Big Data Audit of Banks Based on Fuzzy Set Theory to Evaluate Risk Level

Yilin Bi[1], Yuxin Ouyang[1], Guang Sun[1], Peng Guo[1, 2], Jianjun Zhang[3] and Yijun Ai[1, *]

Abstract: The arrival of big data era has brought new opportunities and challenges to the development of various industries in China. The explosive growth of commercial bank data has brought great pressure on internal audit. The key audit of key products limited to key business areas can no longer meet the needs. It is difficult to find abnormal and exceptional risks only by sampling analysis and static analysis. Exploring the organic integration and business processing methods between big data and bank internal audit, Internal audit work can protect the stable and sustainable development of banks under the new situation. Therefore, based on fuzzy set theory, this paper determines the membership degree of audit data through membership function, and judges the risk level of audit data, and builds a risk level evaluation system. The main features of this paper are as follows. First, it analyzes the necessity of transformation of the bank auditing in the big data environment. The second is to combine the determination of the membership function in the fuzzy set theory with the bank audit analysis, and use the model to calculate the corresponding parameters, thus establishing a risk level assessment system. The third is to propose audit risk assessment recommendations, hoping to help bank audit risk management in the big data environment. There are some shortcomings in this paper. First, the amount of data acquired is not large enough. Second, due to the lack of author' knowledge, there are still some deficiencies in the analysis of audit risk of commercial banks.

Keywords: Bank audit, fuzzy set theory, membership function, risk control.

1 Introduction

With the rapid development of the Internet of things, cloud computing, and mobile Internet, the era of "big data" has arrived. And it has penetrated into all areas of the national economy and society, bringing us new thinking, business and management changes. Auditing is a behavior undertaken by an independent auditing agency and professional auditors to supervise and evaluate the authenticity, validity and benefit of the financial revenue and expenditure and other related economic activities of the audited entity according to law. And the aim is to evaluate the economic responsibilities, maintain the financial discipline, improve the level of operating management and raise the economic benefits. Auditing as an independent economic monitoring, evaluation,

assurance and service activity, the ultimate goal is to achieve a healthy and stable national economic system. Based on the rapid development of information technology and the change of auditing environment, big data auditing has become an important means to improve audit efficiency and quality. Through big data auditing technology, it can quickly and flexibly summarize, analyze, evaluate, monitor and warn various basic data. The realization of the audit inspection from sampling to full coverage has become the direction of comprehensive transformation and deepening of audit work [Zheng (2016)]. The banking industry is one of the earliest industries to use computers. Almost all businesses in China's commercial banks are completed through computer information systems, and a large amount of data is generated and processed every day. However, as a special industry that operates currencies, commercial banks are inherently high-risk and the risks are increasing. The traditional auditing method cannot meet the needs of commercial banks in the audit work under the big data environment. Many violations and potential risks are hidden in the data [Mo (2018); Tao (2014)]. Therefore, commercial banks urgently need to break through traditional auditing methods and models. And it makes big data auditing inevitable.

It is of great significance for banks to use big data for audit work. First, it helps to improve the efficiency and quality of audit work. The audits are not limited by time and space, and limited auditors are freed from cumbersome documentation work. In the auditing system, the voucher and account book are only the database to be checked in the audit project. The query, sampling, analysis and calculation are convenient, which improves the efficiency of the audit work. Second, it helps to promote scientific and standardized audit work. The use of big data for auditing makes the audit process open and the results public, enhances the transparency of the audit work, and aggregates and analyzes the problems existing in the work, so as to facilitate the understanding of the types and areas of risk-prone business and arrange the audit work more specifically. Resource optimization achieves the scientification of the audit plan, while various types of information in the audit process are stored in a standardized format. Third, it can help reduce audit risk and improve the ability of auditors. The use of information systems for auditing can calculate and analyze audit data, and conduct critical review of data in question. At the same time, various audit models can be used to detect data, so that auditors need to master capabilities of data analysis, comprehensive judgment and cross-professional knowledge application to enhance self-audit work ability. Finally, it helps to enhance the audit effectiveness. The bank audit department or China Banking Regulatory Commission (CBRC) can use the audit information system to achieve centralized sharing of audit information, review the audit completion status and basic data, and supervise the process. For some suspicious data, they can be prevented in time [Zhao (2016); Wang (2004)].

 Schönberger pointed out that big data is the source of people's new knowledge and new value [Schönberger (2013)]. The most important part of big data is forecasting. Most of the foreign research on audit risk is to define it and establish some measurement models.

