Evaluation of China's Financial Risk Using Projection Pursuit Clustering Model and Its Positive Research

Jitong Lou, Wengao Lou, and Xiurong Yu

Abstract—**The evaluation index system for China's financial risk (CFR) and its single-index assessment criteria were used to generate sufficient samples. The projection pursuit clustering (PPC) model with a new objective function is applied to evaluate the CFR during 1994 to 2012. Four rules for establishing valid PPC model are put forward in this paper, and the real optimal projection vector coefficients as well as others parameters are obtained by applying artificial bee colony (ABC) algorithm. Positive research shows that PPC model is suited to evaluate CFR, and a valid PPC model as well as six sub-system models for six aspects of CFR is established. The CFR level during 1994 to 2012 generally lies on Proper-safety level except that in 2008 belongs to Risk level. In terms of the six aspects of CFR, the bank risk is the greatest and the second is the risk of economic growth.**

Index Terms—**Artificial bee colony (ABC) algorithm, China's financial risk, evaluation index system, projection pursuit clustering (PPC) model.**

I. Introduction

The financial crisis in one country could have devastating consequences to its financial industries and an adverse influence on other countries. Since the 2008 global financial crisis, governments around the world have been paying much more attention to early warning of financial risks and their control. Scholars in financial field and the financial industry regulatory authority have been making research to evaluate and prevent financial risk [1]. Various traditional models based on statistical methodologies have been used to study these problems abroad. But the situations and conditions for establishing these models vary from one country to another. Therefore, aforementioned models are not suited to assess China's financial risk (CFR). Meanwhile, some models are put forward and widely used in China, e.g. the comprehensive index method named AHP (the analytic hierarchy process), which use Delphi weighting or entropy weighting as well as integrated weighting method [1]-[3], factor analysis (FA) [4]-[5], and the combinations of these methods and neural networks (NN) [6]-[9]. In principle, AHP with Delphi-weighting is a subjective method: its rationality and validity totally depend on the ability and experience ofthe experts invited to judge the importance between every two indexes. Therefore, the objectiveness of AHP method is dubious. FA, an objective weighting-method, is rational and validate merely in the application conditions of large samples (e.g. the ratio of the number of samples to that of indexes is greater than 3-5) and Kaiser-Meyer-Olkin (KMO) is greater than 0.6 [10]. Otherwise, the validity of FA is unasserted-often incorrect and wrong. Recently researches in the field of NN have shown that NN has powerful pattern classification and pattern recognition capacities. However, there are so many factors that can affect the performance of NN. Gorr believes that NN can be more effective in the application conditions that problems have large data set and nonlinear structure [11]. The above-mentioned literatures applying NN do not conform to the application conditions of FA and NN [6]-[9]. Furthermore, NN belongs to the category of data-driven "black box" models; it cannot avoid the consequences of the "garbage in–garbage out" rule [12]-[13]. If the expected outputs of the training data set, from AHP or FA [6]-[10], are not correct, the results of NN are unlikely to firmly hold up. The financial risk evaluation index system is applied to establish robust and valid NN model with its single-index assessment criteria and interpolation values, which function as training data, verification data, and testing data set [14]. However, the modeling process is very complex and a lot of parameters need to be determined by researcher. So, the problem for China's financial risk evaluation and early warning is still unsolved and needs more study.

In fact, to evaluate CFR is to solve a multi-index problem with the help of reasonable mathematical tools. Friedman and Tukey have developed the projection pursuit clustering (PPC) technique [15], a powerful tool being widely and successfully used in many fields such as humanistic and social science, to deal with multi-index problems in China [16]. In this paper, the PPC technique is innovatively and experimentally introduced to evaluate CFR, and more robust and valid results are thus expected.

Manuscript received April 30, 2013; revised July 16, 2013. This work was supported in part by grants from Shanghai Municipal Innovation Training Projects for College Students, Shanghai Knowledge and Resource Service Platform ZF1226, and First-class Cultivated Discipline "Business Administration" and "Business Economics" of Shanghai Municipal Education Commission, Financial "085" Project of Shanghai Business School from Shanghai Municipal Education Commission.

Jitong Lou is with the Shanghai University, Shanghai, 200444, P R China (e-mail: jitonglou@gmail.com).

Wengao Lou and Xiurong Yu are with the Shanghai Business School, Shanghai, 200235, P R China, corresponding author (e-mail:wlou64@126.com, wglou@sbs.edu.cn, yuxr11@163.com).

II. Establishment of the Evaluation Index System for China's Financial Risk

Various evaluation index systems for CFR were put forward in previous studies [1]-[9] and [14]. In order to study the valid and the robust of these methods conveniently, the evaluation index system in [6] and [14] is applied in this paper and consists of six aspects: economic growth risk (EGR), fiscal risk (FR), monetary risk (MR), international payment risk (IPR), stock market risk (SMR) and bank risk

(BR). As shown in Table I, the system is composed of 20 evaluation indexes: GDP growth rate (GDPGR), fixed-asset investment growth rate (FAIGR), financial debt dependency(FDD), debt to GDP ratio (DGDPR), proportion of government revenue to GDP (GR/GDP), inflation rate (IR), proportion of M2 growth rate to GDPGR (M2GR/GDPGR), proportion of credit growth rate to GDPGR (CGR/GDPGR), proportion of current account balance to GDP (CAB/GDP), short-term debt ratio (STDR), proportion of foreign exchange reserves to total imports and exports (FER/TIE), liability ratio (LR), debt service ratio (DSR), foreign debt ratio (FDR), stock P/E ratio (P/E), proportion of total market capitalization to GDP (TMC/GDP), capital adequacy ratio (CAR), non-performing loan ratio (NPLR), deposit to loans ratio (DLR), and medium and long loans ratio (MLLR). China's financial risk situation is reflected by all of these evaluation indexes synthetically.

