Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 31 (2014) 532 – 541

Procedia Computer Science

Information Technology and Quantitative Management (ITQM 2014)

A New Idea of Study on the Influence Factors of Companies' Debt Costs in the Big Data Era

Li Lin[a], Wang Shuang[b], Liu Yifang[c]*, WangShouyang[d]

a School of Economics & Management, Beihang University, 37 College Road, Haidian District, Beijing 100191, China
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, 52 Sanlihe Road, Beijing 100864, China
b School of Statistics and Mathematics, Central University of Finance and Economics, 39 South College Road, Haidian District, Beijing, 100081, China
c School of Economics, Central University of Finance and Economics, 39 South College Road, Haidian District,Beijing 100081, China
d Academy of Mathematics and Systems Science, Chinese Academy of Sciences, 52 Sanlihe Road, Beijing, Beijing 100864, China

Abstract

Under the background of big data era today, once been widely used method – multiple linear regressions can not satisfy people's need to handle big data any more because of its bad characteristics such as multicollinearity, instability, subjectivity in model chosen etc. Contrary to MLR, LASSO method has many good natures. it is stable and can handle multicollinearity and successfully select the best model and do estimation in the same time. LASSO method is an effective improvement of multiple linear regressions. It is a natural change and innovation to introduce LASSO method into the accounting field and use it to deal with the debt costs problems. It helps us join the statistic field and accounting field together step by step. What's more, in order to proof the applicability of LASSO method in dealing with debt costs problems, we take 2301 companies' data from Shanghai and Shenzhen A-share market in 2012 as samples, and chose 18 indexes to verify that the results of LASSO method is scientific, reasonable and accurate. In the end, we compare LASSO method with traditional multiple linear regressions and ridge regression, finding out that LASSO method can not only offer the most accurate prediction but also simplify the model.

Keywords: debt cost, big data, quality of accounting information, corporate governance, LASSO method

* Corresponding author. Tel: +86+13810249357.
E-mail address: liuyifangcufe@163.com.

doi:10.1016/j.procs.2014.05.299

1. Introduction

Capital is fundamental to business survival and normal turnover. There are mainly two ways to raise money for companies, one is debt financing, the other goes to equity financing. According to pecking order theory, companies chose debt financing first, then goes to equity financing. At present, debt financing is the major source of financing for most companies. With debt financing getting more and more popular, how to reduce the costs of debt financing and thus obtain more profit have become most companies' concerning issues.

Nowadays, quantitative research on debt costs mainly use multiple linear regression (MLR) method. Although MLR method is easy to operate, easy to understand, easy to interpreting, and thus has been widely used, it has many defects when used to solve debt cost issues: (1) MLR has poor stability, and regression results are badly dependent on the sample data. We all know that financial data update very fast, if we apply MLR method to analyzing financial data, it is hard for us to get stable model parameter estimation results, thus the results have no meaning. (2) It is subjective when selecting variables into the model according to MLR method. Artificial setting error always exists, and some variables who can explain the model may not be included in the model. (3) MLR method can not satisfy model's strong explanation and high forecasting accuracy in the same time. (4) As we all know that financial index are highly multicollinearity, and MLR method actually can not solve this problem well.

With the development of Internet, the technology of data collection has been greatly improved. Today, collecting data is no more difficult, so how to effectively dig out useful information from data is taking more and more people's attention. However, the increase of the number of data can cause problems of multicollinearity inevitably, so we need a model or a method that can involved all the variables into it and solve multicollinearity problem in the same time.

In the field of statistics, the biased estimation method with variable selecting function can effectively solve these problems above. This method is an improvement of multiple linear regression method in big data background. Specifically, biased estimation method contents two kinds of models, one is called Ridge regression, and the other is called LASSO regression. Both of the two kinds of models improve prediction accuracy by imposing penalties on the coefficients. What's more, LASSO method is a further improvement of Ridge regression. It can overcome Ridge regression's defect that can not get sparse solution, and can select variables automatically thus exclude subjective interference. Under the background of big data, LASSO method has many good characteristics such as high forecasting accuracy, selecting significant variables automatically, computing fast etc. As a conclusion, it is advisable for us to apply LASSO method to the study of influence factors of debt financing costs.

