

基于 LASSO 的企业财务危机预警与关键指标选择

张艳丽^{1a}, 尤晓琳², 强 薇³, 郑雨晗^{1b}, 李钧涛^{1b}

(1. 河南师范大学 a. 财务处; b. 数学与信息科学学院, 河南 新乡 453007;

2. 鹤壁职业技术学院, 河南 鹤壁 458030; 3. 财政部财政科学研究所, 北京 100142)

摘 要:将 LASSO 方法应用于财务危机预警的关键指标选择,并借助 90 家制造业上市公司的财务数据建立了基于 LASSO 方法的财务预警模型.提出的模型不仅实现了两个关键财务指标的自动选择,而且获得了较高的预测精度,这为企业纠正经营方向、改进经营决策和有效配置资源提供了可靠的依据.

关键词:变量选择;LASSO;财务预警;解路;正则化

中图分类号:TP181

文献标志码:A

复杂的经济环境下,企业常常会因为经营管理不善,或是无法适应外部环境变化而致使其生产经营活动陷入一种严重困境,这便是财务危机^[1],也称财务困境.企业财务危机的产生实际上是一个渐进的积累过程,若能在“潜伏期”内对其进行有效地监测、预警和防范,无论对于企业自身的生存发展,还是对于整个市场经济活动的有序进行都具有重要的意义.

20 世纪 30 年代以来,为了更好地监测和预报财务危机,国内外学者借助企业的相关会计资料,提取不同的财务指标,利用各种数学工具和方法,针对财务预警模型展开了大量的研究.1932 年,菲茨·帕特里克(Fitz Patrick)围绕单个财务比率进行研究,建立了最早的单变量财务预警模型^[2].1966 年,威廉·比弗(William Beaver)^[3]沿用这样的思路对单变量预警模型进行了更进一步地完善.然而,为了克服单变量模型单一、不稳定等缺陷,20 世纪 60 年代起,许多学者开始研究多变量财务预警模型,其中,美国学者爱德华·奥特曼(Edward I. Altman)提出的 Z 计分模型(1968)^[4]和 ZETA 模型(1977)^[5],就是其中的代表.随着科学技术的发展和探究的不断深入,多变量预警模型也因使用计量方法的不同,逐步细分为多元线性回归模型、多元非线性回归模型(Logistic 模型)^[6]以及人工神经网络模型^[7]等.而在国内,吴世农与黄世忠(1986)^[8]等人开始撰文介绍财务危机预警,到 2000 年,张玲^[9]首次利用上市公司是否被 ST 作为财务危机发生标志进行研究.

在多变量预警模型的研究过程中,模型估值和预测的精准度主要取决于指标变量的选择和回归参数的确定,这也是构建财务预警模型的关键步骤.回归模型的变量选择问题,主要是对最小二乘估计(OLS)进行改进,例如,子集选择法^[10]和岭回归^[11]都能达到一定的改进效果,但又总会存在不稳定、可解释性不强等缺陷.在此基础上,Robert Tibshirani(1996)则提出了 LASSO(Least absolute shrinkage and selection operator)^[12]方法,在融合两者优势的同时很好地克服了以上缺点,这也引起了统计领域的广泛关注.随后,不少学者针对 LASSO 理论进行了改进.例如,Fan 和 Liu(1999)^[13]提出了缓解 LASSO 压缩偏差的 SCAD 惩罚模型,Bradley Efron(2004)^[14]等人提出的最小角回归算法(Lars),则最终解决了 LASSO 的计算问题,进一步促进了 LASSO 的流行,Zou 和 Hastie(2005)^[15]加入了能够自动变量分组的弹性网络,李钧涛等人(2013)^[16]提出了具有“自适应群体基因选择”性能的部分自适应弹性网络等.

收稿日期:2015-11-30;修回日期:2016-03-23.

基金项目:国家自然科学基金(61203293;61374079);河南省高校科技创新人才支持计划(13HASTIT040);河南省高校青年骨干教师资助计划(2012GGJS-063);河南师范大学骨干教师资助计划.

作者简介(通信作者):李钧涛(1978-),男,河南南阳人,河南师范大学副教授,博士,研究方向为统计机器学习及其应用, E-mail:juntaol@mail@126.com.