Chinese scholars have conducted a lot of research on commercial banks' risk-oriented internal audit [Tie, Wu, Luo et al. (2019); Zhang (2016); Wei (2019)]. Meng et al. [Meng and Zhang (2010)] believes that China's financial audit must learn from foreign experience and take corresponding measures to better maintain financial security [Zhang

(2013)]. The most of the internal audits of commercial banks in China still carry out regulation audit, and there are many aspects to be improved in risk identification, internal control, personnel training and risk communication of commercial banks. Fang [Fang (2012)] pointed out that under the informatization environment, especially the extensive use of computer technology in the banking industries, the fundamental changes have taken place in the auditing environment, audit objectives, audit objects and auditing methods, making the traditional auditing mode of commercial banks unable to meet this changing needs [Shi (2005)], the former auditor of the National Audit Office, first proposed a new concept of data-based auditing. He thinks data-based auditing represents the future of computer auditing in the informatization environment. Wang et al. [Wang and Weng (2006)] conducted a systematic comparative study on the various stages of the development of commercial banks' auditing models, and comprehensively analyzed the auditing risks under different auditing modes. They found that the data audit mode has the largest amount of information and the smallest information asymmetry, and thus the audit risk may be the lowest.

Although the research on big data auditing at home and abroad has never stopped, the audit management information system that has been developed in China has not applied powerful functions to the professional audit work in the bank. At the same time, the use of big data auditing in the banking industry is not widely covered [Bai (2013)]. In the era of big data, while bringing opportunities to banks, it also increases the difficulty of enterprise development. In order to continue to develop healthily, enterprises should make full use of big data technology to effectively integrate data and improve the effectiveness and quality of business management. In view of the problems existing in the internal audit of enterprises in the context of big data, we should put emphasis on internal audit, improve the informatization degree of the audit process and the professional qualities of auditors, so as to improve the quality of internal audit work and promote the sustainable and healthy development of enterprises.

With the explosive growth of internal and external data in banking industry, the contradiction between the limitation of audit resources and the data information of audited entity becomes increasingly prominent. Therefore, big data audit is an inevitable outcome. In the process of rapid changes in organizational structure and continuous innovation of financial services, commercial banks must simultaneously establish and improve risk control systems and corresponding internal control systems to effectively avoid risks [Song (2019)]. The fuzzy clustering method uses the fuzzy set theory to deal with the analysis problem, and has obvious classification effect on the two-state data or polymorphic data with fuzzy features. It can improve the scientificity of the audit and reduce the audit risk.

2 A summary of fuzzy set theory

Fuzzy clustering analysis is a mathematical analysis method suitable for the case where the definition between things is relatively fuzzy. In the general classification process, we tend to classify things with the same characteristics into one class. However, when there is a vague definition between things, it is often impossible to carry out effective classification. This requires the method of a fuzzy theory to classify by softening the

more rigid classification thresholds between things through a specific membership function, resulting in better classification results [Wu (2019)]. Clustering is an important branch of data mining. Considering the mediation of sample generics, fuzzy clustering extends the membership degree of the sample from (0,1) to [0,1]. It can objectively reflect the real world and become the mainstream of cluster analysis research. This is a very common method of multivariate analysis in mathematics. It uses mathematical methods to determine the relationship of samples.

In 1965, Professor Zadeh from the United States first proposed the fuzzy set theory. The definition of fuzzy set is: let universe be U, any mapping of the closed interval U~ [0, 1] $\mu_A \begin{cases} \mu_A : U \to [0,1] \\ U \to \mu_A(u) \end{cases}$ can determine a fuzzy subset A of U. In the formula, $\mu_A(u)$ is the membership function of the fuzzy subset, and μ_A is the membership degree of U for A. Where: if $\mu_A(u)$ is closer to 0, it means that the degree to which u belongs to A is smaller, and vice versa. The methods for representing fuzzy subsets are as follows.

Zadeh representation:

$$A = \frac{A(u_1)}{u_1} + \frac{A(u_2)}{u_2} + \cdots + \frac{A(u_n)}{u_n} \tag{1}$$

Where $\frac{A(u_i)}{u_i}$ represents the membership degree $A(u_i)$ of u_i to fuzzy set A.