2. Literature Review

Currently, both China scholars' and foreign scholars' research conclusions show that the quality of accounting information and corporate governance are two major factors of the impact of companies' debt financing costs.

In terms of the quality of accounting information, foreign scholars such as Sengupta(1998), Leua and Verrecchia(2004) believed that the quality of accounting information was negatively related to the debt financing costs. Richard Lambert, Christian Leuz, Robert E. Verrecchia(2006) found out that the quality of accounting information affects debt costs in two ways, one being high quality of information disclosure directly affecting companies' cash flow, the other being high quality of information disclosure firstly affecting companies' decision making then affecting companies' cash flow indirectly.

China scholars started late on the research of quality of accounting information, but made some progress too. Yu Fusheng and Zhang Min (2007) found that the higher the quality of accounting information disclosure to the public was, the lower the debt costs were. At the same time, the larger the market risk that companies were faced with was, the greater the degree of influence of the quality of accounting information to the debt costs was.

From the technical point, there are some deadly defects in Yu and Zhang's study. No.1, they do not choose debt cost itself as the dependent variable though the data is easy to get, which is not a good choice because using a qualitative indicator (company's credit rating) instead of a quantitative one (debt cost) can loss much of the

information. No.2, multiple linear regression models can only be used in the situation that the dependent variable is continuous. But in Yu and Zhang's paper, it is applied to the qualitative variable, which has violated the classical assumptions, making the model results meaningless. No.3, only 4 independent variables and 3 control variables are into the model. The variables are too less to make sure all related variables are included in the model.

When it comes to the research on corporate governance, foreign scholars Schleifer and Vishny (1997) pointed out that the direct effect company's governance brought about may be the reducing of debt financing costs, which showed good corporate governance could reduce financing costs effectively. China scholar Cui Wei (2008) found that both the proportion of the largest shareholder and board independence were negative related to company's debt costs. Besides, Yao Lijie, Luo Mei and Xia Donglin (2010) also did some meaningful work on debt costs issues, and finally drawn a conclusion the same as Schleifer and Vishny did, which was that good corporate governance could reduce financing costs effectively.

Though the goal of their studies was same, the selected independent variables were different. Cui selected independent variables from three aspects: shareholding structure, board of directors and management incentive. Yao, Luo and Xia thought about three aspects: shareholding structure, board of directors and board of supervisors based on Liu Liguo (2003)'s and Sun Zheng (2006)'s study results. However, the actual indicators on behalf of the same aspect are still different.

Based on the conclusion above of the literature review, we finally select 17 dependent variables from 3 aspects: quality of accounting information, corporate governance and control variables based on the principle of comprehensiveness, objectivity and easy to collect. In China, state-owned enterprises and non-stated-owned enterprises are different in nature. It is essential to distinguish the state-owned enterprises and non-state-owned ones when analyzing the influence factors of debt costs. In conclusion, all enterprises will be divided into two types under the property right, one is state-owned enterprises and the other is non-state-owned enterprises, after that we will apply LASSO method to optimize a model and find out the similarities and differences about the influence factors of debt costs between the two kind of enterprises. Last but not least, we will compare LASSO method with Ridge regression and MLR to proof that under the background of big data, LASSO method is the best one in variable selection, fitting, forecasting accuracy and many other aspects. Our conclusion will show that LASSO method is applicable, reliable and progress when be used to solve companies' debt financing costs influence factors issue under the background of big data.

3. LASSO method

Tibshirani R came up with a new variable selection method called lasso (least absolute shrinkage and selection operator) by the inspiration of ridge regression of Frank (1993) and nonnegative garrote of Bireman (1995). Lasso method uses the absolute value of the model coefficients as the punishment function, so that coefficients who are small enough will automatically be compressed to zero, and in this way both model selection and parameter estimation can be well done. LASSO method is stable and continuous in coefficient estimation, and can successfully overcome multicollinearity. It is good at variable selection, parameter estimation and many other aspects as well.