III. Principle of the Projection Pursuit Clustering Technique

Friedman and Tukey developed the projection pursuit (PP) principle, a powerful tool dealing with high-dimensional problem, to find out the one-dimensional projection direction which has the main characteristics of data [15]-[17]. The characteristics of high-dimensional data then can be analyzed by studying the structure and characteristics of one-dimensional data. So far, PPC model, which is a powerful tool in multi-index clustering or evaluation, has been widely and successfully applied to science, humanity, and engineering fields in China [15]-[17]. PPC model can find out a specific and interesting projection direction to reveal the complex structure of high-dimensional data-usually non-linear as well as abnormal distribution. Furthermore, projection pursuit uses trimmed global measures with the advantage of robustness against outliers. The principles and steps of establishing PPC model are as follows.

TABLE I: The Evaluation Index System for China's Financial Risk and its Single-Index Assessment Criteria

Risk in sub-system	Indexes	Financial risk level			
		Safety	Proper-safety	Risk	High-risk
EGR	GDPGR ($x_1 \geq$)[a]	12	8	4	<4
	FAIGR (x_2)	13-19	10-13 or 19-22	7-10 or 22-25	<7 or >25
FR	FDD ($x_3 \leq$)	20	35	50	>50
	GDPGR ($x_4 \leq$)	15	20	25	>25
	GR/GDP ($x_5 \geq$)	24	20	15	<15
MR	IR ($x_6 \geq$)	3	6	9	>9
	M2GR/GDPGR ($x_7 \geq$)	2	2.5	3	>3
	CGR/GDPGR ($x_8 \geq$)	1.5	2.2	3	>3
IPR	CAB/GDP ($x_9 \geq$)	3	4.5	5	>5 or <0
	STDR ($x_{10} \geq$)	15	25	35	>35
	FER/TIE ($x_{11} \geq$)	50	30	20	<20
	LR ($x_{12} \geq$)	15	25	40	>40
	DSR ($x_{13} \geq$)	10	20	40	>40
	FDR ($x_{14} \geq$)	60	80	100	>100
SMR	P/E ($x_{15} \geq$)	40	60	80	>80
	TMC/GDP ($x_{16} \geq$)	30	60	90	>90
BR	CAR ($x_{17} \geq$)	12	8	4	<4
	NPLR ($x_{18} \geq$)	12	17	22	>22
	DLR ($x_{19} \geq$)	60	75	85	>85
	MLLR ($x_{20} \geq$)	10	15	20	>20

a. The unit of all indexes is % except x_7 and x_8 in value.

A. Data Preprocessing

Denote the original value of the *j*th index of the *i*th sample data by $x_{ij}^{'}$ (where $i=1,2,...,n$, $j=1,2,...,p$, n is the number of sample data and p is the dimension). In order to make the results entirely independent of the ranges of values or the units of measurements, we should bring all values to compatible units with a mean of 0 and standard deviation of 1. For a positive (or benefit) and a negative (or cost) index, the follow transformation has a wide variety of applications,

$$\begin{cases} x_{ij} = (x_{ij}^{'} - \bar{x}_j^{'}) \big/ \sigma_j^{'} \\ x_{ij} = -(x_{ij}^{'} - \bar{x}_j^{'}) \big/ \sigma_j^{'} \end{cases} \tag{1}$$

where $\bar{x}_j^{'}$ and $\sigma_j^{'}$ are the mean value and the root mean square error of the *j*th index.

If the *j*th index is a moderate index (e.g. x_2, with its best value $x_{jbest}^{'}$), we should firstly change it into a negative index by $x_{ij} = \left| x_{ij}^{'} - x_{jbest}^{'} \right|$ and then standardize it by (1).

B. Linear Projection

In principle, projection pursuit is used to analyze the characteristics of high-dimensional data from all directions; projection pursuit aims to expose the hidden characteristics or structure of the high-dimensional original sample data by comparing and searching all low-dimensional projections. Let $\vec{a} = (a_1, a_2, ..., a_p)$ be a p-dimensional unit vector, we can calculate the *i*th sample projection value Z_i,

$$Z_i = \sum_{j=1}^{p} a_j x_{ij} \tag{2}$$

where a_j is the projection vector coefficient of the *j*th index.

The samples can be ranked and classified in terms of their projection values, from the maximum to the minimum, in one-dimensional direction. Also, all the indexes can be ranked and classified according to their optimal projection vector coefficients.

C. Optimization of PPC Model Objective Function

The widely used optimization objective function of PPC model, proposed by Friedman and Tukey, is to maximize the production of standard deviation S_z of the samples' value Z_i and the window density value D_z [15]-[17],

$$Q(a) = \max(S_z \times D_z) \tag{3}$$

where S_z and D_z measure the spread and the "local density" of the samples' projection value in $\vec{a}$ direction.

However, the widely used PPC model does have an annoying problem in practice: how to determine the rational cutoff radius R, which has significant effects on the results? Hitherto most researchers still set R based on their experience or by trial and error method. Friedman and Tukey suggest that R should equal to 10% of S_z [15]. Fu and Zhao state that R should be $(r_{ik})_{\max} + \frac{p}{2} \le R \le 2p$ (where $(r_{ik})_{\max}$ is the maximum value of $r_{ik} = |Z_i - Z_k|$); specially, let R be the samples' dimension p [16]. Meanwhile, Lou and Qiao propose that R should be $(r_{ik})_{\max}/5 \le R \le (r_{ik})_{\max}/3$ [17]. Different R values reveal various characteristics in different projection directions. How can we dig the samples' characteristics out by using PPC model (3) with different R values? In order to solve the annoying problem, we have to put forward another reasoned optimization objective function of PPC model.

