the long run, the international steel price index, industrial added value and fixed asset investment have a significant impact on the comprehensive steel price index. He has established a multiple linear regression model, and adopted the methods of stability test and multiple collinearity test. At present, there are few researches on the comprehensive quantitative relationship among macro economy, supply and demand, labor cost and output value of related industries.

3. Analysis of Steel Price Factors based on Multiple Regression

3.1. Theoretical model

3.1.1. Determine statistical indicators, collect and sort out data, and de seasonalize seasonal data

Through the literature study of steel price, it is concluded that steel price is affected by the macroeconomic situation, supply and demand, cost situation and construction industry [5-8]. Therefore, this paper selects steel price index as the explanatory variable y and six explanatory variables: steel sales volume (10000 tons) x 1, steel production (million tons) x 2, and GDP, RMB 100 million) x

3, secondary industry value-added index (RMB 100 million) x 4, labor cost price index x 5 and total output value of construction industry (RMB 100 million) x 6. In addition, the quarterly data from the fourth quarter of 2014 to the fourth quarter of 2018 are used in this paper. See Table 1 for specific data.

Table 1. Experimental variable data

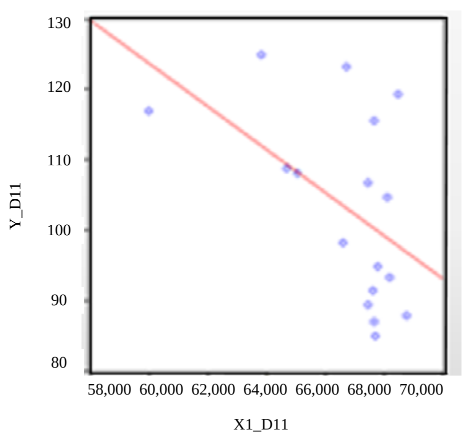
T	Y	X1	X2	GDP	X4	X5	X6
2014Q4	93.2	110152.4	1109.289023	180190.3	107	105.9	176713.40
2015Q1	90.4	25188.0	262.9227557	150593.8	106.4	105.7	28895.36
2015Q2	90.0	54134.6	550.7080366	167874.5	106.2	104.5	72374.43
2015Q3	87.2	82007.2	828.3555556	175803.8	106.0	103.9	117947.30
2015Q4	84.8	109359.1	1107.994934	191720.8	106.3	104.3	180757.47
2016Q1	86.8	25616.9	258.7565657	160967.3	106.0	103.7	30849.97
2016Q2	95.5	54370.9	551.4290061	179878.7	106.4	103.6	77461.75
2016Q3	98.5	80698.7	837.123444	189337.6	106.3	103.3	125791.60
2016Q4	104.5	109988.3	1121.185525	209877.2	106.3	103.4	193566.78
2017Q1	118.0	25508.6	259.4974568	179403.4	106.2	103.6	34188.67
2017Q2	116.4	54299.5	544.6288867	199177.8	106.3	103.5	85871.09
2017Q3	123.5	80870.6	818.5283401	209824.1	105.8	104.0	139259.73
2017Q4	124.8	103106.4	1034.166499	232349.0	105.5	104.5	213953.96
2018Q1	115.6	22344.6	233.4858934	197920.0	106.3	103.5	38691.96
2018Q2	109.5	51899.8	522.6565962	219295.4	106.0	104.6	94790.39
2018Q3	108.4	78845.2	808.6687179	229495.5	105.3	104.8	152326.24
2018Q4	106.6	108958.9	1089.589000	253598.6	105.8	104.8	235085.53

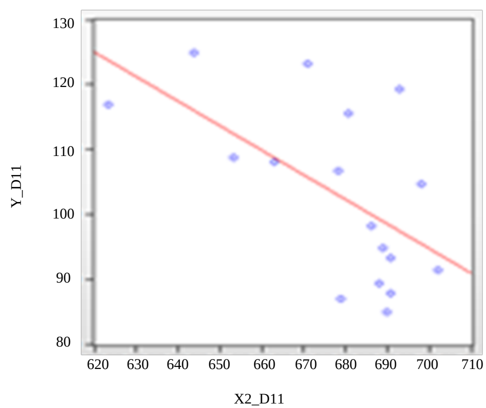
Note: the data are from China Statistical Yearbook, China Bureau of statistics and the official website of China Iron and Steel Industry Association.

