Accounting and Corporate Management (2022)
Clausius Scientific Press, Canada

DOI: 10.23977/acccm.2022.040402
ISSN 2523-5788 Vol. 4 Num. 4

Institutional Investor Heterogeneity on Corporate Governance under the Background of Big Data

Yisui Wu[1,a], Pochang Ko[1,b], Juichan Huang[2,c,*]

[1]*Department of Intelligent Commerce, National Kaohsiung University of Science and Technology, Kaohsiung City, 824004, Taiwan*
[2]*Yango University, Fuzhou, 350015, China*
[a]*I111167101@nkust.edu.tw*, [b]*cobol@nkust.edu.tw*, [c]*wish0718@outlook.com*
**Corresponding author*

Keywords: Big Data, Institutional Investors, Investor Heterogeneity, Corporate Governance

Abstract: With the development of institutional investors, institutional investors have become an active subject in today's capital market. They begin to obtain long-term benefits by improving corporate governance, improving corporate performance and enhancing core competitiveness of listed companies. This study mainly discusses the impact of institutional investors' heterogeneity on corporate governance under the background of big data. The financial data of this study are all from wind database, and statal5.1 software is used for comprehensive operation. In order to ensure the logicality of the whole data research, this study eliminates st and * ST listed companies, companies with various data missing, and finally determines 2043 samples as the object of this study. In this study, ROA is taken as the overall performance index of the company, which shows the overall profit level of the enterprise. In addition, the ratio of research investment to main business income is used as one of the important indicators to evaluate the income intensity. In order to study the relationship between institutional investors and real earnings management activities, this paper uses the method of controlling the shareholding ratio of institutional investors to further explore whether the heterogeneity of institutional investors has different effects. In terms of control variables, we mainly consider several important indicators such as enterprise scale, asset liability ratio, cash flow and equity concentration. The average value of the overall shareholding ratio of institutional investors is 0.1208, and the average shareholding ratio of institutional investors is 12.08%, which is far less than the 50-60% shareholding ratio of mature market. The results show that the proportion of investors holding shares increases with the increase of return on total assets, which means that the higher the proportion of institutional investors, the better the performance of enterprises.

1. Introduction

There is no doubt that in the modern market economy, the main role of the company has become more prominent. However, the modern enterprise system developed from the separation of

company ownership has caused two problems: information asymmetry and agency costs. These two objective constraints constitute a threat to the interests of investors and hinder the normal operation of enterprises. As an important part of improving the corporate governance mechanism, institutional investors participate in corporate governance and promote the development and improvement of the capital market governance mechanism, which is an important area of current theoretical and empirical research.

In addition to a wider range of data analysis and real-time data collection, big data technology can also achieve automatic processing of public opinion related data. Big data technology can classify and sort out information and supplement the missing public opinion data. The data with obvious errors shall be tracked and collected or submitted for identification. This is also the process of reprocessing the data. This will make the information analyzed by public opinion data more accurate and provide relevant support for corporate governance.

The emergence of big data has promoted the development of many industries. Xu L believes that the basic idea of PPDM is to modify data in order to effectively execute data mining algorithms without compromising the security of sensitive information contained in the data. In fact, in the process of data collection, data release, and information (ie data mining results), undesired leakage of sensitive information may also occur. He looks at privacy issues related to data mining from a broader perspective, and studies various methods that can help protect sensitive information. His research lacks practice [1]. Mauro AD analyzes a compelling corpus of industry and academic articles related to big data to discover commonalities between the topics dealt with. He also compiled a survey of existing definitions in order to produce a more solid definition to cover most of the work happening in the field. The main themes of big data are: information, technology, methods and impact. He proposed a new definition of the term, as follows: "Big data is an information asset, which is characterized by such a high quantity, speed and variety, and requires specific technology and analysis methods to be transformed into value." He proposed The formal definition of can enable the concept of big data to develop more coherently, because it only relies on the basic elements of the latest technology and is consistent with the most popular definitions currently in use. This is one of the first structured attempts to build a convincing definition of big data. It also contains original explorations on topics related to library management. His research lacks experimental data [2]. Baccarelli E formalized this paradigm, discussed its most important application opportunities, and outlined the main challenges of real-time energy-efficient management of distributed resources available on mobile devices and Internet-connected data centers. It also includes performance analysis of small-scale prototypes to gain insight into the trade-offs between energy and performance that can be achieved through optimized design of resource management modules. Performance comparisons with some of the most advanced resource managers confirm this discussion. His research lacks innovation [3]. Su Z proposed a novel framework. First, he studied the characteristics and challenges of mobile big data. Then introduced the content-centric network architecture, used to transmit mobile big data in MSN, where each data is composed of interest packets and data packets. Next, by defining the priority of interest packets and data packets, how to choose a proxy node to forward interest packets and relay nodes to send data packets is given. Finally, the simulation results show the performance of our framework with various parameters. His research has no practical significance [4].

The financial data of this study are all from the wind database, and use Stal5.1 software for comprehensive operation. This research mainly uses ROA as the overall performance indicator of the company. This indicator shows the overall profitability of the company. In addition, the ratio of research investment to main business income is used, and this is used as an important factor in evaluating income intensity. One of the indicators. In order to study the relationship between institutional investors and corporate real earnings management activities, the method of controlling

the shareholding ratio of institutional investors is used to further explore whether the heterogeneity of institutional investors has different effects. In terms of control variables, it mainly considers several important indicators of enterprise size, asset-liability ratio, cash flow, and equity concentration.