D. Generation of Sufficient Sample Data

According to Table I, the only three critical-value samples (CVSs) dividing the financial risk (FR) situations into different levels are too few to establish a valid PPC model. So, we should generate sufficient sample data strictly according to Table I. As we have seen from Table I, different FR levels correspond to different ranges of each index. Therefore, we can generate sufficient sample data for various FR levels based on different ranges of the evaluation indexes. If we generate the random-value sample data in the following range of each index: $40 > x_1 \ge 25$, $1.8 > x_1 \ge 1.4$, $5 > x_1 \ge 1.0$, $5 > x_1 \ge 1.0$, $10 > x_1 \ge 5$, $2.0 > x_1 \ge 1.5$, $50 > x_1 \ge 40$, $1.2 > x_1 \ge 0.8$, $35 > x_1 \ge 20$, $50 > x_1 \ge 30$, and $10 > x_1 \ge 8$, the FR level of these sample data certainly belongs to the Proper-safety (PS) level. In this way, sufficient samples belonging to the PS level are thus generated. Sufficient samples belonging to other FR levels can be generated in the same way. After totally 640 samples are generated(160 samples for each level), we use 320 samples of them to establish PPC model in this paper. The other 320 samples are applied to check the effectiveness and robustness of the established PPC model. Denoted as Y_i, the different FR levels - Safety, PS, Risk and High-risk (HR) - can be supposed as values 1, 2, 3, and 4. These values Y_i (theoretical values of different financial risk levels) correspond to the samples' projection values Z_i.

E. New Optimization Objective Function of PPC Model

PPC model should contain as much information about the x_{ij} as possible. In other words, the standard deviation S_z of Z_i should be as great as possible. Meanwhile, the absolute value $|R_{yz}|$ of the correlation coefficient of samples' projection values Z_i and their theoretical FR levels Y_i should be as great as possible, too. According to the above principle, we introduce a new optimization objective function of PPC model as follows,

$$\begin{aligned} &Q(a) = \max(S_z \times |R_{yz}|) \\ &s.t. \sum_{j=1}^{p} a_j^2 = 1, \quad -1 \le a_j \le 1 \end{aligned} \tag{4}$$

where $|R_{yz}|$ is the absolute value of the correlation coefficient of samples' projection values Z_i and their corresponding theoretical FR levels Y_i, i.e.

$$|R_{yz}| = \left| (\textstyle\sum_{i=1}^{n} Z_i Y_i - n\bar{Z}\bar{Y}) \Big/ \left[\left(\sum_{i=1}^{n} Z_i^2 - n\bar{Z}^2 \right) \left(\sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2 \right) \right] \right| \tag{5}$$

where $\bar{Z}$ and $\bar{Y}$ are the mean values of Z_i and Y_i, respectively.

The restrictions upon the optimization objective function (4) are various: consisting of non-linear, equality, and inequality constraints. The optimization objective function itself is nonlinear, high-dimensional, and complex. Therefore, it is very difficult to solve (4) by traditional optimization techniques, and the searching process usually reaches local minimum. In this paper, we apply an artificial bee colony (ABC) algorithm to solve (4).

F. Principle of Artificial Bee Colony Algorithm

Artificial bee colony (ABC) algorithm is one of the most recently introduced swarm-based algorithms [18]-[20]. ABC simulates the intelligent foraging behavior of a honeybee swarm. An important difference between ABC and other swarm intelligence algorithms is that in the ABC algorithm the possible solutions represent food sources, not individuals (honeybees). The results of optimizing a large set of numerical test functions show that the performance of the ABC is better than or similar to those of other population-based algorithms, like genetic algorithm, particle swarm optimization (PSO) algorithm, and differential evolution algorithm, with the advantage of employing fewer control parameters and can be efficiently employed to solve engineering problems with high dimensionality [18]-[20].

In ABC algorithm, the colony of artificial bees contains three groups of bees: employed bees, onlookers, and scouts. The position of a food source represents a possible solution to the optimization problem and the nectar amount (fitness value) of a food source corresponds to the quality (fitness) of the associated solution. For every food source, there is only one employed bee. Every bee colony has scouts that are the colony's explorers. The scouts are characterized by low search costs and a low average in food source quality. Occasionally, the scouts can accidentally discover rich, entirely unknown food sources. The number of the employed bees or the onlooker bees is equal to the number of solutions in the population. Each solution $b_l(l=1,2,...,SN)$ is a D-dimensional vector, where SN denotes the size of population or the number of solutions and D is the number of variables of the problem. An employed bee produces a modification on the solution in her memory depending on the local information and tests the nectar fitness value of the new solution. Provided that the nectar fitness value of the new one is higher than that of the previous one, the bee memorizes the new solution and forgets the old one. Otherwise she keeps the solution of the previous one in her memory. The food source of which the nectar is abandoned by the bees is replaced with a new food source by the scouts. The employed bee of an abandoned food source becomes a scout. An artificial onlooker bee chooses a food source depending on the probability value associated with that food source, p_l , calculated by the following expression,

$$p_l = fit_l \Big/ \sum_{k=1}^{SN} fit_k \tag{6}$$

where fit_l is the fitness value of the solution *l* which is proportional to the nectar fitness value of the food source in the position *l*.

In order to produce a candidate food position from the old one in memory, the ABC uses the following expression,

$$\upsilon_{l,m} = b_{l,m} + \phi_{l,m}(b_{l,m} - b_{q,m}) \tag{7}$$

where $q \in \{1,2,...,SN\}$ and $m \in \{1,2,...,D\}$ are randomly chosen indexes. Although *q* is determined randomly, *q* has to be different from *l*. $\phi_{l,m}$ is a random number between [-1, 1]. A greedy selection mechanism is employed as the selection operation between the old and the candidate one. Provided that a position cannot be further improved through a predetermined number of cycles, the food source is assumed to be abandoned. So, the value of predetermined number of cycles, which is called *"limit for abandonment"* [18]-[20], is an important control parameter of the ABC algorithm. In the ABC, the parameter *limit* is equal to *SN*D*. There are three important control parameters used in the ABC: the number of food sources which is equal to the number of employed or onlooker bees (*SN*), the value of *limit*, and the maximum cycle number (*MCN*). Assume that the abandoned source is b_l and m , then we must replace b_l with a new food source that the scout discovers. This new food source b_l can be defined as follows,

$$b_{l,m} = b_{\min,m} + rand[0,1](b_{\max,m} - b_{\min,m}) \tag{8}$$

In ABC algorithm, while onlookers and employed bees carry out the exploitation process in the search space, the scouts control the exploration process. Detailed pseudo-code of ABC algorithm is given as follow [18]-[20],