本文将LASSO及其衍生的相关算法应用于财务危机预警的变量选择和模型建立进行研究,并借助R软件实现相关财务危机预警模型的构建和检验。

1 LASSO方法

作为一种变量选择方法,LASSO是通过将模型系数的绝对值函数进行复杂度惩罚,最终实现结构风险最小化,防止“过拟合”现象发生。其实质就是借助L1范数作为惩罚求出以下方程的最优解,即实现拟合函数和惩罚函数的最小和:

$$\arg \min \left\{ \sum_{i=1}^n \left(y_i - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}. \quad (1)$$

$$\text{s. t. } \sum_{j=1}^p |\beta_j| \leq t, s \in [0, \infty), \quad (2)$$

其中, y_i 为输出变量; x_{ij} 为预测变量; $\lambda \sum_{j=1}^p |\beta_j|$ 为惩罚函数; $i = 1, 2, 3, \dots, n; j = 1, 2, 3, \dots, p$.

在回归模型中引入LASSO算法,通过对调参数 t 进行控制,可以使回归系数总体变小,即实现回归系数的自动压缩;当 t 取值适当时,该运算能够使一些相关度较低的系数缩小为0,从而轻松得到稀疏解,实现有效的变量选择。

参数 t 的估计通常用Efron和Tibshirani(1993)^[17]提出的交叉验证法。定义 $\hat{\beta}_j$ 为回归模型的最小二乘估计, $s = t / \sum \hat{\beta}_j$ 为正则化参数,则 s 在闭区间 $[0, 1]$ 上取值来估计预测误差:

$$CV(s) = \sum \left(y_i - \sum \beta_j(s) x_{ij} \right)^2. \quad (3)$$

当选定的 s 使该误差达到最小值,即可确定回归参数。

根据参考文献[14]的描述,最小角回归算法(Lars)的过程是一个残差拟合的过程,其原理类似于向前回归,即先选择最相关的变量,进行最小二乘回归,再按照相关系数的绝对值大小顺序依次进行评估和拟合。但相比于向前回归法,Lars在运作过程中,对变量间相关系数的评估更加精细,前进步骤更加严谨。具体算法详见参考文献^[15]。

2 基于LASSO方法的财务风险预警模型

2.1 研究设计

1998年,中国证监会宣布我国证券市场开始对财务状况异常的上市公司股票施行“特别处理”(Special Treatment,简称“ST”),将上市公司分为ST企业和非ST企业(即正常企业),随后沪深交易所将连续两年亏损的企业纳入异常情况,至此ST开始被视为评价一个企业财务状况的一个重要标准。本文选取若干企业的财务数据作为研究样本,将ST企业定义为出现财务危机的企业,将非ST企业定义为暂无财务危机的企业。针对企业财务状况的偿债能力、营运能力、盈利能力、发展能力和现金流量各个方面进行财务原始指标变量的选取,在对多个指标进行处理的过程中,采用LASSO的方法对进行变量的分析和选择,确定影响企业财务危机发生的关键性指标,构建财务预警模型,并对模型的准确性进行检验。

2.2 数据说明

考虑到样本数据的易得性、可比性和代表性,本文以制造业为例,共选取了90家上市公司公开披露的财务数据作为数据样本,其中,45家公司在2011—2013年期间首次被标记ST(即ST公司),45家公司在此期间未被标记ST(即非ST公司)。由于ST公司通常指企业连续两年亏损,所以,本文选取了首次被ST判别(即 T 年)前连续3年的财务数据,拟分别建立 $T-1$ 年, $T-2$ 年, $T-3$ 年的财务预警模型。

作为本文研究样本的90家制造业上市公司的所有财务数据均来自财经类专业网站和讯网(www.hexun.com)。

2.3 财务风险预警指标体系

2.3.1 原始指标体系

财务指标的选择应尽可能的满足全面、普遍、可比等原则。本文根据财务管理的相关理论,在前人研究成果的基础上从以下5个方面进行了原始指标的选取:

- 1) 偿债能力指标:通过偿债能力可以反映企业资金的流动性、负债水平,有利于评价企业的财务状况和财务风险;
- 2) 营运能力指标:营运能力反映的是企业对资产的利用和管理情况,能够帮助评价企业资产的保值和增值;
- 3) 盈利能力指标:利润的获取是企业的主要经营目标,也是企业综合能力的体现,对于盈利能力的评价能在一定程度上反映出企业的财务运营状况;
- 4) 发展能力指标:发展能力反映企业未来的经营,因此是预测企业潜力和发展前景的重要依据;
- 5) 现金流量指标:现金流量是从动态中反映企业在一定会计期间现金和现金等价物的流动情况,通常认为这类指标对于预测企业运转时分有效。

基于以上考虑,本文围绕偿债能力、营运能力、盈利能力、发展能力和现金流量5个方面选取了11项财务指标并绘制成表1,作为对制造业上市公司财务预警模型的初始自变量,这些指标足以全面、有效地反映企业的财务状况。

表1 财务风险预警模型指标体系

	编号	财务指标	定义
偿债能力指标	X ₁	速动比率	速动资产/流动负债
	X ₂	资产负债率	(负债总额/资产总额)×100%
	X ₃	应收账款周转率	赊销收入净额/应收账款平均余额
营运能力指标	X ₄	存货周转率	销售成本/存货平均余额
	X ₅	总资产周转率	销售收入/资产平均总额
盈利能力指标	X ₆	主营收入税后利润率	(净利润/资产平均总额)×100%
	X ₇	总资产收益率	(净利润/销售收入)×100%
发展能力指标	X ₈	主营业务增长率	(本期主营业务收入增长额/上期主营业务收入总额)×100%
	X ₉	总资产增长率	(本年总资产增长额/年初资产总额)×100%
现金流量指标	X ₁₀	每股经营现金流量	经营活动现金流量净额/流通在外的普通股股数
	X ₁₁	现金流量结构比率	经营活动产生的现金净流量/期末流动负债

2.3.2 基于 LASSO 的关键指标选择

借助 R 软件的 lars 程序包,本文利用收集所得的数据进行训练,并估值求解,得到基于 LASSO 方法的回归系数解路径图(详细数据见表2)。

表2 基于 LASSO 的财务风险预警回归系数压缩数据表

V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
0.000 00	0.000 00	0.000 00	0.000 00	0.000 00	0.000 00	0.000 00	0.000 00	0.000 00	0.000 00	0.000 00
0.000 00	0.001 04	0.000 00	0.000 00	0.000 00	0.000 00	0.000 00	0.000 00	0.000 00	0.000 00	0.000 00
0.000 00	0.479 34	0.000 00	0.000 00	0.000 00	0.000 00	-2.013 16	0.000 00	0.000 00	0.000 00	0.000 00
0.000 00	0.484 79	0.000 00	0.000 00	0.000 00	0.000 00	-2.024 30	-0.002 72	0.000 00	0.000 00	0.000 00
0.000 00	0.547 42	0.000 00	0.000 00	-0.018 79	0.000 00	-2.090 80	-0.023 44	0.000 00	0.000 00	0.000 00
-0.006 30	0.628 70	0.000 00	0.000 00	-0.083 22	0.000 00	-2.260 00	-0.074 62	0.000 00	0.000 00	0.000 00
-0.007 44	0.673 34	0.000 00	0.000 00	-0.119 63	-0.123 78	-2.095 61	-0.085 43	0.000 00	0.000 00	0.000 00
-0.009 08	0.705 34	0.000 00	0.000 00	-0.173 97	-0.249 69	-2.072 83	-0.092 63	0.000 00	0.033 64	0.000 00
-0.009 60	0.716 42	0.000 00	0.000 00	-0.184 73	-0.292 72	-2.022 00	-0.100 33	0.011 29	0.037 77	0.000 00
-0.010 23	0.729 11	0.000 06	0.000 00	-0.201 52	-0.348 02	-1.945 94	-0.108 14	0.025 03	0.042 86	0.000 00
-0.012 50	0.783 67	0.000 27	0.000 00	-0.270 93	-0.569 53	-1.622 27	-0.132 72	0.076 39	0.061 05	-0.000 11
-0.014 28	0.821 35	0.000 45	0.001 67	-0.337 92	-0.770 68	-1.311 15	-0.146 87	0.116 83	0.075 28	-0.000 22

从路径图1中可以看出,随着回归系数估值的不断增大,相关指标的系数呈现了不同程度的离散;相反,

当回归系数估值逐渐减小,部分变量的标转化系数被压缩为 0.当变量被压缩为 5 个时,模型的剩余指标分别为 X_1, X_2, X_5, X_7 和 X_8 ,当变量被压缩为 2 个时,模型的剩余指标为 X_2 和 X_7 ,由此可以直观的判断基于 LASSO 财务困境模型挑选出的关键指标.

