基于LightGBM分类的上市公司财务报表舞弊识别研究

Research on Fraud Identification of Financial Statements of Listed Companies Based on LightGBM Classification

学生姓名：王淙烨

指导教师：孙景翠

所在院系：经济管理学院

所学专业：会计学

研究方向：会计学

东 北 农 业 大 学

中国·哈尔滨

2020 年 5 月

摘要

会计舞弊对于会计信息使用者、企业管理者、市场监督者有着十分重大的威胁。会计舞弊识别是保证会计信息质量、健全我国宏观体系的重要手段之一。本文使用机器学习模型构建，研究上市公司财务报表舞弊识别的方法。

首先，本文分析舞弊研究现状、动因、2010 年至 2018 年整体舞弊情况与分行业舞弊情况，其次，本文选取 2010 年至 2018 年沪深两市全部 A 股上市公司，并能够精确计算 163 个财务指标的公司交集作为样本，以公司年度审计报告中审计意见作为分类标准，使用 Wilcoxon 秩和检验、MANOVA 多元方差分析、Pearson 相关性分析、因子分析等手段剔除、构建新加权的 24 个指标；然后，使用上述指标构成数据训练 LightGBM 分类模型；最后，评估整体模型与指标。

结果显示，上市公司财务舞弊具有较强行业聚集性；加权综合构成每股指标与现金流分析指标对判断财务舞弊具有较强先导性；模型具有较好识别效果。

关键词： 财务分析　会计舞弊　分类识别　机器学习　LightGBM 模型

Research on Fraud Identification of Financial Statements of Listed Companies Based on LightGBM Classification

Abstract

Accounting fraud poses a significant threat to accounting information users, business managers, and market monitors. Accounting fraud identification is one of the important means to ensure the quality of accounting information and improve Chinese macro system. Based on a machine learning model, this paper constructs a method for identifying fraud in financial statements of listed companies.

First of all, this article analyzes the status quo and causes of fraud research, overall fraud and sub-industry fraud from 2010 to 2018. Secondly, this article covers all A-share listed companies in Shanghai and Shenzhen from 2010 to 2018, and it can accurately calculate 163 financial statements. The company intersection of the indicators is used as a sample, and the audit opinions in the company's annual audit report are used as the classification criteria. Wilcoxon rank-sum test, MANOVA multivariate analysis of variance, Pearson correlation analysis, and factor analysis are used to remove and construct the newly weighted 24 indicators. The LightGBM classification model was trained by the dataset. Finally, the overall model and indicators are evaluated.

The results show that listed companies' financial fraud has strong industry agglomeration, the weighted comprehensive composition of the indicators per share and cash flow analysis indicators have a strong lead in judging financial fraud, the model has a good recognition effect.

Key words: Financial Irregularity　Accounting Fraud　Classification Recognition　Machine Learning　LightGBM Model

目录

摘要 I
Abstract II
1 前言 1
1.1 本研究的目的与意义 1
1.2 国内外研究文献综述 2
1.3 本研究的主要内容 4
2 相关概念综述 4
2.1 财务舞弊概念与表现方式 4
2.2 LightGBM 算法 5
3 样本选取与分析 7
3.1 样本选取 7
3.2 舞弊分析 8
3.3 样本确定 8
4 财务舞弊识别分类构建 10
4.1 特征变量的选取 10
4.2 分类器建立 17
4.3 结果及分析 27
5 结论 32
参考文献 35
致谢 36
附录 A 37

1 前言

1.1 本研究的目的与意义

1.1.1 本研究的目的

财务报告是指企业对过去期间发生的经营活动，经营状况，和特定时点财务成果的高度概括，与各项经济业务事项产生后形成的对外报告。财务报告为投资者、债权人等会计信息使用者，企业内部高级管理人员、会计等利益相关者，审计署、银保监会、金融监督管理局等监管部门提供所需的必要信息，而财务报告的准确性关乎国家健全宏观调控体系的根本。但自 1995 年以来，随着资本市场的不断创新与发展，上市公司管理层与治理层往往会因其信息不对称等优势，影响外部投资人进行不恰当的决策，而财务报表的粉饰与舞弊成为了其牟取不当利益的主要手段。在扩大经营规模，资本增资等利益驱使、大股东借助家族信托公司向海外转移公司资产、公司管理层与治理层操纵股价以及为掩盖公司经营困境及避免退市和特别处理等目的的驱使下，通过调整会计政策、未对重大事项进行披露、进行重大且广泛地关联方交易等问题频发。

2020 年 4 月 3 日，瑞幸咖啡公司自曝 COO 伪造销售额高达 22 亿元人民币。此公司从 2017 年 10 月瑞幸开设中国第一家门店，到 2019 年 5 月在美国纳斯达克交易所上市，且其市值超过 82 亿仅用时 18 个月。作为对标星巴克咖啡的商业新星，瑞幸在其 Pre-IPO 时便获得 1.5 亿美元融资，其中不乏世界知名投资公司为其背书，包括黑石头投资、新加坡 GIC、贝莱德等。但过快增长也给投资人带来了疑惑，2020 年 1 月 30 日浑水调研公司收到指控瑞幸咖啡财务造假 89 页的匿名做空报告，报告称其调动 92 名全职和 1418 名兼职人员进行实地监控，记录了 981 个工作日门店流量，覆盖 620 家门店 100%营业时间。该报告提供 5 个确凿证据与 6 个危险信号。其中，瑞幸夸大了其每件商品净售价至少 12.3%，门店层面的亏损高达 24.7%-28%，排除免费产品，实际的销售价格是上市价格的 46%，而不是管理层声称的 55%；其夸大了其在 2019 年第三季度的广告费用 150%以上，有可能将其夸大的广告费回收，以增加收入和门店层级的利润；其在 2019 年第三季度来自“其他产品”的收入贡献仅为 6%左右，近 400%的膨胀率。

上述近期发生事件能够体现，企业粉饰报表方法层出不穷，这不仅对财务报告的使用者和潜在使用者用重大而广泛的影响，还对整个经济市场造成了冲击，进而影响社会资源的合理分配与国家的宏观经济调控。由此可见，有效甄别上市公司财务舞弊也成为了亟待解决的实际问题之一。

1.1.2 本研究的意义

企业财务舞弊主要具体包括以下三点影响：首先，误导投资者。投资者在进行投资前，必然选择了解上市公司的财务状况与发展前景，而大多数投资者，尤其是距企业较远的中小投资者，无法进行实地考察，在一定程度上，极度依赖于财务报告来估计与验

证其投资效果。但会计舞弊必然导致虚假会计信息，而大多数没有经验的投资者无法凭借一般经验判断企业会计信息虚假程度，进而无法获知企业真实经营情况，这将重大而广泛的导致其决策与判断，从而使其蒙受重大的经济损失。其次损害债权人利益。企业为获取充足资金用以发展与生产，会通过债券市场发行债券，向银行或其他金融机构进行贷款，和赊账付款等方式向其他企业购买相应产品与服务。企业的会计舞弊与信息虚假会切实对相关资质审核与款项权益产生实质上的冲击。最后，降低市场出清效率。在出清的市场中不应该有超额需求，但会计舞弊可能导致企业虚增资源，控制产品利润，进而使得市场资源配置功能失灵，无法使社会总效率达到最大值。

综上所述，加强会计监管与会计信息识别判断是极其重要的。对投资者与债权人而言，如果能提供一种相对有效的识别模型，将虚假报告信息对投资者影响降到最低限度，从大规模的财务数据分析中抽离出来，则能最大程度对其知情权提供重要保障；对监管机构来讲，构建一个识别财务舞弊的模型，能够及时发现与披露出具财务异常报告的上市公司，避免因其固有的滞后性导致市场竞争公平性存在漏洞，切实保护投资者与债权人权益；对审计从业者来说，通过与现行发展的大数据与机器学习模型，互相印证、理论创新，能够辅助识别虚假会计信息，提高其审计效率、准确性，降低误判风险，开源节流提供一定帮助与参考。

近几年，大数据与人工智能在不断完善与发展，随着其在交叉学科的试验与应用，越来越多的工业界从业人员开发出其在传统理论框架之外的全新应用。希望在应用实际理论的同时，更多考虑其在低等重复性工作上的辅助性与替代性，真正帮助财务工作者避免因密集重复性劳动而造成的沉没成本。

1.2 国内外研究文献综述

1.2.1 国外研究文献综述

Tommie W Singleton（2006）提出舞弊冰山理论，又称二因素理论。该理论冰山上端部分主要包括：企业组织结构、控制方法、公司政治、人员管理和制度、财务状况等，是其发生的表象；与此同时，其海平面下端是真正的危机，属于舞弊行为方面，主要包括：实施主体的价值取向、道德水准、受教育程度、周围环境影响等，是其发生的本质。冰山下端部分相比于上端更具主观能动性。该理论显示，对于企业而言，相比于组织架构方面，个体行为带来的舞弊风险更大，与此同时应该将内部管理的注意力集中于此。

Albrecht W Steve（2014）提出舞弊三角理论，又称三因素理论。该理论将舞弊动机诠释为三个因素：（1）压力：大多数人需要某种形式的压力才能实施犯罪行为。这种压力不一定需要外部观察者理解，但确实需要存在。压力可能包括金钱问题，赌博债务，酗酒或吸毒成瘾，压倒性的医疗费用。贪婪也可能成为一种压力，但通常需要与不公正联系在一起。（2）机会：必须有实施该行为的机会。在欺诈的情况下，通常会出现一种暂时的情况，即有机会实施该行为而没有很大的机会被抓住。未积极开展预防欺诈工作的公司可能会向符合欺诈三角条件的所有三个标准的个人提供反复的机会。(3)合理化：一个人要犯下不道德行为的心态是合理化之一。个人设法证明他或她将要做什么。有些

人可能认为他们只是要借用被盗的货物，或者他们比从其窃取的“大”公司中更需要资金。

G. Jack Bologna（1993）提出 GONE 理论，又称四因素理论。该理论指出舞弊由四个因素构成即贪婪、机会、需求与揭露,这四个因素共同作用，决定财务舞弊的概率。

G. Jack Bologna（1995）提出风险因子理论，该理论在 GONE 的理论基础上继续优化，将舞弊动因分为个人风险因子和一般风险因子。个人风险因子主要是个人自身的素质等不稳定因素，而一般风险因子则主要是企业环境等稳定性因素。当两者条件均达到舞弊要求后，舞弊就会发生。

Kleinman Gary（1999）抽取 173 家标准审计意见公司和 61 家续经营疑虑保留意见公司构成数据框，提取 6 个表外非财务指标对审计意见类型进行预测。结果指出，对标准审计意见的样本准确率为 100%，对保留意见样本准确率达 96.7%。

JW Lin（2003）以 Logistic 回归为基准，对比人工神经网络（ANN）与模糊神经网络（FNN）。结果显示 FNN 优于以前的研究报告的大多数统计模型和人工神经网络，与基准 Logit 模型相比，它的性能也令人满意，尤其是在欺诈案件的预测中。

Erkki K.Laitinen（2010）用 1992 年至 1994 年的 37 家上市公司开发了基于财务报表信息的模型，以识别合格的审计报告。文献显示，审计报告的资格主要与低盈利能力，高负债和低（负）增长有关；获得资格的可能性更大，公司的成长越慢，资产负债表中的权益份额就越低，员工人数越少。该模型准确率为 62%。

Ngai（2011）分析了 1997 年至 2008 年之间发表的有关该主题的 49 篇期刊文章，并将其分为四类金融欺诈：银行欺诈，保险欺诈，证券和商品欺诈以及其他相关的金融欺诈，和六类数据挖掘技术：分类，回归，聚类，预测，离群值检测和可视化。

Fernández-Gámez MA（2015）选取从 2008 年至 2010 年 447 家西班牙上市公司作为样本，构建多层感知器（MLP）和概率神经网（PNN）模型，结果显示 MLP 在训练样本的准确率高达 99.40%，在测试样本的准确率达到 98.10%；PNN 在训练样本的准确率高达 85.60%，在测试样本的准确率达到 81.70%。好于现有分类平均准确率 80%。

1.2.2 国内财务舞弊研究

方军雄（2003）通过独立样本 t 检验、线性概率模型、Logistic 模型统计财务舞弊鉴别最重要的几个财务指标：应收账款，应收账款周转率，资产负债率，速动比率，主营业务税金及附加比率，管理费用，和销售费用率，模型 LMP 误判率回判结果为 24.18%，预测结果为 33.33%；Logistic 模型误判率回判结果为 35.16%，预测结果为 27.78%。

田金玉（2010）随机抽取 2007 年沪深两市 94 家上市公司作为测试集，抽取 30 家上市公司作为训练集，计算 7 大类共 20 个指标，通过 BP 神经网络与线性回归模型进行分类，最终结果：建模样本正确判定率为 82.81%，检验样本正确判定率为 83.33%。

夏明（2015）选取沪深两所 2004 年至 2011 年选取舞弊样本 37 个，非舞弊样本 49 个运用主成分分析与 BP、RBF、RBF-BP 神经网络分类训练舞弊识别模型，最终发现 RBF-BP 模型训练效果最好，平均准确率为 85.0%，最好准确率为 87.5%，最差准确率为 81.25%。

刘佳进（2019）选取 2008 年至 2015 年 A 股企业作为训练集，2016 年 A 股企业作为测试集，以 1:1 的舞弊与非舞弊比例选取 95 家制造业上市公司，7 家与制造业相关的

其他行业上市公司，算计 4 大类共 18 个特征作为指标，采用 Matlab2015b 神经网络工具箱中 TRAINLM 自带的训练函数进行训练。结果显示：舞弊组训练集准确率为 90%，测试集为 85%；正常组训练集准确率为 88%，测试集准确率为 77%。

1.3 本研究的主要内容

本研究共分为五个部分：

第一部分为前言，主要介绍研究目的、研究意义、以及通过查阅国内外文献对于现今会计舞弊识别的研究作以梳理。

第二部分为相关概念综述，主要介绍了财务舞弊概念、财务舞弊表现方式、以及 LightGBM 算法的一些理论，本文将依据上述理论基础进行研究。

第三部分为样本选取与分析，通过国泰安数据库选取 2010 年至 2018 年沪深 A 股全部上市公司作为样本进行选取与分析。

第四部分为财务舞弊识别分类器构建，主要介绍研究过程，包括特征变量的选取与确定，模型参数优化调整与训练，模型评价与变量分析。

最后部分为结论，对整体研究进行归纳总结，提出意见和建议。

2 相关概念综述

2.1 财务舞弊概念与表现方式

财务舞弊一般指故意操纵财务报表以造成公司财务状况的虚假表象。此外，它涉及雇员，会计师或组织本身误导投资者和股东。公司可以通过夸大其收入，不记录费用以及虚报资产和负债来伪造其财务报表，例如安然丑闻便是历史上最著名的财务舞弊案例之一。常见会计舞弊表现方式主要包括以下三种：