- Initialize the population of solutions $b_l(l=1,2,...,SN)$
- Evaluate the population
- cycle =1
- repeat
- Produce new solutions υ_l for the employed bees by using (7) and evaluate them
- Apply the greedy selection process for the employed bees
- Calculate the probability values p_l for the solutions b_l by using (6)
- Produce the new solutions υ_l for the onlookers from the solutions b_l selected depending on p_l and evaluate them
- Apply the greedy selection process for the onlookers
- Determine the abandoned solution for the scout, if exists, and replace it with a new randomly produced solution b_l by using (8)
- Memorize the best solution achieved so far
- cycle = cycle+1
- until cycle = *MCN*

G. Rules for Judging Whether the Optimizing Searching Process Reaches the Real Global Optimal

Although there are more than ten different swarm intelligence optimization algorithms applied to optimize the PPC model objective function, the question about how to judge whether the searching process reaches the real global optimal or not is stills unsolved. After concerning the characteristics of the objective function, Z_i , and $\vec{a}$, we thus put forward the rules from both theoretical and practical perspectives as follows:

Rule 1: If the *j*th index is standardized with positive mode and its coefficient is a_j , then the coefficient will be $-a_j$ when the *j*th index is standardized with negative mode. But the values of $|R_{yz}|$, S_z and $Q(a)$ all remain unchanged.

Rule 2: If $x_{ij}(i=1,2,...,n)$ is a constant, then its coefficient a_j must be equal to zero.

Rule 3: If $x_{ij} = x_{ik}(j \neq k)$, then the coefficient $a_j = a_k$.

Rule 4: If $\vec{a}$ is the optimal projection vector, then $-\vec{a}$ must be the optimal one as well.

Based on above rules, we can assume dummy indexes $x_{i(p+1)} \equiv 1$, $x_{i(p+2)} = x_{ip}$, and $x_{i(p+3)} = 1 - x_{ik}$. If the coefficients $a_{p+1} = 0$, $a_{p+2} = a_p$, and $a_{p+3} = -a_k$, we can, almost undoubtedly, judge that the optimizing searching process does reach the real global optimal. Additionally, in that case, the program developed by authors and the applied optimal parameters such as *SN* and *MCN* are valid and reliable. Now, we can delete the two dummy indexes $x_{i(p+2)}$ and $x_{i(p+3)}$, and then rerun the program to obtain the real optimal projection vector $\vec{a}$, its coefficients $a_1, a_2, ..., a_p$, the value of Z_i , $|R_{yz}|$, S_z , and $Q(a)$. In practice, if we change the standardization mode of odd indexes, then we can absolutely judge that the optimizing searching process has reached the real global optimal point under the circumstance that not only the coefficients of those indexes become the opposite, but also all the values of Z_i, $|R_{yz}|$, S_z and $Q(a)$ remain unchanged. Otherwise, the optimizing searching process hasn't reached the real global optimal.

H. Analysis of the CFR Level and Indexes' Importance

According to the samples' projection values Z_i, we can analyze the CFR situation and its variation tendency with time. According to the absolute value and the sign of each index' coefficient, we are able to analyze and judge the importance and the characteristics (being positive or negative index) of each index. Furthermore, we can rank them in order.

IV. Positive Research

A. To Establish PPC Model for Evaluating China's Financial Risk level

The computer program is developed by applying MATLAB 7.0 and the ABC algorithm. We input the 320 sample data for establishing (SDE) PPC model into the program. We adjust the ABC algorithm parameters such as *MCN* and *SN*, change the standardization mode of odd indexes and check whether the projection vector coefficients of odd indexes equal to their opposite value. Finally, according to the above rules, we can confidently judge that the optimization searching process reaches the real global optimal, and thus we can obtain the optimal projection vector $\vec{a}$ and its coefficients, the values of $|R_{yz}|$, S_z, and $Q(a)$. The optimization results with different standardization mode of odd indexes are shown in Table II (Case 1 and 2). Till now, we obtain the optimization results: $\vec{a}$ = (0.2164, 0.1914, 0.2333, 0.2294, 0.2304, 0.2106, 0.2284, 0.2332, 0.2204, 0.2286, 0.2034, 0.2308, 0.2258, 0.2328, 0.2325, 0.2336, 0.2326, 0.2244, 0.2150, 0. 2133), S_z =4.1719, $|R_{yz}|$ =0.9923, and $Q(a)$ =4.1398. The calculated coefficient of each index shows that the importance of index almost equals to each other.

According to the optimization results, we calculate the samples' projection values Z_i by (2) (there are so many values that we cannot list them here). The coupling relationship between the projection values Z_i and their corresponding theoretical values Y_i is shown in Fig. 1.