2. Corporate Governance

2.1 Big Data

Big data stream mobile computing is proposed as a paradigm that relies on the integration of broadband Internet mobile networks and real-time mobile cloud computing [5-6]. It aims to promote the rise of new self-configuring integrated computing communication platforms to offload and process big data streams acquired by mobile/wireless devices with limited resources in real time [7]. The era of big data has begun, and organizations in all industries are investing heavily in big data initiatives. From previous studies, we know that only one investment can not produce a competitive advantage [8-9]. Instead, companies need to create capabilities that competitors cannot match. Using the company's resource-based theory and the latest research on big data, various resources have been identified, which together constitute a big data analysis (BDA) capability and create a tool to measure the company's BDA capability. Test the relationship between BDA capabilities and company performance. Evidence shows that BDA capabilities can bring excellent company performance [10-11].

However, a large amount of data also brings us into the era of big data. In the new era, new data collection, transmission and processing technologies are needed [12-13].To ensure ubiquitous data collection, the scale of the mobile eHealth network must be expanded [14]. Big data provides a lot of opportunities for mobile network operators to improve service quality. First according to the framework of big data-driven (BDD) mobile network optimization [15-16]. Then, the characteristics of big data collected not only from user devices but also from mobile networks are good breakthroughs. In addition, from the perspective of network optimization, for the feasibility of the framework, some user cases about the application of the proposed framework in improving network performance should be given. With the integration of emerging fifth-generation (5G) mobile networks and big data analysis, the quality of our daily mobile life is expected to be greatly improved. This research mainly discusses corporate governance issues from the perspective of big data. Of course, it will also further discuss corporate performance and corporate profit and loss issues [17]. The research process is shown in Figure 1.

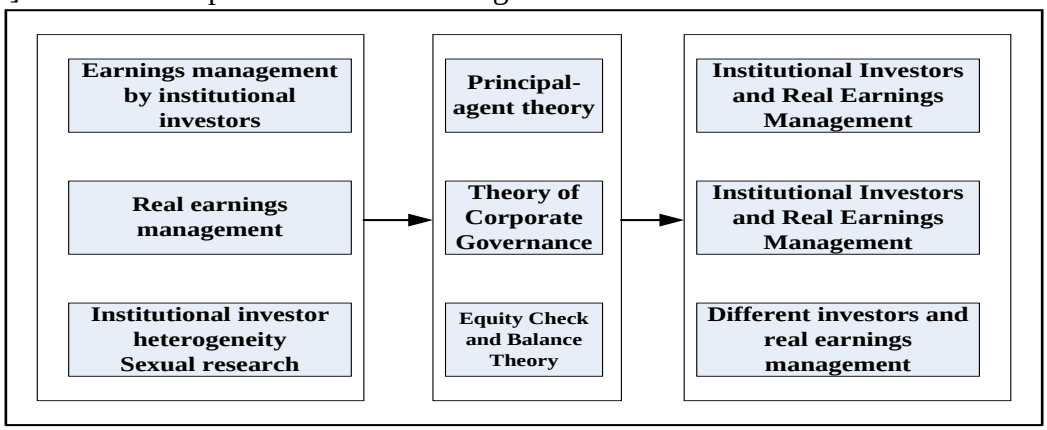

Figure 1: Research process

2.2 Heterogeneity of Institutional Investors

Excluding the first phase of institutional investor income, the income obtained in each period can be expressed as $\beta V_3 - C$. The overall benefits of institutional investors participating in corporate governance are:

$$R_1 = (\beta V_1 - C) + (\beta V_3 - C) * \delta + \ldots + (\beta V_3 - C) * \delta^n \quad (1)$$

Can be replaced with:

$$R_1 = (\beta V_1 - C) + (\beta V_3 - C) * \frac{\delta(1-\delta^n)}{1-\delta} \quad (2)$$

If the institutional investors choose not to supervise, the management must choose not to work hard, then the income obtained by the institutional investors in each period is βV_2, and the total income obtained by the institutional investors is:

$$R_2 = \beta V_2 + \beta V_2 * \delta + \ldots + \beta V_2 * \delta^n \quad (3)$$

Which is:

$$R_2 = \beta V_2 (1-\delta^n)/(1-\delta) \quad (4)$$

When $R1 = R2$ and the period n tends to infinity, then:

$$(\beta V_1 - C) + (\beta V_3 - C) * \frac{\delta(1-\delta^n)}{1-\delta} = \beta V_2 (1-\delta^n)/(1-\delta) \quad (5)$$

The critical value for deciding whether to supervise institutional investors can be obtained δ^*:

$$\delta^* = (\beta V_2 - \beta V_1 + C)/(\beta V_3 - \beta V_1) \le \delta \quad (6)$$

According to the basic explanation of the above game theory, δ represents the actual cooperation enthusiasm of the participants, that is, the larger, the stronger the enthusiasm of institutional investors to invest [18-19].

2.3 Heterogeneity Test Model

Drawing lessons from the intermediary effect test method of three-step regression, the following intermediary effect test model was constructed to test the proposed research hypothesis.

$$ROA_1 = \alpha_0 + \alpha_1 ins + \alpha_2 Control + \varepsilon \quad (7)$$

$$ROA_2 = \alpha_0 + \alpha_1 ins + \alpha_2 \varepsilon \quad (8)$$

$$RD = \alpha_0 + \alpha_1 inst + \alpha_2 Control + \varepsilon \quad (9)$$

$$ROA_3 = \alpha_0 + \alpha_1 inst + \alpha_2 RD + \alpha_3 Control + \varepsilon \quad (10)$$

Among them, ROA_1, ROA_2 can be used to study the impact of investors as a whole and heterogeneous institutional investors on corporate performance [20]. ROA_2, ROA_3, RD use the method of stepwise testing of regression coefficients into three steps to test the mediating effect of R&D investment [21-22].