When using seasonal data to build a model, it will be accompanied by seasonal fluctuations. Therefore, before the model estimation, we need to use Eviews to de seasonalize the data, and then we can get the seasonal experimental variable data (sequence name: variable name □ D11, hereinafter all the variable names for writing convenience are variable names □ D11).

3.1.2. Determining the mathematical relationship between variables

Using Eviews software to analyze the correlation graph of experimental variables x1, X2, X3, x4, X5, X6, y, and get the mathematical relationship between experimental variables.





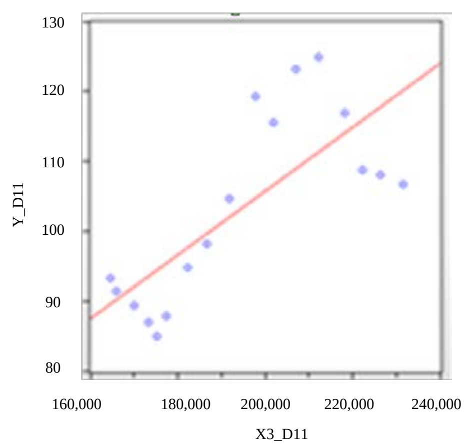


Figure 1. Correlation diagram of Y and X1, X2, X3

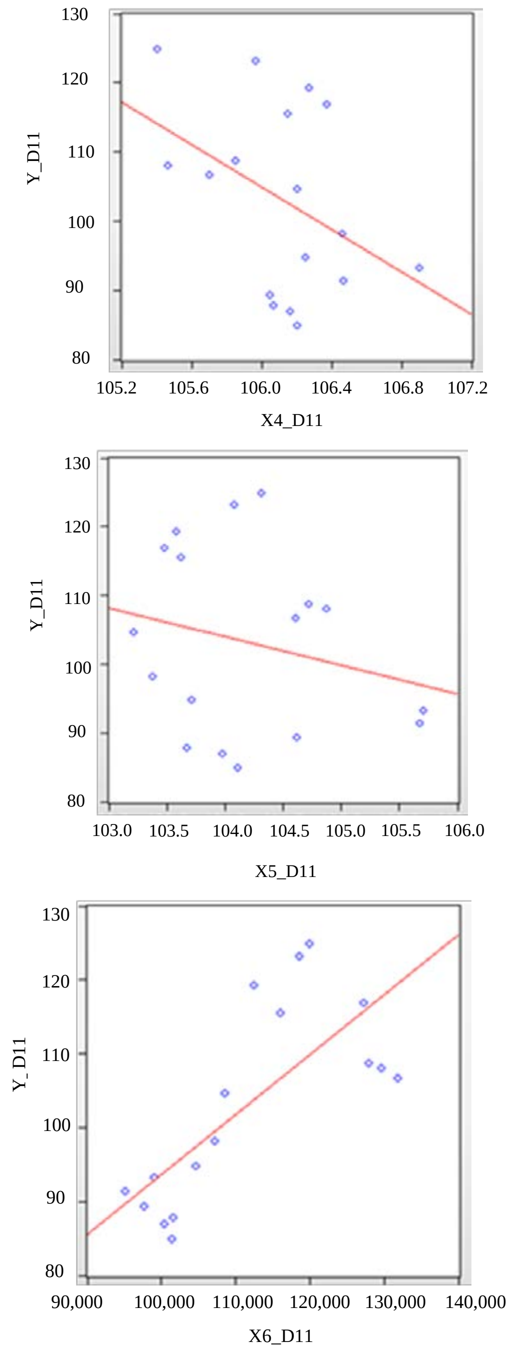


Figure 2. Correlation diagram of Y and X4, X5, X6

From the analysis of the correlation graph between the explanatory variables and the explained variables, we can see that: the steel price index y has a linear relationship with the secondary industry value-added index X4 and labor cost price index X5; the steel price index y has a strong linear relationship with steel sales x1, steel output X2, GDP X3 and total construction output X6.