Fig. 1 shows that the coupling relationship, for each i, obviously obeys to logistic curve and can be described as:

$$Y_i = \frac{4.5}{1+\exp(c_0 + c_1 Z_i)} \tag{9}$$

We can obtain the optimal c_0 and c_1 by minimizing the following function:

$$Q(c) = \min \sum_{i=1}^{n} \left[Y_i - \frac{4.5}{1+\exp(c_0+c_1 Z_i)} \right]^2 \tag{10}$$

TABLE II: THE OPTIMIZATION RESULTS WITH DIFFERENT STANDARDIZATION MODE OF ODD INDEXES FOR SDE AND SDC

Index x_i	**Case 1**[b]	**Case 2**	**Case 3**	**Case 4**
x_1	0.2164	0.2164	0.2144	0.2144
x_2	0.1914	-0.1914	0.1919	-0.1919
x_3	0.2333	0.2333	0.2326	0.2326
x_4	0.2294	-0.2294	0.2309	-0.2309
x_5	0.2304	0.2304	0.2303	0.2303
x_6	0.2106	-0.2106	0.2110	-0.2110
x_7	0.2284	0.2284	0.2301	0.2301
x_8	0.2332	-0.2332	0.2321	-0.2321
x_9	0.2204	0.2204	0.2193	0.2193
x_{10}	0.2286	-0.2286	0.2289	-0.2289
x_{11}	0.2034	0.2034	0.2038	0.2038
x_{12}	0.2308	-0.2308	0.2319	-0.2319
x_{13}	0.2258	0.2258	0.2259	0.2259
x_{14}	0.2328	-0.2328	0.2318	-0.2318
x_{15}	0.2325	0.2325	0.2314	0.2314
x_{16}	0.2336	-0.2336	0.2328	-0.2328
x_{17}	0.2326	0.2326	0.2330	0.2330
x_{18}	0.2244	-0.2244	0.2262	-0.2262
x_{19}	0.2150	0.2150	0.2157	0.2157
x_{20}	0.2133	-0.2133	0.2125	-0.2125
x_{21} [a]	0.0000	0.0000	0.0000	0.0000
c_0	-0.3163	-0.3163	-0.3147	-0.3147
c_1	-0.2961	-0.2961	-0.2951	-0.2951
S_z	4.1719	4.1719	4.1706	4.1706
$\lvert R_{yz}\rvert$	0.9923	0.9923	0.9930	0.9930
$Q(a)$	4.1398	4.1398	4.1413	4.1413
$Q(c)$	1.7944	1.7944	1.7843	1.7843

[a] $x_{21} \equiv 1$, be constant; [b] Case 1 and 2, odd indexes with different standardization mode using 320 SDE; Case 3 and 4, odd indexes with different standardization mode using 320 SDC.

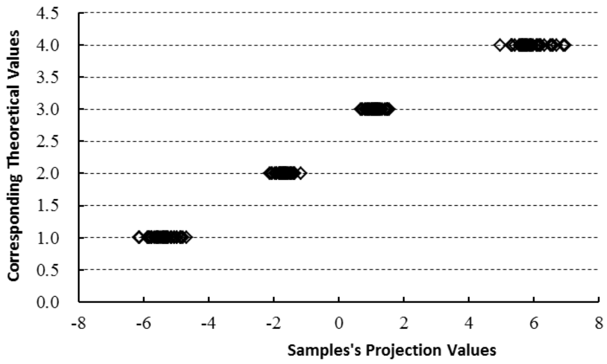

Fig. 1. The coupling relationship between samples' projection values Z_i of PPC model and their responding theoretical values Y_i

where the constant 4.5 is the theoretical maximum value of Y_i. We also apply the ABC algorithm to obtain the optimal values: c_0=-0.3163, c_1=-0.2961, and $Q(c)$=1.7944 (shown in Table II, Case 1 and 2). The calculated values Y_i by (9) are called the risk score, namely, the actual risk situation of China financial industry.

B. To Check the Reliability and Performance of PPC Model

We rerun the program with the other 320 sample data for checking (SDC) the reliability of the established PPC model. The optimal results-projection vector $\vec{a}$ as well as its coefficients, the values of S_z, $|R_{yz}|$, $Q(a)$, c_0, and c_1 - are all shown in Table II (Case 3 and 4). We compared these optimization results with the previous obtained results using 320 SDE: the maximum absolute errors of the optimization results are all less than 0.003. Such small errors show that the above-established PPC model of China's financial risk is quite reliable and validate. It also shows that the PPC model with the new optimization objective function and the ABC algorithm are robust and suitable for studying and analyzing the China's financial risk.

In order to analyze the performance of the established PPC model, we should calculate the main model's performance indicators: mean absolute error (MAE), mean absolute relative error (MARE, %), maximum error ($E_{\max}$), maximum relative error ($RE_{\max}$, %), and the distribution of absolute error (AE). The values of these model's performance indicators are calculated and shown in Table III. Refer to Table III, we can surely conclude that the PPC model is effective and validate not only for SDE but for SDC.

C. To Calculate the Critical-Value Samples' Outputs and Distinguish China's Financial Risk Level

After inputting the three critical-value sample data into the above-established PPC model, we can obtain their corresponding outputs: 1.6131, 2.5207, and 3.3596. Therefore, as shown in Table IV, the output ranges of the synthesized PPC model of China's financial risk - (0, 1.6131], (1.6131, 2.5207], (2.5207, 3.3596], and (3.3596, 4.5000]- correspond to the four financial risk levels: Safety, Proper-safety, Risk and High-risk, respectively. In the same way, we can obtain the output ranges of the six sub-systems corresponding to the different four financial risk levels, which are also shown in Table IV. By studying these output ranges, we can analyze and judge that, in every year, to which risk level China's financial industry and each of its sub-system belongs.

D. To Analyze and Judge China's Financial Risk Level

The financial risk usually has one-year incubation, so we should use operation data of China's financial industry in1993 to predict the financial risk level in 1994. By inputting the collected operation data - from [6], [14], Almanac of China

Finance and Banking (2009-2012), Finance Yearbook of China (2009-2012), and China Statistical Yearbook (2009-2012) - into the PPC model, we obtain the outputs of the PPC model describing China's financial risk levels during 1994-2012, which are shown in Table V. According to the output ranges of different financial risk level in Table IV, we can judge China's financial risk level during 1994-2012 and display them in Table V. The outputs of sub-systems' PPC models and the sub-systems' risk level of China financialindustry during 1994-2012 are calculated and shown in Fig. 2.