（1）夸大收入

如果公司夸大其收入，则可能会造成会计欺诈。假设公司实际上处于亏损状态并且没有产生足够的收入。为了掩盖这种情况，该公司可能声称自己在财务报表中产生的收入要比实际产生的更多。在其声明中，该公司的利润将被夸大。如果公司夸大了其收入，将推高公司的股价并制造虚假的财务状况。

（2）未记录费用

当公司不记录其费用时，会发生另一种会计欺诈。该公司的净利润被夸大了，而其成本却在低调的损益表。这种类型的会计欺诈对公司获得多少净收入产生了错误的印象。实际上，它可能正在赔钱。

（3）漏报资产和负债

当公司夸大其资产或低估其负债时，会发生另一种形式的会计欺诈。例如，一家公司可能夸大其流动资产而低估了其流动负债。这种欺诈行为代表了公司的短期流动性。

假设一家公司的流动资产为 100 万元，其流动负债为 500 万元。如果公司夸大了其流动资产而又低估了其流动负债，那就是在歪曲其流动性。公司可以声明其拥有 500 万

元的流动资产和 50 万元的流动负债。然后，潜在的投资者将相信公司拥有足够的流动资产来支付其所有负债。

但财务舞弊很难被指控，因为财务舞弊需要意图，但很难证明有某种意图的财务舞弊。为了进行财务舞弊，公司必须故意伪造财务记录。考虑一个做出估计的公司，该估计必须在以后进行修订。因为错误不是故意的，所以没有发生会计欺诈。现在，假设一家上市公司的首席执行官在知情的情况下对公司的前景进行了虚假陈述。证券交易所可能会指控该首席执行官舞弊。但是，这不是财务舞弊，因为没有伪造财务记录。

2.2 LightGBM 算法

2.2.1 机器学习简介

随着人工智能与大数据在其实际应用的效率与精度的提高，越来越多的模型被进行跨学科交叉实验。本质上，机器学习是集统计学与概率论等综合学科，经过严谨的算法设计，并依赖数据与计算机来提取数据信息，从现有已知数据推测未知信息，让计算机自主学习的自动化分析与总结模式。

本文使用的 LightGBM 算法本质上在执行分类任务，即根据具体问题构建特定分类规则或分类器将样本进行类别划分。分类过程主要由三部分构成，一是训练过程，将已经标记好的训练集输入训练模型，使其自动进行训练以达到可接受的分类标准；二是测试过程，将已经标记好的测试集输入已训练成功的模型后，产生的测试结果与标记的标签经由某种评价方式进行评估；三是分类过程，将未知数据输入已训练成功的模型，输出预测后的标签值。

2.2.2 LightGBM 相关理论与算法

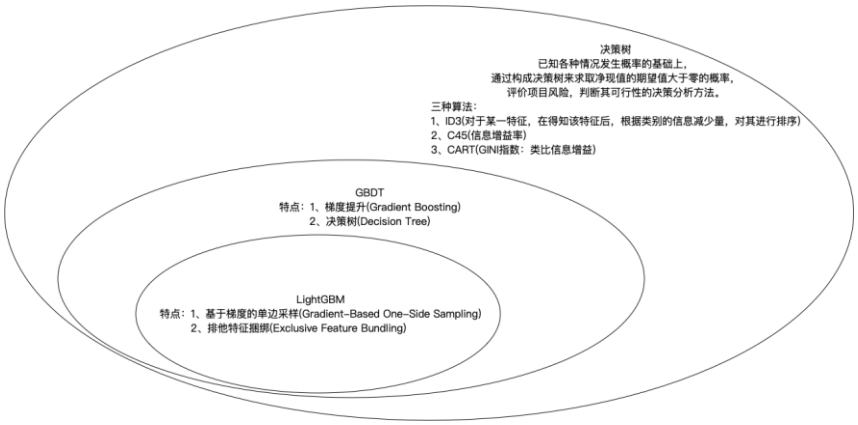

图 2.1 LightGBM 韦恩图

2.2.2.1 决策树

由图 2.1 能够看出，LightGBM 算法是基于决策树理论的改进。决策树是运用最为广泛的机器学习算法之一。决策树为树结构，其每个非叶节点表示一个特征属性；每个内部节点有一个判断条件，输出是与否，并按照输出值选择输出分支；将上述步骤不断循

环，进行判断与分配，直到到达叶子节点输出样本类别。决策树示意图如图所示：

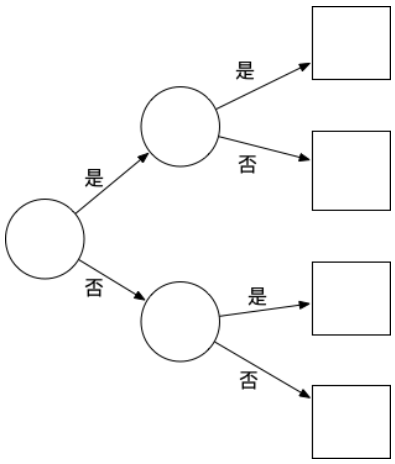

图 2.2 决策树示意图

2.2.2.2 LightGBM 算法

LightGBM 是一种新的梯度增强树框架，在 2017 年由微软亚洲研究院 DMTK 团队研发。此模型效率高且扩展性强，可以支持许多不同的算法，包括：GBDT，GBRT，GBM 和 MART。由于其完全贪婪的树增长方法以及基于直方图的内存和计算优化，LightGBM 被证明比梯度增强树(GBDT)现有实现快几倍。它还具有基于 DMTK 框架的完整的分布式训练解决方案。LightGBM 的分布式版本仅用一两个小时即可完成对 Criteo 数据集的 CTR 预测器的训练，该数据集包含 16 台计算机，其中包含具有 67 个特征的 17 亿条观测。

LightGBM 作出以下了三点优化：一是基于梯度单侧采样算法，此算法可基于梯度对实例进行下采样。我们知道，梯度小的实例训练得好（训练误差小），梯度大的实例训练不足。通过只关注梯度大的实例而放弃梯度小的实例得到一个简单的下采样，但这将改变数据的分布。简而言之，此算法保留梯度大的实例，而对梯度小的实例进行随机抽样；二是带深度限制的 Leaf-wise 叶子生长策略，当其他算法水平生长树时，LightGBM 垂直生长树，这意味着 LightGBM 在叶子上生长树，而其他算法在水平上生长。为了生长，它将选择具有最大损失的叶子。当生长同一片叶子时，与逐级算法相比，逐叶算法可以减少更多的损失，示意图如图 2.3 所示；三是直方图差加速，叶子的直方图可以由父节点直方图与兄弟直方图做差得到。由此算法得出，仅需遍历直方图的 k 个组数构成的统计量，无需遍历全部树。

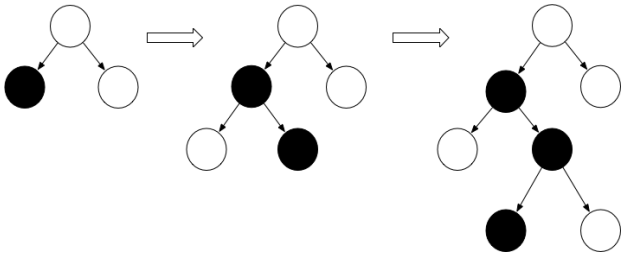

图 2.3 Leaf-wise 示意图

3 样本选取与分析

3.1 样本选取

由于 2007 年新会计准则的颁布与实施，上市公司将所有者权益变动表从会计报表附注中单独，使得此部分数据更具结构化，故考虑使用 2007 年后的数据进行分析，同时综合考虑样本大小与计算机处理模型速度，选取 2010 年至 2018 年国泰安财务指标分析数据库中持续经营企业作为样本。

财务舞弊的样本依据选取国泰安财务报告审计意见数据库对应数据作为解释变量，选取 2010 年至 2018 年审计意见为对应响应变量，同时对于财务舞弊样本的选取主要有以下三点：

第一，只分析年报数据。因为月报表、季度报表、中期报表、和临时报表不能完全反映企业的情况，所以需要去除上述报告出具非标准无保留意见的情况，只标记年报出具非标准无保留意见；

第二，只选取沪深两市 A 股上市公司。因为需要进行结构化转换，数据需要统一标准，且沪深两市 A 股上市公司所构成数据样本量足够对模型进行训练；

第三，只区分标准无保留审计意见与非标准无保留审计意见。因为审计结论统共分为无保留意见、保留意见、无法表示意见和否定意见，其中无保留意见中有标准无保留意见和非标准无保留意见。

对于模型的训练与预测中，二分类模型效果明显好于多分类，且出具几种意见有审计人员主观判断成分，为提高准确性选取无明显交集的无保留审计意见与非标准无保留审计意见分类方式，以此尽量控制响应变量的准确程度，用以减小样本不准确带来研究结果的负面影响。

3.2 舞弊分析

对于沪深 A 股上市公司 2010 年至 2018 年的审计意见进行统计并可视化，结论如下图所示：

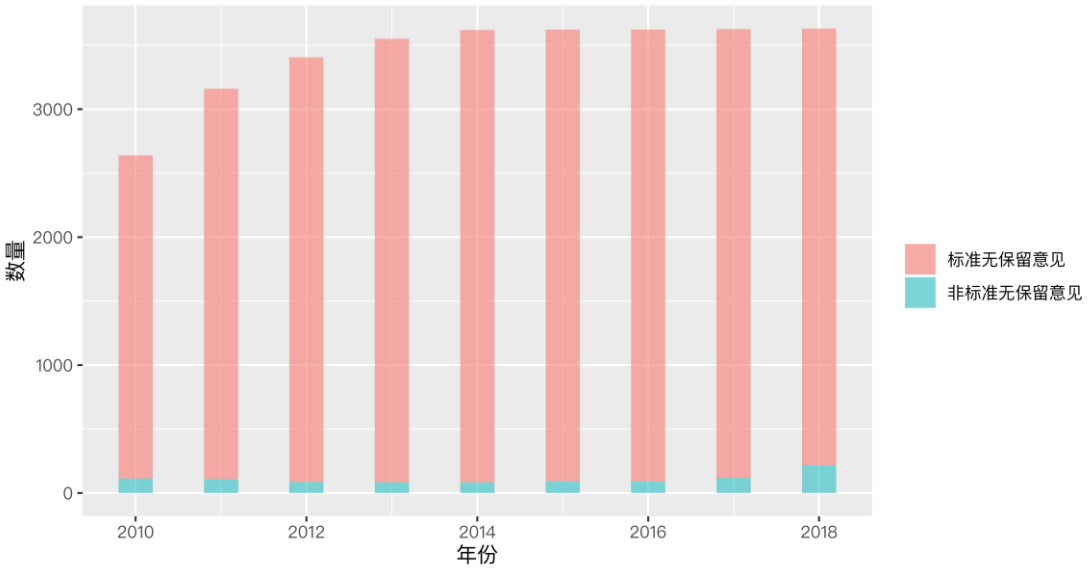

图 3.1 样本审计意见分类

可以发现，被出具非标准无保留意见的公司数量在逐年增加。这是由于大部分被披露公司财务舞弊并非单一情况，存在连续多年财务异常情况，企业的财务异常必将导致接下来几年的财务异常，进而造成财务舞弊被披露的现象。而之所以起初财务异常未被披露，是因为当代金融工具发展极其迅速，大多数审计人员缺乏必要的经验与相关知识；公司首席执行官、财务管理人员、会计等故意欺骗审计人员；监管存在一定滞后性，往往会在企业财务存在舞弊的几年后才会被审查。

3.3 样本确定

由图 3.1 可知，样本中共有 999 个财务舞弊样本，占总样本的 3.24%，其中 2010 年共 114 个，占当年总样本量的 4.32%；2011 年共 108 个，占当年总样本量的 3.42%；2012 年共 88 个，占当年总样本量的 2.58%；2013 年共 84 个，占当年总样本量的 2.37%；2014 年共 85 个，占当年总样本量的 2.35%；2015 年共 91 个，占当年总样本量的 2.51%；2016 年共 91 个，占当年总样本量的 2.51%；2017 年共 119 个，占当年总样本量的 3.28%；2018 年共 219 个，占当年总样本量的 6.03%。

