Predictability of Investment Behavior Based on Personal Characteristics about China's Individual Investors

Qiujun Lan [1*], Xuqing Xu [1], Xingye Hu [2]

[1] Business School of Hunan University, Hunan, Changsha, CHINA
[2] Jixiang Life Insurance Company Limited, Hunan, Changsha, CHINA

ABSTRACT

This paper analyzes the relationship between individual investor's characteristics and behaviors by rank correlation analysis and logistic regression based on more than 20,000 samples of China' individual investors nationwide collected through online questionnaires. And we select some critical characteristics as input variables to predict investor's behaviors. The results show that investors' personal characteristics and their behaviors are closely linked, and the data mining models inputting an investor's personal characteristics can predict his investment behaviors effectively. These results would help related financial organizations to optimize their customer management, service quality and cost or profit control.

Keywords: individual investors, personal characteristics, investment behavior, data mining

INTRODUCTION

With the rapid development of China's comprehensive national strength and family wealth, China's financial market has obtained a rapid development and become more and more mature. The financial activities have penetrated into many aspects of the economy and society, and the vitality of financial investors has been enhanced (Wang, 2012). Generally, institutions are great majority of investors in the countries or regions with developed financial markets. However, the individual investors in China account for nearly 80% (Guo, 2013). Facing such a huge and important customer group, how to effectively improve their services, developing customers and decrease management cost is a key problem for financial servers like stock brokers. Hence, having real insight into investors' characteristics, behaviors and customer segmentation, providing personalized services became the necessary measures to solve the problems (Cao, 2009). Investors' individual characteristics are easy to obtain by financial servers, but investors' behaviors are not easy to observe. So, the financial servers are all interested in the answers to these questions as what kind of behaviors do the individual investors have in China's financial market? How are they specifically affected by the personal characteristics of the investors? And whether individual characteristics can predict individual behaviors? These answers are very important to help organizations making decisions.

RELATED WORK

Scholars like Kahneman and Tversky (1979) had established and developed the behavioral finance theory, since then analysis and researches on the investors' characteristics have begun to attract attention. According to the view of Pompian (2012), studies on behavioral finance can be divided into two types: one is Behavioral Finance Macro, whose research objects are usually institutional investors with more influence; the other one is Behavioral Finance Micro, which mainly studies on some psychological bias and behavioral characteristics of the individual investors. Although the former is the mainstream of research, the latter also gradually attracts some scholars' attention, and many interesting phenomena have been found.

For example, after conducting the investigation to 100 investors by using Myers-Briggs personality test list and questionnaire, Pompian and Longo (2004) found there were striking differences among the individual investors with different personality characteristics in terms of such aspects as preference of investment type, choice of

Contribution of this paper to the literature

- This study find that investors' personal characteristics are strong predictive to their behaviors.
- When using the selected significant personal characteristics as input variables, we could construct effective predictive models based on data mining methods, which would provide a better understanding of the individual investors and explore the irrational behaviors of investors in the financial market in China.
- This study could not only provide decision information for investor education, marketing and service personalization, but also provide clues to strengthen customers 'cost control, quantitative management and risk control, etc.

information channel and transaction behavior. Wen et al. (2014) examined the relation between investors' risk preference and return on stock market, found that investors become risk averse when they gain and risk seeking when they lose and the extent of risk aversion in gains and that of risk-seeking in losses were different. Hira and Loibl (2008) paid special attention to differences of investment behavior caused by the gender, and they found that the gender had an impact on the acquisition channel of investment information, and at the same time it also influenced the risk taking level of different the individual investors. Barnea et al. (2010) used some twins investors' investment records that is very difficult to obtain to discuss the relations among individual investor's characteristics, market participation habit and capital investment distribution behaviors, finding that one third of investment behavior differences can be explained by individual genetic characteristics. According to the different attitudes and decision-making behaviors of the individual investors, Clark-Murphy and Soutar (2005) divided the samples into four clear categories, finding that each category's individual investors has different features in investment preference and target choice. Based on these results, some financial services and educational institutions have gained advices and measures that are more targeted.

At the early stage of China's stock market, Peng et al. (1995) made an investigation and study on investment behavior features and personality psychology of stock investors in Shanghai and found that the personal characteristics influencing stock investment performance include character, ability, social and economic environment, etc. Since then, Yan (2002) made a questionnaire survey on the individual investors and institutional investors of 126 sales departments in Jiangsu Province, aiming at understanding the factors that affect their investment behaviors, such as their composition situations, psychological qualities, investment techniques, as well as politics, economies, policies, information etc. Wang et al. (2003) hold that the individual investors who are able to effectively master market information and have an advantage over others on investment knowledge will be more likely to gain profits. Li et al. (2002), Tan et al. (2006) studied the features of specific investment behaviors on excessive trading, and considered that those excessive trading are common among individual investors, but the definition of excessive trading is ambiguous. Pei et al. (2005) made a research on the investment behavior characteristics of the individual investors, arguing that Chinese individual investors not only have a cognitive behavioral deviation in general sense, but also have localization deviation.