3.2. Estimation of model parameters

Based on the correlation analysis of the above experimental variables, it is found that the steel price index y has a linear relationship with the secondary industry value-added index X4 and labor cost price index X5, and has a strong linear relationship with steel sales x1, steel output X2, GDP X3 and total construction output X6. Therefore, the multivariate linear model is established as follows:

$$y_i = b_0 + b_1x_{1i} + b_2x_{2i} + b_3x_{3i} + b_4x_{4i} + b_5x_{5i} + b_6x_{6i} + u_i$$

Eviews was used for multiple regression, and the regression results were as follows. $\hat{y}_i = -4291.429 + 0.006939x_1 - 1.161455x_2 + 0.005198x_3 + 39.42917x_4 + 4.433068x_5 - 0.008395x_6$

3.2.1. Marginal analysis

Regression coefficient estimates $\hat{b}_1$ = 0.006939 > 0, $\hat{b}_2$ = -1.161455 < 0, $\hat{b}_3$ = 0.005198 > 0, $\hat{b}_4$ = 39.42917 > 0, $\hat{b}_5$ = 4.433068 > 0, $\hat{b}_6$ = -0.008395 < 0, It shows that the steel price index changes in the same direction with the steel sales volume, GDP, added value index of the secondary industry and labor cost price index; the steel price index changes in the opposite direction with the steel output and the total output value of the construction industry, and the steel price index increases by 0.006939 for every 10000 tons of steel sales volume increase and 0.006939 for every 1 million tons of steel output increase Price index decreased by 1.161455; steel price index increased by 0.005198 for every 100 million Yuan increase in GDP; steel price index increased by 39.42917 for every 1 increase in value-added index of secondary industry; steel price index increased by 4.433068 for every 1 increase in labor cost price index; steel price index decreased by 0.008395 for every 100 million Yuan increase in total output value of construction industry.

3.2.2. Standard error test of regression equation

S. E. = 8.453995, indicating that the average error between the actual observation point of steel price index and its estimated value is 8.453995.

3.2.3. Goodness of fit test

R2 = 0.753364 indicates that the model has a high fitting degree to the sample data or 75.3364% of the explained variable steel price index y can be explained by steel sales x1, steel production X2, GDP X3, secondary industry added value index x4, labor cost price index X5 and total construction industry output value X6.

3.2.4. F test

From the common influence of all variables, on the significance level of 0.05, f = 5.090939> $F_\alpha(k,n-k-1) = F_{0.05}(6,17-6-1)$ =3.22, indicating that the steel sales volume, steel output, GDP, secondary industry added value index, labor cost price index, and the total output value of the construction industry have significant influence on the steel price index. It can be seen from P = 0.012119 that the regression model is significant.

3.2.5. T test

From the influence of a single variable, at a significance level of 0.05, $\left|t(\hat{b}_1)\right|$ = | 1.777811 | < $t_{0.025}(10)$ = 2.2281; $\left|t(\hat{b}_2)\right|$ = | -2.192689 | < $t_{0.025}(10)$ = 2.2281; $\left|t(\hat{b}_5)\right|$ = | 1.125706 | < $t_{0.025}(10)$ = 2.2281, indicating steel sales, steel production and labor costs The price index has no significant effect on the steel price index; $\left|t(\hat{b}_3)\right|$ = | 2.607962 |> $t_{0.025}(10)$ = 2.2281; $\left|t(\hat{b}_4)\right|$ = | 2.470757 |> $t_{0.025}(10)$ = 2.2281; $\left|t(\hat{b}_6)\right|$ = | -2.387081 |> $t_{0.025}(10)$ = 2.2281, indicating the GDP, the value added index of the secondary industry and the construction industry The gross output value has a significant impact on the steel price index.

Through the test of multivariate model, we can preliminarily judge that there may be multicollinearity in the model.

3.3. Test of model

The initial model was tested by multicollinearity test, autocorrelation test and heteroscedasticity test, and the optimal model was obtained by comparison.

3.3.1. Multicollinearity test

In this paper, the multiple collinearity of the model is determined by the auxiliary regression model test and variance expansion factor analysis. The statistical table of auxiliary regression model parameters for each explanatory variable is as follows.