Dummy variable (inst) of the stability of institutional investors: Taking the two dimensions of time and industry as the standard, it is divided according to the size and stability of the proportion of institutional investors in a certain industry. The formula is:

$$SD_{i,t} = \frac{INS_{i,t}}{STD\left(INS_{i,t-3}, INS_{i,t-2}, INS_{i,t-1}\right)} \tag{11}$$

$$INST_{i,t} = \begin{cases} 1, SD_{i,t} \geq MEDIAN_i \\ 0 \end{cases} \tag{12}$$

Among them, $SD_{i,t}$ is the company's institutional shareholding ratio in t [23].

Measurement model:

$$ROA_{i,t} = \alpha + \beta_1 INST_{i,t-1} + \beta_2 Size_{i,t-1} + \beta_3 Lev_{i,t-1} + \sum Industry + \sum Year \tag{13}$$

For an institutional investor k, the total assets of stocks bought and sold cumulatively are:

$$CR\_buy_{k,t} = \sum_{i=1}^{N_k} \left| S_{k,t} P_{k,t} - S_{k,t-1} P_{k,t-1} - S_{k,t} \Delta P_{k,t} \right| \tag{14}$$

$$CR\_sell_{k,t} = \sum_{i=1}^{N_k} \left| S_{k,t} P_{k,t} - S_{k,t-1} P_{k,t-1}, S_{k,t+1} \Delta P_{k,t+1} \right| \tag{15}$$

Among them, $CR\_buy_{k,t}$ represents buying, and $CR\_sell_{k,t}$ represents total selling[24-25]. Then, it is necessary to calculate the turnover rate of each institutional investor CR:

$$CR_{k,t} = \frac{\min\left(CR\_buy_{k,t}, CR\_sell_{k,t}\right)}{\sum_{i=1}^{N_k} \frac{S_{k,t} P_{k,t} + S_{k,t-1} P_{k,t-1}}{2}} \tag{16}$$

Calculate the average turnover rate based on the turnover rate of institutional investors in the past year:

$$AVG\_CR_k = \frac{1}{2}\left(CR_{k,t}, CR_{k,t-1}\right) \tag{17}$$

Finally, institutional investors are divided into three groups according to the value of AVG_CR_k, and the highest group is recorded as short-term institutional investors [26]. The lowest group is recorded as long-term institutional investors [27].

3. Corporate Governance Experiment

3.1 Data Source and Sample Selection

At present, listed companies have not clearly announced the actual total amount of R&D investment, and there is no uniform disclosure standard. Some listed companies clearly divide capitalized expenditures and expensed expenditures, so there may be some misunderstandings in understanding the content and knowledge involved in "development expenditure" items, and the actual expense items disclosed by some companies are not clear enough. Therefore, at the same

time as the target, you can choose the project to manually collect R&D expenditure data. The data is not true and comprehensive, which may affect the actual empirical results. In contrast, Shenzhen listed companies disclose more research and development expenses, including the amount of research expenses and development expenses. At the same time, R&D investment is industry-sensitive, and most companies that disclose R&D investment data are concentrated in manufacturing and information technology services. In summary, the final sample selected for this study is companies in the manufacturing and information technology service industries.

3.2 Variable Design

(1) Main research variables

1) Enterprise performance variables

Through the introduction of the previous article, we can know that there are various indicators to measure the company's operating performance, including net profit, ROA, ROE, etc. These measurement methods have certain uniqueness, and various data and materials are easily obtained, and the calculation is relatively simple. While studying, this article mainly uses ROA as the company's overall performance indicator. This indicator shows the overall profitability of the company, and the information contained in it is more specific. The higher the ROA, it means that the company's overall performance development level is higher.

2) R&D investment variables

This study mainly adopts the ratio of research input and main business income, and uses this as one of the important indicators to evaluate income intensity.

(2) Explaining variables

Research the relationship between institutional investors and corporate real earnings management activities, and whether the heterogeneity of institutional investors has different effects. The selected explanatory variables are as follows.

1) In Hypothesis 1, the shareholding ratio of institutional investors is INST-T. There are two methods for calculating the shareholding ratio of institutional investors:

a. Institutional investor’s shareholding ratio = number of institutional investors’ shareholding at the end of the third quarter/total number of shares at the end of the third quarter

b. Institutional investors’ shareholding ratio = the number of institutional investors’ shareholding at the end of the year/total number of shares at the end of the year. Due to the use of the end-of-year institutional investor shareholding ratio for regression, serious endogenous problems will arise.

Therefore, the institutional investor holding ratios in this study all use the data at the end of the third quarter, while other variables are all year-end data. The shareholding ratios of other institutional investors in the following text are also data at the end of the third quarter.

2) In Hypothesis 2, the institutional investors holding less than 6% are represented by INST-L, 6%-30% are represented by INST-M, and those greater than 30% are represented by INST-H.

3) In hypothesis 3, the shareholding ratio of different types of institutional investors is represented by INST-N, where N-1,2,3,4,5, 1 represents securities investment funds, 2 represents social security funds, 3 represents QFII, and 4 Represents insurance funds, and 5 represents securities companies.

3.3 Control Variables

The specific definition of control variables is shown in Table 1.

1) Enterprise size (Size): Shows the total assets of the enterprise as a whole. According to the general law of development, large enterprises and small enterprises have a higher level of overall technological innovation development. Moreover, the larger the scale of the development of large

enterprises, the sufficient internal funds of the enterprise do not have certain cash risks. Moreover, the construction of various management systems within the company is relatively complete, and the company has a strong ability to resist risks, so there are fewer behaviors in the company to reduce R&D investment.

2) Asset-liability ratio (lev): A high asset-liability ratio is a hindering factor for companies to carry out various businesses, and may lead to corporate bankruptcy. The debt-to-asset ratio is one of the important financial indicators of enterprise development, and there is an inseparable relationship between it and the actual performance development of the enterprise.

3) Cash flow (cash): Cash flow is one of the important components of corporate financial indicators, which directly determines the level of corporate performance. The enterprise has more cash flow, the overall development strength of the enterprise is strong, and the performance level is higher. The ratio of net cash flow from business operations to total operating income is one of the important indicators for evaluating cash flow.