Throughout the available literature, the following awareness can be formed: (1) The study on investors rapidly draws attention with the establishment of behavioral financial theory; (2) Obviously, the study on institutional investors gains more attention than that of the individual investors, and even lots of literature generally mix individual investors and institutional investors; (3)The personal characteristics involved in the literature are relatively extensive, and the investment behaviors are also relatively general without a clear defined scope; (4) Most of the researchers observe the personal characteristics and the influence bought by investors' investment behaviors from the aspect of investment effect, but ignore the relationship between personal characteristics and behaviors; (5) The sample size of most research is within 2,000, and the sample subjects are concentrated in a certain specific securities company or area, without extensive representativeness, which makes it difficult to have an overall description of the characteristics of the individual investors in China at present. (6) Some research are conducted from the aspect of psychological cognition and character trait, which needs professional and complex psychological tests, tracking survey, etc.. Because few samples can be obtained, it is more difficult to be popularized and applied in a relatively large scale. Therefore, in order to understand the individual and behavior characteristics of investors in China and the relation between these two items, it is necessary to sort out and analyze again systematically on the basis of referring to the literature results above, and conduct the research based on more extensive investigation.

DATA

Questionnaire

At present, there is no recognized system about types of personal characteristics of investors. As mentioned above, current research on personal characteristics of investors are involved in quite extensive factors, such as

gender, age, genetic factors, personality, psychological quality, knowledge, experience, emotion, cognition, economic income, social status, occupations, and regions. Among them, some characteristics have relatively vague concepts, which are not strictly defined and often have intersection or even contain relationships with each other. In view of this, we put forward four principles to filter personal characteristics, "common", "easy to measure", "stable" and "relevant", which are taken as a selecting basis of study objects for investors' personal characteristics in this paper.

Among them, "common" means the personal characteristics are often involved in daily life, and meanwhile often discussed in literature. For example, the personal characteristics such as gender, age and income are commonly concerned; "easy to measure" means the selected personal characteristics should be relatively easy to measure or describe, and it's not necessary to use complex means to obtain them so as not to relate privacy. For example, characteristics such as genetic factors, personality and psychological quality need special means to be obtained, and they are hard to be operated and applied in practice; "stable" means the personal characteristics don't change significantly in a short period and they have a weak subjectivity. For example, the emotional characteristics of individuals such as pleasure, anger, sorrow and joy vary as time goes by and they have a bad stability, so they won't be considered; "relevant" means the selected personal characteristics should have the potential of influencing investment decision-making. For personal characteristics which are obviously irrelevant, like height and weight, should be excluded. According to the above principles and referring to available researches (Shi et al., 2006; Xu et al., 2008; Li et al., 2010; Li et al., 2011), this paper finally determines 6 common personal characteristics as study objects, included age, occupations, income ranges, professional knowledge levels, investment experience and residential regions. No matter for the need of service, marketing, supervision or research, these personal characteristics all have strong operability.

Similar to personal characteristics, "investor behavior" is indeed also a broad and fuzzy concept. Most of the available literature only focused on investors' cognition and preference about risk and mentioned some investor behaviors generally and vaguely. In fact, what investor behavior emphasizes is investors' specific behavior during the process of investment decision-making, so it can be defined and screened on this basis. Although most of individual investors will not deliberately pay attention to and structure their own behavior, investors will naturally or half unconsciously follow such a framework composed of four stages from the view of decision-making theory. These four stages are decision-making preparation, decision making, execution and feedback. The main tasks in the stage of decision-making preparation include evaluating self-ability, determining investment goals and searching information; the most important tasks in the decision making stage are choosing investment directions and products as well as determining of investment scale and allocation proportions; the decision execution stage includes determining of transaction time and specific trading operations; the feedback stage is to evaluate and rethink the previous decision. Based on this viewpoint and referring to available studies (Yang et al., 2011; Liu, 2011; Tan and Chen, 2012; She, 2012; Zhang et al., 2012; Chang, 2016; Wang, 2016; Huang et al., 2014), we finally decide to investigate the following 6 categories of specific investor behaviors: channel selection of investment information, investment variety choice, investment scale choice, speed of making decision, transaction frequency and satisfaction evaluation of investment result. All of them are major behaviors of investors in different phases. According to the above personal characteristics and behavioral variables, the questionnaire design follows the principle of simple and clearness, easiness prior to hardness, clear framework and anonymity. It includes questions for measuring the validity and consistency of questionnaires. There are 28 questions in all, and they should be done within about 15 minutes. **Table 1** briefly lists key questions and options relevant to the above personal characteristics and behavioral variables.