Let the coefficients of LASSO model be vector β , $\beta = (\beta_1, \beta_2, ..., \beta_p)^T$, and the corresponding loss function be $L(\beta)$, then we can write down the punishment likelihood function as follows:

$$L(\beta) + \sum_{i=1}^{p} p_{\lambda_i}(|\beta_i|) \tag{1}$$

Specifically, when $L(\beta) = (y - X^T\beta)^2$, $p_{\lambda_i}(|\beta_i|) = \lambda |\beta_i|^q$, expression (1) represents the punishment function of ridge regression if q=2, and represents the punishment function of LASSO method if q=1.

For linear model $y = X^T\beta + e$,

Where $y = (y_1, y_2, ..., y_n)^T, X = (x_1, x_2, ..., x_p), x_j = (x_{1j}, x_{2j}, ..., x_{nj})^T, j = 1, 2, ...p$

The specific form of LASSO model can be expressed as:

$$\hat{\beta} = \arg\min_{\beta}(y - X^T\beta)^2, s.t. \sum_{i=1}^{d} |\beta_i| \le t \tag{2}$$

Expression (2) is the same as

$$\hat{\beta} = \arg\min_{\beta} \{(y - X^T\beta)^2 + \lambda \sum_{i=1}^{d} |\beta_i|\} \tag{3}$$

Where t and λ have a one to one relationship.

When taking forecasting as a main purpose, PE (predicting error) is always used to measure the pros and cons of LASSO model. The smaller PE is, the higher LASSO model prediction accuracy is. Compared with MLR method, LASSO fitting results can improve the prediction accuracy greatly by the cost of losing part of information, which in fact leads to a better result.

In general, samples are randomly divided into train set and test set. Samples in train set are used to select model and estimate coefficients, and others in the test set are used to compute PE according to the coefficient estimators. Lastly, the best model will be picked out based on the PE value.

The prediction error can be expressed as

$$PE = E(y - \hat{y})^2 = E(y - E(y \mid x))^2 + E(E(y \mid x) - \hat{y})^2 \tag{4}$$

The first term in expression (4) represents the inherent error of the system, which can not be eliminated. The second one in expression (4) is the error caused by the model fitting, called model error, whose size reflects the pros and cons of different models.

4. Empirical Study

4.1. variable and data description

Variables and their descriptions are shown in Table 1.

All samples come from Shanghai and Shenzhen A share listed companies in China, 2012. We have excluded the followed samples: (1) All 39 financial listed companies. The borrowing behavior and financial statements of these companies are special. In order to reflect general rules, we exclude all financial listed companies. (2) 9 companies with missing data. (3) 111 companies who belongs to S, ST, *ST, S*ST, SST listed companies in 2012. Finally, we get 2013 samples, where 946 are state-owned enterprises and 1355 are non-state-owned enterprises. The data respectively come from Xenophon database, GTA database and Wind information, and some of the data are obtained by manual calculation.

Table 1 Variables and data description

	code	name	descriptions
dependent variable	DEBT	debt costs	finance costs / total debt
quality of accounting information	AUDIT	audit opinion	dummy variable, 1 for standard unqualified audit opinion, 0 for others
corporate governance	DIRNUM	size of board of directors	the number of board of directors
	IDNUM	independence of board of directors	the number of independent directors
	SUPNUM	size of board of supervisors	the number of board of supervisors
	MANAGE	management incentive	executive shareholding / total shares
	H	ownership concentration	square of the proportion of the largest shareholder
	ROE	ROE	net profit / (total assets – total debts)
control variables	FIXAST	fixed assets level	total net fixed assets / total assets
	ROA	ROA	net profit / total assets
	SIZE	scale of company	ln (total assets at the end of year)
	LEV	debt level	total debts / total assets
	CFO	own capital adequacy	net cash flow from operating activities / total assets
	EBIT	EBIT	net profit + financial costs + income tax expense
	LLOAN	long-term debts	long-term debts / total assets
	NTCOV	interest coverage	（net profit + financial costs + income tax expense）/finance costs
	CR	current ratio	current assets / current debts
	GROWTH	main business revenue growth	increase of the main business revenue / main business revenue at the beginning of the year
	CASH	capital requirements	dummy variable，if the net cash flow from operating activities plus the net cash flow form investing activities less than or equal to 0, CASH will be 1. otherwise CASH will be 0.