Ordered representation:

$$A = \{(u_1, \ A(u_1)), \ (u_2, \ A(u_2)) \dots (u_n, A(u_n))\} \tag{2}$$

Vector representation:

$$A = \{A(u_1), A(u_2) \dots A(u_n)\} \tag{3}$$

If the universe U is an infinite set, then the fuzzy set on it can be expressed as $A = \int_{u \in U} \frac{A(u)}{u}$. Since the fuzzy set can represent the fuzzy concept object well, the fuzzy set becomes a tool for the user to solve the problem with fuzziness in practical applications [Tang and Song (2019); Mao (2009); Dou, Yuan and Liu (2011)].

The membership function, also known as the attribute function, is an important concept in fuzzy mathematics and a mathematical tool for expressing fuzzy sets. In order to get the membership of the elements in the universe, you need to take a suitable value from 0 and l to represent the membership of the element. The most characteristic feature of the membership in the fuzzy set is that it is not directly represented by 0 or l, but the number of memberships is represented by the number between 0 and l. This membership degree can well reflect the degree of closeness between the element and the fuzzy set.

At present, the common methods of determining membership degree include intuition, fuzzy statistics and fuzzy distribution to determine the membership degree of the element for the fuzzy set. The following is an overview of several methods [Chen and Liu (2014); Wu (2013); Xing (2013); Li, Li and Dan (2002)].

The intuition method is that people understand some fuzzy concepts according to their own practical experience, or establish a membership function for fuzzy concepts based on common sense. It is often used to describe things that people are familiar with but have objective fuzziness in their daily lives. Intuitive methods can also solve such problems

well when evaluation indicators have difficulty in obtaining data. Although the intuitive method is simple and intuitive, the degree of membership is greatly disturbed by the outside world, as different people may establish different membership functions according to their own practice and the degree of understanding of things.

The fuzzy statistics method means that based on the investigation statistics, analyze and calculate the occurrence frequency of elements, then make a quotient between the occurrence frequency and the total investigation frequency, and the final result is the membership degree of elements to the fuzzy set. Fuzzy statistics method can not only objectively reflect the membership degree of the element in the universe, but also has a certain theoretical basis. When the number of statistics is large enough, the frequency of the element is always stable at a certain number, so the fuzzy statistical method is also a common method for determining the degree of membership.

The fuzzy distribution method is the membership function of the fuzzy set on the real number set, which takes the real number set as the universe. The fuzzy distributions often used in practice include the rectangular distribution, Gaussian distribution, Cauchy distribution and so on.

3 Implementation steps and case analysis

Big data processing usually includes getting data, cleaning data, management data, analysis, and visualization (see Fig. 1). Getting data refers to the data required for the experiment from the professional data organization, the National Bureau of Statistics, the internal enterprise and the Internet; cleaning data refers to the cleaning of the acquired data source and elimination of unnecessary, incorrect and invalid data; management data means the processing of cleaned data according to business needs; and on this basis, the analysis and research; finally, the results are visualized. So that it can be displayed more clearly and intuitively, and maximize the use of data value.
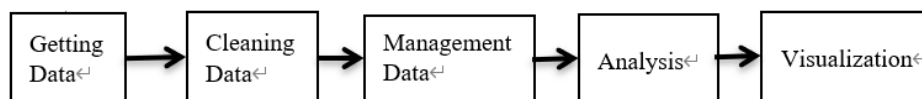

Figure 1: Data processing flow

In this paper, the bank is audited by fuzzy set theory, and the fuzzy membership method is used to obtain the specific membership degree, and the blunt classification threshold is softened to obtain better classification effect. The research content is whether the fluctuation of operating income of China Construction Bank is reasonable. First of all, through the on-site, letter, SMS and other means to ask the bank staff with relevant financial knowledge about their views on the "reasonable fluctuations" of the bank's operating income, and get the data needed for the experiment. After screening and summarizing the statistical data, there is no need to clean up or operate the relevant data. The membership degree can be calculated directly through the membership function for statistical analysis. Finally, the experimental results can be obtained and visualized.

In the bank audit process, analytical audit procedures are often used. By analyzing and comparing the relationships between information or calculating the relevant ratios, the method obtains audit evidence, determines auditing emphasis, and supports audit

conclusions. Analytical procedures focus on the rationality of information. For example, compare the data in this year's financial statements with last year to find unreasonable data fluctuations. The more unreasonable fluctuations are, the higher the risk level. However, for the definition of the scope of the two concepts of "reasonable fluctuation" and "irrational fluctuation", auditors often cannot give an accurate range of values, so it is impossible to clearly classify the obtained volatility. In order to better judge the risk level, the risk is evaluated by the fuzzy statistical method that calculates of the membership degree in the fuzzy set theory, and the degree of membership of the data for a certain interval is obtained, and the data is characterized, so that the data can be more accurately defined and classified.