4) Ownership concentration (top): By consulting a large number of domestic and foreign documents, we can know that the ownership concentration system may affect corporate performance in practice.

5) Turnover of total assets (turn): The higher the turnover rate, and the effective use of internal assets, it shows the strength of enterprise development. Enterprises should speed up their total asset turnover rate as much as possible, and increase their own profitability during the development process. The turnover rate of total assets gradually increases with the increase in the utilization rate of corporate assets. All in all, the total assets of an enterprise are an important factor affecting enterprise performance.

6) Annual dummy variable (Y): The market macro-economy changes at any time. Under different years and development environments, the market macro-economy may change at any time, which directly affects corporate performance. Therefore, enterprises must pay close attention to the development of annual dummy variables while developing, and set them as dummy variables.

7) Industry dummy variables (I): The actual corporate performance levels of different industries are obviously different, and industry (IND) dummy variables are introduced.

Table 1: Specific definitions of control variables

Variable type	Variable name	Variable symbol	Variable description
Explained variable	Return on total assets	Roa1	Net profit/total assets, the next period
Explanatory variables	Institutional investors' overall shareholding ratio	Ins	Total shareholding ratio of institutional investors
Mediating variable	R&D intensity	Lev	R&D expenditure/operating income
Control variable	Company Size	Cash	Ln (total assets)
	Financial leverage	Top	Total liabilities/total assets
	Total asset turnover	Variable symbol	Income/total assets
	Year	Roa1	Annual variable
	Industry	I	Industry variables

4. Corporate Governance Analysis

4.1 Variable Statistics

The descriptive statistical results of the main variables are shown in Table 2. During the research

sample period selected in this study, the average value of the listed company performance variable ROA is 0.04, and the standard deviation is 0.05. The mean of TobinQ is 2.14, the median is 1.70, and the standard deviation is 1.35. The results of the descriptive statistical analysis of the main variables are shown in Figure 2. The semiparametric regression method of Levinsohn and Petrin estimates that the average TFP of the company is 9.39 and the standard deviation is 1.5, which means that there is a significant difference between the production efficiency of listed companies. The average value of return on total assets (roa1) is 0.0692, the maximum value is 0.445, and the minimum value is -0.5. The difference between the maximum value and the minimum value is very large, and the development is unstable, which means that various enterprises in my country The actual operating conditions are quite different. The average value of R&D intensity (rd) is 0.0460, the minimum value is close to 0, and the maximum value is 0.247. The actual R&D expenditures of different enterprises are obviously different. The average shareholding ratio (ins) of institutional investors is 0.395, which means that the average shareholding ratio of institutional investors in the selected sample has reached 39.5%. The maximum value is 0.881, the minimum value tends to 0, and the standard deviation is 0.238, indicating that the scale of institutional investors has increased greatly in recent years and has become an important force that cannot be ignored in the secondary market. At the same time, the shareholding ratios of institutional investors of different companies also show obvious differences. A high probability of this is because China has implemented a Qualified Foreign Institutional Investor (QFII) system for a short period of time, and the government has stricter approvals for its entry into the capital market. This has led to a small number of listed companies holding shares and investment with domestic institutions. There is a large gap in the number of participants.

Table 2: Descriptive statistical results of main variables

Vaniable	obseIvation	Mean	D	Min	Median	Max
ROA	22740	0.0400	0.0500	-0.150	0.0400	0.190
TFP-LP	22740	2.140	1.350	0.900	1.700	8.760
TFP -ACF	5278	9.390	1.500	0.0000	9.480	12.14
UDINS	5278	17.02	2.360	0.0000	17.22	20.38
DINS	22740	0.0700	0.0900	0.0000	0.0500	0.550
Domesic	22740	0.0300	0.0400	0.0000	0.0200	0.170
QFII	22740	0.0400	0.0800	0.0000	0.0200	0.500
Size	22740	0.0000	0.0000	0.0000	0.0000	0.0300
Lev	22732	22 .0400	1.2800	19.7500	21.8600	26.0000
Indepen	22732	1.4e+10	8.4e+10	3. 1e+06	3.1e+09	5.6e+1 2

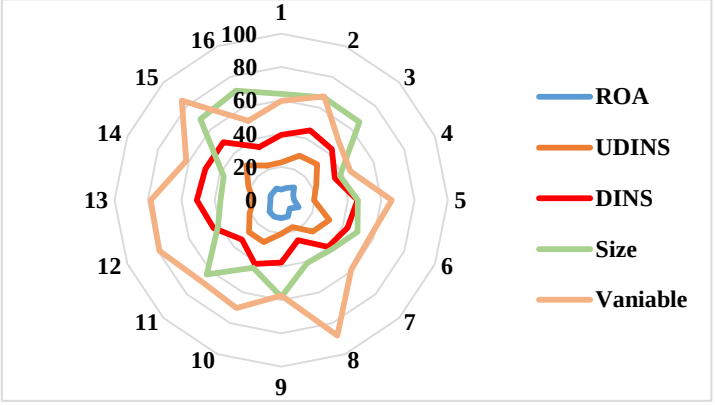

Figure 2: Results of descriptive statistical analysis of main variables

Table 3: Comparison of the low shareholding ratio group of institutional investors and the high shareholding ratio group of institutional investors

Variable symbol	Institutional investors low shareholding ratio group		Institutional investors high shareholding ratio group		T test
	Sample size	Mean	Sample size	Mean	
ROA	10867	0.0349	10873	0.0488	-13.014
TobinQ	10867	2.1996	10873	2.4048	-2.0642

Table 4: Comparison of the low shareholding ratio of independent institutional investors and the high shareholding ratio of independent institutional investors

