Volume 3 No.2, February 2013 ISSN 2224-3577

International Journal of Science and Technology

http://www.ejournalofsciences.org

Discriminate Analysis in the Futures Price Forecasting Application

Xing Yu, Taotang, Junhua Jia
Department of Mathematics & Applied Mathematics of Hunan University of humanities, Science and technology

ABSTRACT

Fisher method is based on the analysis of variance to set up a better able to distinguish all general linear discriminate method. From more than one overall in extraction with P index sample observation data, by means of analysis of variance of the ideological structure of linear discriminate function. This paper forecasts the futures market price trend. First of all, select relevant variables, while discrimination, while selecting, and select the strongest discriminate ability variables, and using data discriminate analysis discriminate function can be obtained, and further to the discriminate function to do the test, we get the fitting degree is up to 79.3%, therefore, discriminate function has better forecast.

Keywords: *Discriminate analysis; Futures price forecasting; Discriminate function*

1. THE ESTABLISHMENT OF DISCRIMINATE FUNCTION OF FUTURES PRICE

Suppose that there are k populations $G_1, G_2, \cdots, G_k$, with mean vector $\bar{\mu}_1, \bar{\mu}_2, \cdots \bar{\mu}_k$ and covariance matrix $\sum_1, \sum_2, \cdots, \sum_k$, the samples are $x_{i1}, x_{i2}, \cdots, x_{in_1}$, where $i = 1, 2, \cdots, k$, $n_1 + n_2 + \cdots + n_k = n$;

The i^{th} population's sample mean is

$$\bar{X}_i = \frac{1}{n_i}\sum_{t=1}^{n_i} x_{it} \tag{1}$$

Integrated sample mean vector

$$\hat{X}_i = \frac{1}{n}\sum_{i=1}^{k} n_i \bar{X}_i \tag{2}$$

The i^{th} total sum of deviation square within sample

$$v_i = \sum_{t=1}^{n_i} (X_{it} - \bar{X}_i)(X_{it} - \bar{X}_i)' \tag{3}$$

Integrated within group variance

$$E = \sum_{i=1}^{k} v_i \tag{4}$$

Square deviation between groups

$$V = \sum_{i=1}^{k} n_i (\hat{X}_i - \bar{X}_i)(\hat{X}_i - \bar{X}_i)' \tag{5}$$

$$Y(x) = c_1 x_1 + c_2 x_2 + \cdots + c_k x_k \tag{6}$$

$$V_{iy}\sum_{i=1}^{n_i}(Y_{it} - \bar{Y}_i)^2 = \sum_{i=1}^{n_i}(Y_{it} - \bar{Y}_i)(Y_{it} - \bar{Y}_i)' = C'V_iC \tag{7}$$

$$E_0 = \sum_{i=1}^{k} V_{iy} = C'EC \tag{8}$$

$$B_0 = \sum_{i=1}^{k} n_i (\bar{Y}_i - \bar{Y})(\bar{Y}_i - \bar{Y})' = C'BC \tag{9}$$

From the above results if the discriminate analysis is valid, then all the samples of the linear combination is

$$Y(x) = c_1 x_1 + c_2 x_2 + \cdots + c_k x_k \tag{10}$$

It satisfies that group within the square deviation is small, squares are large, so the maximum of

$$\square^2(c) = \frac{B_0}{E_0} = \frac{C'BC}{C'EC} = \max \tag{11}$$

is related to the maximum characteristic root.

So we can get the Fisher sample discriminate function is

$$\hat{Y}_1(x) = \hat{c}_{11} x_1 + \cdots + \hat{c}_{p1} x_p \tag{12}$$

When the group number is too large, the more indexes are needed to be discussed, that is it is not enough for one discriminate function. Then we should to find the second one, and so on.

The maximum of $\Delta^2(C) = \frac{B_0}{E_0} = \frac{C'BC}{C'EC}$ is related to the second large characteristic root, whose feature vector consists of the second discriminate function coefficient

$$C_2 = (c_{12}, \cdots, c_{p2})'$$

$$\hat{Y}_2(x) = \hat{c}_{12} x_1 + \cdots + \hat{c}_{p2} x_p \tag{13}$$

Similarly, it finds m linear functions.

2. DETERMINE THE TYPES OF FUTURE PRICE

Suppose that the closed price difference of future at i is cp_i

(1) When $cp_i > \theta$, to determine the future price is rising.

(2) When $|cp_i| < \theta$, to determine the future price is flat.

(3) When $cp_i < -\theta$, to determine the future price is falling ;

where θ is determination coefficient .