Take the volatility of operating income of China Construction Bank in 2011-2018 as an example (see Tab. 1). The total operating income of 2011-2018 was obtained from the annual financial statements of China Construction Bank, and the annual operating income was compared with the operating income of the previous year. The volatility was calculated to be 16.03%, 10.39%, 12.16%, 6.09%, -0.02%, 2.74%, 5.99%. Since there is no certain threshold, it is impossible to effectively judge whether the fluctuation of the growth or decrease of operating income is "reasonable fluctuation" or "irrational fluctuation", except for subjective judgments such as relevant experience and regulations. So as to carry out the experiment.

Table 1: Volatility of operating income of China Construction Bank in 2011-2018

year	2011	2012	2013	2014	2015	2016	2017	2018
Operating income (10 thousand yuan)	397,090	460,746	508,608	570,470	605,197	605,090	621,659	658,891
Volatility		16.03%	10.39%	12.16%	6.09%	-0.02%	2.74%	5.99%

The 140 bank staff used their experience and relevant financial knowledge to give a range of reasonable volatility of bank operating income, resulting in the following set of data (see Tab. 2).

Convert the "reasonable volatility" interval statistics data into a scatter plot as follows (see Fig. 2). It can be seen from the figure that the values are more scattered and inconsistent, indicating that people have different views on the "reasonable volatility" interval, which is a subjective concept, mostly depending on relevant experience and various regulations, and there is no specific value range.

The volatility of China Construction Bank from 2011 to 2012 is now selected as $\mu_0 = 16.03\%$. According to the definition, the membership function of "reasonable volatility" is:

$$\mu = \frac{\text{Interval number including } 16.03\%}{\text{Number of visitors } (n)}$$

(4)

μ is used as the approximate value of 16.03% to the membership degree of "reasonable fluctuation", and the calculation results are as follows (see Tab. 3).

The membership frequency is represented by line graph, and the stability of 16.03% to "reasonable fluctuation" membership frequency is obtained (see Fig. 3).

Table 2: "Rational Volatility" Interval Statistics

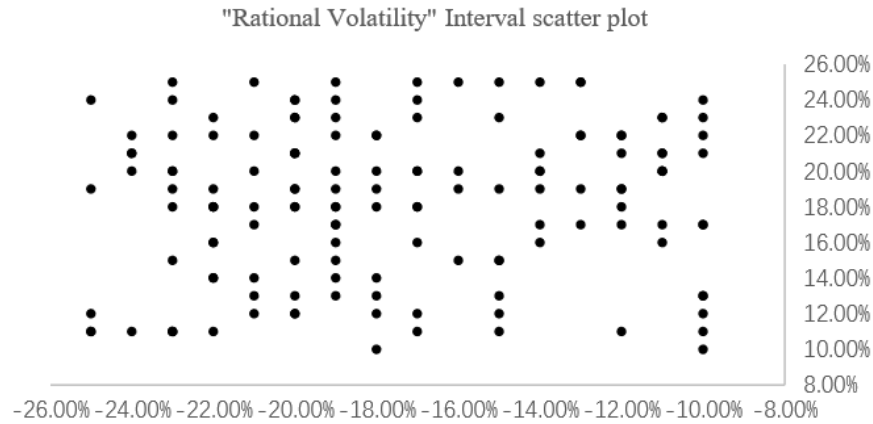| "Rational Volatility" Interval Statistics					
(-12%,12%)	(-11%,24%)	(-10%,18%)	(-24%,16%)	(-22%,18%)	(-12%,12%)
(-18%,11%)	(-12%,19%)	(-20%,21%)	(-14%,24%)	(-17%,21%)	(-19%,14%)
(-21%,19%)	(-23%,23%)	(-15%,15%)	(-24%,19%)	(-25%,14%)	(-10%,11%)
(-22%,10%)	(-23%,16%)	(-13%,10%)	(-18%,24%)	(-24%,20%)	(-18%,12%)
(-15%,22%)	(-16%,23%)	(-12%,23%)	(-12%,21%)	(-20%,18%)	(-18%,13%)
(-22%,13%)	(-15%,13%)	(-13%,23%)	(-19%,21%)	(-11%,22%)	(-14%,11%)
(-15%,17%)	(-24%,18%)	(-14%,14%)	(-11%,19%)	(-13%,24%)	(-23%,14%)
(-16%,16%)	(-20%,13%)	(-15%,10%)	(-15%,25%)	(-13%,11%)	(-10%,10%)
(-22%,22%)	(-15%,12%)	(-19%,10%)	(-18%,17%)	(-13%,18%)	(-24%,21%)
(-11%,25%)	(-15%,21%)	(-22%,21%)	(-12%,10%)	(-18%,18%)	(-25%,14%)
(-17%,20%)	(-10%,16%)	(-15%,23%)	(-14%,11%)	(-14%,19%)	(-12%,12%)
(-16%,17%)	(-16%,19%)	(-23%,14%)	(-17%,22%)	(-20%,16%)	(-19%,19%)
(-17%,13%)	(-18%,13%)	(-12%,19%)	(-24%,20%)	(-10%,20%)	(-22%,23%)
(-17%,25%)	(-23%,11%)	(-17%,12%)	(-22%,17%)	(-23%,10%)	(-14%,24%)
(-23%,16%)	(-19%,20%)	(-20%,25%)	(-24%,25%)	(-10%,22%)	(-20%,23%)
(-19%,11%)	(-25%,16%)	(-16%,19%)	(-12%,24%)	(-22%,17%)	(-14%,11%)
(-15%,19%)	(-12%,20%)	(-14%,16%)	(-20%,10%)	(-10%,24%)	(-24%,10%)
(-11%,14%)	(-15%,11%)	(-12%,15%)	(-24%,18%)	(-18%,15%)	(-10%,24%)
(-21%,20%)	(-17%,19%)	(-17%,14%)	(-10%,25%)	(-20%,11%)	(-12%,20%)
(-22%,21%)	(-22%,12%)	(-11%,11%)	(-11%,12%)	(-18%,11%)	(-13%,12%)
(-21%,10%)	(-15%,14%)	(-14%,13%)	(-21%,18%)	(-10%,17%)	
(-11%,22%)	(-23%,13%)	(-12%,12%)	(-25%,24%)	(-14%,24%)	
(-16%,13%)	(-18%,13%)	(-10%,24%)	(-18%,13%)	(-13%,10%)	
(-12%,15%)	(-13%,20%)	(-19%,19%)	(-13%,23%)	(-15%,13%)	