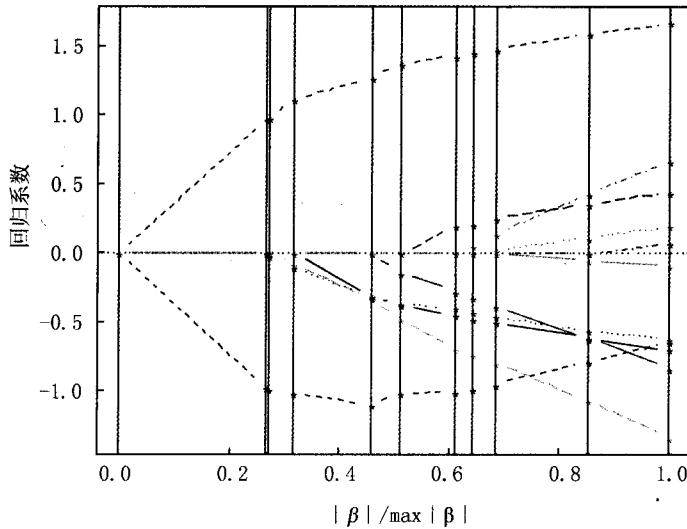


图1 基于LASSO的财务风险预警回归系数解路径图

另外,上文已提到,LASSO 的求解过程实际上是通过调节系数的控制,实现回归系数的自动压缩,按照变量的相关度大小的顺序逐步将系数压缩为 0,表 2 即反映了这一过程.因此,也可根据 LASSO 算法的变量移动顺序判断关键指标的重要性.

表 3 即为基于 T-1, T-2, T-3 3 组样本的 LASSO 算法移动顺序.

表 3 LASSO 变量运行顺序表

模型编号	LASSO 移动顺序(变量编号)
T-1	2→7→9→5→3→11→6→4→1→10→8
T-2	7→2→5→9→10→4→8→1→6→11→3
T-3	2→7→8→5→1→6→10→9→3→11→4

根据表 3 可以明显看出,3 次拟合过程中,变量 X_2 和 X_7 始终优先被挑选,这也意味这两个变量所代表的指标就是 LASSO 挑选出的财务预警最关键指标,其中, X_2 代表资产负债率, X_7 代表总资产收益率.结合图 1,资产负债率越大,财务风险越大;总资产收益率越小,财务风险越大,这完全符合财务管理理论中的实际情况.

2.3.3 财务预警模型的构建

基于上述 LASSO 方法,首先对样本输出变量 y 进行二分法处理,设定第 T 年时 ST 公司的输出值为 1;非 ST 公司输出值为 0;利用样本集中的数据进行 10-折交叉验证,选取均方误差最小时的 λ 值,作为惩罚因子,并将此时选出的指标变量和输出变量 y 作为建模变量代入模型进行拟合. T-3 年数据为例,利用 10-折交叉检验法进行调节参数 λ 的估计,得到如下 CV 曲线(图 2).

取交叉检验均方误差最小时的 λ 值作为惩罚参数.经过多次运行并观测,均出现图 2 类似的检验结果,根据 BIC 准则由此可以判断 λ 值基本稳定在 0.3 附近.