TABLE III: COMPARISON OF MAIN MODEL'S PERFORMANCE INDICATORS BETWEEN SDE AND SDC

sample	MAE	MARE	$E_{\max}$	$RE_{\max}$	AE ≤ 0.05	AE ≤ 0.1	AE ≤ 0.2	AE ≤ 0.25
SDE	0.060	3.35	0.217	18.2	51%[a]	81%	99%	100%
SDC	0.059	3.25	0.223	18.7	50%	84%	99%	100%

[a]The ratio of the number of samples whose AE less than 0.05 to the total number of samples is 51%.

TABLE IV: THE OUTPUT RANGES OF PPC MODELS FOR VARIOUS FINANCIAL RISK LEVEL

PPC model	Financial risk level			
	Safety	Proper-safety	Risk	High-risk
EGR ≤	2.541	2.594	2.647	> 2.647
FR ≤	2.441	2.596	2.757	> 2.757
MR ≤	2.451	2.588	2.729	> 2.729
IPR ≤	2.312	2.588	2.865	> 2.865
SMR ≤	2.480	2.599	2.717	> 2.717
BR ≤	2.397	2.571	2.728	> 2.728
Synthesized ≤	1.613	2.521	3.360	>3.360

TABLE V: PPC MODEL OUTPUTS AND CHINA'S FINANCIAL RISK LEVEL DURING 1994~2013 BASED ON DIFFERENT MODELING METHODS

	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
Outputs	2.275	1.958	1.808	1.895	2.153	2.211	2.218	2.063	2.193	2.148
FR level	II	II	II	II	II	II	II	II	II	II
[6]	I	I	I	II	II	III	III	III	IV	IV
[14]	II	II	II	II	II	II	II	II	II	II
FA	II	II	I	II	II	II	II	II	II	II
EWTOPSIS	II	II	II	II	II	II	II	II	II	II

	2004	2005	2006	2007	2008	2009[a]	2010	2011	2012
Outputs	2.236	2.040	2.042	2.208	2.635	2.214	2.360	2.226	2.169
FR level	II	II	II	II	III	II	II	II	II
[6]	III	IV	IV	IV	IV				
[14]	II	II	II	II	III	II	III	II	II
FA	II	I	I	I	I	I	I	I	I
EWTOPSIS	II	II	II	II	II	II	II	II	II

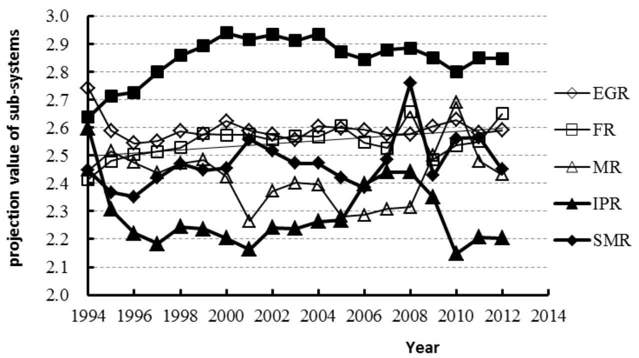

Fig. 2. The output of six sub-systems' PPC models during 1994-2012

During 1994-2012, China's financial risk generally lies on the Proper-safety level except for 2008. The Financial risk situation varies from different years (as shown in Table V): the lowest is in 1997, and the highest is in 2008. The financial risk situation in the period of 2008-2012 is higher than that in any other periods.

In terms of the six sub-systems, the bank risk is the highest and lies on the Risk level for 17 years. The financial risk situation of the economic growth is next to that of the bank risk. The financial risk situation of the stock market is the lowest. The financial risk situation of the two sub-systems of the international payment and the monetary is safe except that 2010 belongs to Risk level.

E. To Compare the Results Using PPC Model and Using Other Modeling Methods

During 1994-2012, the evaluation, which uses PPC model established in this paper, of the risk level of China's financial industry is good agreement with that in [14] applying artificial neural networks (ANNs) modeling technique and generating efficient sample data. Meanwhile, during 1994-2008, the risk level using PPC model is quite different from that in [6] applying factor analysis modeling technique (FA) and critical-value samples. We also apply FA to model using the collected data during 1993-2011, and we discover that the KMO=0.352 and the ratio of the number of sample data to the number of indexes is 1.1 - the application conditions of applying FA are not satisfied with and the results is not surely reliable and validate. The synthesized factor score (*FS*) is calculated as follows,

$$\begin{aligned} FS = {} & 0.041x_1 + 0.081x_2 - 0.023x_3 + 0.016x_4 + 0.047x_5 \\ & -0.013x_6 + 0.044x_7 + 0.037x_8 + 0.060x_9 - 0.029x_{10} \\ & +0.041x_{11} + 0.065x_{12} + 0.076x_{13} + 0.063x_{14} + 0.098x_{15} \\ & +0.044x_{16} + 0.052x_{17} + 0.059x_{18} - 0.025x_{19} - 0.034x_{20} \end{aligned}$$

The *FS* formula shows that the indexes x_3, x_6, x_{10}, x_{19}, and x_{20} are negative. In fact, there is a contradiction: these indexes are all standardized by (2) and definitely positive. In other words, if the application conditions of using FA cannot be satisfied with, the results will be wrong and unreliable. During 1994-2012, the China's financial risk level using FA is also shown in Table V. The result is quite different from that using PPC model and ANNs.

As one of the multi-index system engineering decision-making methods, the combination method of information entropy-weighting and TOPSIS (technique for order preference by similarity to ideal solution) (EWTOPSIS) is widely used in practice [1], [21]-[22]. So, we apply the EWTOPSIS to model China's financial risk based on the collected data during 1993-2011 and the three critical-value sample data. According to the modeling principles, we obtain the entropy-weight ω_j = (0.0260, 0.0573, 0.0338, 0.0455, 0.0347, 0.0673, 0.0563, 0.0591, 0.0520, 0.0806, 0.0268, 0.0555, 0.0694, 0.0610, 0.0501, 0.0552, 0.0407, 0.0536, 0.0277, 0. 0473). The calculated results in Table V show that the risk level of China's financial industry during 1994-2012 generally belongs to Proper-safety level. The financial risk level in 2008 is higher than that in any other years, and the lowest one is in 1997. The results using EWTOPSIS are almost, except the year 2008, in good agreement with that using PPC model in this paper. They also agree with the results calculated by artificial neural networks [14] except for 2008 and 2010.