3. MODEL SOLVING

3.1 . Choose the variances

The standard of choosing variances is:

i. Closely relating the purpose of discriminate analysis.
ii. Reflecting to sentence the class variable characteristics.
iii. Different research object value has obvious difference.

According to Zhengzhou and Shanghai futures market transaction data as follows: The former closed, Former settlement price, opening price, ceiling price, bottom price, Closing price, Settlement price, Ups and downs 1, Ups and downs 2, turnover, Transaction amount, Positions. All these variables into discriminate analysis using a futures rose or fell or flat.

Have the above variables from the futures market random choose a certain number of trading time medium futures contract in the operation of the SPSS to the result. Because of variables and with some relevant variables, So cut out variables such as Closing price, Ups and downs 1 and Ups and downs 2, Using the variable data such as The former closed, Former settlement price, opening price, ceiling price, bottom price, Settlement price, turnover, Transaction amount, Positions to Analyze its category in the day(Up or down or flat).

The θ is a coefficient of determination, it based on expert opinion as one-thousandth of the futures Current Prices.

3.2 Discriminate Function

Discriminate analysis of 152 sets of data on the two contracts following form:

Table 1: Eigenvalue

function	eigenvalue	variance %	Accumulation %	Canonical correlation
1	2.096[a]	94.5	94.5	.823
2	.121[a]	5.5	100.0	.329

2 canonical discriminate function used in the analysis

Table 3:Classification Function Coefficients

	VAR00010		
	1.00	2.00	3.00
VAR00001	.148	.200	.249
VAR00002	.084	.061	.057
VAR00003	-.282	-.254	-.247
VAR00004	.022	.028	.025
VAR00005	-.027	-.023	-.028
VAR00006	.507	.438	.396
VAR00007	-.006	-.006	-.006
VAR00009	.002	.002	.002
(Constant)	-3845.831	-3821.336	-3847.560

Fisher's linear discriminate function

Table 4: Standardized Canonical Discriminate Function Coefficients

	function	
	1	2
VAR00001	-7.548	-.759
VAR00002	2.012	2.638
VAR00003	-2.686	-3.234
VAR00004	-.214	-1.240
VAR00005	.115	-1.182
VAR00006	8.508	4.056
VAR00007	-.508	.218
VAR00009	.609	.087

The stone figure reveals that correct classification probability at 79.3%.

$$f_1 = -7.548x_1 + 2.012x_2 - 2.686x_3 - 0.214x_4 + 0.115x_5 + 8.508x_6 - 0.508x_7 + 0.609x_9$$
$$f_2 = -0.759x_1 + 2.638x_2 - 3.234x_3 - 1.240x_4 - 1.182x_5 + 4.056x_6 + 0.218x_7 + 0.087x_9$$

Substitution data $(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_9)$,According to the values of discriminate function f_1, f_2, Category belong to larger values.

4.EXAMINATION OF THE EFFECT OF DISCRIMINATE ANALYSIS

The testing result is as table 5

Table-5 :Classification results [b,c]

		VAR00010	predicted group membership			Total
			1.00	2.00	3.00	
Initial	Count	1.00	23	9	0	32
		2.00	0	17	0	17
		3.00	2	7	29	38
	%	1.00	71.9	28.1	.0	100.0
		2.00	.0	100.0	.0	100.0
		3.00	5.3	18.4	76.3	100.0
Cross-validation	Count	1.00	22	10	0	32
		2.00	0	15	2	17
		3.00	2	7	29	38
	%	1.00	68.8	31.3	.0	100.0
		2.00	.0	88.2	11.8	100.0
		3.00	5.3	18.4	76.3	100.0

a. Only for analysis in the case of cross-validation. At cross-validation, each case is classified in accordance with all other cases derived function from the outside of the case.
b. Correctly classified 79.3% in the case of the initial packet.
c. Correctly classified 75.9% in the case of cross-validation packet.

According to the data of Table 5, we can see that the probability of calculated discriminate correct is 79.3% and the discriminate result is good and the discriminate function is also good.

ACKNOWLEDGMENT

This research was partially supported by Hunan University research study and innovative pilot projects and Education Topics of Hunan Province. (NO 12C0749); Hunan Institute of science and humanities teaching reform subject（RKJGZ1210）;

REFERENCES

[1]. Minglian Han,Fisher multi-class discriminant analysis method application in forecasting of futures price[J].Journal of Statistics Education,2007,(11):25-28.

[2]. Zhihang Chen,Gansheng Cheng.Attribute recognition method and its application in futures forecasting.Systems engineering theory and Practice,1999,99(6):90-93.

[3]. Jie Zhu,Forecasting of futures price - BP neural network model, mathematics in practice and theory,2000,30(4):416-419.