对相关参数进行估计后,即可构建财务预警模型.为保证实验结果的稳定性,本文对 90 家制造业上市公司样本进行 5 次随机分组,每次均选取 76 家企业(包括 38 家 ST 企业和 38 家非 ST 企业)作为训练样本,剩余的 14 家企业(包括 7 家 ST 企业和 7 家非 ST 企业)作为测试样本,分别建立基于 LASSO 的 T-1, T-2, T-3 回归模型,如下:

T-1 回归模型:

$$y_i = 0.2015674X_{2,T-1} - 0.1015907X_{7,T-1}.$$

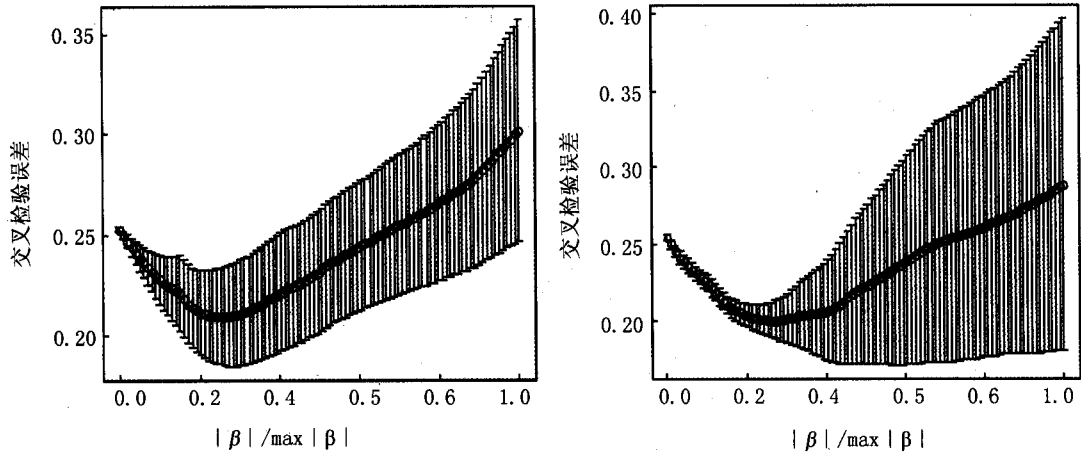


图2 不同惩罚参数下的10-折交叉检验估计

T-2 回归模型:

$$y_i = 0.020\ 578\ 17 X_{2,i,T-2} - 0.808\ 697\ 55 X_{7,i,T-2}.$$

T-3 回归模型:

$$y_i = 0.074\ 156\ 4 X_{2,i,T-3} - 0.307\ 753\ 4 X_{7,i,T-3}$$

2.3.4 财务预警模型的检验结果

利用测试集分别对 5 个随机分组产生的财务预警模型进行预测准确性检验. 测试结果如表 4 所示.

表 4 测试结果

年度	模型	准确率			
		ST 误判情况	非 ST 误判情况	平均准确率/%	综合准确率/%
T-1	模型一	1/7	2/7	78.57	88.57
	模型二	1/7	0/7	92.86	
	模型三	0/7	0/7	100.00	
	模型四	0/7	2/7	85.71	
	模型五	2/7	0/7	85.71	
T-2	模型一	1/7	1/7	85.71	92.86
	模型二	0/7	0/7	100.00	
	模型三	0/7	0/7	100.00	
	模型四	0/7	2/7	85.71	
	模型五	0/7	1/7	92.86	
T-3	模型一	2/7	1/7	78.57	78.57
	模型二	2/7	1/7	78.57	
	模型三	2/7	1/7	78.57	
	模型四	1/7	1/7	85.71	
	模型五	1/7	3/7	71.43	

从测试结果来看, 基于 LASSO 的财务预警模型对于样本检验的准确率均超过了 70%, 将 5 次随机分组建立的 T-1, T-2, T-3 年共 15 个预警模型的综合准确率进行平均, 得到本文基于 LASSO 的财务预警模型预测的平均准确率为 86.66%. 其中 T-1 年, T-2 年的模型预测准确率普遍较高, T-3 年的模型预测准确率相对稍低, 这也基本符合以往的模型测试经验.

3 结 论

本文针对财务危机预警中财务指标选择的关键问题, 利用我国制造业上市公司的财务数据, 建立了基于 LASSO 的财务预警模型. 借助 LASSO 方法不仅实现了财务指标体系的精简, 而且基于该方法建立的财务预警模型能够有效地进行财务危机预警, 且在可操作性和计算量方面比以往其他模型更有优势.

