



Correlation Analysis on Increasing and Decreasing of Futures Price

XingYu , Wei Hu , Zhenyan Li , Tao Tang

Department of Mathematics & Applied Mathematics of Hunan University of humanities, Science and technology

ABSTRACT

To hold the mixed futures directly decide whether people can better profit in the futures trading of futures price, so it is more and more important to study the situation of futures price. In view of this situation, this article studies the degree of correlation and its classification of eight kinds of material which take the date provided by the Shanghai futures exchange for promise, it true and reliable response to these eight kinds of material change, and for the investors to continue futures operations provided a basis.

Firstly, according to various materials of different contracts unified planning, through data processing and using means of various materials, all contracts of various materials represent a unified (Table 2). Then using of statistical methods in clustering analysis method to classify eight kinds of materials, combined with K means clustering algorithm, and using of SPSS and Matlab software to get the classification results as follows:

The first class	AL、 AU、 FU、 RB
Second kinds	CU
Third kinds	RU、 WR、 ZN

Keywords: mean value method, cluster analysis, K means clustering algorithm, SPSS software, Matlab software

1. INTRODUCTION

In China, future, as a kind of financial derivatives, is one of the effective tools for investors to avoid risks. Currently, futures has become many investors' investment tool, using futures can dodge the stock price fluctuation risks, but the futures development in our country in a relatively short time, futures breed is less and the main agricultural products, participants of futures market are mostly of hedging. In this case, people can not give a positive answer whether the futures market can really play a role in evading risk until China's futures market and spot market data were verified,. In view of this situation, futures investment futures subject matter from the degree of correlation between to analyze some inner connections so as to avoid risk better. In subject matter between various factors, price range is always affects every investor heart weapon, how to derive the correlation between them is particularly important. According to the 2010 Shanghai futures exchange data is the main source of subject matter, the futures price level of correlation between

study and classification, as part of the whole, the change rule, for investors to provide the very good reference.

2. RESEARCH METHODS

2.1 Data processing

As futures trading data is huge and some data must be handled accordingly to use software programming, such as text data must be converted into numerical data. This paper selects Shanghai futures exchange's data^[1] mainly because the Shanghai is an international city, the data are more representative, also on the research results have certain popularization. This paper mainly choose aluminum, gold, copper, oil, steel, rubber, wire, zinc eight kinds of materials in all contracts in 2010 price data to analysis and compare. Owing to the large number of each contract data, in order to facilitate research, using the average method to process data, the processed data are given in the following table (table 2):



Table 2: The Data after Processing

Material type Contract	AL	AU	CU	FU	RB	RU	WR	ZN
201001	72		32		10.6	111	11.75	-95.5
201002	-41.85	-0.43444	-254.81		-6.3		-9.57	-190.19
201003	-19.35	-0.22	-89.57	5.58	-2.33	-14.67	4.34	-84.13
201004	-21.76	0.32667	26.47	3.48	6.34	41.54	9.48	-53.9
201005	-39.94	0.39351	-79.66	4.22	-7.44	-28.09	-2.47	-82.92
201006	-27.29	0.01255	-79.17	-0.77	-7.21	-9.72	-2.26	-71.42
201007	-21.09	0.16057	-68.06	-2.53	-7.37	-20.74	-0.76	-56.51
201008	-14.24	0.12317	-28.94	-2.07	-4.32	-11.03	-1.61	-38.28
201009	-18.93	0.13742	-7.05	-1.98	-1.58	7.75	-0.7	-25.55
201010	-5.21	0.36543	36.1	-2.49	-2.22	40.81	0.62	-11.42
201011	-9.37	0.34191	28.55	-2.44	-2.01	58.58	1.16	-14.25
201012	-11.32	0.23817	27.57	-3.14	0.76		-0.84	-20.11
201101	79.39	0.18254	316.34	-2.58	22.91	171.23	30.57	83.66
201102	87.72	0.29068	351.65		22.89		-3.14	-0.94
201103	-3.8	0.43973	391.79	23.12	24.17	208.37	52.09	7.94
201104	-5.38	0.32915	457.07	34.92	28.32	208.08	107.16	123.46
201105	115.52	0.42798	518.56	-0.76	31.64	268.4	1.53	140.49
201106	21.78	2.41737	574.66	3.09	36.26	306.13	10.11	60.23
201107	8.13	6.08532	684.91	4.08	43.54	354.73	21.64	196.19
201108	245.42	3.77923	822.09	7.98	57.34	406.32	-8.71	38.33
201109	14.85	0.45108	248.21	71	69.61	584.78	280.13	27.5
201110	423.9	0.19735	205.47	117.6	89.29	685.83	35.83	352.82
201111	3.64	1.252	2137.35	146.77	3.88	1191.55	782.5	624.09
201112	14.09	1.43286	471	453.92	-30		15.5	267.5

2.2 The Theory of Cluster Analysis^[2]

Cluster analysis is a kind of multifactor classification problems in quantitative methods. The basic principle is based on the sample itself, using mathematical method according to some similarity or dissimilarity index, the quantitative determination of the relations between samples, and according to the relation degree of clustering. The common clustering analysis methods, such as system clustering method, dynamic clustering method, fuzzy clustering method.

2.2.1 Clustering Elements of the Data Processing

Assuming there are m objects of cluster, each object has a cluster of elements. They correspond to the elements of data

to be used in clustering analysis, commonly used clustering elements of the data processing methods are as follows.