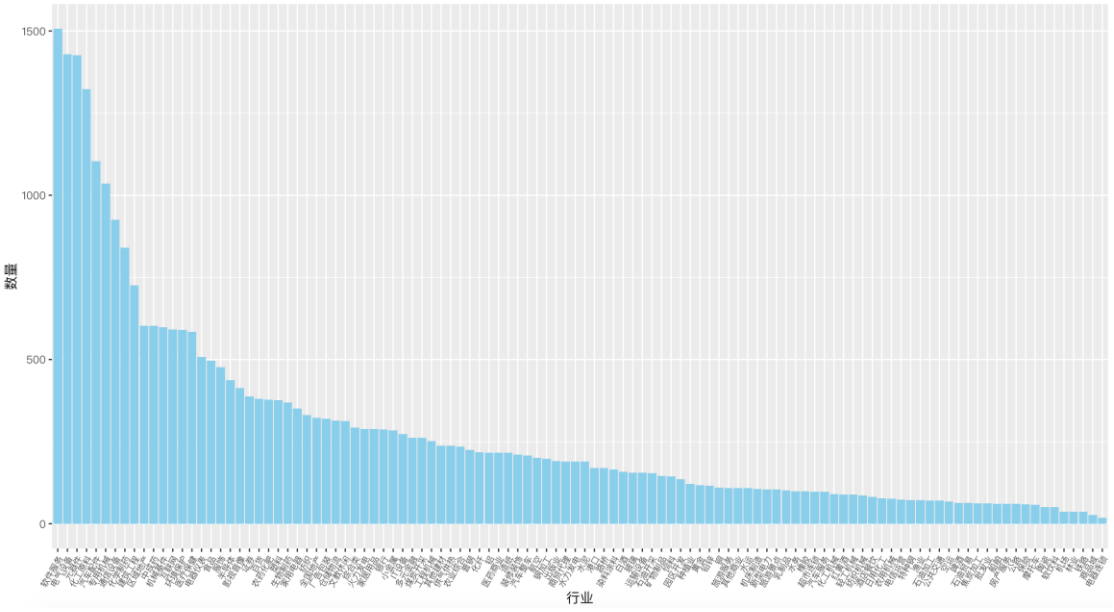

图 3.2 样本中各行业比例分布图

由图 3.2 可以看出样本中个行业公司占比分布，其中软件服务类企业数据占比最高，约占总体的 4.88%，电器连锁类企业数据占比最低，约占总体的 0.058%。

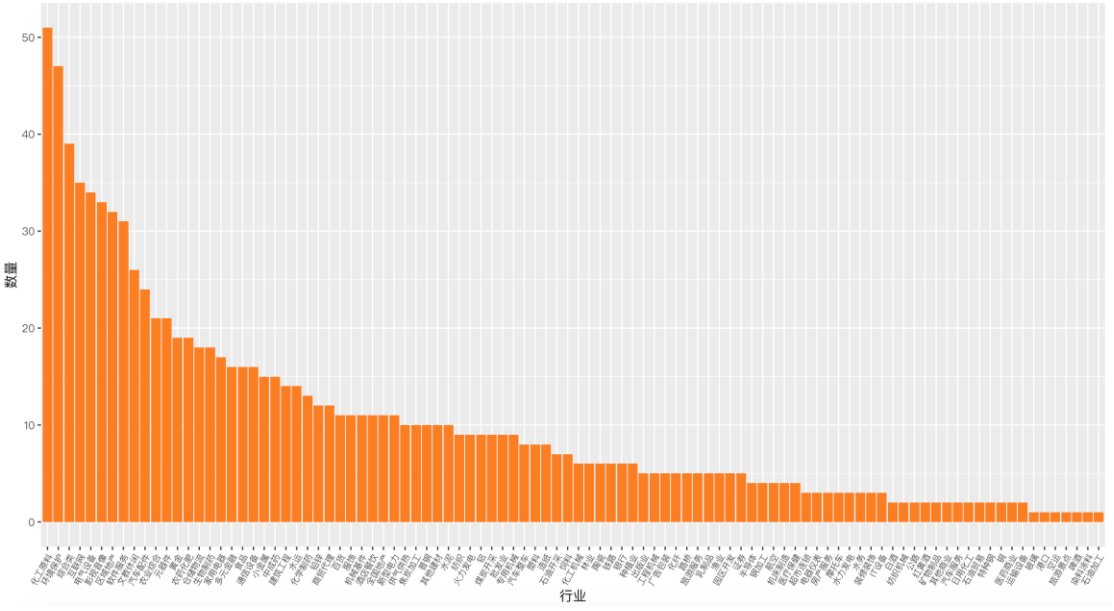

图 3.3 非标准无保留意见中各行业比例分布图

由图 3.3 可以看非标准无保留意见中个行业公司占比分布，其中化工原料类企业数据占比最高，约占总体的 5.11%，电器连锁类企业数据占比最低，约占总体的 0.058%。

4 财务舞弊识别分类构建

4.1 特征变量的选取

4.1.1 特征变量选取原则

（1）最小冗余性：解释变量与响应特征实际信息存在实际误差，若冗余度越高，则数据完整性越低，应选取冗余度小的解释变量；

（2）最大关联性：有些解释变量与响应目标关联性低，而有些解释变量与与响应目标的研究有较大的关联性，解释变量的相关性越高，越应优先选择；

（3）可加性：解释变量必须为非虚拟变量，相加后有经济学上的实际意义，且对于所需统计学变换，特定的“加法”和该变换的顺序可颠倒而不影响结果；

（4）发散性：若一个特征不发散，则此特征为共性特征或极相似，说明此特征代表性低，则应考虑剔除。

4.1.2 变量初步选取

在文献综述相关分析中，变量的选取主要来源于两部分：第一部分为公司管理机构体系；第二部分为公司财务报表所列示财务信息经计算得出的财务指标。公司管理结构涉及公司战略与目标，董事会的集体决议，和职业经理人的主观判断等。上述指标无法以同一框架进行指标定量计算，且量化后置信度无法保证，不符合选取的可操作性。综上考虑，本次研究变量选取财务指标，根据国泰安数据库中对财务指标的计算，从指标中剔除一些冗余指标，共得到 7 类共 163 个指标，见附表 A1。

4.1.3 特征变量确定

本次样本采样全部 A 股上市公司数据，对于 2010 年至 2018 年，共 9 年的数据取交集，共包括 29938 个观测。根据附表 A1 所列指标显示，每个观测具有 163 个特征。因此，从以上样本和变量的确定可以计算出，此数据框共有 4879894 个数据。由此，可以显著地判断出，此次数据量较大。但同时也存在一些对于总体的研究没有重大意义的特征，或特征之间存在相互依赖关系，故需要对总体特征进行剔除。本文采用过滤型特征约简作为管道进行初步处理，即先将数据框全部特征属性作为输入，再将输出后的运用过滤型特征约简后的特征子集作为输入应用于 LightGBM 分类器的构建。

（1）正态性检验

添加哑变量$y \in \{0,1\}$，将所有数据按审计报告出具标准无保留意见为 0，非标准无保留意见为 1 进行分类，对所有观测的标准类和非标准类所构成的全部解释变量进行 L-illie 正态性检验。利用 R 语言的 nortest 包进行检验，设置显著性水平为 90%来判断变

量是否符合正态性假设。

从附表 A5 可以看出，所有解释变量的 D 统计量均较小，其假设检验量 p 值均小于 0.10，结果极其显著。故舍弃H_0假设，即全部解释变量均不符合正态性假设。所以不能使用以正态性为假设条件的单样本 T 检验等参数检验进行指标的初步筛选。由此，本文对解释变量进行非参数检验。本文选用 Wilcoxon 秩和检验对全部解释变量进行检验，设置显著性水平为 99%来判断真实位置偏移不等于 0 与否，结果如附表 A3 所示。

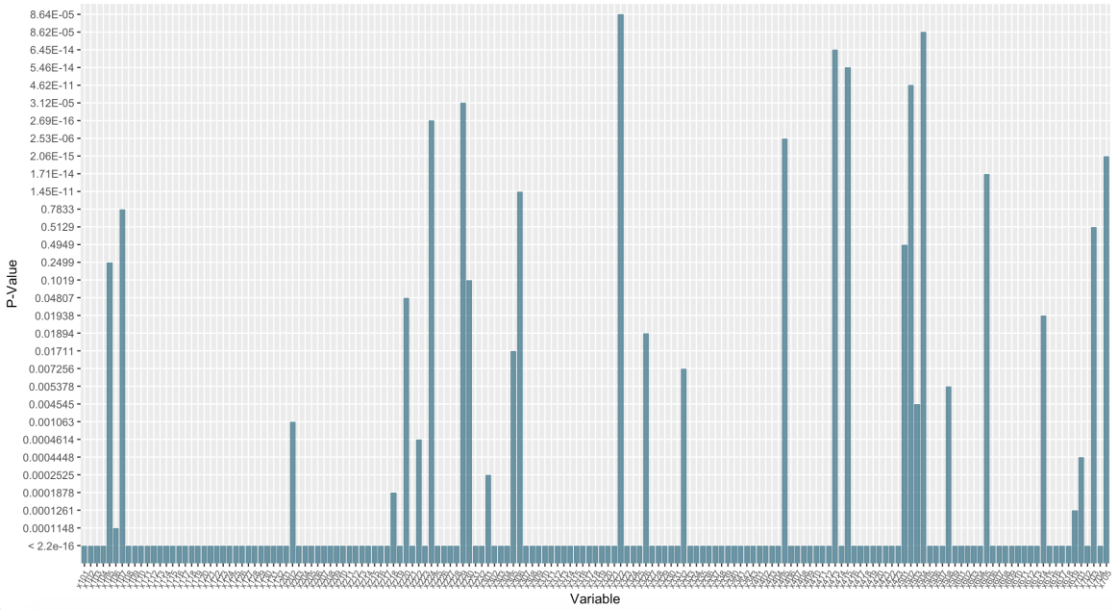

图 4.1 Wilcoxon 秩和检验

从图 4.1 可以显著地判断，X_{105}、X_{107}、X_{220}、X_{230}、X_{305}、X_{326}、X_{501}、X_{614}、X_{703}未通过假设检验，即上述解释变量对于分类响应变量的相关性为 0 的概率极大，故应剔除上述指标。

（2）多元方差分析

上述分析未考虑两个独立总体对于其解释变量的影响，由此本文采用 MANOVA 模型对剩余解释变量进行检验，从而筛选出相关性较强的指标。由于分类响应变量为哑变量，故两个总体Y_0~$N(0,0)$, Y_1~$N(1,0)$均成立，满足方差分析的正态性和方差齐性假设条件。

表 4.1 方差分析

变量	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
X_{101}	1	0.5	0.47	19.168	1.20E-05	***
X_{102}	1	2	1.99	81.177	< 2e-16	***
X_{103}	1	0	0.04	1.633	0.201287	
X_{104}	1	0.3	• 0.26	10.524	0.00118	**

续表：

变量	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
X_{106}	1	0	0.01	0.351	0.553299	
X_{108}	1	0.6	0.58	23.646	1.16E-06	***
X_{109}	1	0.2	0.18	7.202	0.007287	**
X_{110}	1	0.2	0.23	9.183	0.002444	**
X_{111}	1	0.3	0.35	14.202	0.000164	***
X_{112}	1	0.1	0.05	2.159	0.141755	
X_{113}	1	0	0.01	0.292	0.589218	
X_{114}	1	2.8	2.79	113.869	< 2e-16	***
X_{115}	1	9.6	9.59	390.942	< 2e-16	***
X_{116}	1	4.4	4.4	179.349	< 2e-16	***
X_{117}	1	1.2	1.16	47.262	6.33E-12	***
X_{118}	1	0	0	0.003	0.959972	
X_{119}	1	0	0.02	0.624	0.429709	
X_{120}	1	0.1	0.05	2.077	0.149559	
X_{121}	1	1.7	1.72	70.243	< 2e-16	***
X_{122}	1	0.1	0.09	3.732	0.053397	.
X_{123}	1	0.1	0.13	5.146	0.023304	*
X_{124}	1	0	0.03	1.34	0.246993	
X_{125}	1	0	0.01	0.42	0.517155	
X_{126}	1	0	0.01	0.245	0.620896	
X_{127}	1	0.6	0.6	24.537	7.33E-07	***
X_{128}	1	0.4	0.38	15.323	9.08E-05	***
X_{129}	1	0.2	0.19	7.555	0.005988	**
X_{130}	1	0	0.03	1.321	0.25035	
X_{131}	1	1	1	40.621	1.88E-10	***
X_{132}	1	1.2	1.18	48.1	4.13E-12	***
X_{201}	1	7.4	7.4	301.573	< 2e-16	***
X_{202}	1	0.2	0.24	9.626	0.00192	**
X_{203}	1	0	0.01	0.601	0.438129	
X_{204}	1	2.6	2.56	104.321	< 2e-16	***
X_{205}	1	4.8	4.77	194.449	< 2e-16	***
X_{206}	1	0.3	0.3	12.314	0.00045	***
X_{207}	1	0.3	0.34	14.016	0.000182	***
X_{208}	1	0.7	0.73	29.66	5.19E-08	***

续表：

变量	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
X_{209}	1	1.7	1.74	70.805	< 2e-16	***
X_{210}	1	9.8	9.82	400.222	< 2e-16	***
X_{211}	1	1	1.03	42.059	8.99E-11	***
X_{214}	1	1	1.05	42.615	6.77E-11	***
X_{215}	1	2.9	2.93	119.365	< 2e-16	***
X_{216}	1	1.1	1.08	44.116	3.15E-11	***
X_{217}	1	1.8	1.85	75.34	< 2e-16	***
X_{218}	1	0.1	0.14	5.653	0.017428	*
X_{219}	1	3.1	3.05	124.347	< 2e-16	***
X_{220}	1	0.6	0.6	24.602	7.09E-07	***
X_{221}	1	0	0.03	1.378	0.240437	
X_{222}	1	0	0.02	0.719	0.396591	
X_{223}	1	0.1	0.1	3.988	0.045832	*
X_{224}	1	0.6	0.63	25.607	4.21E-07	***
X_{225}	1	0	0	0.015	0.903491	
X_{226}	1	1.1	1.07	43.589	4.12E-11	***
X_{227}	1	0.4	0.37	15.053	0.000105	***
X_{228}	1	2.1	2.05	83.6	< 2e-16	***
X_{229}	1	2.1	2.14	87.028	< 2e-16	***
X_{231}	1	0	0.01	0.549	0.458771	
X_{232}	1	0	0.01	0.282	0.595294	
X_{301}	1	0.1	0.08	3.092	0.078668	.
X_{302}	1	0	0	0.071	0.789277	
X_{303}	1	1.2	1.18	48.1	4.13E-12	***
X_{304}	1	0.5	0.48	19.389	1.07E-05	***
X_{306}	1	0.3	0.29	11.692	0.000629	***
X_{307}	1	0.1	0.06	2.322	0.127595	
X_{308}	1	0.1	0.05	2.227	0.13565	
X_{309}	1	0.6	0.58	23.445	1.29E-06	***
X_{310}	1	2	2	81.463	< 2e-16	***
X_{312}	1	1.1	1.09	44.304	2.86E-11	***
X_{313}	1	17.5	17.52	714.002	< 2e-16	***
X_{314}	1	3.7	3.7	150.962	< 2e-16	***
X_{315}	1	18.7	18.73	763.334	< 2e-16	***

续表：

变量	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
X_{317}	1	0	0	0.001	0.97374	
X_{318}	1	0.5	0.46	18.676	1.55E-05	***
X_{319}	1	0	0.01	0.301	0.583299	
X_{320}	1	0	0	0.023	0.879861	
X_{322}	1	0.1	0.09	3.535	0.060107	.
X_{324}	1	1.1	1.14	46.647	8.66E-12	***
X_{325}	1	1.3	1.33	54.173	1.88E-13	***
X_{327}	1	1.1	1.11	45.142	1.87E-11	***
X_{328}	1	1.3	1.32	53.808	2.27E-13	***
X_{329}	1	0.3	0.28	11.594	0.000663	***
X_{330}	1	0	0	0.079	0.778033	
X_{332}	1	0.4	0.39	15.939	6.56E-05	***
X_{333}	1	0	0.01	0.273	0.601597	
X_{334}	1	0.7	0.72	29.48	5.69E-08	***
X_{335}	1	0.1	0.09	3.641	0.056397	.
X_{338}	1	0.5	0.47	19.078	1.26E-05	***
X_{339}	1	1.9	1.92	78.137	< 2e-16	***
X_{340}	1	3.2	3.23	131.569	< 2e-16	***
X_{341}	1	0	0.03	1.125	0.288941	
X_{342}	1	0.1	0.08	3.395	0.065402	.
X_{343}	1	0	0.02	1.014	0.313878	
X_{401}	1	0.1	0.06	2.344	0.125772	
X_{402}	1	0	0.01	0.265	0.606733	
X_{403}	1	0.5	0.47	18.967	1.33E-05	***
X_{404}	1	0	0.02	0.638	0.424286	
X_{405}	1	0	0	0.003	0.95473	
X_{406}	1	0.2	0.2	8.126	0.004366	**
X_{407}	1	0.1	0.13	5.158	0.023152	*
X_{408}	1	0.1	0.06	2.536	0.111321	
X_{409}	1	0.3	0.26	10.703	0.001071	**
X_{410}	1	0.6	0.59	24.138	9.01E-07	***
X_{411}	1	0	0	0.174	0.676304	
X_{412}	1	0	0.04	1.512	0.218886	
X_{413}	1	0	0	0.108	0.74232	