Table 1. Questions and Options of Key Variables

Variable	Question	Option
Age	How old?	25 or less, 26 to 30, 31 ~ 40 and 41 ~ 50 and 51 ~ 60, 60 or more.
Gender	What is your gender?	Man, woman.
Profession	Current profession?	Government agencies, public institutions, corporate enterprises, private individuals, students, freelance, and other.
Living region	Current region?	Provinces and autonomous regions of China.
Income	Current average monthly income levels?	Below 2000, 2000~5000, 5001~8000, 8001~12000, 12001~16000, 16001~20000, 20001~25000, 25000 more.
Expertise knowledge	Knowledge of investment?	Knows little expertise, only read five or less relevant professional books, read five or more specialized books, financial related professional background.
Investment experience	Years engaged in securities investment?	Less than 2 years, 2~5, 6~10, 11~15, more than 15 years.
Information channel	What is the most dependent on the information channel?	Financial websites, Shares forums, specialized software, celebrity blogs, the company's official website, QQ, microblogging.
Investment scale	What is the proportion of the funds used to invest in your assets?	0 ~ 20, 21 ~ 40 and 41 ~ 60 and 61 ~ 80 and 81 ~ 100.
Investment variety	What are the main areas of investment choices?	Stocks, funds, bonds, futures, precious metals, foreign exchange, and other.
Decision-making speed	How long it will take to identify the authenticity of statements of investment?	A few minutes, 1~2hours, 2 hours or more.
Transaction frequency	How many times of transactions per year?	Less than 5 times, 5~10 times, 11~20 times, 21~35 times, 36~50 times, more than 50 times.
Satisfaction evaluation of investment results	Are you satisfied with the investment results?	Yes, no.

According to *the Statistical Report of Development Status of China Internet Network* released by China Internet Network Information Centre (CNNIC), by December 2013, the number of Chinese cyber citizens had reached 618 million. These Chinese cyber citizens have a relatively distribution in aspects such as income, occupation, gender and education background. Network clients have been commissioned to place orders universally by investors. It is an effective and economic way to publish questionnaires via Internet. We pay to have the "Sojump" released the questionnaire link nationally through its channels. At the same time, we also mobilize all forces by ourselves to issue the questionnaire link to all investors by forums, QQ groups and WeChat groups and even ask our relatives and friends to forward it. In order to observe the stability of the result, the publishment and collection of questionnaires are carried out in three stages. The first stage is from January 7th, 2013 to January 31th, 2013. The second stage starts from December 8th, 2013 and ends by February 8th, 2014. The third stage lasts from November 14th, 2014 to December 1st, 2014.

Data Description Statistics

For the questionnaire survey, 20234 samples have been collected in all by the deadline. 19972 samples come from independent IP while 262 samples are from non-independent IP. There are 19872 valid questionnaires in the end, and the number of samples in three phases is respectively 2912, 7456 and 9504.

Table 2 describes the geographical distribution of survey samples. It is told in the table that samples are mainly from six developed provinces referring to Guangdong, Shanghai, Beijing, Shandong, Jiangsu, Zhejiang, which already account for 55.92% as a whole. Nevertheless, the distribution rather conforms to the basic situation of current economic development in China. The GDP index of all provinces in *2013 Provincial Annual Statistics*, released by National Bureau of Statistics, showed that the GDP gross of six provinces, with the top 6 largest surveyed population, had accounted for 40.42% in national one either. Generally speaking, since the eastern region is considered as the developed area, the total wealth of people living there is always more than that of any other areas. Accordingly, people of these areas tend to be engaged more in investment activities; when it comes to the middle and west areas, the economic development lags behind the eastern developed areas. The wealth per capita is lower and the disposable income is fewer, which can therefore explain why people involving in the investment activity are fewer than ones from the eastern developed areas. The survey data samples conform to these characteristics.

Table 2. Geographical distribution of survey objects

Region	Percent	Region	Percent	Region	Percent
Guangdong	15.29%	Guangxi	3.08%	Yunnan	0.99%
Shanghai	11.06%	Tianjin	2.98%	Guizhou	0.52%
Shandong	8.25%	Anhui	2.87%	Xinjiang	0.47%
Beijing	7.60%	Liaoning	2.86%	Neimenggu	0.45%
Zhejiang	7.14%	Hunan	2.02%	Hainan	0.25%
Jiangsu	6.57%	Shanxi	1.88%	Gansu	0.22%
Hebei	4.38%	Chongqing	1.77%	Ningxia	0.15%
Hubei	3.57%	Jiangxi	1.76%	Qinghai	0.04%
Fujian	3.37%	Shanxi	1.52%	Unknown	0.22%
Sichuan	3.27%	Heilongjiang	1.18%	—	—
Henan	3.25%	Jilin	1.02%	—	—