Table 2. Statistical table of auxiliary regression model parameters

Model	R2	F Statistic	Adjoint probability of F	VIF	TOL
X1=f(X2,x3,x4,x5,x6)	0.939734	34.30488	0.000002	16.59310	0.060226
X2=f(X1,x3,x4,x5,x6)	0.963655	58.33158	0.000000	27.51410	0.036345
X3=f(X1,x2,x4,x5,x6)	0.997723	964.1609	0.000000	439.1744	0.002277
X4=f(X1,x2,x3,x5,x6)	0.872139	15.00618	0.000136	7.820993	0.127861
X5=f(X1,x2,x3,x4,x6)	0.489377	2.108459	0.000000	1.958392	0.510623
X6=f(x1,X2,x3,x4,x5)	0.997617	921.14	0.000000	419.6391	0.002383

In the table, most of the variance expansion factor (VIF) is greater than 10 and most of the tolerance (TOL) is less than 0.1, indicating that the model has serious multicollinearity.

3.3.2. Autocorrelation test and generalized difference correction

In this paper, the partial autocorrelation coefficient method is used to test the autocorrelation of the model, and Eviews software is used to test the partial autocorrelation coefficient of the model. If the lag period is 12, the autocorrelation coefficient and partial autocorrelation coefficient of residual e_t and $e_{t-1}, e_{t-2}, \cdots e_{t-12}$ in each period will be obtained, as shown in the figure below.

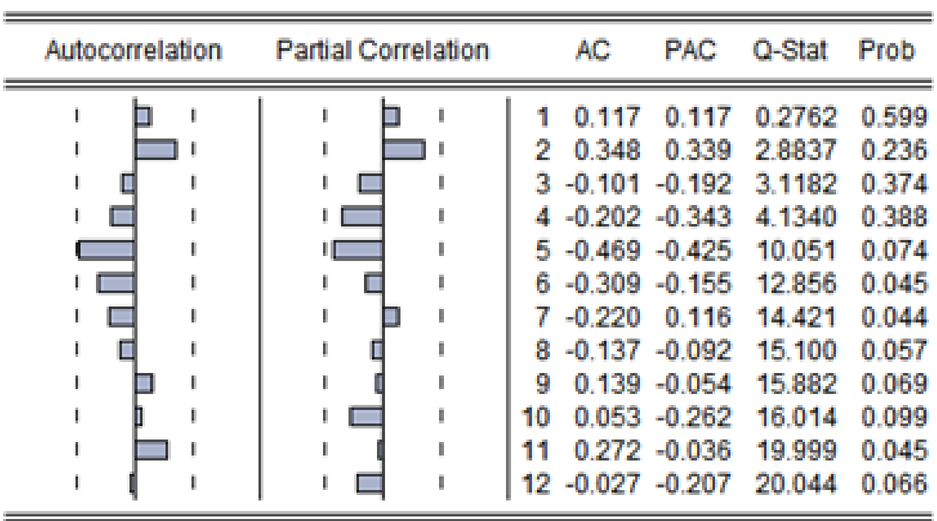


Figure 3. Partial correlation coefficient graph

It can be seen from the above figure that the histogram of PAC in the late lag phase is all in the dotted line, but some p values of Q statistics are greater than 0.05, indicating that the regression model may have autocorrelation.

In this paper, the generalized difference method is used to modify the model and eliminate autocorrelation. The above test shows that there is no first-order and second-order autocorrelation in the model, so the conclusion of t-statistics and F-statistics in the model estimated by the least square method is not credible, so the generalized difference method is needed to modify the model. By directly adding AR (1), AR (2), AR (1) and AR (2) in the LS command to detect the autocorrelation of the model, and comparing with the test results in the original model, the results are shown in the table below.

Table 3. Generalized difference correction results

AR	X1	X2	X3	X4	X5	X6	Ar(1)	Ar(2)	R^2
Ar(2)	0.0112 (3.6766)	-1.6923 (-4.2019)	0.0077 (4.5874)	53.6853 (3.9547)	9.7389 (3.0714)	-0.0128 (-4.4266)		1.0000 (1.3143)	0.8690
Ar(1) Ar(2)	0.0186 (11.8239)	-2.6890 (-13.6563)	0.0119 (15.1110)	85.4971 (15.4112)	14.4750 (10.5940)	-0.0200 (-14.1762)	-1.4883 (-2.5726)	-0.4883 (-1.3471)	0.9349
Ar(1)	0.0015 (0.6828)	-0.3643 (0.9132)	0.0015 (1.1977)	8.3983 (0.8292)	3.2814 (0.6210)	-0.0022 (-1.0410)	0.8403 (2.4375)		0.8604

Through the F-test, R2 and t-test of the model, it is found that the generalized difference method with AR (1) and AR (2) model is the best.