In terms of the importance of indexes, the ratio of ω_{10} to ω_1 is 3.1; the 10th index is much more important than the 1st index. The result is different from that using PPC and ANNs.

V. Conclusions

The PPC model with the new optimization objective function $\max(S_z * |R_{yz}|)$ is applied to evaluate China's financial risk level and overcome the drawback of traditional PPC model: the model is greatly influenced by the cutoff radius R. Some rules for building up a reliable and valid PPC model are put forward in this paper and then applied to judge whether the optimization searching process does reach the real global optimal or not. The ABC algorithm is used to optimize the objective functions and obtain the real optimal projection vector as well as its coefficients a_j, samples' projection values Z_i, the values of S_z, $|R_{yz}|$, $Q(a)$, c_0, c_1, and $Q(c)$. Therefore, the actual risk level of China's financial industry during 1994-2012 can be evaluated.

The evaluation index system for China's financial risk and its single-index assessment criteria are used to generate sufficient samples. The positive research shows that the above-mentioned rules are suitable and feasible in practice, and a reliable and valid PPC model for evaluating China's financial risk situation is built up in this paper. The China's financial risk level during 1994-2012 lies on Proper-safety level generally, except in 2008 lying on Risk situation. The PPC sub-models of the six financial risk sub-systems are also built up in this paper. In terms of the six financial risk sub-systems, the bank risk is the greatest, and the economic growth risk comes the next. The results also indicate that China's financial risk is not serious and under proper controlled. In other words, the proactive fiscal policy and the prudent monetary policy that China's government have been implementing for years played a positive and effective role in maintaining China's financial operation stability. Furthermore, it also shows that bank industry should improve her operating ability, try her best to take measures to improve the management and efficiency of fund, and raise her capability of competition and risk protection.

References

[1] L. Tang, *China's Financial Security Report: Early Warning and Risk Elimination*, Beijing: Red Flag Press, 2009, ch1, 5, pp. 1-10, 183-231.

[2] C. Wu, "Study on financial risk warning index system in our country," *Technoeconomics and Management Research*, no. 1, pp. 19-24, Jan. 2011.

[3] Y. Shen and Z. Zhang, "Establishment research on precaution system of Chinese financial security," *Journal of ShanXi Finance and Economics University*, vol. 29, no. 10, pp. 89-94, Oct. 2007.

[4] Y. Hu and H. Gao, "Factor analysis: China's financial risk and its countermeasures," *Journal of Guizhou College of Finance and Economic*, vol. 101, no. 6, pp. 10-13, Dce. 2002.

[5] Y. Hong, "The construction and empirical analysis of nonparametric early-warning system of fiscal risk based on risk factor method and AHP," *Journal of Guangdong University of Business Studies*, vol. 26, no. 6, pp. 12-23, Nov. 2011.

[6] Q. Chen, Y. Xue, and L. Xiao, "Early warning of financial risk: indicator, mechanism and empirical research," *Journal of Shanghai University (Social Sciences)*, vol. 16, no. 5, pp. 127-144, Sep. 2009.
[7] M. Zhang and S. Cong, "Research on Chinese fiscal risk nonlinear early warning system," *Economic Management Journal*, vol. 31, no. 5, pp. 147-153, May 2009.
[8] Y. Hu, H. Gao, and J. Xu, "BP artificial neural network model: a new visual angle of the financial risk early-warning," *West Forun*, no. 1, pp. 68-71, Feb. 2002.
[9] M. Li, "Research on early warning system of financial risk in China -based on K-means clustering algorithm and BP neural network," *Journal of Central University of Finance and Economics*, vol. 32, no. 10, pp. 25-30, Oct. 2012.
[10] A. Field, *Discovering Statistics using SPSS*, London: SAGE Publications, Ltd, 2009, ch. 17, pp. 627-685.
[11] L. Gorr, "Research prospective on neural network forecasting," *International Journal of Forecasting*, vol. 164, no.10, pp. 1-4, June 1994.
[12] K. Yale, "Preparing the right data diet for training neural networks," *IEEE Spectrum*, vol. 34, no. 3, pp. 64-66, Mar. 1997.
[13] S. Papadokonstantakisa, S. Macheferb, K. Schnitzleinb, and A. I. Lygerosa, "Variable selection and data pre-processing in NN modelling of complex chemical processes," *Computers and Chemical Engineering*, vol. 29, no. 7, pp. 1647-1659, June 2005.
[14] W. Lou and L. Qiao, "Early warning model of financial risks and empirical study based on neural network," *Finance Forum*, vol. 16, no. 11, pp. 52-61, Nov. 2011.
[15] J. H. Friedman and J. W. Tukey, "A projection pursuit algorithm for exploratory data analysis," *IEEE Trans. Computer*, vol. C-23, no. 9, pp. 881-890, Sep. 1974.
[16] Q. Fu and X. Zhao, *Principle and Application of Projection Pursuit Model*, Beijing, P. R. China: Science Press, 2006, ch. 3-4, pp. 29-119.
[17] W. Lou and L. Qiao. (August 2012). Projection pursuit clustering modeling applying multi-agent genetic algorithm and positive research. *Computer Engineering and Applications*. [Online]. Available: http://www.cnki.net/kcms/detail/11.2127.TP.20120801.1653.025.html
[18] D. Karaboga and B. Akay, "A comparative study of artificial bee colony algorithm," *Applied Mathematics and Computation*, vol. 214, no. 1, pp. 108-132, Aug. 2009.
[19] D. Karaboga, and B. Basturk, "On the performance of artificial bee colony (ABC) algorithm," *Applied Soft Computing*, vol. 8, no. 1, pp. 687-697, Jan. 2008.
[20] S. Yang, *Swarm Intelligent and Simulating Calculations - Matlab Implement*, Beijing, P. R. China: Publishing House of Electrionics Industry, 2012, ch. 11, pp. 236-252.
[21] R. V. Rao, *Decision Making in the Manufacturing Environment Using Graph Theory and Fuzzy Multiple Attribute Decision Making Methods*, London, U.K.: Springer-Verlag, 2007, ch. 3, pp. 27-41.
[22] W. Lou, G. Wang, and G. Feng, "Evaluation of travel security early warnings by applying TOPSIS with information entropy weighting and positive research," *Tourism Tribune*, vol. 28, no. 4, pp. 66-74, Aug. 2013.