Table 3. Personal characteristic distribution of investors

	Percentage in questionnaire（%）				
	Age	Profession	Knowledge	Experience	Income
1	19.49%	4.89%	19.76%	32.50%	11.04%
2	30.87%	17.80%	30.62%	38.38%	33.77%
3	37.00%	47.56%	34.45%	21.70%	29.32%
4	10.28%	12.64%	15.17%	5.23%	13.64%
5	1.95%	10.69%	—	2.18%	4.77%
6	0.42%	4.77%	—	—	4.33%
7	—	0.57%	—	—	2.13%
8	—	1.07%	—	—	1.00%

Note: the numbers of each option are represented in the following details:
Age: 25 or less, 26 to 30, 31 ~ 40 and 41 ~ 50 and 51 ~ 60, 60 or more.
Profession: government agencies, public institutions, corporate enterprises, private individuals, students, freelance, and other. Knowledge: knows little expertise, only read five or less relevant professional books, read five or more specialized books, financial related professional background.
Experience: less than 2 years, 2~5, 6~10, 11~15, more than 15 years.
Income: below 2000, 2000~5000,5001~8000,8001~12000,12001~16000,16001~20000,20001~25000,25000 more.

Table 3 lists the descriptive statistics of personal characteristic variation. The group below 40 years old is 87.4% from the age level. The characteristics of this age structure is somewhat different from those in the field survey on the securities before, which reveals older age structure (Li et al., 2010). This reflects, compared with the beginning of the development of securities market, that the young and middle-aged people possess more fortune and have the aspiration and ability of financial investment. By profession, the main force is from enterprises, public institutions, private and individual businesses, especially of which people from enterprises are 47.6%. On the earnings aspect, investors with the monthly earnings of 2000-8000 Yuan account for 63%, which shows that medium income group is the majority. In addition, 30.42% of the respondents only have read not up to 5 books related to financial investment. The percentage of the people who were educated professionally is the least. Most of the investors have comparatively short investment experience. Investors with experience of less than 5 years account for 71%. All these characteristics conform with the ones which were showed in the *2013 Survey Report of Individual Investors* by Fenghua Finance and Economics, that China's individual investors are mainly ordinary clerks from enterprises and public institutions, and it appears that they have inadequate investment horizon, and that the families with medium and low income account for higher percentage, and the knowledge level of the investors is not high.

Table 4 lists the details of the investors' behaviors. From it we can know that finance and economics websites are the main channel for individual investors to get the information from. The second channel is the securities forums. This shows that most investors rely on formal media information. On the investment scale level, the proportion is basically controlled less than 40% of their discretionary assets, which shows that most investors are relatively cautious in investment, as their investing proportion is not more than half of their assets. On the transaction frequency level, people who have 10 or less transactions in a year (less than once a month) account for 55%, but the active investors who have more than 20 transactions account for 20% as well. On the level of investment varieties, most investors, focusing on traditional stock and fund, account for 70%. But other investment types and channels, including Internet financial products and non-break-even investment from banks, are developing rapidly. Compared with *Shenzhen Stock Exchange 2013 Survey Report of Individual Investor Situations*, the report shows that the company bulletin and securities forums are the main sources for individual investors to obtain main information and their investment capital occupies a low proportion of their total household assets. The investment

Table 4. Investor Behavior Distribution Statistics

	Percentage in questionnaire（%）				
	Scale	**Variety**	**Frequency**	**Speed**	**Channel**
1	26.41%	38.16%	21.67%	26.44%	45.71%
2	37.03%	31.96%	33.32%	52.18%	13.85%
3	24.31%	6.22%	24.56%	21.37%	10.76%
4	10.00%	4.86%	10.75%	—	10.13%
5	2.25%	3.86%	3.98%	—	8.41%
6	—	1.64%	5.72%	—	7.33%
7	—	13.30%	—	—	3.81%

Note: the numbers of each option are represented in the following details:
Scale: 0 ~ 20, 21 ~ 40 and 41 ~ 60 and 61 ~ 80 and 81 ~ 100.
Variety: Stocks, funds, bonds, futures, precious metals, foreign exchange, and other.
Frequency: less than 5 times, 5~10 times, 11~20 times, 21~35 times, 36~50 times, more than 50 times.
Speed: a few minutes, 1~2hours, 2 hours or more.
Channel: finance and economics websites, securities forums, specialized software, celebrity blogs, the company's official website, QQ, microblog.

types are diversifying. The investment types are increasing apart from stock. All these conclusions are also basically verified in this survey data.

In all, the above survey sample conforms to the individual investors' situations in the current China financial market summed up and reflected by other authorities, so it can be regarded as a representative sample for the individual investors in China financial market.