从实验的结果来看,虽然LASSO从5个财务方面共11个财务指标中只挑选出了两个关键指标—— X_2 (资产负债率)和 X_7 (总资产收益率),但基于此构建的模型的预测精准率已达到理想效果,基本实现了以往学者利用其他模型进行财务预警的准确度,说明基于LASSO的财务预警模型关键指标选择是有效而精准的,具有一定的实际应用价值。

参 考 文 献

- [1] 荆 新,王化成,刘俊彦. 财务管理学[M]. 北京:中国人民大学出版社,2012.
- [2] Fitz Patrick P J. A Comparison of the Ratios of Successful Industrial Enterprises with Those of Failed companies[J]. *Analyse Molecular Do Gene Wwox*,1932(3):598-605.
- [3] Beaver W H. Financial Ratios as Predictors of Failure[J]. *Journal of Accounting Research*,1966(4):71-102.
- [4] Altman E I. Financial Ratios, Discriminate Analysis and the Prediction of Corporate Bankruptcy[J]. *Journal of Finance*,1968(23):589-609.
- [5] Altman E I, Haldeman R, Narayanan P. Zeta analysis: A new model to identify bankruptcy risk of corporations[J]. *Journal of Banking and Finance*,1977,18(1):29-54.
- [6] Ohlson J A. Financial Ratios and the Probabilistic Prediction of Bankruptcy[J]. *Journal of Accounting Research*,1981,18(1):109-131.
- [7] 王冰灿. 基于人工神经网络模型的寿险公司财务风险预警研究[D]. 杭州:浙江财经大学,2014.
- [8] 吴世农,黄世忠. 企业破产的分析指标和预测模型[J]. *中国经济问题*,1986(6):18-20.
- [9] 张 玲. 财务危机预警分析判别模型[J]. *数量经济技术经济研究*,2000(3):51-53.
- [10] John G H, Kohavi R, Peleger K. Irrelevant features and the subset selection problem[C]. *Proc of the 11th International Conference on Machine Learning*, New Brunswick,1994.
- [11] Hastie T, Taylor J, Tibshirani R. Forward stagewise regression and the monotone Lasso[J]. *Electronic Journal of Statistic*,2007(1):1-29.
- [12] Tibshirani R J. Regression Shrinkage and Selection via the Lasso[J]. *Journal of the Royal Statistical Society, Series B*,1996,58(1):267-288.
- [13] Bunea F, Tsybakov A, Wegkamp M, et al. Aggregation for Gaussian Regression[J]. *Annals of Statistics*, 2007,35(4):1674-1697.
- [14] Efron B. Least Angle Regression[J]. *The Annals of Statistics*,2004,32(2):408-414.
- [15] Zou H, Hastie T. Regularization and variable selection via the elastic net[J]. *Journal of the Royal Statistical Society*,2005,67(2):301-320.
- [16] Li J, Jia Y, Zhao Z. Partly Adaptive Elastic Net and Its Application to Microarray Classification[J]. *Neural Computing and Applications*,2013(22):1193-1200.
- [17] Efron B, Tibshirani R. An Introduction to the Bootstrap[J]. *Journal of the American Statistical Association*,1993,89(428):49-54.

LASSO-based Early Warning of Enterprise's Financial Crisis and the Selection of Key Indicators

ZHANG Yanli^{1a}, YOU Xiaolin², QIANG Wei³, ZHENG Yuhan^{1b}, LI Juntao^{1b}

(1. a. The Finance Department; b. School of Mathematics and Information Science, Henan Normal University, Xixiang 453007, China; 2. Hebi Vocational and Technical College, Hebi 458030, China; 3. Research Institute for Fiscal Science, Ministry of Finance, Beijing 100142, China)

Abstract: This article applies lasso method to the key indicators choice of financial crisis warning, and establishes a financial warning model based on LASSO method with the financial data of 90 manufacturing listed companies. The proposed model not only achieves the automatic selection of two key financial indicators, but also obtains the higher prediction accuracy, which provides the reliable foundation for the enterprise to correct the management direction, improve the management decision and allocate resources efficiently.

Keywords: variable selection; LASSO; financial warning; solution path; regularization