续表：

变量	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
X_{415}	1	0	0.01	0.262	0.608491	
X_{416}	1	0	0	0.122	0.727006	
X_{417}	1	0.1	0.13	5.312	0.021189	*
X_{418}	1	0	0.05	1.864	0.172229	
X_{419}	1	0.1	0.08	3.15	0.075937	.
X_{420}	1	0.2	0.22	9.15	0.00249	**
X_{421}	1	0	0	0.062	0.802798	
X_{422}	1	0	0	0	0.992913	
X_{423}	1	0.2	0.18	7.294	0.006922	**
X_{502}	1	3.5	3.5	142.586	< 2e-16	***
X_{503}	1	0	0.05	1.886	0.169642	
X_{504}	1	0	0	0.101	0.750608	
X_{505}	1	0.9	0.9	36.528	1.52E-09	***
X_{506}	1	2.6	2.62	106.644	< 2e-16	***
X_{507}	1	0.1	0.14	5.739	0.0166	*
X_{508}	1	0	0.02	0.688	0.406871	
X_{509}	1	0	0	0.177	0.673864	
X_{601}	1	43.7	43.69	1780.723	< 2e-16	***
X_{602}	1	1.6	1.64	66.963	2.88E-16	***
X_{603}	1	0	0	0.033	0.855978	
X_{604}	1	0.1	0.14	5.767	0.016332	*
X_{605}	1	0	0.03	1.137	0.286195	
X_{606}	1	0.8	0.81	33.011	9.25E-09	***
X_{607}	1	5.6	5.61	228.774	< 2e-16	***
X_{608}	1	12	12.04	490.836	< 2e-16	***
X_{609}	1	0	0.04	1.59	0.207316	
X_{610}	1	0.7	0.66	26.899	2.16E-07	***
X_{611}	1	0.1	0.07	2.909	0.088107	.
X_{612}	1	0.1	0.1	3.925	0.047593	*
X_{613}	1	3.1	3.12	127.044	< 2e-16	***
X_{615}	1	0.2	0.16	6.522	0.01066	*
X_{616}	1	4.2	4.16	169.364	< 2e-16	***
X_{617}	1	0.2	0.17	6.815	0.009045	**
X_{618}	1	0	0.04	1.53	0.216166	

续表：

变量	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
X_{619}	1	0	0	0.18	0.671537	
X_{701}	1	0	0	0.142	0.706536	
X_{702}	1	0.1	0.15	6.024	0.014117	*
X_{703}	1	0.6	0.59	23.86	1.04E-06	***
X_{704}	1	0	0	0.125	0.724187	
X_{705}	1	0.3	0.28	11.41	0.000731	***
X_{613}	1	3.1	3.12	127.044	< 2e-16	***
Residuals	29792	731.0	0.02			
Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**						

从表 4.1 多因素方差分析的结果可以看出，对于解释变量的显著性检验中，p 值在 0.05 至 0.1 之间的有X_{122}、X_{302}、X_{322}、X_{335}、X_{342}、X_{419}、X_{611}；p 值在 0.01 至 0.05 之间的有X_{123}、X_{218}、X_{224}、X_{407}、X_{417}、X_{507}、X_{604}、X_{612}、X_{615}、X_{702}；p 值在 0.01 至 0.001 之间的有X_{104}、X_{109}、X_{110}、X_{202}、X_{129}、X_{617}、X_{406}、X_{409}、X_{420}、X_{423}；p 值小于 0.001 的有X_{101}、X_{102}、X_{108}、X_{111}、X_{114}、X_{115}、X_{116}、X_{117}、X_{121}、X_{127}、X_{128}、X_{131}、X_{132}、X_{201}、X_{204}、X_{205}、X_{206}、X_{207}、X_{208}、X_{209}、X_{210}、X_{211}、X_{214}、X_{215}、X_{216}、X_{217}、X_{219}、X_{221}、X_{225}、X_{227}、X_{228}、X_{229}、X_{231}、X_{304}、X_{306}、X_{309}、X_{310}、X_{312}、X_{313}、X_{314}、X_{315}、X_{318}、X_{324}、X_{325}、X_{327}、X_{328}、X_{329}、X_{332}、X_{334}、X_{338}、X_{339}、X_{340}、X_{403}、X_{410}、X_{502}、X_{505}、X_{506}、X_{601}、X_{602}、X_{606}、X_{607}、X_{608}、X_{610}、X_{613}、X_{616}、X_{703}、X_{705}，共有 94 个指标通过假设检验。

（3）因子分析

由于通过前面的方差分析共有 94 个财务指标通过假设检验，在进行因子分析前，需要对 94 个指标 2814172 个样本组成的数据进行 KMO 检验和 Bartlett 球形检验，以验证是否适合做因子分析。KMO 检验的假设条件是变量组成的矩阵为正定矩阵，即矩阵的特征值需要大于 0，但是经过查阅文献可以了解到当特征值极小也可能导致非正定，即特征值需要远大于 0，因此这一条件要求各个变量之间不存在高度相关关系。但是因子分析主要目的是通过变量之间的关系提取公因子，所以两两相关的变量不需要进行剔除，主要是剔除某个与其他两个以上变量存在高度相关的变量。

4.2 分类器建立

4.2.1 前期处理

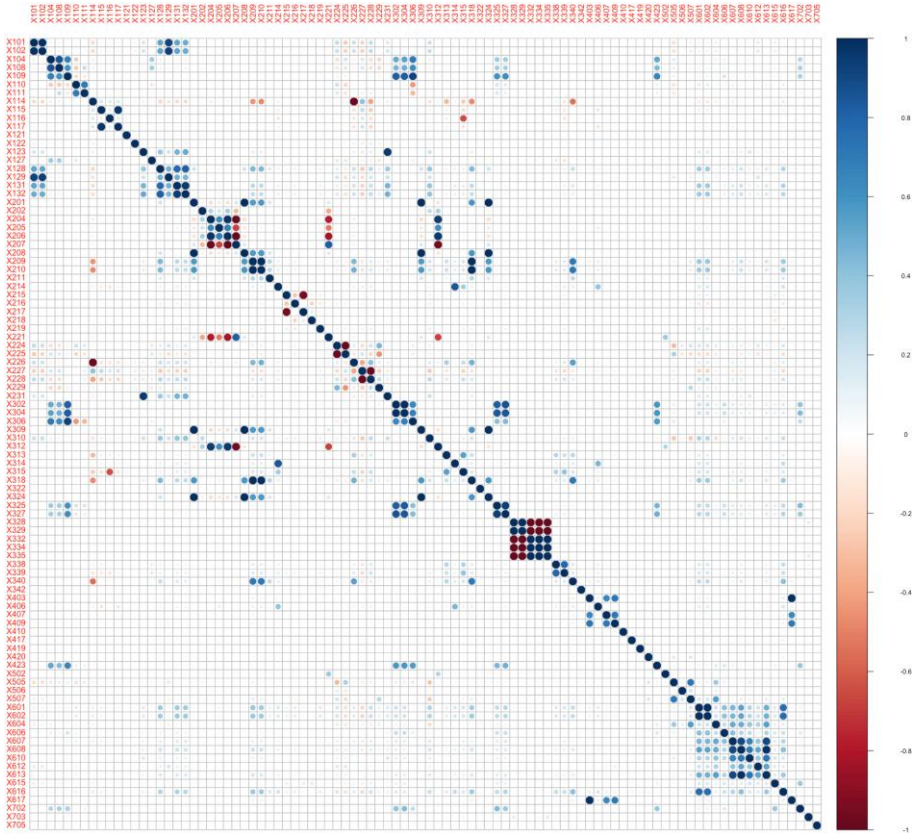

图 4.2 解释变量 Pearson 相关系数

本文以±0.95 为阈值筛选指标，通过图 4.2 可以计算出，X_{101}与X_{102}、X_{114}与X_{226}、X_{115}与X_{117}、X_{231}与X_{123}、X_{309}与X_{201}，X_{208}，X_{324}、X_{207}与X_{204}，X_{206}，X_{312}、X_{309}与X_{208}，X_{324}、X_{210}与X_{209}，X_{318}、X_{215}与X_{217}、X_{224}与X_{225}、X_{227}与X_{228}、X_{304}与X_{302}、X_{309}与X_{324}、X_{328}与X_{329}，X_{332}，X_{334}，X_{335}、X_{617}与X_{403}、X_{601}与X_{602}、X_{613}与X_{608}，共 7 对指标均高度相关。所以，本文剔除X_{102}、X_{226}、X_{117}、X_{123}、X_{201}、X_{208}、X_{324}、X_{204}、X_{206}、X_{312}、X_{208}、X_{324}、X_{209}、X_{318}、X_{217}、X_{225}、X_{228}、X_{302}、X_{324}、X_{329}、X_{332}、X_{334}、X_{335}、X_{403}、X_{602}、X_{608}，共 26 个指标。因此对余下的 71 个解释变量进行因子分析。在做因子分析前先对解释变量进行 KMO 检验和 Bartlett 球形检验，用以验证处理后数据是否适合做因子分析。

表 4.2 KMO 检验和 Bartlett 球形检验

Kaiser-Meyer-Olkin Factor Adequacy	
Overall MSA	0.64
Bartlett Test of Sphericity	
Chisq	7209006
p-value	0
Df	4371

通过其他文献得出结论，KMO 值一般大于 0.5 表明比较适合做因子分析，选取解释变量的检验值为 0.64，表明这 71 个财务指标适合基于此方法提取公共因子，即因子分析通过了该样本做特征选择的前提假设，而 Bartlett 球形检验的 p-value 小于 0.001，远小于 0.05 也表明本文筛选后的财务指标适用于此种方法。

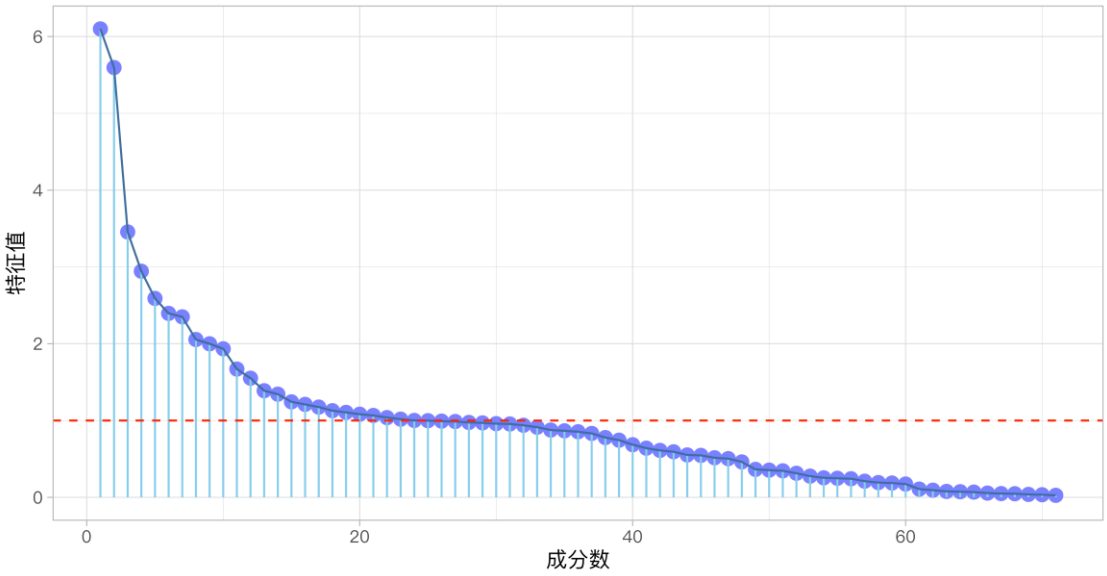

图 4.3 财务指标碎石图

比例表示该公因子所反映指标的反应程度，累积比例表示多个公因子反映指标累积的反应程度。从附表 A4 所示结果中选取特征值大于 1 的公因子，从图 4.3 可以看出，共有 24 个成分选中，这 24 个成分的累积比例达到 68.22%，也就是说这 24 个因子可以反映原指标 68.22%的信息，经过显著性检验的 71 个财务指标可以利用 24 个公因子来表示。但是，初始的载荷矩阵中各个指标的系数之间不存在明显的差异，因此需要对矩阵进行旋转，使得新组成的因子能够对部分指标进行较好的表达。

从附表 A6 正交旋转后因子荷载矩阵得出，新形成的 24 个公因子由原始财务指标构成，本文将根据上述表格将公因子按各指标所占权重重新定义并分类。通过表格将新组成的因子分别以PA_1-PA_{24}来表示，可以看出新指标PA_1中X_{304}、X_{306}、X_{306}、X_{325}、X_{327} 占比较高，因此PA_1反应的是企业的偿债能力；PA_2中X_{604}、X_{606}、X_{607}、X_{610}、X_{612}、X_{613} 具有正的并且大的荷载，因此PA_2体现了企业的每股指标；PA_3中X_{128}、X_{131}、X_{132}具有较大的正荷载，因此PA_3体现了企业的偿债能力；PA_4中X_{207}、X_{221}具有正的并且大的荷载，因此PA_4体现了企业的产权结构；PA_5中X_{407}、X_{409}具有正的并且大的荷载，因此PA_5体现了企业的成长能力；PA_6中X_{314}具有正的并且大的荷载，因此PA_6体现了企业的产权结构；PA_7中X_{313}、X_{340}具有正的并且大的荷载，因此PA_7体现了企业的盈利能力；PA_8中X_{505}、X_{507}具有正的并且大的荷载，因此PA_8体现了企业的经营能力；PA_9中X_{110}、X_{111}具有正的并且大的荷载，因此PA_9体现了企业的偿债能力；PA_{10}中X_{101}、X_{129}具有正的并且大的荷载，因此PA_{10}体现了企业的经营能力；PA_{11}中X_{338}、X_{339}具有正的并且大的荷载，因此PA_{11}体现了企业的盈利能力；PA_{12}中X_{104}、X_{108}、X_{127}具有正的并且大的荷载，因此PA_{12}体现了企业的偿债能力；PA_{13}中X_{315}具有正的并且大的荷载，因此PA_{13}体现了企业的盈利能力；PA_{14}中X_{224}、X_{229}具有正的并且大的荷载，因此PA_{14}体现了企业的产权结构；PA_{15}中X_{309}具有正的并且大的荷载，因此PA_{15}体现了企业的盈利能力；PA_{16}中X_{601}、X_{616}具有正的并且大的荷载，因此PA_{16}体现了企业的每股指标；PA_{17}中X_{215}、X_{218}、X_{219}具有正的并且大的荷载，因此PA_{17}体现了企业的产权结构；PA_{18}中X_{114}、X_{115}、X_{116}、X_{121}具有正的并且大的荷载，因此PA_{18}体现了企业的产权结构；PA_{19}中X_{615}具有正的并且大的荷载，因此PA_{19}体现了企业的每股指标；PA_{20}中X_{502}具有正的并且大的荷载，因此PA_{20}体现了企业的经营能力；PA_{21}中X_{328}具有正的并且大的荷载，因此PA_{21}体现了企业的盈利能力；PA_{22}中X_{410}具有较大的并且正的荷载，因此PA_{22}体现了企业的成长能力；PA_{23}中X_{417}、X_{419}具有较大的并且正的荷载，因此PA_{23}体现了企业的成长能力；PA_{24}中X_{705}具有较大的并且正的荷载，因此PA_{24}体现了企业的现金流量分析。将原指标变换后的新变量及其计算公式整理如下表：