PREDICTABILITY ANALYSIS OF INVESTMENT BEHAVIORS

The Rank Correlation Analysis

For preliminary understanding the basic correlation between two sets of variables, first of all, the sample data of the three stages are summed up, and then the correlations of each variable pair (i.e. a characteristic variable and a behavioral variable) between the two groups of variables are analyzed one by one. When analyzing, an orderly conversion is conducted to the value of each variable. Specifically, ages are converted to 1~6 from small to large, incomes are converted to 1~8 from low to high, professional knowledge levels are converted to 1~4 from low to high, scales of investment are converted to 1~5 from low to high, investment experience is converted to 1~5 from few to many, transaction frequencies are converted to 1~6 from few to many, investment varieties are divided into 1~7 according to the level of recognition, information-acquisition channels are converted to 1~8 according to formal and informal, and information-decision speeds are converted to 1~3 from slow to fast. Because they are all ordinal variables, the Spearman rank correlation method is used, and the result is shown in **Table 5**.

Seen from **Table 5**, most of the p values of each variable pair between the two sets of variables are less than 0.0001, and all are correlative in case of a significant level of 5%. Therefore, the personal characteristics of the investors generally have a significant impact on the behavior of investors.

Specifically, for example, there is a significant positive relationship between "transaction frequency" and investors' "investment experience", indicating that more experienced investors have more sensitive reflection on information during the market operation, and make more adjustments, which is consistent with the research conclusion that "experienced investors are generally over-confidence, thus leading to frequent transactions" by reference (Shi et al., 2006). Moreover, "investment varieties" and "ages" present a negative correlation, reflecting

Table 5. Spearman rank correlation coefficient between personal characteristic variables and behavioral variables

	Scale	**Variety**	**Frequency**	**Speed**	**Channel**
Age	0.1223	-0.233	0.2684	-0.084	-0.075
	(<.000)	(<.000)	(<.000)	(<.000)	(<.000)
Profession	-0.120	0.0786	-0.157	0.0253	0.0227
	(<.000)	(<.000)	(<.000)	(0.019)	(0.036)
Knowledge	0.2338	-0.178	0.2676	-0.071	-0.050
	(<.000)	(<.000)	(<.000)	(<.000)	(<.000)
Experience	0.3384	-0.319	0.5126	-0.084	-0.047
	(<.000)	(<.000)	(<.000)	(<.000)	(<.000)
Income	0.3029	-0.223	0.4131	-0.055	-0.071
	(<.000)	(<.000)	(<.000)	(<.000)	(<.000)

Note: the p value of the t test is indicated in parentheses, and in the case of a significant level of 5%, it is significant when the value is less than 0.05.

that the investors with older age are more inclined to choose common investment varieties with lower risk, which is consistent with the intuitive feel that investors are more conservative with age. While "investment experience" and "the scale of investment" show a positive correlation, indicating that experienced investors have confidence in the operation of more money, and are more willing to take the money accounting more proportion of their incomes for investment. In addition, investors' "information decision-making speeds" show a negative correlation with the few factors including investors' "age", "knowledge", "income", which is also consistent with the intuitive feel that when individual investors are making judgments, the investors with more experience and higher knowledge level think more comprehensively and have lower blindness.

Logistic Regression Analysis

The above rank correlation analysis above only investigates the correlation between one characteristic variable and one behavioral variable each time, but in fact, the behavior performance of the investors is related to their comprehensive quality. And this kind of relation is complex, so it is difficult to be depicted by a simple linear model.

In fact, in addition to the personal characteristics, the factors that affect the behavior of the individual still include a variety of other random ones, and the impact of personal characteristics on the investment behavior is mainly showed on the probability. Logistic regression is a typical data mining method which is especially suitable for probability estimate of categorical data. Assuming that the probability of a behavioral variable (Y) is P, then note that:

$$logit(P) = log(\frac{P}{1-P}) \tag{1}$$

Among them, log is natural logarithm. Express $logit(P)$ as the function of independent variable (individual feature) X:

$$logit(P) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \tag{2}$$

This is the Logistic regression. Maximum likelihood method is usually adopted for its parameter estimation, and chi-squared statistic is usually adopted for the test of the significance of models and parameters. This paper utilizes stepwise selecting method during regression. Detailed analysis is based on software SAS 9.3 to calculate the cumulative logistic probability function in all levels of the multi-class dependent logistic function:

$$Logit(P|Y_i < h) = \beta_h + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$
$$(i = 1,2,3,4,5 \; ; \; h = 2,3,\dots,n) \tag{3}$$

Among them, Y_i is the representative of the i kind of investor behavior, n represents the investor behavior has n values. It can be converted to:

$$P(Y_i < h) = \frac{exp(\beta_h + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)}{1 + exp(\beta_h + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)}$$
$$(i = 1,2,3,4,5 \; ; \; h = 2,3,\dots,n) \tag{4}$$

It is further returns the probability for each value of investors' behavior variables:

$$P(Y_i = h) = \begin{cases} P(Y_i < h+1) - P(Y_i < h), & 2 \leq h < n \\ 1 - P(Y_i < h), & h = n \end{cases} \tag{5}$$

In this way, the probability of each value of each variable on the investment behavior can be expressed as a function of personal characteristics variables. In addition, in order to investigate the stability of the results, we have specifically analyzed the data collected from the three stages. **Table 6** shows the results.