Variable symbol	Independent institutional investor group with low shareholding ratio		Independent institutional investor group with low shareholding ratio		T test
	Sample size	Mean	Sample size	Mean	
ROA	10869	0.0301	10871	0.0536	-22.1356
TobinQ	10869	2.2097	10871	2.395	-1.8624

Table 5: Comparison of non-independent institutional investors with low shareholding ratio group and high shareholding ratio group

Variable symbol	Independent institutional investor group with low shareholding ratio		Independent institutional investor group with low shareholding ratio		T test
	Sample size	Mean	Sample size	Mean	
ROA	10869	0.0452	10871	0.0385	6.2742
TobinQ	10869	2.2797	10871	2.3248	-0.4538

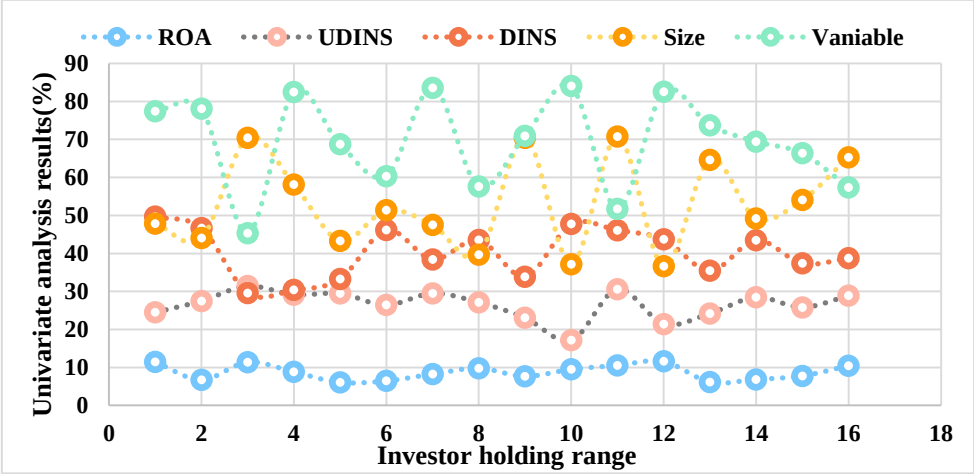

Figure 3: A univariate analysis of the relationship between institutional investor holdings and corporate performance

Before empirical regression analysis, the relationship between institutional investor holdings and corporate performance is tested from the perspective of univariate analysis. According to whether it is greater than the median value of the research sample as the grouping standard, the comparison between the low and high shareholding ratio groups of institutional investors is shown in Table 3.The comparison between the low shareholding ratio and high shareholding ratio of independent institutional investors is as follows As shown in Table 4, the comparison between the low shareholding ratio group and the high shareholding ratio group of non-independent institutional investors is shown in Table 5. We can see that companies with a significant share of institutional investors perform better. Figure 3 shows the relationship between institutional investor holdings and

corporate performance from a univariate analysis. From the descriptive statistical results of the main variables, the average and median of the real earnings management activity RM are -0.331 and 0.0892, respectively, while the maximum and minimum are 5.7371 and -9.5931, respectively, which indicates the existence of listed companies in China Two-way real earnings management activities include both downward adjustment of profits and upward adjustments to increase profits, but the scope of upward real earnings management activities is greater than that of downward real earnings management activities.

4.2 Institutional Investor Holdings on Corporate Performance

The results of panel data fixed effect estimation are shown in Table 6. The results of panel data fixed effect estimation analysis are shown in Figure 4. It can be found from Table 6 that the average of the overall shareholding ratio of institutional investors is 0.1208, and the overall average institutional investors hold 12.08% of the shares of listed companies, which is far less than the 50-60% shareholding ratio of mature markets. In addition, it is found that the relationship between institutional investor's shareholding and corporate performance is not only statistically significant but also economically significant. Specifically, if the institutional investor's shareholding ratio (Instit) increases by one standard deviation, the company Financial performance ROA will increase by 0.7%, while company market performance TobinQ will increase by 18%. The shareholding ratio of investors gradually increases with the increase in the return on total assets, which means the shareholding ratio of institutional investors The higher the enterprise performance, the better. The dummy variable of institutional investor stability is positively correlated with the return on total assets. Compared with transactional institutional investors, stable institutional investors have a stronger positive effect. Other related variables are kept within the range of 0.5.

Table 6: Results of panel data fixed effects estimation

	ROA1	TobinQ1	ROA2	TobinQ2	ROA3	TobinQ3
Size	-0.0060	-0.6850	-0.0051	0.6710	-0.0060	-0.6850
	(-4.50)	(-19.58)	(-3.93)	(-19.20)	(-4.49)	(-19.56)
Lev	-0.0237	0.5010	-0.0269	0.4450	-0.0237	0.5000
	(-4.48)	(4.78)	(-5.17)	(4.26)	(-4.49)	(4.77)
Indepen	-0.0037	-0.0317	-0.0037	0.0317	-0.0037	-0.0324
	(-1.63)	(-0.61)	(-1.64)	(-0.61)	(-1.63)	(-0.62)