表 4.3 因子分析各变量成分表

变量类别	变量名称	变量构成	包含指标名称
1、偿债能力	PA_3	$PA_3 = 0.783 \times X_{128} + 0.873 \times X_{131} + 0.891 \times X_{132} + 0.68 \times X_{231} + 0.465 \times X_{310}$	息税折旧摊销前利润/负债合计 营业利润 / 流动负债 营业利润 / 负债合计 经营活动产生的现金流量净额/流动负债
	PA_9	$PA_9 = 0.876 \times X_{110} + 0.909 \times X_{111}$	销售毛利率 营运资金 营运流动资本
	PA_{10}	$PA_{10} = 0.849 \times X_{101} + 0.84 \times X_{129}$	流动比率 货币资金 / 流动负债
	PA_{12}	$PA_{12} = 0.662 \times X_{104} + 0.734 \times X_{108} + 0.703 \times X_{127}$	利息费用 净债务 长期债务与营运资金比率
	PA_{18}	$PA_{18} = 0.325 \times X_{114} + 0.646 \times X_{115} + 0.099 \times X_{116} + 0.196 \times X_{121} + 0.558 \times X_{227}$	资产负债率 权益乘数 权益乘数(杜邦分析) 有形资产/带息债务 带息债务/全部投入资本

续表：

变量类别	变量名称	变量构成	包含指标名称
2、产权结构	PA_4	$PA_4 = 0.927 \times X_{207} + 0.854 \times X_{221}$	营业利润/营业总收入 经营活动产生的现金流量净额/营业收入
	PA_6	$PA_6 = 0.896 \times X_{214} + 0.926 \times X_{314} + 0.672 \times X_{406}$	平均净资产收益率(增发条件) 加权平均净资产收益率 净资产收益率(摊薄)同比增长率(%)
	PA_{14}	$PA_{14} = 0.157 \times X_{122} + 0.803 \times X_{224} + 0.758 \times X_{229}$	有形资产/净债务 流动资产/总资产 流动负债/负债合计
	PA_{17}	$PA_{17} = 0.693 \times X_{215} + 0.283 \times X_{218} + 0.373 \times X_{219}$	经营活动净收益/利润总额 所得税/利润总额 扣除非经常损益后的净利润/净利润

续表：

变量类别	变量名称	变量构成	包含指标名称
	PA_1	$PA_1 = 0.879 \times X_{109} + 0.936 \times X_{304} + 0.745 \times X_{306} + 0.852 \times X_{325} + 0.856 \times X_{327} + 0.721 \times X_{423} + 0.502 \times X_{702}$	有形资产 经营活动净收益 折旧与摊销 经营活动单季度净收益 扣除非经常损益后的单季度净利润 研发费用 股权自由现金流量
	PA_7	$PA_7 = 0.532 \times X_{313} + 0.836 \times X_{340}$	净资产收益率 总资产净利润(单季度)
3、盈利能力			净资产收益率(单季度)
	PA_{11}	$PA_{11} = 0.927 \times X_{338} + 0.887 \times X_{339}$	净资产单季度收益率(扣除非经常损益)
	PA_{13}	$PA_{13} = 0.855 \times X_{315}$	净资产收益率(扣除非经常损益)
	PA_{15}	$PA_{15} = 0.689 \times X_{210} + 0.516 \times X_{211} + 0.867 \times X_{309}$	总资产净利润 投入资本回报率 销售净利率
	PA_{21}	$PA_{21} = 0.761 \times X_{328}$	销售净利率(单季度)

续表：

变量类别	变量名称	变量构成	包含指标名称
4、成长能力	PA_5	$PA_5 = 0.878 \times X_{407} + 0.91 \times X_{409} + 0.874 \times X_{617}$	每股净资产相对年初增长率(%) 归属母公司的股东权益相对年初增长率(%) 基本每股收益同比增长率(%)
	PA_{22}	$PA_{22} = 0.207 \times X_{202} + 0.653 \times X_{410} + 0.413 \times X_{506}$	销售费用/营业总收入 营业总收入同比增长率(%) 固定资产周转率
	PA_{23}	$PA_{23} = 0.279 \times X_{342} + 0.621 \times X_{417} + 0.675 \times X_{419}$	价值变动净收益 / 利润总额(单季度) 营业利润环比增长率(%)(单季度) 净利润环比增长率(%)(单季度)
5、经营能力	PA_8	$PA_8 = 0.323 \times X_{420} + 0.854 \times X_{505} + 0.877 \times X_{507}$	归属母公司股东的净利润同比增长率(%)(单季度) 流动资产周转率 总资产周转率
	PA_{20}	$PA_{20} = 0.116 \times X_{205} + 0.941 \times X_{502}$	资产减值损失/营业总收入 应收账款周转天数

续表：

变量类别	变量名称	变量构成	包含指标名称
6、每股指标	PA_2	PA_2 $=0.613\times X_{604}$ $+0.401\times X_{606}$ $+0.893\times X_{607}$ $+0.803\times X_{610}$ $+0.631\times X_{612}$ $+0.93\times X_{613}$	每股营业收入 每股盈余公积 每股未分配利润 每股经营活动产生的现金流量净额 每股现金流量净额 每股息税前利润 基本每股收益
	PA_{16}	PA_{16} $=0.504\times X_{601}$ $+0.543\times X_{616}$ $+0.025\times X_{703}$	每股收益(单季度) 销售商品提供劳务收到的现金 / 营业收入(单季度)
	PA_{19}	PA_{19} $=0.224\times X_{216}$ $+0.761\times X_{615}$	价值变动净收益/利润总额 每股股东自由现金流量
7、现金流分析	PA_{24}	PA_{24} $=0.545\times X_{322}$ $+0.587\times X_{705}$	经营活动产生的现金流量净额 / 营业利润 经营活动产生的现金流量净额 / 经营活动净收益(单季度)

4.2.2 贝叶斯优化调整模型参数

本文运用 LightGBM 模型对上市公司财务报表舞弊进行识别研究，此模型的具体原理已在第 2 章给出，此处对于该具体模型进行参数调整，以提高准确率及时间空间效率。此次 LightGBM 模型主要通过表 4.4 中的参数来实现算法的优化与控制：

表 4.4 LightGBM 参数设置

参数	作用
min_child_weight	子节点中样本权重之和，用于降低模型复杂度
colsample_bytree	特征采样，用于对样本随机采样
max_depth	定义树最大深度，用于降低模型复杂度
subsample	对样本随机采样，用于降低过拟合
reg_alpha	L1 正则化参数，用于降低过拟合
reg_lambda	L2 正则化参数，用于降低过拟合
min_child_samples	子节点中样本采样，用于调整树深度

本次使用 Python 的 bayes_opt 库对表 4.4 中参数进行优化，以 roc-auc 为评价指标进行迭代，具体结果如附表 A7 所示，可视化结果如下图所示：

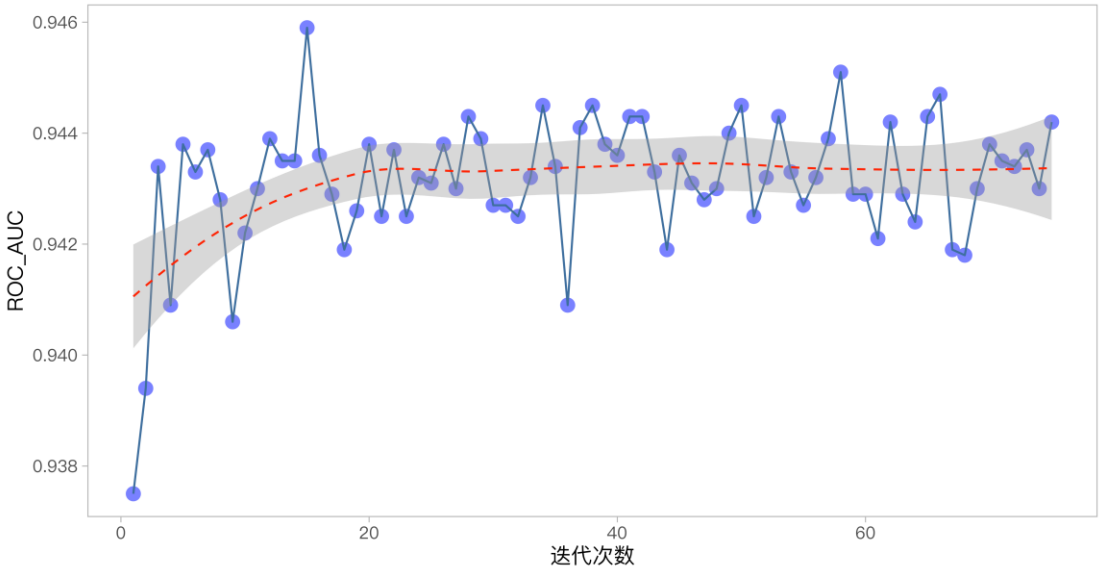

图 4.4 贝叶斯优化迭代结果

根据图 4.4 中可以明显看出第 15 次迭代使评估指标 roc-auc 达到近 75 次的最大大值，为 94.59%。之后出现若干极值点，但之后出现的极大值均小于最大值，且随着极值点的增大，极值随定值上下波动。故选取第 15 次迭代参数以用于分类器的参数调整中。参数如下表所示：

表 4.5 调整后分类器参数

参数名称	数值
colsample_bytree	0.9192087192313044
max_depth	14.519322397355996
min_child_samples	29.133726881249675
min_child_weight	3.031832773419597
reg_alpha	0.45886331874822717
reg_lambda	0.5739141154282311
subsample	0.5953917054767892

4.2.3 模型过拟合检验

为避免模型出现过拟合现象，现根据学习曲线检视模型过拟合。学习曲线如下图所示：

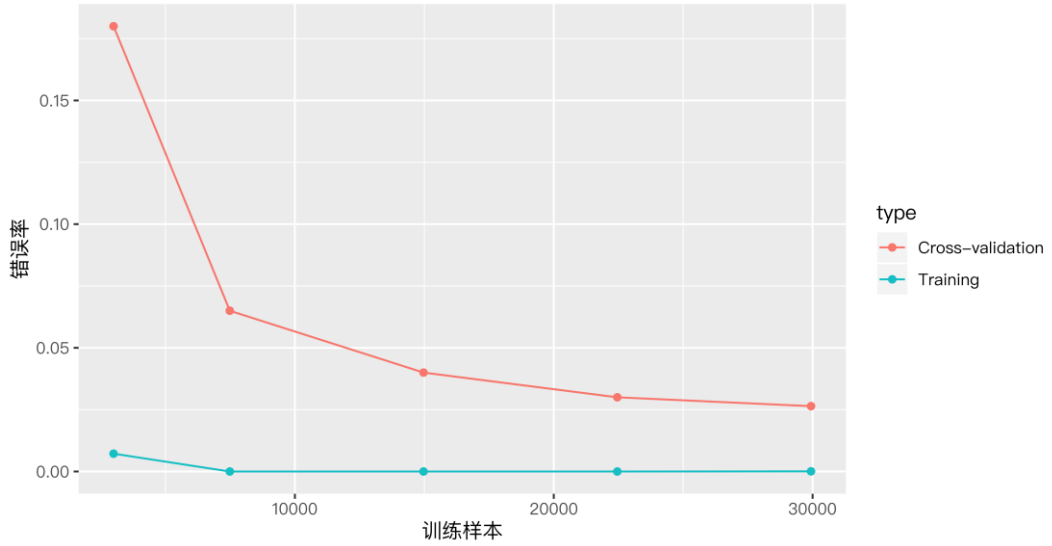

图 4.5 模型学习曲线

由图 4.5 很直观显示模型学习进度，观察样本由小到大的学习曲线变化，采用 K 折交叉验证$cv=10$，选择准确率检视模型效能，样本由小到大分成 5 轮检视学习曲线(10%，25%，50%，75%，100%)。可以看出对于测试集的错误率在逐步降低，未发生过拟合情况，说明模型可用。