Y_D11=-9976.930+0.018559*X1_D11-2.689030*X2_D11+0.011920*X3_D11+85.49710*X4_D11+14.47495*X5_D11-0.019994*X6_D11+[AR(1)=-1.488339,AR(2)=-0.488339]

R2=0.934863 F=14.35217 DW=2.227252

After adjustment, the economic significance of the model is reasonable, and the adjustable determinable coefficient $\bar{R}^2$ is improved to 0.869725, which indicates that the model fits the sample data well; the F statistic is 14.35217, and its adjoint probability is 0.000537, which is close to zero, which indicates that x1, X2, X3, x4, X5 and X6 together have significant influence on the variables to be interpreted, and the overall linear relationship of the model is significant; the regression coefficients of each explanatory variable are all significant Remarkable. The output results show that AR (1) is -1.488339, the regression coefficient is significant, which indicates that there is a first-order autocorrelation AR (2) is -0.488339 and the t-test of the regression coefficient is not signifcant, which indicates that there is no second-order autocorrelation in the regression model.

Re test the partial autocorrelation coefficient of the model estimated by the generalized difference method, use Eviews software to test the partial autocorrelation coefficient of the model, and select the lag period as 12, then the autocorrelation coefficient and partial autocorrelation coefficient of each period of the residual and will be obtained, as shown in the figure below.

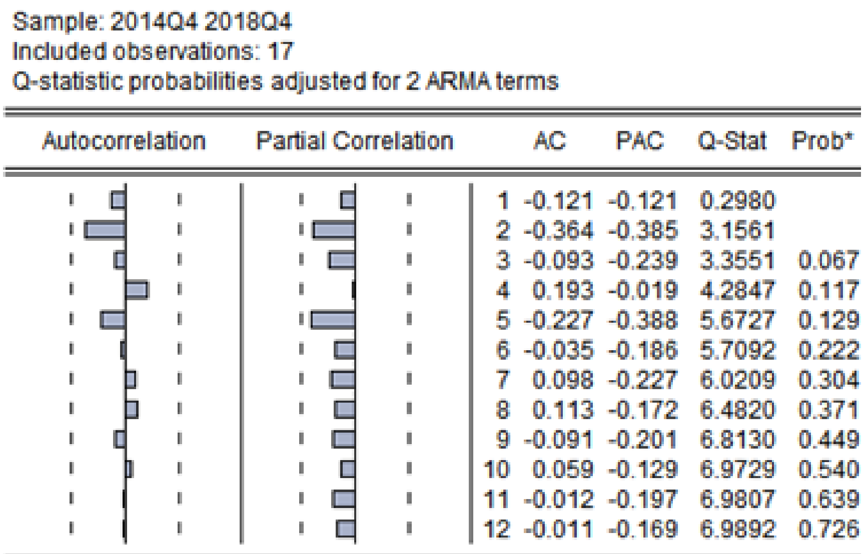


Figure 4. Partial correlation coefficient of the model after generalized difference correction

The histogram of PAC is in the dotted line, and the p value of Q statistic is greater than 0.05, indicating that there is no autocorrelation in the regression model.

3.3.3. Test the heteroscedasticity of the model modified by generalized difference

In this paper, the heteroscedasticity test of the model after generalized difference is carried out with the test of breusch pagan Godfrey. The test result is: LM = NR2 = $3.162290 < \chi^2_\alpha(k) = \chi^2_{0.05}(6) = 12.592$, and the accompanying probability prob (LM) is 0.7882, which is greater than the given significance level = 0.05. It is considered that there is no heteroscedasticity in the regression model.

3.4. Model comparison

Through multiple regression analysis [9] and its measurement test correction, the following models are obtained for comparison, as shown in the table below.