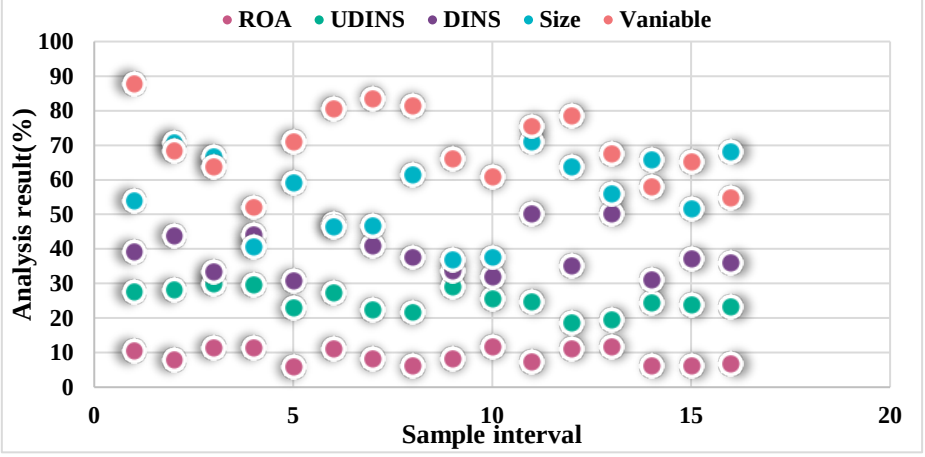

Figure 4: Panel data fixed effect estimation analysis results

Table 7 shows the high shareholding ratio group of independent institutional investors and the

high shareholding group of non-independent institutional investors. Table 8 shows the group with low shareholding ratio of domestic institutional investors and the group with high shareholding ratio of domestic institutional investors. The independent variables of the model only include the overall shareholding ratio (ins) of the institution's shareholding. Through analysis, it can be known that the institution's shareholding (ins) and corporate performance (roa1) show a positive development relationship. (a=0.0381, p<0.01), which means that if the institution's shareholding increases, the actual performance level of the company will also increase significantly. The conclusion of hypothesis 1 is verified. In the regression line again, it is also found that the correlation coefficient between cash strength (cash) and operating performance (roa1) is 0.052, and the two show a positive development relationship, which means that the company's cash flow increases with the gradual increase in operating capabilities. Asset-liability ratio (lev) and operating performance (ROA) show a negative development relationship, but the relationship between the two is not obvious. Asset turnover rate (turm) and operating performance (roa1) are significantly positively correlated, indicating that asset operating efficiency is an important factor affecting corporate performance. There is a close relationship between top ownership and corporate performance, with a correlation coefficient of 0.14. This fully demonstrates that the ownership structure is also a key factor affecting the level of company performance. In some companies with low equity concentration, the high concentration of equity may control agency costs and directly affect company performance. According to the domestic institutional investor holdings (Domestic) and company performance,which means that domestic institutional investors can improve corporate performance by participating in investment companies. QFII holdings have not yet been able to play the role of "value creator", and its incentive function to improve company performance still needs to be strengthened.

Table7: Independent institutional investors high shareholding ratio group and non-independent institutional investors high shareholding group

Variable symbol	Independent institutional investor high shareholding ratio group		Non-independent institutional investor high shareholding group		T test
	Sample size	Mean	Sample size	Mean	
ROA	10871	0.0536	10871	0.0385	14.2746
TobinQ	10871	2.2797	10871	2.3248	-0.7528

Table 8: Domestic institutional investors with low shareholding ratio group and domestic institutional investors with high shareholding ratio group

Variable symbol	Independent institutional investor high shareholding ratio group		Non-independent institutional investor high shareholding group		T test
	Sample size	Mean	Sample size	Mean	
ROA	10871	0.0353	10870	0.0484	12.1854
TobinQ	10871	2.2001	10870	2.4044	-2.0552

4.3 F Test and Hausman Test Analysis

When the majority of shareholders' desire to control the company is dashed and shareholders give up the right to participate in corporate governance, it provides a hotbed and opportunity for managers' "inside control". At this time, large institutional investors show its importance.The strength and status of institutional investors determine that they have a natural advantage over individual investors in this respect, and their cooperation has increased energy far more than thousands, tens of thousands of individual small shareholders. Therefore, institutional investors can

alleviate the internalization of supervision costs and free-riding problems of small and medium investors in corporate governance by using their financial advantages and professional talents. The result of F test is shown in Figure 5. The results of Hausman test are shown in Figure 6. In order to select the optimal model, this study performed F test and Hausman test in turn: the F test result after regression showed P=0.0000, strongly rejecting the null hypothesis that the variance is not significantly different, that is, the fixed effect model is better than the mixed regression model, and Hausman The test results show that P=0.0000. The test results are shown in Table 9.

Table 9: Test results

Model checking	Null hypothesis	Test result
F test	There is no significant difference in variance	Prob> F = 0.0000
Hausman test	Support random effects	Prob>chi2 = 0.0000