4.3 结果及分析

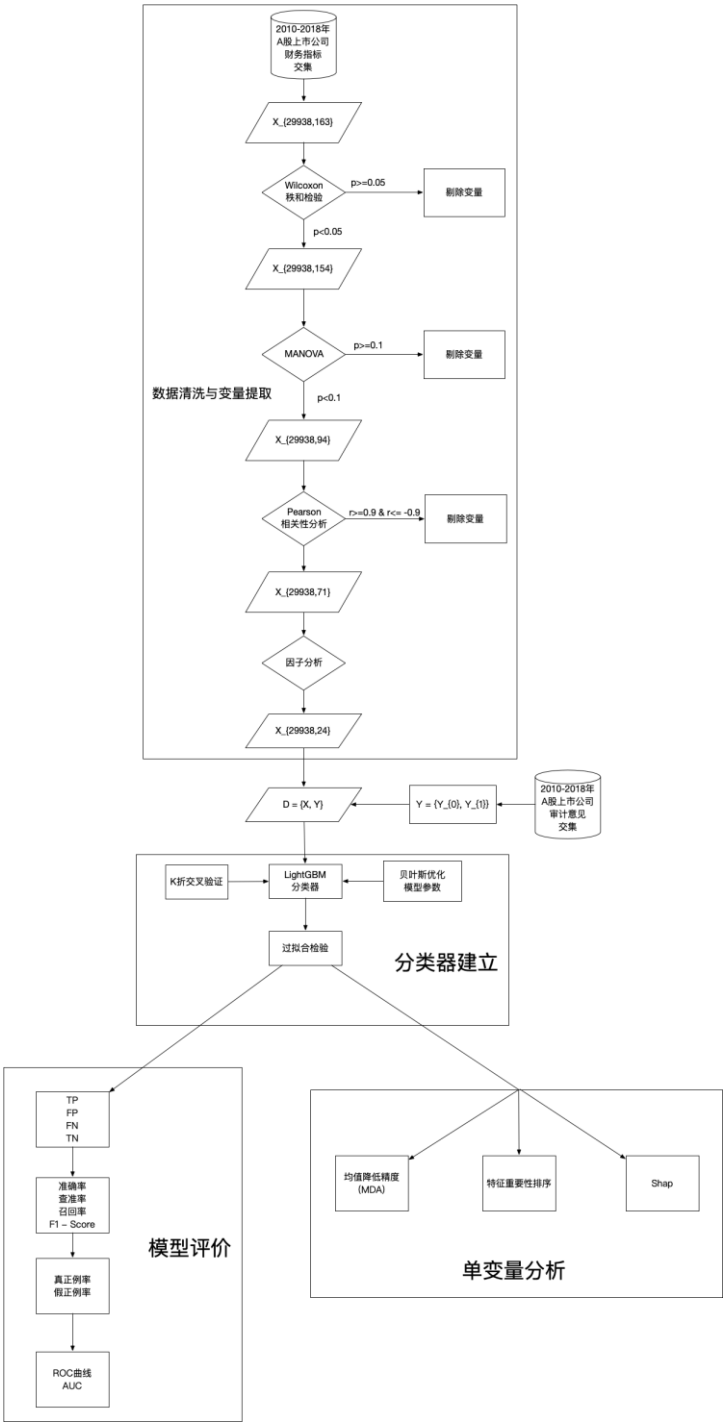

图 4.6 模型流程图

4.3.1 模型结构

根据第 2 章给出的原理，LightGBM 模型基于决策树算法。本次分类器训练后构建出决策树的可视化结果如下所示：

图 4.7 LightGBM 决策树可视化

4.3.2 模型评价

如图 4.6 所示，数据框内数据经过 Wilcoxon 秩和检验、多元方差分析、与因子分析后筛选出指标作为模型特征属性，然后将其对应样本的审计意见按标准无保留意见与否进行分类处理。模型训练完毕后，计算预测正确的正类个数 TP，预测错误的正类个数 FP，预测错误的负类个数 FN，预测正确的负类个数 TN，如下表所示：

表 4.6 LightGBM 判别预测准确率

观测值		预测值		
		异常类型		准确率（%）
		$Y_i = 0$	$Y_i = 1$	
异常类型	$Y_i = 0$	$TP = 28947$	$FP = 0$	100
	$Y_i = 1$	$FN = 14$	$TN = 977$	98.59
整体准确率		$TPR = 99.95$	$FPR = 0$	99.95

通过对样本进行上述计算可以发现，最终 LightGBM 分类器的准确率为：

$$ACC = \frac{TP + TN}{P + N} = 99.95\%$$

查准率为：

$$PRE = \frac{TP}{TP + FP} = 100\%$$

召回率为：

$$REC = \frac{TP}{TP + FN} = 98.59\%$$

现计算查准率与召回率的调和平均数 F1 值：

$$F_1 - Score = \frac{2}{\frac{1}{Precison} + \frac{1}{Recall}} = 99.47\%$$

其中，TP 为预测正确的正类个数，FP 为预测错误的正类个数，FN 为预测错误的负类个数，TN 为预测正确的负类个数。

以上检验指标说明 LightGBM 分类模型对于两种样本识别均较好，错误率在可接受范围内。现对数据框进行随机抽样，抽取训练集:测试集 = 8:2，根据不同分类阈值和每个样本属于正类或负类概率大小变化画出接收者操作特征曲线（ROC）：

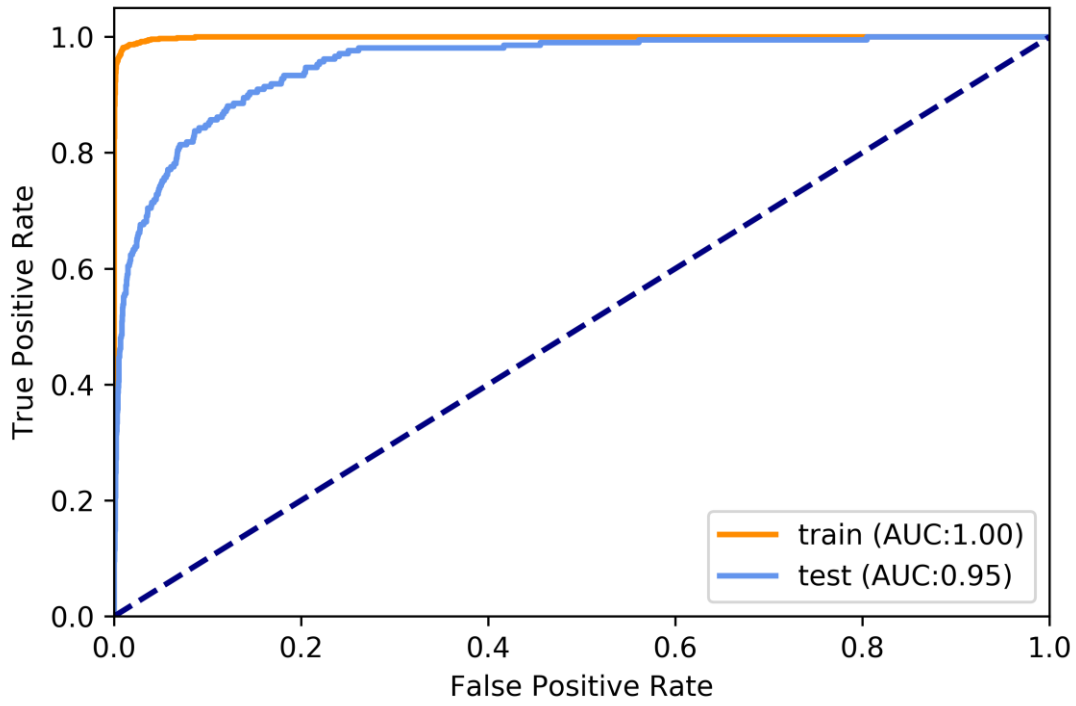

图 4.8 ROC-AUC 曲线

ROC-AUC 曲线是在各种阈值设置下针对分类问题的性能度量。ROC 是概率曲线，AUC 表示可分离性的程度或度量。它告诉我们多少模型能够区分类。AUC 越高，模型在将 0 预测为 0 和将 1 预测为 1 时越好。以此类推，AUC 越高，该模型在区分舞弊和非舞弊的样本方面表现越好。出色模型的 AUC 接近 1，这意味着它具有良好的可分离性度量。较差的模型的 AUC 接近于 0，这意味着它的可分离性度量最差。实际上，这意味着模型正在偏离结果，将其预测 0 为 1，1 为 0。当 AUC 为 0.5 时，表示模型没有任何类别分离能力。

4.3.3 变量分析

4.3.3.1 均值降低精度

均值降低精度又称排列重要性（MDA），通过查看特征不可用时，评价指标得分降低值来衡量该特征的重要性。为此，该算法从数据集中删除特征，重新训练估算器并检查评价指标数值。本次训练使用 Python 的 eli5 库，估算指标选用 F1 值，误差线以标准差形式给出，结果如下图所示：

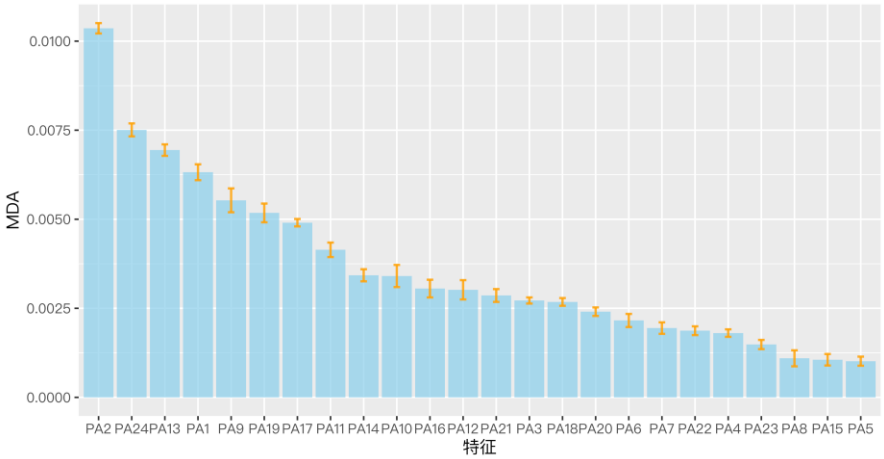

图 4.9 均值降低精度

4.3.3.2 特征重要性

在模型的训练与使用中，特征重要性默认体现为该特征在模型中使用的次数，该特征使用次数越高，则其越重要，反之亦然。特征在模型中的使用次数如下图所示：

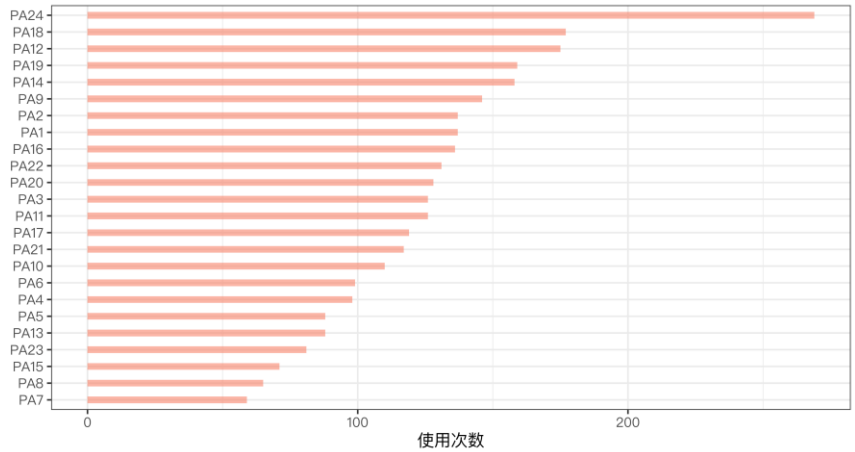

图 4.10 特征重要性排序

4.3.3.3 边际贡献加权平均值

对于机器学习模型的特征值的影响，可以采用边际贡献加权平均值（*Shap*）来表示，具体算法请参阅参考文献。本次计算采用 Python 的 Shap 库，模型全部特征的*Shap*解释能力如下图所示：

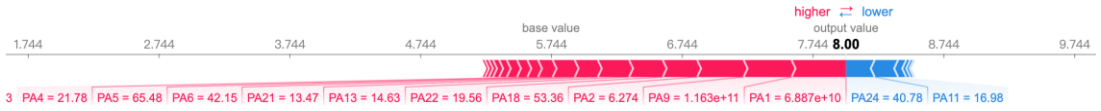

图 4.11 特征 Shape 值

其中，每个*Shap*值是一个箭头，正值为增加预测，负值为减少预测，基线为平均预测值8.00，指标全部为正，即全部指标均能增加对预测的影响。

在 4.3.3.2 章中，特征重要性由训练过程中，使用特征次数越高，特征重要性越高，反之。在本节采用*Shap*值重新计算特征重要程度，从不同角度综合评价特征。对特征重要度降序排序并进行可视化，特征*Shap*值重要性如下图所示：

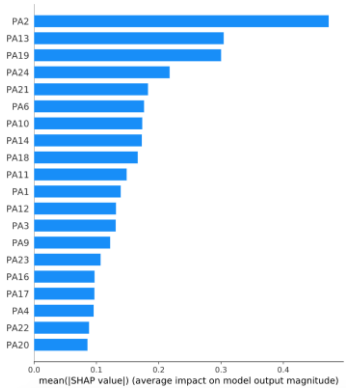

图 4.12 特征 Shape 值重要性排序

结果显示，*PA2*指标是最重要的特征，平均将预测绝对值改变6.18%。

下图为*Shap*摘要图，摘要图结合了特征重要度和特征的影响，其每点都是对应特征和观测的*Shap*值，*Shap*值影响其所在x轴位置，特征影响其所在y轴位置，颜色代表特征值从小到大，重叠点在y轴方向上抖动，因此，此图可以了解每个特征*Shap*值的分布：

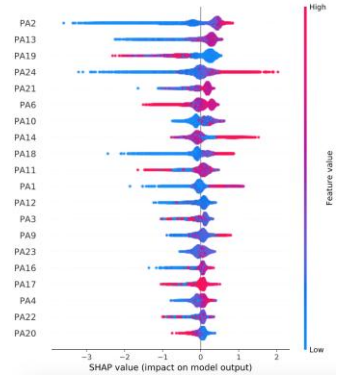

图 4.13 特征 Shape 值分布

5 结论

本文通过对我国 2010 年至 2018 年沪深 A 股上市公司财务舞弊的分析，反映出上市公司存在财务舞弊状况,发现有 999 家上市公司财务报表存在各种问题,并利用 LightGBM 机器学习模型对上述上市公司进行识别。通过研究发现：（1）上市公司财务舞弊具有很强的行业聚集性。由于三去一降一补,供给侧结构性改革等政策因素与市场主导等问题，在研究样本中制造业上市公司，尤其是化石原材料等产业链上游企业是财务舞弊的重灾区，而批发零售、房地产与高科技信息技术等行业则是由于市场估值过高导致管理层有舞弊动机；（2）根据变量重要性分析得出，加权综合构成的每股指标与现金流分析指标对于判断财务舞弊具有较强的先导性。企业在进行财务舞弊时必将通过粉饰财务报表影响会计信息使用者进行判断，进而影响到各财务指标，虚假的人为干预会计操作在企业现金流与每股息税前利润中体现较为明显，这也与现实情况比较相符。但为避免动作被发现，企业会尽力粉饰上述两项，但无论企业如何操作，其真实情况会在本文分析的 24 种综合加权指标中有所体现；（3）模型具有较好的识别效果，简单易行。本文采取 Wilcoxon 秩和检验、MANOVA 多元方差分析、因子分析等手段人工构造出 24 个相互独立的特征,避免多重共线性对模型精度带来的影响。并用贝叶斯优化调参提高模型精确度，对 2010 年至 2018 年 29938 条观测进行分类，采用 10 重交叉验证对模型的识别性能进行评判。结果显示:准确率平均达到 99.95%,查准率平均高达 100%,召回率平均为 98.59%，F1 值达到 99.47%，训练集 AUC 值为 1.0，测试集 AUC 值为 0.95。