Table 6. Logistic regression model results summary table

Dependent variable	Stage	Overall Evaluation of Model		Model parameter test				
		likelihood chi square ratio test	Score test	Age	Profession	Knowledge	Experience	Income
Scale	Stage 1	196.626	185.009	0.085	0.011	-0.128	-0.307	-0.106
		(<.0001)***	(<.0001)***	(0.0095)***	(0.7129)	(<.0001)***	(<.0001)***	(0.0008)***
	Stage 2	621.189	552.651	0.105	0.068	-0.128	-0.266	-0.197
		(<.0001)***	(<.0001)***	(<.0001)***	(0.0001)***	(<.0001)***	(<.0001)***	(.0001)***
	Stage 3	757.834	691.597	0.102	0.059	-0.113	-0.272	-0.184
		(<.0001)***	(<.0001)***	(<.0001)***	(0.0002)***	(<.0001)***	(<.0001)***	(<.0001)***
Variety	Stage 1	147.215	134.896	0.113	0.038	0.050	0.214	0.124
		(<.0001)***	(<.0001)***	(0.0011)***	(0.2221)	(0.0950)**	(<.0001)***	(0.0002)***
	Stage 2	527.741	489.548	0.119	0.005	0.130	0.233	0.064
		(<.0001)***	(<.0001)***	(<.0001)***	(0.7628)	(<.0001)***	(<.0001)***	(0.0015)***
	Stage 3	515.744	475.453	0.104	0.015	0.082	0.215	0.075
		(<.0001)***	(<.0001)***	(<.0001)***	(0.3320)	(<.0001)***	(<.0001)***	(<.0001)***
Frequency	Stage 1	608.482	495.124	-0.057	0.140	-0.240	-0.471	-0.209
		(<.0001)***	(<.0001)***	(0.0956)	(<.0001)***	(<.0001)***	(<.0001)***	(<.0001)***
	Stage 2	1246.240	1051.758	-0.001	0.052	-0.152	-0.396	-0.244
		(<.0001)***	(<.0001)***	(0.9689)	(0.0033)***	(<.0001)***	(<.0001)***	(<.0001)***
	Stage 3	1620.405	1347.931	0.008	0.071	-0.107	-0.481	-0.183
		(<.0001)***	(<.0001)***	-0.645	(<.0001)***	(<.0001)***	(<.0001)***	(<.0001)***
Speed	Stage 1	23.265	22.142	0.065	0.007	0.047	0.094	-0.064
		(0.0003)***	(0.0005)***	(0.051)***	(0.8188)	(0.1118)	(0.0057)***	(0.0499)***
	Stage 2	71.869	68.764	0.065	0.005	0.083	0.034	0.040
		(<.0001)***	(<.0001)***	(0.0019)***	(0.7712)	(<.0001)***	(0.1437)	(0.0504)***
	Stage 3	52.424	50.570	0.087	0.015	0.047	0.008	0.020
		(<.0001)***	(<.0001)***	(<.0001)***	(0.3522)	(0.0038)***	(0.6756)	(0.2452)
Channel	Stage 1	63.431	61.005	0.154	-0.023	0.095	-0.084	0.130
		(<.0001)***	(<.0001)***	(<.0001)***	(0.4600)	(0.002)***	(0.0606)	(0.0002)***
	Stage 2	33.518	33.064	0.078	-0.013	0.026	-0.058	0.061
		(<.0001)***	(<.0001)***	(0.0002)***	(0.4485)	(0.0162)***	(0.0094)***	(0.0025)***
	Stage 3	60.280	57.715	0.074	-0.008	0.0693	0.002	0.018
		(<.0001)***	(<.0001)***	(<.0001)***	(0.5955)	(<.0001)***	(0.9182)	(0.2859)

Note: 1. Numbers without brackets in the table model parameter test series represent the values of the standardized regression coefficients.
2. Values with brackets in the table are the significance test statistic P values under 95% significance level, and they are considered as significant when smaller than 0.05 and with *** as marker.

It can be seen from the table that in three different data gathering stages, the Logistic models of investors' behavior and personal characteristics pass the significance test as a whole, suggesting that the personal characteristics and behavior probability of individual investors have strong connections. And in those three different stages, the regression coefficient of each independent variable is also basically the same, fully indicating that this relationship has a relatively strong stability. Based on the regression coefficients, the probability of occurrence of corresponding behavior characteristics under different individual features can be calculated.

Further analyzing each variation coefficient, it is not hard to find that there is a connection between the "investment scale" of individual investors and their age, profession, professional knowledge levels, investment experiences and income level, among which the investment experiences, income level and professional knowledge levels have the largest absolute values of coefficient. This is easy to explain: income level reflects the financial strength of investors, whose risk preference and risk tolerance capacity will change with the increase of knowledge and experience.