Figure 2: "Rational Volatility" Interval scatter plot

It can be seen that 16.03% to the membership frequency of "reasonable fluctuation" is stable around 0.56, so

$$\mu_{reasonable\ fluctuation}(16.03\%)=0.56 \tag{5}$$

Table 3: Approximate value of membership of "16.03%"

n	10	20	30	40	50	60	70	80	90	100	110	120	130	140
Membership times	6	10	17	23	26	33	39	45	52	57	62	65	70	74
Membership frequencies	0.6	0.5	0.56	0.58	0.52	0.55	0.56	0.56	0.58	0.57	0.56	0.54	0.54	0.53

That is, 16.03% to the membership frequency of "reasonable fluctuation" is 0.56, which means that the proportion defined as "reasonable" (0.56) is larger than the proportion defined as "unreasonable" (0.44), so the operating income volatility rate of 16.03% of China Construction Bank in 2011-2012 is "reasonable volatility".
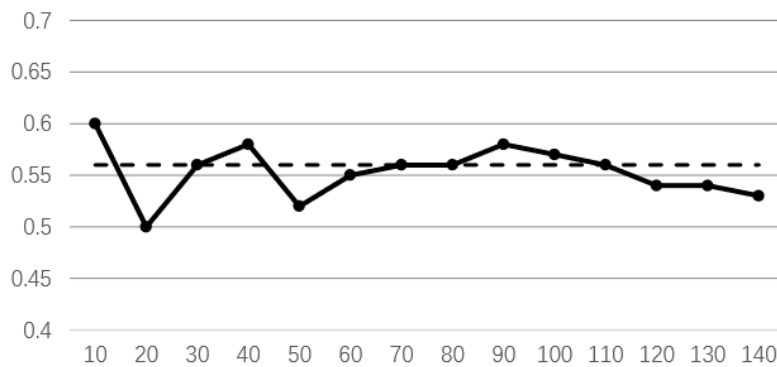

Figure 3: Stability of 16.03% to "reasonable fluctuation" membership frequency

The fuzzy statistics method can minimize the impact caused by the different subjective cognition among auditors, make the data more accurate and standardized, and ensure the objectivity and credibility of the conclusion of the analytical audit procedure. With this method, the membership function of all account data in financial statements can be calculated, and all account data in financial statements can be analyzed, so as to establish a complete risk level evaluation system.