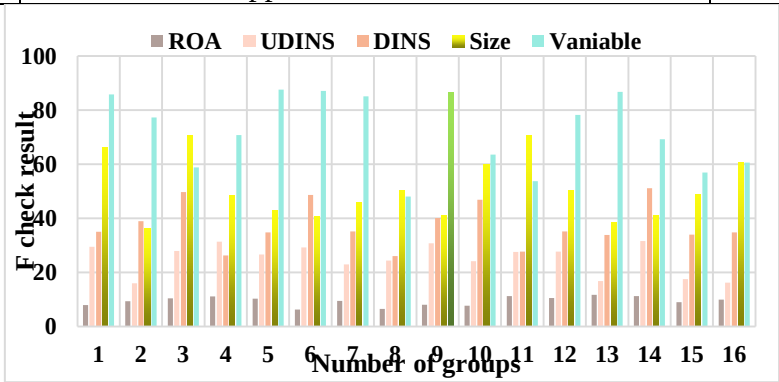

Figure 5: F test results

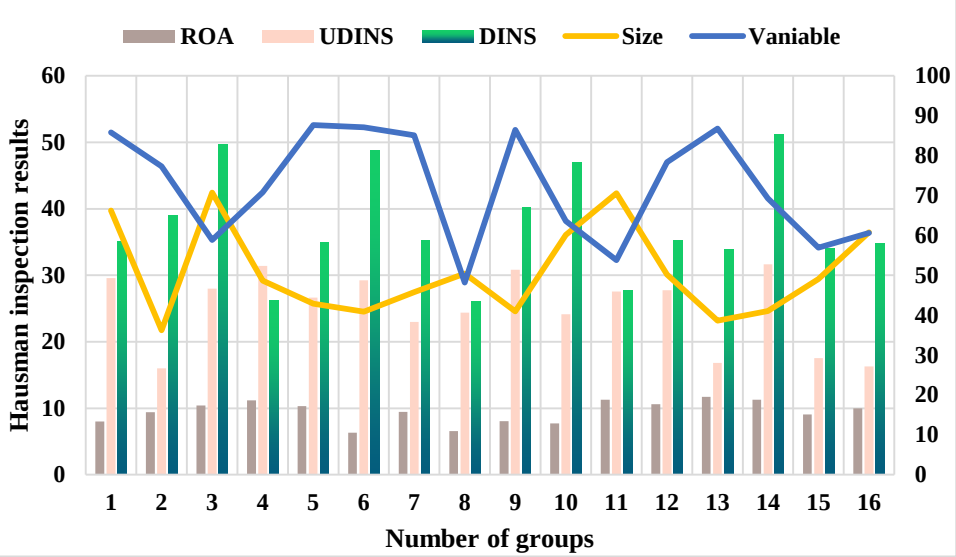

Figure 6: Hausman test results

This simple way of making decisions through sampling analysis and empirical judgment is no longer applicable, and the data obtained through the overall data analysis is more convincing. Therefore, it has become an inevitable trend that big data decision-making replaces empirical decision-making. If an enterprise wants to achieve greater improvement, it must trust the facts based on the overall analysis rather than the judgment of any elite or interest group. The process of using big data technology to collect and analyze all data is inseparable from the use of sensor technology

and the Internet of Things. In the process of using big data technology, the Internet of Things can collect and update data in real time, thus ensuring that some time-sensitive information can be used in time, so that decisions can be made based on the latest situation. In this process, enterprise decision-making will be transformed from static to dynamic, because real-time update, from data acquisition to data analysis to decision-making time will be greatly shortened, and the efficiency of decision-making will also be greatly improved. Of course, decisions made based on big data include more employee data because of the broad scope of the data. This is essentially the increase in employee participation. When companies no longer rely solely on their own experience to make decisions or only consider the interests of a certain class, the results of the decision will become more reasonable and fair. Decisions made in this way can be more supported by employees, and it is easier to get a positive response. In this way, it is easier for companies to use big data for governance innovation to improve governance. All in all, the application of big data technology is conducive to companies to more effectively analyze information resources in the fields of politics, economy, culture, etc., and to provide more reliable support for corporate decision-making. The use of big data technology when companies make decisions will make decisions more It fits reality, conforms to public opinion, is more responsive, fairer, and enterprises make more scientific decisions. Change the independent variable of the model to the dummy variable (inst) of institutional investor stability, and the regression results are shown in Table 10. The relationship between corporate governance time and revenue is shown in Figure 7. It can be seen from the regression results that corporate performance (roa1) is significantly positively affected by the dummy variable (a=0.0113, p<0.01). It shows that prudent institutional investors can effectively promote corporate performance, while transactional institutional investors cannot. The hypothesis holds. Similarly, in this regression, we found that corporate cash strength (cash), equity concentration (top), asset turnover (turn) are still significantly positively correlated with corporate performance (roal).

Table 10: Regression results

Model / Variable	Model 1	Model 2
Size	-0.0269	-0.0209
	(0.00288)	(0.00313)
Lev	-0.00895	-0.00329
	(0.0158)	(0.0158)
Cash	0.0520	0.0487
	(0.0131)	(0.0131)

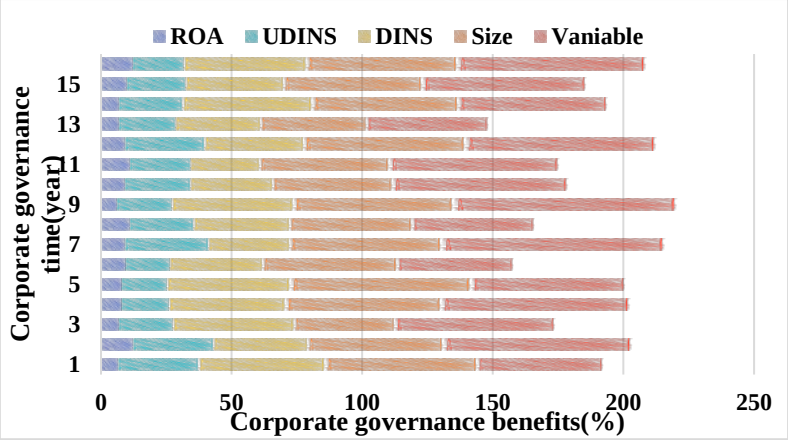

Figure 7: The relationship between corporate governance time and revenue

5. Conclusions

According to the general law of development, large enterprises and small enterprises have a higher level of overall technological innovation development. Moreover, the larger the scale of the development of large enterprises, the sufficient internal funds of the enterprise do not have certain cash risks. Moreover, the construction of various management systems within the company is relatively complete, and the company has a strong ability to resist risks, so there are fewer behaviors in the company to reduce R&D investment. A high debt-to-asset ratio is a hindering factor for companies to carry out various businesses, and may lead to corporate bankruptcy. The debt-to-asset ratio is one of the important financial indicators of enterprise development, and there is an inseparable relationship between it and the actual performance development of the enterprise.

Cash flow is one of the important components of corporate financial indicators, which directly determines the level of corporate performance. The enterprise has more cash flow, the overall development strength of the enterprise is strong, and the performance level is higher. The ratio of net cash flow from business operations to total operating income is one of the important indicators for evaluating cash flow.The higher the turnover rate, and the effective use of internal assets, demonstrates the strength of enterprise development. Enterprises should speed up their total asset turnover rate as much as possible, and increase their own profitability during the development process. The turnover rate of total assets gradually increases with the increase in the utilization rate of corporate assets. All in all, the total assets of an enterprise are an important factor affecting enterprise performance.