综合考虑，本文存在以下 3 点不足：（1）对于财务舞弊分类的缺乏。由于需要保证模型的精确度与泛化能力，本文将多分类问题改为二分类问题，未能考虑各舞弊情况带来不同广泛性与重大性影响；（2）模型参数的统计推断。由于机器学习模型非传统意义的统计学模型，无法找出一个确定的统计量，对其参数作假设检验等严格的数学分析，针对这一问题本文以计算机与工程学界公允的 4 种指标对模型整体进行评价，并用 3 种方法评估变量的重要性程度；（3）没有考虑定性数据。本文研究财务舞弊更多选用定量数据，没有考虑公司治理结构、审计报告中自然语言部分、关于公司的新闻等定性数据分析。若将上述部分囊括于模型中，需要考虑 NLP 算法的构建与应用。但经前期小规模测试后发现此类非结构化数据难以进行统一的数据清洗与处理，故考虑剔除该类数据。

本文依照第 1.2 章文献研究方向,提出了 3 点问题,并做出了进一步的研究与改进：（1）模型分类结果差。可以看出定量研究多采用经典统计学模型，其模型大多严格遵循数学先期假设条件，虽然其可解释程度、因果关系判别、有效度优于当今在计算机领域使用较多的深度学习等模型，但囿于其本身经典假设与严格证明，使其更关心模型方法背后的数学正确性，导致理论虽十分优美，但大多统计学模型均对实际问题做假设与简化，有时在实际问题中并不成立，欠缺对于实际应用的价值；（2）指标过少致使有偏估计。虽然有偏估计在某些环境下显著优于无偏估计，但从随机抽样的角度分析，有偏估计会使样本数据分布偏离总体数据分布，进而造成其重要特征被计算在误差中，最后产生模型效果不佳的问题。但特征变量过多则会导致维度灾难的产生，由此引起过拟合现

象。由此可见，这也是上述研究需要改进之处；（3）样本数据量过少。众所周知，假阳性与高估的效应量是小样本实验很自然的结果。从长远看（即无数次抽样），随着样本量的增加，抽样分布呈现更窄的趋势。这意味着从长远看，样本数据量越大，其估计越精确，反之亦然。另外，小样本量的信度与效度会使相关估计带来不确定性，导致效应很可能无法被重复,也即构建出的模型不能够对同分布的样本外数据进行效果较高的分类，不具有泛化性,不能够很好的反应总体特征,不能够对之后的真实情况作出指导性判断。

本文针对上述提出的 3 点问题提出以下 3 点创新：（1）使用机器学习模型。本文使用 LightGBM 机器学习模型作为分类器对样本进行训练，经上述研究发现模型分类效果有了显著地提升，进而使得模型在实际工业应用与模型给出结果解释可靠程度均大幅提升；（2）尽最大可能选取财务指标。本文从国泰安财务指标分析数据库中选取 2010 年至 2018 年沪深两市全部 A 股上市公司，并能够精确计算 7 类 163 个财务指标的公司交集作为样本特征值,选取国泰安财务报告审计意见数据库对应数据作为标签初步组成数据框。在尝试 Lillie 正态性检验时，发现指标全部未通过，后调整方向选用非参数检验的 Wilcoxon 秩和检验进行初步筛选，并依次使用 MANOVA 多元方差分析、Pearson 相关性分析、因子分析等降维手段，使用正交旋转后因子荷载矩阵构建出 7 类 24 个新指标，尽可能在经典统计学框架下将多重共线的影响降至最低，以此真实反映总体特征；（3）选取 9 年沪深 A 股全部符合标准公司作为样本。本文选取共 29,938 条观测，173 种特征，1 种标签，合计 5,209,212 个数据进行分析。对于样本中缺失较少指标的观测没有进行剔除，而使用 KNN 最邻近算法分行业对缺失值进行填补，尽最大可能确保数据没有发生较大偏移，不影响其准确程度。因本人使用的计算机性能与模型复杂度等多种因素限制，此数据量为在没有搭建 Hadoop 等分布式系统中能够自动优化分析的最优数据量。除解决以上问题外，基于 LightGBM 机器学习模型最大优点为简单易行，使用者只需对获得的会计信息按财务管理中规定公式进行标准化计算，模型选取管道可以自动处理指标后传递给分类器；企业可以按此流程由后续数据的输入自动学习，自动分类结果，并自动给出相应解释，对企业自身进行自适应，减少使用者依据个人筛选指标经验给出较为主观的指标选择判断，以此来辅助审计从业人员与监督管理人员进行初步判断，从而缩减其工作量。

通过对于上市公司财务报表舞弊识别的研究与模型构建，本文主要有以下意见和建议：一是加强公司外部监督。为健全我国宏观调控体系，防范行业性、系统性风险，稳定市场预期,应作好事先监督工作。但由于财务舞弊需要专业人士通过一定时间来揭露，因此需要加强外部监督，同时提高专业人员的识别能力。随着科技的不断发展与提高，监管部门也应辅助使用科技手段，更多应用大数据与人工智能等方案的便利性，辅助进行识别与评判。对于发生会计舞弊的企业，应强化管理、加大制裁、严厉打击、绝不容忍。二是强化公司内部控制。企业应加强内部管理、深化落实内部控制，做到事前、事中、事后内部留有痕迹，采取风险管控与职责分离措施，真正做到提高经营效率，充分有效获取与使用资源，达到既定管理目标，不能让内部控制流于形式化，懒作为，不作为等情况发生。

综上所述，财务舞弊研究对于企业与国家都是具有重要意义的，随着企业信息管理

系统的不断完善，互联网、大数据与人工智能的不断发展，企业财务信息会更全面化、透明化，与此同时也会在整个管理与运营体系中形成全面的财务信息库。而企业会计电算化的发展与手工化的去除降低了人工造假的概率，同时也方便会计信息使用者评判企业发展，预估企业前景。但由于数据挖掘与其相关产业发展，数据量的暴增与无用信息的倍增也会给财务舞弊研究带来新的困难与挑战。

参考文献

1.敖世友.审计学.四川大学出版社，2017.
2.方军雄.我国上市公司财务欺诈鉴别的实证研究.上市公司，2003（04）.
3.顾宁生.基于 LVQ 神经网络的财务舞弊识别模型实证研究.价值工程，2009（10）.
4.贺颖.基于偏最小二乘法——支持向量机的上市公司财务舞弊识别模型研究.石河子大学，2010.
5.李安.基于机器学习 LightGBM 和异质集成学习方法的新闻分类.电子制作，2019（04）.
6.刘君.基于概率神经网络的财务舞弊识别模型.哈尔滨商业大学学报（社会科学版），2006（03）.
7.刘佳进.优化神经网络输入指标提高财报舞弊识别效果.南京理工大学学报(社会科学版)，2019（32）.
8.刘红云.高级心理统计.中国人民大学出版社，2019.
9.南东亮.基于消息队列的 LightGBM 超参数优化.计算机工程与科学，2019（08）.
10.田金玉.基于 BP 神经网络的上市公司审计意见预测模型.财会月刊，2010（03）.
11.王继美.上市公司财务报表舞弊特征的识别研究.山东农业大学，2014.
12.魏绒.基于数据挖掘的上市公司财务舞弊识别模式研究.西安石油大学，2013.
13.夏明.基于神经网络组合模型的会计舞弊识别.统计与决策，2015（16）.
14.谢勇.基于 Xgboost 和 LightGBM 算法预测住房月租金的应用分析.计算机应用与软件，2019（09）.
15.张丽红.上市公司财务舞弊识别实证分析.财会通讯，2010（32）.
16.ALBRECHT W S.Iconic.Fraud Triangle Endures.Fraud Magazine，2014（29）.
17.BREIMAN L.Random Forests.Machine Learning，Springer，2001（45）.
18.FAWCETT T.An Introduction to ROC Analysis.Pattern Recognition Letters，El-sevier，2006（27）.
19.FERNÁNDEZ-GÁMEZ M.Integrating Corporate Governance and Financial Variables for the Identification of Qualified Audit Opinions with Neural Networks.Neural Computing and Applications，Springer，2016（27）.
20.HILL B D.Sequential Kaiser-Meyer-Olkin Procedure as an Alternative for De-termining the Number of Factors in Common-Factor Analysis: A Monte Carlo Simulation.Oklahoma State University，2011.

致谢

感谢孙景翠老师对本论文的指导，感谢东北农业大学四年的教育培养，感谢父母多年的陪伴与鼓励。

附录 A

表 A1 财务指标分类表

序号	指标类别	指标名称	字母	计算公式
1		流动比率	X_{101}^{1}	流动资产/流动负债
2		速动比率	X_{102}	速动资产/流动负债
3		保守速动比率	X_{103}	保守速动资产[2]/流动负债
4		利息费用	X_{104}	利息费用
5		无息流动负债	X_{105}	无息流动负债
6		无息非流动负债	X_{106}	无息非流动负债
7		带息债务	X_{107}	带息债务
8		净债务	X_{108}	净债务
9		有形资产[3]	X_{109}	有形资产
10	1、偿债能力	营运资金[4]	X_{110}	营运资金
11		营运流动资本	X_{111}	营运流动资本
12		全部投入资本	X_{112}	全部投入资本
13		留存收益	X_{113}	留存收益
14		资产负债率	X_{114}	负债总额/资产总额
15		权益乘数	X_{115}	资产总额/股东权益总额
16		权益乘数[4]	X_{116}	净资产收益率/总资产净利率
17		产权比率	X_{117}	负债总额/所有者权益总额

[1]表格中字母列的下标与指标类别的序号对应

[2]保守速动资产=现金+短期证券+应收账款净额

[3]此有形资产=资产总额-无形资产-商誉净值

[4]此权益乘数为杜邦分析法中拆分指标

续表：

序号	指标类别	指标名称	字母	计算公式
18		归属于母公司的股东权益/负债合计	X_{118}	归属于母公司的股东权益/负债合计
19		归属于母公司的股东权益/带息债务	X_{119}	归属于母公司的股东权益/带息债务
20		有形资产/负债合计	X_{120}	有形资产/负债合计
21		有形资产/带息债务	X_{121}	有形资产/带息债务
22		有形资产/净债务	X_{122}	有形资产/净债务
23		经营活动产生的现金流量净额/负债合计	X_{123}	经营活动产生的现金流量净额/负债合计
24		经营活动产生的现金流量净额/带息债务	X_{124}	经营活动产生的现金流量净额/带息债务
25		经营活动产生的现金流量净额/净债务	X_{125}	经营活动产生的现金流量净额/净债务
26		已获利息倍数	X_{126}	$EBIT$[6]/利息费用
27		长期债务与营运资金比率	X_{127}	长期债务/营运资金
28		息税折旧摊销前利润/负债合计	X_{128}	息税折旧摊销前利润/负债合计
29		货币资金 / 流动负债	X_{129}	货币资金 / 流动负债
30		货币资金 / 带息流动负债	X_{130}	货币资金 / 带息流动负债
31		营业利润 / 流动负债	X_{131}	营业利润 / 流动负债
32		营业利润 / 负债合计	X_{132}	营业利润 / 负债合计
33		净利润/营业总收入	X_{201}	净利润/营业总收入
34		销售费用/营业总收入	X_{202}	销售费用/营业总收入
35	2、产权结构	管理费用/营业总收入	X_{203}	管理费用/营业总收入
36		财务费用/营业总收入	X_{204}	财务费用/营业总收入
37		资产减值损失/营业总收入	X_{205}	资产减值损失/营业总收入

[5]营运资金=流动资产-流动负债

[6]EBIT＝净利润＋所得税＋利息

续表：

序号	指标类别	指标名称	字母	计算公式
38		营业总成本/营业总收入	X_{206}	营业总成本/营业总收入
39		营业利润率	X_{207}	营业利润/营业总收入
40		息税前利润[7]/营业总收入	X_{208}	息税前利润/营业总收入
41		总资产报酬率	X_{209}	总资产报酬率
42		总资产净利润	X_{210}	总资产净利润
43		投入资本回报率	X_{211}	投入资本回报率
44		年化净资产收益率	X_{212}	年化净资产收益率
45		年化总资产报酬率	X_{213}	年化总资产报酬率
46		平均净资产收益率[8]	X_{214}	净利润/平均股东权益
47		经营活动净收益占比	X_{215}	经营活动净收益/利润总额
48		价值变动净收益占比	X_{216}	价值变动净收益/利润总额
49		营业外收支占比	X_{217}	营业外收支净额/利润总额
50		所得税占比	X_{218}	所得税/利润总额
51		净利润占比	X_{219}	扣除非经常损益后的净利润/净利润
52		销售商品提供劳务收到的现金占比	X_{220}	销售商品提供劳务收到的现金/营业收入
53		经营活动产生的现金流量净额占比	X_{221}	经营活动产生的现金流量净额/营业收入
54		经营活动产生的现金流量净额占比	X_{222}	经营活动产生的现金流量净额/经营活动净收益
55		资本支出占比	X_{223}	资本支出/折旧和摊销

[7]息税前利润=净利润＋所得税费用＋财务费用

[8]此为股票增发条件

续表：

序号	指标类别	指标名称	字母	计算公式
56		流动资产占比	X_{224}	流动资产/总资产
57		非流动资产占比	X_{225}	非流动资产/总资产
58		有形资产占比	X_{226}	有形资产/总资产
59		带息债务占比	X_{227}	带息债务/全部投入资本
60		归属于母公司的股东权益占比	X_{228}	归属于母公司的股东权益/全部投入资本
61		流动负债占比	X_{229}	流动负债/负债合计
62		非流动负债占比	X_{230}	非流动负债/负债合计
63		经营活动产生的现金流量净额占比	X_{231}	经营活动产生的现金流量净额/流动负债
64		固定资产合计	X_{232}	固定资产合计
65		非经常性损益	X_{301}	非经常性损益
66		扣除非经常性损益后的净利润	X_{302}	扣除非经常性损益后的净利润
67		毛利	X_{303}	毛利
68		经营活动净收益	X_{304}	经营活动净收益
69		价值变动净收益	X_{305}	价值变动净收益
70		折旧与摊销	X_{306}	折旧与摊销
71		息税前利润	X_{307}	息税前利润
72	3、盈利能力	息税折旧摊销前利润	X_{308}	息税折旧摊销前利润
73		销售净利率	X_{309}	净利润/销售收入
74		销售毛利率	X_{310}	毛利润/销售收入
75		销售成本率	X_{311}	销售成本/销售收入
76		销售期间费用率	X_{312}	期间费用[9]/销售收入
77		净资产收益率	X_{313}	净利润/股东权益
78		加权平均净资产收益率	X_{314}	净利润/平均所有者权益