"Variety of investment" is connected with investors' age, investment experiences, professional knowledge levels and income level, among which age and investment experiences have the largest absolute values of regression coefficient. Reasonable explanation maybe: investors become more familiar with the characteristics of investment products with the increase of their age, experience and knowledge, and their varieties of investment tend to grow larger.

"Transaction frequency" is most related to profession, professional knowledge levels, investment experiences and income level, especially "investment experiences". This shows that in the current stage, the behavior of individual investors in the market is mainly driven by subjective factors, and their decisions by experience. Meanwhile, it also indicates that the more experience investors have, the more likely it would be for them to adjust their investment strategy on their own.

As for "information channel", age and professional knowledge levels are the major factors, especially age. And it suggests that young investors are more open to accept new information channels while older investors are more dependent on traditional ones.

"Policy-decision speed" is obviously subject to the age, and other items of personal features show a kind of non-uniform characteristics and their relations also are unstable.

Especially it can be found, "professional knowledge levels", "investment experience" and "income level" these three features have the most significant influence on all types of investment behaviors, whose individual characteristic factors are worthy of being paid special attention to. Therefore, in the marketing management, the financial services can provide personalized service to these groups, and it can reduce the operating costs while improving the work efficiency.

Results from Data Mining

Data mining method is a kind of method which is based on inductive statistics. It belongs to a kind of data driven method, which is especially suitable for finding the hidden, complex and nonlinear models of the data (Lan et al., 2003). It has been widely used in financial analysis. We carry out data mining analysis based on IBM Modeler 15.0. Five classic and widely used classification models: C5.0 (decision-making tree), C&R (classification and regression tree), BP (neural network) and SVM (Supportive Vector Machine) and NB (Naive Bayes) are carried out. The data collected from the three stages are used as training set, test set and validation set. This not only can view the predictive effect, but also can understand the stability of the models in different periods. **Table 7** is the summary of results.

Table 7. Data Mining Model Result Summary Table

	C5.0	C&R	BP	NB	SVM
Scale	63.37%	63.43%	63.00%	64.14%	63.37%
	Experience > Income > Knowledge > Professional > Age				
Variety	72.18%	72.16%	71.96%	72.08%	72.80%
	Experience > Knowledge > Professional > Income > Age				
Frequency	70.37%	69.95%	70.37%	69.58%	70.00%
	Experience > Income > Knowledge > Professional > Age				
Speed	79.54%	79.54%	79.54%	79.50%	80.01%
	Knowledge > Experience > Income > Professional > Age				
Channel	50.26%	50.26%	50.26%	50.63%	51.00%
	Knowledge > Experience > Income > Professional > Age				

Note: the first row in the table represents the validation accuracy of the data mining model; the second row of the table content indicates the different importance of the personal characteristics.

From **Table 7**, we can see that by personal characteristics of individual investors, to predict the accuracy of their investment behavior has reached a good level. Among them, through the personal characteristics to predict the accuracy of investment varieties, transaction frequency and decision-making speed are all more than 70%. Hence, it can be considered that personal characteristics of individual investors can be used to predict the investment behaviors directly. At the same time, it can be seen in the personal characteristics of individual investors, these factors like investment experience of individual investors, professional knowledge levels, income level have a greater influence on the prediction of the investment behaviors, and the occupation of investors, age factors have less influence on the prediction ability, this conclusion is also consistent with Logistic regression analysis results.

CONCLUSION

In this paper, we used the questionnaires data on individual investors 'knowledge level, economic income, education background, family status, investment experience, and some investments behaviors, etc. to study the predictability of investment behaviors based on investors' personal characteristics of individual in China. No matter what analysis methods such as rank correlation analysis, logistic regression, and some other data mining techniques are applied, We can find that investors' personal characteristics are strong predictive to their behaviors. More detail, such three personal characteristics as "professional knowledge levels", "investment experience" and "income level" have the most significant predictability on all types of investment behaviors. When using the selected significant personal characteristics as input variables, we could construct effective predictive models based on data mining methods, which would provide a better understanding of the individual investors and explore the irrational behaviors of investors in the financial market in China. Furthermore, those results could not only provide decision information for investor education, marketing and service personalization, but also provide clues to strengthen customers' cost control, quantitative management and risk control, etc.

ACKNOWLEDGEMENTS

REFERENCES

Barnea, A., Cronqvist, H., & Siegel, S. (2010). Nature or nurture: What determines investor behavior? *Journal of Financial Economics, 98*(3), 583-604.a

Cao, J. (2009). A preliminary study on the development mode of securities brokerage business. *Market Weekly (Disquisition Edition), 07*, 96-97+72.