(1) total standardization

$$x'_{ij} = \frac{x_{ij}}{\sum_{i=1}^m x_{ij}} \quad (i = 1, 2, \dots, m, j = 1, 2, \dots, n) \quad \text{and}$$

$$\sum_{i=1}^m x'_{ij} = 1, j = 1, 2, \dots, n$$

(2) The standard deviation standard

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (i = 1, 2, \dots, m, j = 1, 2, \dots, n)$$



$$\bar{x}_j = \frac{1}{m} \sum_{i=1}^m x_{ij}, s_j = \sqrt{\frac{1}{m} \sum_{i=1}^m (x_{ij} - \bar{x}_j)^2}$$

(3) The maximum value of standardization

$$x'_{ij} = \frac{x_{ij}}{\max_i \{x_{ij}\}} \quad (i = 1, 2 \dots m, j = 1, 2 \dots n)$$

After the standardization of the new data, various elements of the maximum value of 1, each of the remaining value is less than 1.

(1) The poor standardization

$$x_{ij} = \frac{x_{ij} - \min \{x_{ij}\}}{\max \{x_{ij}\} - \min \{x_{ij}\}} \quad (i = 1, 2 \dots m, j = 1, 2 \dots n)$$

After the standardization of the new data, various elements of the maximum value of 1, the minimum is 0, the rest of the values were in 0 and 1 between.

2.2.2 Distance Calculation

The distance is between things difference measurement, difference is bigger then the similarity distance is smaller, so the system cluster analysis based on distance of absolute value.

① Distance of absolute value $d_{ij} = \sum_{i=1}^n |x_{ik} - x_{jk}|$

② Euclidean distance $d_{ij} = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2}$

③ Mingke popovski distance $d_{ij} = \left[\sum_{k=1}^n |x_{ik} - x_{jk}|^p \right]^{\frac{1}{p}}$

Choose a different distance, the clustering results will vary. On the geographical division and classification, often using several distance calculation, comparison, select a more appropriate distance to cluster.

3. K MEAN CLUSTERING ALGORITHM ^[3]

3.1 The Principle of K-means Algorithm

The algorithm first randomly from the data set K as the initial clustering centers, second to calculates each sample to the cluster in the distance and then to send the samples to the nearest the clustering center of the class. Calculating mean of

formation data objects of every cluster to get new clustering center, if two adjacent of the cluster center without any changes, it illustrated with sample adjustment, clustering criterion function convergence. This algorithm is a characteristic in each iteration will examine each sample classification is correct. If incorrect, it is necessary to adjust, after adjustment in all samples, then to modify the cluster center into the next iteration. If the iteration of the algorithm, all the samples were correctly classified, there would be no adjusted, the clustering center also won't have any change, which marks the clustering criterion function convergence, thus algorithm is ending.

3.2 K-means Clustering Algorithm of the General Steps

- 1) The arbitrary choice of K objects from the N data object as the initial cluster center;
- 2) According to each cluster objects average (central object), to calculate the distance of each object to the center of the object; then according to the minimum distance to the corresponding object classification;
- 3) To calculate each (changed) cluster mean (the central object);
- 4) Cycle (2) to (3) until each cluster will not change.

4. RESEARCH RESULTS AND ANALYSIS

Based on researching the certainly data, It will be good for classify the eight materials by using cluster analysis. Using the above principle, using SPSS software to carry out a series of operations, the classification results are shown in table 3:

Cluster Membership

Case Number	Material type	Cluster	Distance
1	AL	1	355.020
2	AU	1	222.646
3	CU	2	0.00
4	FU	1	375.679
5	RB	1	172.735
6	RU	3	693.206
7	WR	3	480.475
8	ZN	3	449.619



In order to test the reliability of the results, further using K means clustering algorithm was tested, using Matlab programming, K mean clustering algorithm get the following results shown in table 4:

Table 4: Results, K means Clustering Algorithm

The first class	AL、AU、FU、RB
Second kinds	CU
Third kinds	RU、WR、ZN

Comparison analysis the above shows that the classification result is more reasonable and with some persuasion.

5. CONCLUSION

In this paper, using the method of cluster analysis is a good solution to the futures price of the subject matter and therefore classified which provide basis for investors to select, that is in the AL big growth rate, we can think AU, FU, RB also is in growth, investors can apply its laws to control, to gain considerable profits. But this classification also exist certain defects, because K means clustering algorithm itself has certain limitations and the data processing also has certain insufficiency, there are some room for improvement, such as the algorithm can use the ISODATA algorithm, not only the algorithm flexibility is relatively high but also the application is more wide. Overall, the result of classification is still relatively reasonable, good response to

the futures objects change the degree of correlation between and more practical.

ACKNOWLEDGMENT

This research was partially supported by Hunan University students research study and innovative pilot projects ; Education Topics of Hunan Province (Hunan teaching [2012 401); Teaching reform subject of Hunan university of humanities, science and technology (RKJGZ1210);A Project Supported by Scientific Research Fund of Hunan Provincial Education Department (12C0749).

REFERENCES

- [1]. Shanghai futures exchange data:
<http://www.shfe.com.cn/statements/download.html>
- [2]. Cluster analysis methods:
http://wenku.baidu.com/view/519a4f287375_a417866f_8fb7.html
- [3]. K means clustering algorithm:
http://wenku.baidu.com/view/e_bea2721aaea998_fcc220e3c.html II.
- [4]. Ou Chenwei 《K- mean clustering algorithm research and improvement》,Changsha University of Science and Technology 2011-3-1
- [5]. K mean clustering algorithm Matlab source:
<http://zhidao.baidu.com/>