Table 4. Comparison of model results

Model	R^2	F statistic	Adjoint probability of F	T test significant
y=f(x1,X2,x3,x4,x5,x6)	0.753364	5.090939	0.012119	Not obvious
y=f(X1,x2,x3,x4,x5,x6,ar(2))	0.869035	0.767173	0.002326	Significantly
y=f(X1,x2,x3,x4,x5,x6,ar(1),ar(2)）	0.934863	0.869725	0.000537	Significantly
y=f(X1,x2,x3,x4,x5,x6,ar(1))	0.860412	0.751843	0.003038	Not obvious

Through the comparison of models, y = f (x1, X2, X3, x4, X5, X6, AR (1), AR (2)) is selected as the optimal model. Marginal analysis: estimated value of regression coefficient $\hat{b}_1$ = 0.018559 > 0, $\hat{b}_2$ = -2.689030 < 0, $\hat{b}_3$ = 0.0111920 > 0, $\hat{b}_4$ = 85.49710 > 0, $\hat{b}_5$ = 14.47495 > 0, $\hat{b}_6$ = -0.01994 < 0, it shows that the steel price index changes in the same direction with the steel sales volume, GDP, added value index of the second industry and labor cost price index; the steel price index changes in the opposite direction with the steel output and the total output value of the construction industry, and the steel price index increases by 0.018559 for every 10000 tons of steel sales volume increase and 0.018559 for every 1 million tons of steel output increase Price index decreased by 2.689030, steel price index increased by 0.0111920 for every 100 million Yuan of GDP increase, steel price index increased by 85.49710 for every 1 of secondary industry increase, steel price index increased by 14.47495 for every 1 of labor cost price index increase, and steel price index decreased by 0.01994 for every 100 million Yuan of construction industry increase.

4. Conclusion

Through the quantitative analysis of the econometric model, the iron and steel price index model is obtained, which can analyze and predict the price of iron and steel; quantitatively reflect the influence degree of the explanatory variables, so as to control the change of the iron and steel price index; the government can control the national economic development by controlling the macroeconomic indicators, affecting the iron and steel price. According to the optimal steel price index model, this paper gives the following suggestions [10-15]: optimize the industrial structure, strengthen the analysis of macroeconomic situation and optimize the import and export volume of steel industry.

5. Acknowledgments

Anhui Province teaching and research project "Research on discipline competition in the context of big data to cultivate innovative ability of new economic management talents" (Project No.: 2018jyxm1305).
Anhui University of Finance and economics teaching research project: (acxkjsjy201803zd; acjyb2018006).

References

[1] Zhou Yaobang, Wang Min. Empirical analysis and research on influencing factors of steel price. Modern Economic Information. 2018, (18), 336-340.
[2] Su Ying, Zhao Shuting. Econometric analysis of factors influencing the price of China's steel industry. Times Finance. 2012, (33), 76.
[3] Xu Yaping, Wang Tian. An empirical analysis on the influencing factors of steel price in China. Price Theory and Practice. 2013, (08), 58-59.
[4] Li Yujun. Analysis of influencing factors of steel price in China. Modern Economic Information. 2016, (34), 308.
[5] Wu Guoyue. An empirical study on the influencing factors of China's steel price. Modern Economic Information. 2016, (13), 350.
[6] Li Sainan, Chai Zhihui. A study on the mechanism of production fluctuation of agricultural products in China and Countermeasures. China Price. 2016, 000(001), 29-31.
[7] Wang Chang. The formation mechanism and preventive measures of stripping on the surface of hot rolled coil. China Metallurgy. 2018, 28(7), 46-49.
[8] Liu Chao. Analysis of China's steel price trend. Market Modernization. 2005, (14), 132.
[9] Wu Xiaodong. Research on steel price formation mechanism in imperfect competition market. China Steel. 2010, (12), 13-16.
[10] Yu Yizhan. Analysis on the risk of Chinese enterprises carrying out overseas business activities -- Taking steel enterprises as an example. China Business Theory. 2019, (21), 90-91.
[11] Chen Shengyong, Sun shiqi. Chinese Characteristics, Formation Mechanism and Governance Countermeasures of Overcapacity— Taking Steel Industry as an Example since 1996. Social Sciences in Nanjing. 2013, (05), 13-20.

[12] Chen Xiangyu. Case study on cost control of Hegang group in the context of supply side reform. Hebei Enterprise. 2019, (11), 44-45.

[13] Lu Minfeng, Luan Chengkai. Steel market financing risk mechanism and measures. Journal of Tianjin University of Commerce. 2012, 032(006), 3-8.

[14] Liu Shaoji. Agricultural product price formation mechanism, fluctuation effects and stable countermeasures. Commercial Time. 2014, 000(025), 24-26.

[15] Wang Jue. Analysis of typical financial management problems of iron and steel enterprises from the perspective of supply side Reform. Taxation. 2019, (28), 123-125.