Chang, T., & Hui, W. (2016). A Multi Criteria Group Decision-making Model for Teacher Evaluation in Higher Education Based on Cloud Model and Decision Tree. *Eurasia Journal of Mathematics, Science & Technology Education, 12*(5).

Clark-Murphy, M., & Soutar, G. (2005). Individual investor preferences: a segmentation analysis. *The Journal of Behavioral Finance, 6*(1), 6-14.

Guo, S. (2013). The speech and answers in the economist's summit. Retrieved on 2013-01-16 from http:// finance. sina.com.cn/stock/y/20130116/141814304762.shtml (Chinese)

Hira, T. K., & Loibl, C. (2008). *Gender differences in investment behavior*. New York: Springer.

Huang, M., Zhou, X., & Wang, C. (2014). Are More Professional Customers Less Loyalty? An Empirical Study of Paradox of Customer Expertise of Fund Investors. *Nankai Business Review*, (1), 105-112. (Chinese)

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, 263-291.

Lan, Q., Ma, C., Wu, J., Gan, G. (2003). Application of Data Mining in Finance Analysis. *Management Reviews, 05*, 57-62+64. (Chinese)

Li, D., Li, X., & Xiao, B. (2011). Investors experience, reinforcement learning and securities market stability. *Social Sciences in Nanjing, 01*, 43-48+54. (Chinese)

Li, X., Wang, J., & Fu, H. (2002). The empirical study of individual securities investors' trade behavior in China. *Economic Research Journal, 11*(54), 3. (Chinese)

Li, Y., Gao, R., & Du, R. (2010). A research of investors characteristics and portfolio of geopolitical effect. *Modern Management Science, 11*, 13-15. (Chinese)

Liu, T. (2011). The different information of reference group on impact differences of financial investors behavior——Based on the perspective of the theory of prospect. *Southwestern university of finance and economics*. (Chinese)

Pei, P., & Zhang, Y. (2005). The empirical test of investors cognitive deviation in China. *Management World*, (12), 12-22. (Chinese)

Peng, X., & Wang, X. (1995). A study of the Shanghai stock holders' investing-behaviorand personality traits. *Psychological Science*, (2), 94-98. (Chinese)

Pompian, M. (2012). *Behavioral finance and wealth management: how to build optimal portfolios that account for investor biases*. New Jersey: John Wiley & Sons.

Pompian, M. M., & Longo, J. M. (2004). A new paradigm for practical application of behavioral finance: creating investment programs based on personality type and gender to produce better investment outcomes. *The Journal of Wealth Management, 7*(2), 9-15.

She, G. (2012). An empirical study on the influence factors of family financial asset allocation. *Central South University*. (Chinese)

Shi, K., Fan, H., & Xu, J. (2006). Research on risk perceptions of Chinese stock investors. *Journal of Management Sciences in China, 8*(6), 74-82. (Chinese)

Tan, S. (2013). Feedback effect, learning and overconfidence behavior of individual investors in China. *Economic Theory and Business Management, 11*, 71-79. (Chinese)

Tan, S., & Chen, Y. (2012). Can the investment experience improve the income of the investors? *Journal of Financial Research*, (5), 164-177. (Chinese)

Tan, S., & Wang, Y. (2006). Do these curities investors excessive trading?——based on the research of data a Chinese securities brokerage house. *Economic Research Journal, 41*(10), 83-95. (Chinese)

Wang, H., Lin, N., & Wang, X. (2012). The present situation and the development of financial market in China. *Modern Economic Information, 20*, 6+11. (Chinese)

Wang, L., Zheng, X., & Shi, J. (2003). Chinese securities investors' investment behavior and personality. *Psychological Science, 26*(1), 24-27. (Chinese)

Wang, Y. (2016). A Study on Kinmen Resident's Perception of Tourism Development and Culture Heritage Impact. *Eurasia Journal of Mathematics, Science & Technology Education, 12*(12).

Wen, F., He, Z., Dai, Z., & Yang, X. (2014). Characteristics of investors' risk preference for stock markets. *Economic Computation & Economic Cybernetics Studies & Research, 3*(48), 235-254.

Xu, T., & Chen, L. (2008). An empirical analysis of the individual investors age characteristics and its preference of dividend. *Economic Tribune, 18*, 106 -108+112. (Chinese)

Yan, B. (2002). An analysis of the investment behavior biases of securities investors in China——a survey for investors in jiangsu. *Modern Management Science, 2*, 009. (Chinese)

Yang, X., Zhu, J., & Jin, X. (2011). Investment and Happiness: An Empirical Study of Individual Investors in China. *Journal of Zhejiang University (Humanities and Social Sciences), 41*(2), 42-51. (Chinese)

Zhang, C., Liu, C., & Liao, L. (2012). Applying the logistic model to the rishy asset allocation: An empirical study on the questionnaires of the mutual fund investors. *Journal of Tsinghua University (Science and Technology), 08*, 1142-1149. (Chinese)

http://www.ejmste.com