4 Conclusion

Based on the collected data, we calculate the membership degree of the volatility of operating revenue to the "reasonable fluctuation", and judge whether the fluctuation of operating revenue belongs to the "reasonable" range. The more unreasonable data is, the higher the risk is, which means that bank needs to take risk prevention. The experimental results show that from 2011 to 2012, the growth rate of operating revenue is 16.03% and the membership degree of "reasonable fluctuation" is 0.56. The growth rate is within the range of reasonable fluctuation and does not constitute a risk factor. The experimental data comes from 140 bank staff, and the amount of data is relatively small. In the next research process, we can appropriately increase the amount of data, so as to obtain a more stable function membership degree. At the same time, it is not limited to a certain research factor, and research on various aspects of bank auditing makes the results more credible.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

Bai, T. (2013): Strategic thinking on the development of internal auditing in banking industry in the age of big data. *Financial Times*, vol. 9.

Chen, H. P.; Liu, Y. L. (2014): Application of fuzzy mathematics in auditing inherent risk assessment. *Friends of Accounting*, vol. 29, pp. 77-80.

Dou, Y. D.; Yuan, Y. B.; Liu, Y. (2011): Research on credit risk early warning of commercial bank–Based on AHP weight variable fuzzy model. *Technoeconomics & Management Research*, vol. 12, pp. 82-87.

Fang, T. (2012): Analysis of commercial bank data audit mode. *Finance Economy*, vol.3, pp. 77-79.

Li, Z. Y.; Li, J. N.; Dan, Z. Q. (2002): Application of fuzzy mathematics in audit risk assessment. *Journal of Shijiazhuang Railway University*, vol. 1, pp. 70-72.

Mao, W. S. (2009): *Application Research of Fuzzy Comprehensive Evaluation Model in Audit Risk Assessment*. Hunan University.

Meng, Y.; Zhang, J. (2010): On implementation mechanism of national audit "immune system". *Journal of Central University of Finance & Economics*, vol. 5, pp. 81-84.

Mo, X. M. (2018): Discussion on bank internal audit in the age of big data. *Modern Economic Information*, vol. 7.

Schönberger, V. M.; Cukier, K. (2013): Big data: a revolution that will transform how we live, work and think. *Mifflin Harcourt.*

Song, J. J. (2019): Auditing strategy based on big dat. *Technology and Economic Guide*, vol. 26, pp. 220-223.

Tang, B. Z.; Song, S. C. (2019): Multi-dimensional poverty measurement of urban households based on fuzzy set theory. *Statistics and Decision*, vol. 9, pp. 81-85.

Tao, X. Q. (2014): Discussion on the internal auditing work of banks in the age of big data. *Modern Commerce*, vol. 30, pp. 189-190.

Tie, X. Y.; Wu, Y.; Luo, W. L.; Xu, C. Y. (2019): Research on the misconduct of auditing procedures in the background of big data. *Cooperative Economy and Technology*, vol. 22, pp. 150-151.

Wang, W. Q. (2004): Audit risk and relevance research. *Southern Finance*, vol. 12, pp. 14-17.

Wang, Y. H.; Weng, H. Y. (2006): A comparative study on the auditing mode of commercial banks. *Communication of Finance and Accounting (Academy Version)*, vol. 10, pp. 24-27.

Wei, D. (2019): Research on audit risk of commercial banks in big data environment. I*ndustry and Technology Forum*, vol. 16, pp. 219-220.

Wu, J. G. (2019): Application of fuzzy clustering method in atmospheric environmental quality assessment. *Environment and Development*, vol. 3, pp. 18-20.

Wu, G. L. (2013): Research on evaluation method of audit quality based on fuzzy mathematics. *Technology Information*, vol. 13, pp. 247.

Xing, F. Y. (2013): Research on application of analytic hierarchy process and fuzzy mathematics in risk-oriented auditing. *Henan Social Sciences*, vol. 1, pp. 44-47.

Zhang, C. (2013): Internal audit and risk management in commercial banks. *China Investment*, vol. 9, pp. 156-159.

Zhang, L. (2016): *Research on Audit Risk Analysis and Control of Commercial Banks in Big Data Environment.* Southwestern University of Finance and Economics.

Zhao, L. C. (2016): Bank internal audit analysis based on big data. *National Business*, vol. 30, pp. 68-69.

Zheng, Z. H. (2016): Bank audit thinking under big data. *China Finance*, vol. 14, pp. 52-53.