The financial data of this study are all from the wind database, and use Stal5.1 software for comprehensive operation. This research mainly uses ROA as the overall performance indicator of the company. This indicator shows the overall profitability of the company. In addition, the ratio of research investment to main business income is used, and this is used as an important factor in evaluating income intensity. In order to study the relationship between institutional investors and corporate real earnings management activities, the method of controlling the shareholding ratio of institutional investors is used to further explore whether the heterogeneity of institutional investors has different effects. In terms of control variables, it mainly considers several important indicators of enterprise size, asset-liability ratio, cash flow, and equity concentration.

References

[1] Xu L , Jiang C , Wang J , et al. Information Security in Big Data: Privacy and Data Mining[J]. IEEE Access, 2017, 2(2):1149-1176.

[2] Mauro A D , Greco M , Grimaldi M . A formal definition of Big Data based on its essential features[J]. Library Review, 2016, 65(3):122-135.

[3] Baccarelli E , Cordeschi N , Mei A , et al. Energy-efficient dynamic traffic offloading and reconfiguration of networked data centers for big data stream mobile computing: review, challenges, and a case study[J]. Computers & Chemical Engineering, 2016, 91(2):182-194.

[4] Su Z , Xu Q , Qi Q . Big data in mobile social networks: a QoE-oriented framework[J]. IEEE Network, 2016, 30(1):52-57.

[5] Obermeyer Z , Emanuel E J . Predicting the Future - Big Data, Machine Learning, and Clinical Medicine.[J]. N Engl J Med, 2016, 375(13):1216-1219.

[6] Chen M , Ma Y , Song J , et al. Smart Clothing: Connecting Human with Clouds and Big Data for Sustainable Health Monitoring[J]. Mobile Networks & Applications, 2016, 21(5):1-21.

[7] Lei Y , Jia F , Lin J , et al. An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data[J]. IEEE Transactions on Industrial Electronics, 2016, 63(5):3137-3147.

[8] Zheng K , Yang Z , Zhang K , et al. Big data-driven optimization for mobile networks toward 5G[J]. IEEE Network, 2016, 30(1):44-51.

[9] Rathore M M U , Paul A , Ahmad A , et al. Real-Time Big Data Analytical Architecture for Remote Sensing Application[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2017,

8(10):4610-4621.
[10] Athey S . [Special Issue Perspective] Beyond prediction: Using big data for policy problems[J]. ence, 2017, 355(6324):483-485.
[11] Wang K , Shao Y , Shu L , et al. Mobile big data fault-tolerant processing for ehealth networks[J]. IEEE Network, 2016, 30(1):36-42.
[12] Gligorijevi? V , Malod-Dognin N , Pr?Ulj N . Integrative methods for analyzing big data in precision medicine[J]. Proteomics, 2016, 16(5):741-758.
[13] Zhang Y . GroRec: A Group-Centric Intelligent Recommender System Integrating Social, Mobile and Big Data Technologies[J]. IEEE Transactions on Services Computing, 2016, 9(5):786-795.
[14] Kune R , Konugurthi P K , Agarwal A , et al. The anatomy of big data computing[J]. Software Practice & Experience, 2016, 46(1):79-105.
[15] Gupta M , George J F . Toward the Development of a Big Data Analytics Capability[J]. Information & Management, 2016, 53(8):1049-1064.
[16] Whyte J , Stasis A , Lindkvist C . Managing change in the delivery of complex projects: Configuration management, asset information and 'big data'[J]. International Journal of Project Management, 2016, 34(2):339-351.
[17] Wu J , Guo S , Li J , et al. Big Data Meet Green Challenges: Big Data Toward Green Applications[J]. IEEE Systems Journal, 2016, 10(3):888-900.
[18] Hossain M S , Muhammad G , Alhamid M F , et al. Audio-Visual Emotion Recognition Using Big Data Towards 5G[J]. Mobile Networks and Applications, 2016, 21(5):753-763.
[19] Zhong R Y , Lan S , Xu C , et al. Visualization of RFID-enabled shopfloor logistics Big Data in Cloud Manufacturing[J]. The International Journal of Advanced Manufacturing Technology, 2016, 84(1-4):5-16.
[20] Stevens H . Big Data, Little Data, No Data: Scholarship in the Networked World[J]. Journal of the Association for Information ence & Technology, 2016, 67(3):751–753.
[21] Wu J , Guo S , Li J , et al. Big Data Meet Green Challenges: Greening Big Data[J]. IEEE Systems Journal, 2016, 10(3):873-887.
[22] Xu Z , Frankwick G L , Ramirez E . Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective[J]. Journal of Business Research, 2016, 69(5):1562-1566.
[23] Jiang P , Winkley J , Zhao C , et al. An Intelligent Information Forwarder for Healthcare Big Data Systems With Distributed Wearable Sensors[J]. IEEE Systems Journal, 2016, 10(3):1147-1159.
[24] Xu W , Zhou H , Cheng N , et al. Internet of Vehicles in Big Data Era[J]. IEEE/CAA Journal of Automatica Sinica, 2018, 5(1):19-35.
[25] Paul A , Ahmad A , Rathore M M , et al. Smartbuddy: defining human behaviors using big data analytics in social internet of things[J]. IEEE Wireless Communications, 2016, 23(5):68-74.
[26] Chi M , Plaza A , Jón Atli Benediktsson, et al. Big Data for Remote Sensing: Challenges and Opportunities[J]. Proceedings of the IEEE, 2016, 104(11):2207-2219.
[27] George G , Osinga E C , Lavie D , et al. Big data and data science methods for management research[J]. Academy of Management Journal, 2016, 59(5):1493-1507.