Actually we should involve more probably related variables into consideration to show that we are trying to solve a big data problem with LASSO, but we do not because the pages are limited and 17 variables*2301 samples can also tell something about big data.

4.2. Descriptive statistical analysis

Fig. 1 to Fig. 6 show that for state-owned enterprises, the debt cost is positively correlated with the fixed assets levels, own liquidity and long-term debt level and is negatively correlated with the scale of the company, ROE and the interest coverage ratio.

Positively relative:

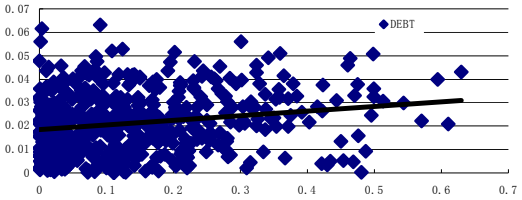

Fig. 1 debt against long-term debt

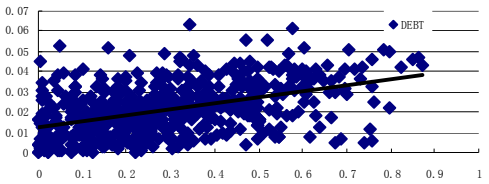

Fig. 2 debt against fixed assets level

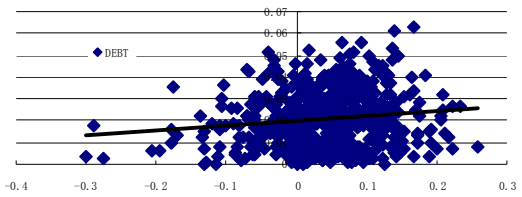

Fig. 3 debt against own liquidity

Negatively relative:

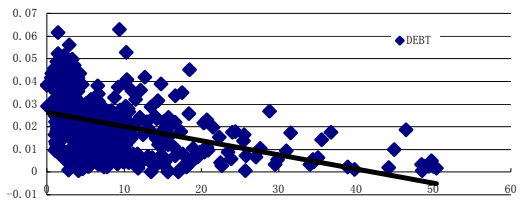

Fig.4 debt against interest coverage ratio

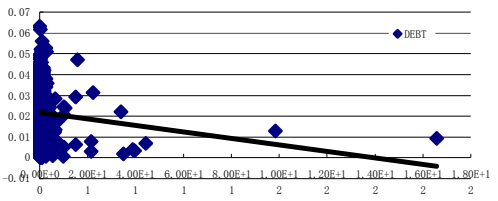

Fig. 5 debt against the scale of company

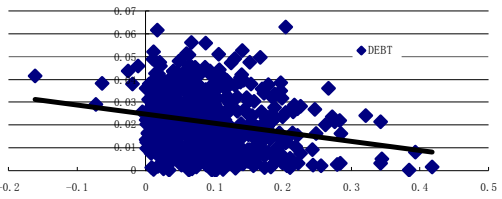

Fig. 6 debt against ROE

Fig. 7 to Fig.11 show that own liquidity and fixed assets level are positively related to the debt costs. ROE, the scale of company and interest coverage ratio are negatively related to the debt costs.

Positively relative:

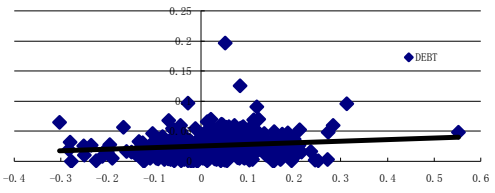

Fig. 7 debt against own liquidity

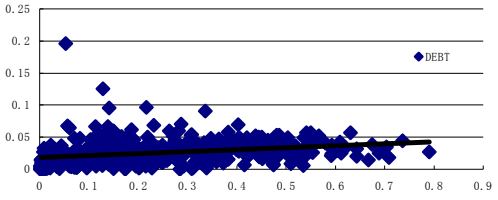

Fig. 8 debt against fixed assets level

Negatively relative:

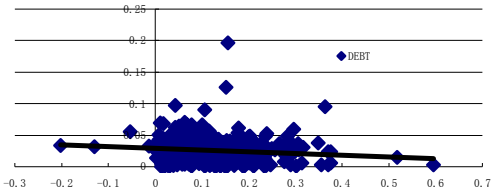

Fig. 9 debt against ROE

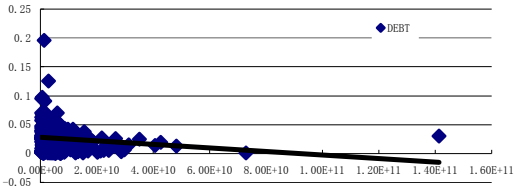

Fig. 10 debt against the scale of company

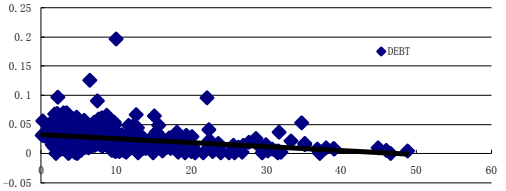

Fig. 11 debt against interest coverage ratio

4.3. Modeling and result analysis

First of all, draw 70% of the entire samples as the training samples using simple random sampling without replacement. The remaining 30% are as testing samples. Then optimize objective expression (3), and estimate the shrinking paths (Fig. 12 and Fig. 13) of both state-owned enterprises and non-state-owned enterprises. Last, chose the best model whose PE value is smallest. The estimating results of the best model are listed in Table 4.

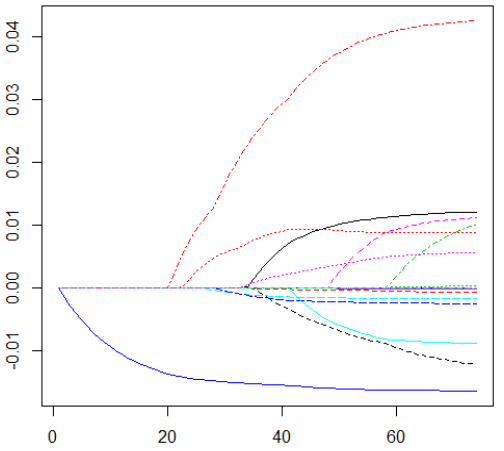

Fig. 12 LASSO coefficient shrinking paths of state-owned enterprises

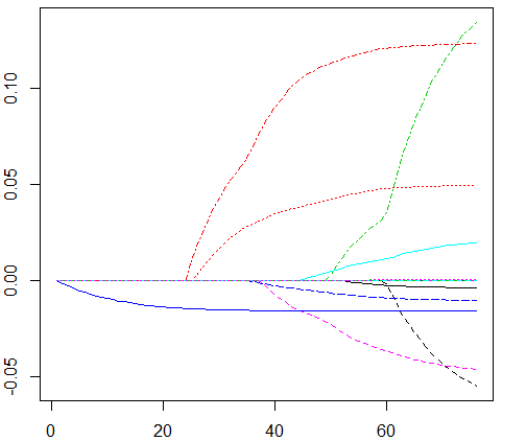

Fig. 13 LASSO coefficient shrinking paths of non-state-owned enterprises

Note: In Fig 3.1, the corresponding variables of the shrinking paths from top to bottom are: LLOAN, FIXAST, AUDIT, ROA, LEV, CASH, IDNUM, MANAGE, NTCOV, EBIT, SUPNUM, GROWTH, H, DIRNUM, SIZE, CFO, ROE, CR.
In Fig 3.2, the corresponding variables of the shrinking paths from top to bottom are: LLOAN, FIXAST, CASH, ROE, DIRNUM, IDNUM, NTCOV, EBIT, SUPNUM, GROWTH, MANAGE, H, AUDIT, SIZE, CFO, CR, LEV, ROA.