Jitong Lou is born on June 22nd, 1992 in Shanghai City, P. R. China. He has been in No.2 High School of China East Normal University (No. 2 HSCENU), which is one of the best high schools in China, from Sept.1st, 2007 to July 1st, 2010. Now he is a rising senior student majoring in applied mathematics of Shanghai University, located at No.99 Shangda Road, Shanghai City, P. R. China. As far as the academic performance is concerned, he is a candidate of Bachelor of Science and the most outstanding undergraduate student in College of Science. In 2012, he registered for the summer session of Cornell University, Ithaca, USA. He experienced the academic life in top American university and got excellent grades in all of the courses he took. He has started his scientific career at No. 2 HSCENU since 2008, and he continues his researcher's work in mathematical modeling and innovation in Shanghai University. In 2008, he did the research of the program named Handwritten Number Recognition based on Artificial Neural Networks, and earned the prize of Shanghai Intel Science and Technology Innovation Contest. During the period of undergraduate study, he is the author/co-author of 2 scientific publications in applied mathematics, financial modeling, and data mining using projection pursuit technique. There are some about his selected publications: "Generalization of Goodman Formula", *Commun. Appl. Math. Comput.*, vol. 26, no. 4, pp. 355-359, 2012. "Financial Risk Evaluation of Chinese Commercial Banks Using Projection Pursuit Clustering", *2013 10th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD'13)* (to be accepted and published in Aug. 2013). Besides the achievements mentioned above, he is acting as a leader of the program, which is supported by Shanghai Municipal Education Commission, about the research of artificial bee colony (ABC) algorithm and projection pursuit clustering (PPC) model. He has also successfully taken part in some influential academic contests: he won first class prize in Mathematics Contest in Modeling and second prize in National Mathematical Modeling Contest.

Wengao Lou is born in Zhejiang province, P. R. China, in 1964. In 1985 he has graduated from Xi'an Jiaotong University and received his Bachelor's degree in mechanical engineering. In 1991 he obtained a Master's degree of mechanical engineering at Shanghai University of Technology (at present Shanghai University) and in 1995 has become an associate professor at Shanghai Fisheries University (at present Shanghai Ocean University, SHOU). In 2002 he has become FULL PROFESSOR majoring in Computer Science and Technology at University of Shanghai for Science and Technology (USST). In 2005 he has earned a Ph. D of engineering, majoring in civil engineering (applications of computer modeling) at Tongji University, Shanghai, P. R. China. He has started his scientific career at Shanghai Ocean University of computer simulation since 1985, later continuing his researcher's and teacher's work in computer-aided design and manufacturing (CAD/CAM), focusing on the computer simulation in mechanical engineering. Since 1999 he continues his researcher's and teacher's work in data mining technology, focusing on the artificial neural networks, projection pursuit, multi-attribute decision-making and comprehensive evaluation at SHOU and USST. Since 2009 until now he has started his career at Shanghai Business School (SBS) and be FULL PROFESSOR and Vice-President of SBS, located in Shanghai, at 2271# W. Zhongshan Rd., Shanghai 200235, P. R. China. He is the author/co-author of more than 100 scientific and technical publications in modeling of mechanical engineering, culture industry, printing and publishing, business economics and trade, management science and engineering. Dr. Lou is a senior member of China management science society, and be a peer reviewer of J. Agric. Food Chem., Journal of Guangdong University of Business Studies, Chinese Journal of Management Science, et al. She is honored with the title of "Shanghai Outstanding Young Teachers in Colleges and Universities" by Shanghai Municipal Government in 1995, and awarded the title of "Shanghai Outstanding Educators" in 1997. In 2009 he is awarded the title of "Shanghai High Level Teaching Achievement" and "China National Packaging Corporation Science and Technology Award". Since 2010 he has become the "Grading-three Full Professor" at SBS.

Xiurong Yu is born on Sep. 3th, 1973 in Henan province, P. R. China. She has studied at Liaoning University, Shenyang, Liaoning province, P. R. China from Sep. 2006 to Jun. 2009 and earned Ph. D of finance; studied at East China Normal University, Shanghai, P. R. China from Sep.1999 to June 2002 and got Master degree of international economy. She has been in Henan Normal University, Xinxiang, Henan province, P. R. China from Sep.1991 to Jun.1995 and earned B.E. in Political Education. Now she is an ASSOCIATE PROFESSOR of finance and Chair of finance department of Shanghai Business School, located at 2271# W. Zhongshan Rd., Shanghai, P. R. China. There are some about her selected publications: *Research of Historical Changes and Functional Evolution of an international financial center,* Beijing :China Financial Publishing House, June 2011; The Functional evolution of an international financial center and Its Implications, *Journal of Henan Normal University (Philosophy and Social Science)* vol.38,no.2,pp.91-95,Mar. 2011; Study of Functional construction of Shanghai international financial center， *Macroeconomics Research,* vol.137,no.4,pp.68-72,Apr.2010. She has been paying attention to research about financial market and financial risk. Prof. Yu is a member of Shanghai Supervisor Board of finance education since 2011. She is honored with the title of an "Outstanding Young Teacher" by Shanghai Municipal Education Commission in 2009.