Table 2 final regression results of state and non-state owned enterprises

independent variables	code	regression coefficients	
		state	non-state
constant	CONSTANT	0.06149	0.18787
quality of accounting information	AUDIT	-0.00021	-0.00019
corporate governance	DIRNUM	-0.00003	——
	IDNUM	——	0.00018
	SUPNUM	——	——
	MANAGE	-0.00101	-0.00011
	H	-0.00002	-0.00018
	ROE	——	——
control variables	FIXAST	0.00723	0.04395
	ROA	——	0.01193
	SIZE	-0.00114	-0.00722
	LEV	——	0.00662
	CFO	——	-0.00271
	EBIT	——	——
	LLOAN	0.02268	0.11539
	NTCOV	——	——
	CR	-0.01518	-0.01587
	GROWTH	——	——
	CASH	0.00056	——
total variables into model (except constant)		9/18	11/18

Note: "——" represents 0.

Quality of accounting information. According to the regression results in Table 2, quality of accounting information has a significant impact on the debt costs both of state-owned enterprises and non-stated-owned enterprises. Regression coefficient is negative indicates that standard unqualified audit opinion can effectively reduce company's debt costs, which is consistent with normal experience. We notice that the absolute value of regression coefficient of state-owned enterprises is larger than it of non-state-owned enterprises, which indicates that the quality of accounting information has a greater degree of influence on state-owned enterprises compared with non-state-owned enterprises.

Corporate governance. For state-owned enterprises, the variables who have a significant impact on debt costs are size of the board of directors, management incentive and ownership concentration, respectively. Regression coefficients of the three variables are all negative, which indicate that a sound corporate governance structure, appropriate management shareholding proportion can bring lower debt costs for state-owned enterprises.

For non-state-owned enterprises, the variables who influence debt costs go to the independent of the board of directors, management incentive and ownership concentration. The regression coefficients of management

incentive and ownership concentration are both negative, which indicates that a sound corporate governance structure, reasonable management shareholding proportion also play an important role in reducing the debt costs of non-state-owned enterprises.

Control variables. Variables who affect both state-owned and non-state-owned enterprises are fixed assets level, the scale of company, long-term debt and current ratio.

For state-owned enterprises, in addition to these 4 variables, capital requirements also have an impact on debt costs. Regression coefficients of the scale of company and current ratio are negative and regression coefficients of the fixed assets level and long-term debts are positive.

For non-state-owned enterprises, ROA, debt levels and own capital adequacy also have an impact on debt costs besides the 4 common variables above. Among them, the scale of company, ROA, own capital adequacy and current ratio have negative regression coefficients and fixed assets level, long-term debts and debt level have positive regression coefficients.

4.4. Effectiveness analysis of LASSO method

We also compare the effectiveness between LASSO, Ridge regression and MLR. To ensure the scientific and reliability of the results, three methods are conducted 100 times respectively, and the average of prediction error, standard deviation of prediction error and the average number of variables into model are all calculated and shown in Table 3.

Table 3 results of effectiveness analysis

	state-owned enterprises			non-state-owned enterprises		
methods	average of PE	std of PE	average number of variables into model	average of PE	std of PE	average number of variables into model
MLR	0.097%	0.100%	17	23.107%	6.517%	17
Ridge	0.072%	0.040%	17	0.342%	0.100%	17
LASSO	0.059%	0.020%	9	0.329%	0.074%	11

Prediction error (PE). As we can see in Table 5, the PE value of LASSO method is just 0.08% lower than it of MLR for state-owned enterprises. However, the average prediction error obtained by MLR is far higher 22.778% than it obtained by LASSO method for non-state-owned enterprises, showing that LASSO method owns quite higher prediction accuracy than MLR.

Model stability. According to Table 5, the standard deviation of prediction error of LASSO method is the smallest. The standard deviation of prediction error of MLR is the largest. It indicates that the model stability of LASSO is better than it of MLR.

Variable selection. It is only LASSO method that successfully achieves variable selection function, who selected 9 and 11 variables from all 18 variables in the situation of keeping the smallest average prediction error and remaining the best stability.

In a conclusion, it is scientific and reasonable to introduce LASSO method to the study of debt cost issues as it has the best characteristics.

5. Conclusions

We have drawn 3 conclusions based on the research as follows:

(1) LASSO method is scientific and reliable. With big data, LASSO method has many good characteristics and is a kind of effective improvement for OLS method. This method has provided a now approach and new ideas of the study of debt costs, which changes the debt cost issue research to be more scientific and reasonable.

(2) Factors that affect state owned and non-state owned enterprises' debt costs are different, we need to be cautious when analyzing. On the one hand, factors influence state-owned and non-state-owned enterprises are different. On the other hand, the influence degree of the same factors on debt costs of difference types of enterprises is different. Therefore, it is necessary to distinguish state-owned and non-state-owned enterprises when analyzing debt cost issues in order to refraining reference to other companies blindly.

(3) It is necessary to pay enough attention on control variables as some of them have significant influences on debt costs. As we can see from the empirical results, the absolute values of the control variable regression coefficients are generally larger than the independent variable regression coefficients, which indicate that the degree of control variables affecting debt costs is relatively greater than that of dependent variables doing. So we need pay more attention to the control variables.

References

1. Chenglei Leng, Yi Lin and Grace Wahba, A Note on the LASSO and Related Procedures in Model Selection, Statistica Sinica, 16 (2006), 1273-1284.
2. Bradley Efron, Trevor Hastie, Iain Johnstone and Robert Tibshirani. Least angle regression, The Annals of Statistics, 32 (2004), 407-451.
3. Tibshirani Robert, Regression shrinkage and selection via the lasso, J.Royal. Statist. Soc. Ser. B, 58 (1996), 267-288.
4. Cao Fang and Zhu Yongzhong, Based on Multi-linearity LASSO Method, Journal of Jiangnan University (Natural Science Edition), 11 (1) (2012), 87-90.
5. Partha Sengupta, Corporate Disclosure Quality and the Cost of Debt, The Accounting Review, 73 (4) (1998), 459-474.
6. Ronald C. Anderson, Sattar A. Mansi and David M. Reeb, Board characteristics, accounting report integrity and the cost of debt, Journal of Accounting and Economics, 37 (2004), 315-342.
7. Valeri Nikolaev and Laurence van Lent, The endogeneity bias in the relation between cost-of-debt capital and corporate disclosure policy, European Accounting Review, 14 (4) (2005), 677-724.
8. Botosan, C.A. and Plumlee, M.A., A re-examination of disclosure quality and the expected cost of capital, Journal of Accounting Research, 40 (1-3) (2002), 21-40.
9. Diamond, D. and Verrecchia, R., Disclosure, liquidity and the cost of capital, Journal of Finance, 46 (1991), 1325-1360.
10. Chen Xiaoshuang, Information Disclosure Quality and the Impact on Debt Financing: Evidence from Shenzhen Stock Exchange A Share Market, Zhongshan University Master Thesis, (2008).
11. Huyiming and Tang Songlian, Audit, Information Transparency and Bank Loan Interest Rate, Auditing Research, 6 (2007), 74-85.
12. Yu Fusheng and Zhang Min, Corporate Disclosure Quality and Debt Cost, Audit & Economy Research, 22 (5) (2007), 93-96.
13. Yao Lijie, Luo Mei and Xia Donglin, Corporate Governance and Bank Loan Financing, Accounting Research, 8 (2010), 55-96.
14. Liu Peipei, Empirical Research on Corporate Governance and Debt Costs -- Based on Small Listed Companies' Data, Business China, 11 (2010), 9-12.
15. Tan Yue and Ma Xiaoxia, Linkage between Size of Cooperation and Debt Financing on the Background of China's Finance Deepening, Communication of Finance and Accounting, 1 (2011), 142-144.
16. Cui Wei, Corporate Governance and Cost of Debt – Empirical Evidence from Listed Firms in Shenzhen Stock Exchange, Communication of Finance and Accounting, 1 (2008), 73-76.
17. Wu Xiafeng, Correlation Analysis on Corporate Governance, Audit Opinion and Debt Costs, Industrial Technology and Economy, 28 (11) (2009), 151-154.
18. Liu Yunguo, Wu Xiaomeng and Jiangtao, Ownership, Debt Financing and Accounting Conservatism: Empirical Evidence from Chinese Listed Companies, Accounting Research, 1 (2010), 43-50.
19. Yu Fusheng, Zhang Min, Jiang Fuxiu and Ren Mengjie, Does Corporate Governance Influence Firm's Financing Risk? Accounting Research, 10 (2008), 52-59.