[9]期间费用=管理费用＋销售费用＋财务费用

[10]此净资产收益率扣除加权损失

续表：

序号	指标类别	指标名称	字母	计算公式
79		净资产收益率[10]	X_{315}	净利润/股东权益
80		年化总资产净利率	X_{316}	净利润/平均资产总额[11]
81		扣除财务费用前营业利润	X_{317}	扣除财务费用前营业利润
82		总资产净利率[12]	X_{318}	净资产收益率/权益乘数
83		非营业利润	X_{319}	非营业利润
84		营业利润占比	X_{320}	营业利润 / 利润总额
85		非营业利润占比	X_{321}	非营业利润 / 利润总额
86		经营活动产生的现金流量净额 / 营业利润	X_{322}	经营活动产生的现金流量净额 / 营业利润
87		年化投入资本回报率	X_{323}	年化投入资本回报率
88		利润总额 / 营业收入	X_{324}	利润总额 / 营业收入
89		经营活动单季度净收益	X_{325}	经营活动单季度净收益
90		价值变动单季度净收益	X_{326}	价值变动单季度净收益
91		扣除非经常损益后的单季度净利润	X_{327}	扣除非经常损益后的单季度净利润
92		销售净利率[13]	X_{328}	销售净利率
93		销售毛利率[13]	X_{329}	销售毛利率
94		销售期间费用率[13]	X_{330}	销售期间费用率

[11]平均资产总额=(年初资产总额+年末资产总额)/2

[12]此总资产净利率为杜邦分析拆分指标

[13]上述指标均为单季度指标

续表：

序号	指标类别	指标名称	字母	计算公式
95		净利润占比[13]	X_{331}	净利润 / 营业总收入
96		销售费用占比[13]	X_{332}	销售费用 / 营业总收入
97		管理费用占比[13]	X_{333}	管理费用 / 营业总收入
98		财务费用占比[13]	X_{334}	财务费用 / 营业总收入
99		资产减值损失占比[13]	X_{335}	资产减值损失 / 营业总收入
100		营业总成本占比[13]	X_{336}	营业总成本 / 营业总收入
101		营业利润占比[13]	X_{337}	营业利润 / 营业总收入
102		净资产收益率[13]	X_{338}	净利润/股东权益
103		净资产收益率[14]	X_{339}	净资产单季度收益率（扣除非经常损益）
104		总资产净利润率[13]	X_{340}	总资产净利润
105		经营活动净收益占比[13]	X_{341}	经营活动净收益 / 利润总额
106		价值变动净收益占比[13]	X_{342}	价值变动净收益 / 利润总额
107		扣除非经常损益后的净利润占比[13]	X_{343}	扣除非经常损益后的净利润 / 净利润
		营业利润同比增长率	X_{401}	Δ营业利润/上期营业利润

[14]此净资产单季度收益率扣除非经常损益

续表：

序号	指标类别	指标名称	字母	计算公式
108		营业利润同比增长率	X_{401}	Δ营业利润/上期营业利润
109		利润总额同比增长率	X_{402}	Δ利润总额/上期利润总额
110		归属母公司股东的净利润同比增长率	X_{403}	Δ归属母公司股东的净利润/上期归属母公司股东的净利润
111		归属母公司股东的净利润同比增长率[15]	X_{404}	Δ归属母公司股东的净利润/上期归属母公司股东的净利润-Δ非经常损益/期初非经常损益
112		经营活动净现金流量同比增长率	X_{405}	Δ经营活动净现金流量/上期经营活动净现金流量
113	4、成长能力	净资产收益率同比增长率[16]	X_{406}	Δ净利润/本期期末净资产
114		每股净资产相对年初增长率	X_{407}	Δ每股净资产/本期期初每股净资产
115		资产总计相对年初增长率	X_{408}	Δ资产总额/本期期初资产总额
116		归属母公司的股东权益相对年初增长率	X_{409}	Δ归属母公司的股东权益/本期期初归属母公司的股东权益
117		营业总收入同比增长率	X_{410}	Δ营业总收入/本期期初营业总收入
118		营业收入同比增长率	X_{411}	Δ营业收入/本期期初营业收入
119		营业总收入同比增长率[13]	X_{412}	Δ营业总收入/本期期初营业总收入

[15]此归属母公司股东的净利润同比增长率扣除非经常损益

[16]此指标为摊薄净资产收益率，强调年末状况，为静态指标

续表：

序号	指标类别	指标名称	字母	计算公式
120		营业总收入环比增长率[13]	X_{413}	Δ营业总收入/上季度营业总收入
121		营业收入同比增长率[13]	X_{414}	Δ营业收入/上期营业收入
122		营业收入环比增长率[13]	X_{415}	Δ营业收入/上季度营业收入
123		营业利润同比增长率[13]	X_{416}	Δ营业利润/上期营业利润
124		营业利润环比增长率[13]	X_{417}	Δ营业利润/上季度营业利润
125		净利润同比增长率[13]	X_{418}	Δ净利润/上期净利润
126		净利润环比增长率[13]	X_{419}	Δ净利润/上季度净利润
127		归属母公司股东的净利润同比增长率[13]	X_{420}	Δ归属母公司股东的净利润/上期归属母公司股东的净利润
128		归属母公司股东的净利润环比增长率[13]	X_{421}	Δ归属母公司股东的净利润/上季度归属母公司股东的净利润
129		净资产同比增长率	X_{422}	Δ净资产/上期净资产
130		研发费用	X_{423}	研发费用
131		存货周转天数	X_{501}	存货周转天数
132		应收账款周转天数	X_{502}	应收账款周转天数
133	5、经营能力	存货周转率	X_{503}	营业成本/存货净额平均余额
134		应收账款周转率	X_{504}	营业成本/应付账款净额平均余额

续表：

序号	指标类别	指标名称	字母	计算公式
135		流动资产周转率	X_{505}	营业收入/流动资产平均净额
136		固定资产周转率	X_{506}	营业收入/固定资产平均净额
137		总资产周转率	X_{507}	营业收入/平均资产总额
138		营业周期	X_{508}	营业周期
139		固定资产合计周转率	X_{509}	销售收入/固定资产净值合计
140		基本每股收益	X_{601}	基本每股收益
141		稀释每股收益	X_{602}	稀释每股收益
142		每股营业总收入	X_{603}	每股营业总收入
143		每股营业收入	X_{604}	每股营业收入
144		每股资本公积	X_{605}	每股资本公积
145		每股盈余公积	X_{606}	每股盈余公积
146		每股未分配利润	X_{607}	每股未分配利润
147		期末摊薄每股收益	X_{608}	期末摊薄每股收益
148		每股净资产	X_{609}	每股净资产
149	6、每股指标	每股经营活动产生的现金流量净额	X_{610}	经营活动现金净流量/总股本
150		每股留存收益[17]	X_{611}	留存收益/总股本
151		每股现金流量净额	X_{612}	现金流量净额/总股本
152		每股息税前利润	X_{613}	息税前利润/总股本
153		每股企业自由现金流量[18]	X_{614}	企业自由现金流量/总股本

[17]留存收益=盈余公积+未分配利润

[18]企业自由现金流量=息税前利润+折旧-所得税-资本性支出-营运资本净增加

[19]股权自由现金流量=(利润总额+利息费用)×(1-税率)-净投资-税后利息费用+债务净增加

续表：

序号	指标类别	指标名称	字母	计算公式
154		每股股权自由现金流量[19]	X_{615}	股权自由现金流量/总股本
155		每股收益[13]	X_{616}	每股收益
156		每股收益同比增长率	X_{617}	Δ每股收益/上期每股收益
157		每股收益同比增长率[20]	X_{618}	Δ每股收益/上期每股收益
158		每股经营活动产生的现金流量净额同比增长率	X_{619}	Δ经营活动产生的现金流量净额×总股本/上期经营活动产生的现金流量净额
159	7、现金流分析	企业自由现金流量[18]	X_{701}	企业自由现金流量
160		股权自由现金流量[19]	X_{702}	股权自由现金流量
161		销售商品提供劳务现金占比[13]	X_{703}	销售商品提供劳务收到的现金 / 营业收入
162		经营活动产生的现金流量净额占比[13]	X_{704}	经营活动产生的现金流量净额 / 营业收入
163		经营活动产生的现金流量净额占比[13]	X_{705}	经营活动产生的现金流量净额 / 经营活动净收益

[20]此指标为稀释每股收益同比增长率，即假设公司存在的上述可能转化为上市公司股权的工具都在当期全部转换为普通股股份后计算的每股收益

表 A2 2010 年至 2018 年各行业财务舞弊个数统计表

行业类别	数量总计	财务舞弊个数	占行业个数比（%）	占财务舞弊比（%）
软件服务	1507	31	2.06	3.10
电气设备	1429	34	2.38	3.40
元器件	1426	21	1.47	2.10
化工原料	1323	51	3.85	5.11
汽车配件	1104	24	2.17	2.40
专用机械	1036	9	0.87	0.90
通信设备	925	16	1.73	1.60
化学制药	840	13	1.55	1.30
建筑工程	725	14	1.93	1.40
中成药	602	15	2.49	1.50
区域地产	602	32	5.32	3.20
机械基件	598	11	1.84	1.10
互联网	591	35	5.92	3.50
环境保护	590	47	7.97	4.70
医疗保健	584	4	0.68	0.40
电器仪表	508	3	0.59	0.30
食品	497	16	3.22	1.60
服饰	477	11	2.31	1.10
半导体	437	4	0.92	0.40
影视音像	591	35	5.92	3.50
证券	590	47	7.97	4.70
百货	584	4	0.68	0.40
农药化肥	508	3	0.59	0.30
塑料	497	16	3.22	1.60
生物制药	477	11	2.31	1.10
家用电器	437	4	0.92	0.40
纺织	413	33	7.99	3.30
全国地产	387	5	1.29	0.50
广告包装	381	11	2.89	1.10
仓储物流	377	19	5.04	1.90
文教休闲	376	8	2.13	0.80
综合类	369	18	4.88	1.80
火力发电	351	17	4.84	1.70
家居用品	331	9	2.72	0.90

续表：

行业类别	数量总计	财务舞弊个数	占行业个数比（%）	占财务舞弊比（%）
银行	322	11	3.42	1.10
小金属	319	5	1.57	0.50
IT 设备	314	18	5.73	1.80
多元金融	313	26	8.31	2.60
煤炭开采	292	39	13.36	3.90
工程机械	288	9	3.13	0.90
其他建材	288	0	0.00	0.00
供气供热	287	6	2.09	0.60
农业综合	284	15	5.28	1.50
普钢	273	3	1.10	0.30
化纤	261	16	6.13	1.60
医药商业	216	2	0.93	0.20
造纸	216	8	3.70	0.80
铝	216	9	4.17	0.90
装修装饰	210	3	1.43	0.30
汽车整车	207	8	3.86	0.80
航空	200	4	2.00	0.40
钢加工	198	4	2.02	0.40
出版业	190	5	2.63	0.50
商贸代理	189	12	6.35	1.20
水泥	189	10	5.29	1.00
水力发电	189	3	1.59	0.30
港口	169	1	0.59	0.10
路桥	169	5	2.96	0.50
染料涂料	165	1	0.61	0.10
白酒	158	2	1.27	0.20
运输设备	155	2	1.29	0.20
玻璃	155	1	0.65	0.10
石油开采	154	7	4.55	0.70
矿物制品	145	2	1.38	0.20
饲料	144	7	4.86	0.70
园区开发	135	5	3.70	0.50
种植业	122	6	4.92	0.60
黄金	117	19	16.24	1.90

续表：

行业类别	数量总计	财务舞弊个数	占行业个数比（%）	占财务舞弊比（%）
铅锌	116	12	10.34	1.20
铜	110	2	1.82	0.20
水运	108	14	12.96	1.40
其他商业	108	2	1.85	0.20
旅游服务	108	5	4.63	0.50
机床制造	106	4	3.77	0.40
新型电力	105	11	10.48	1.10
旅游景点	104	1	0.96	0.10
乳制品	101	5	4.95	0.50
水务	99	3	3.03	0.30
橡胶	99	0	0.00	0.00
汽车服务	98	2	2.04	0.20
超市连锁	98	3	3.06	0.30
化工机械	90	6	6.67	0.60
红黄酒	89	2	2.25	0.20
轻工机械	89	0	0.00	0.00
纺织机械	86	2	2.33	0.20
酒店餐饮	82	11	13.41	1.10
日用化工	78	2	2.56	0.20
农用机械	76	0	0.00	0.00
电信运营	73	0	0.00	0.00
特种钢	72	2	2.78	0.20
渔业	72	5	6.94	0.50
石油加工	71	1	1.41	0.10
公共交通	70	0	0.00	0.00
空运	67	1	1.49	0.10
啤酒	63	1	1.59	0.10
石油贸易	63	2	3.17	0.20
批发业	62	9	14.52	0.90
焦炭加工	62	10	16.13	1.00
船舶	61	0	0.00	0.00
公路	60	2	3.33	0.20
房产服务	60	3	5.00	0.30
保险	59	0	0.00	0.00

续表：

行业类别	数量总计	财务舞弊个数	占行业个数比（%）	占财务舞弊比（%）
摩托车	58	3	5.17	0.30
陶瓷	51	6	11.76	0.60
软饮料	50	0	0.00	0.00
林业	36	6	16.67	0.60
铁路	36	6	16.67	0.60
机场	36	0	0.00	0.00
商品城	27	0	0.00	0.00
电器连锁	18	0	0.00	0.00

表 A3 解释变量 Wilcoxon 秩和检验

变量	W	p-value
X_{101}	20403000	< 2.2e-16
X_{102}	19543000	< 2.2e-16
X_{103}	19830000	< 2.2e-16
X_{104}	11428000	< 2.2e-16
X_{105}	14453000	0.2499
X_{106}	13123000	0.0001148
X_{107}	14220000	0.7833
X_{108}	11143000	< 2.2e-16
X_{109}	20511000	< 2.2e-16
X_{110}	20004000	< 2.2e-16
X_{111}	17682000	< 2.2e-16
X_{112}	17967000	< 2.2e-16
X_{113}	23490000	< 2.2e-16
X_{114}	7722700	< 2.2e-16
X_{115}	7769300	< 2.2e-16
X_{116}	8139700	< 2.2e-16
X_{117}	8242000	< 2.2e-16
X_{118}	20532000	< 2.2e-16
X_{119}	18929000	< 2.2e-16
X_{120}	20575000	< 2.2e-16
X_{121}	19670000	< 2.2e-16
X_{122}	21917000	< 2.2e-16
X_{123}	20249000	< 2.2e-16
X_{124}	20618000	< 2.2e-16
X_{125}	20759000	< 2.2e-16
X_{126}	23563000	< 2.2e-16
X_{127}	9538000	< 2.2e-16
X_{128}	23095000	< 2.2e-16
X_{129}	20192000	< 2.2e-16
X_{130}	19032000	< 2.2e-16
X_{131}	24136000	< 2.2e-16
X_{132}	24131000	< 2.2e-16
X_{201}	21634000	< 2.2e-16
X_{202}	15017000	0.001063
X_{203}	9709100	< 2.2e-16

续表：

变量	W	p-value
X_{204}	8617800	< 2.2e-16
X_{205}	8111000	< 2.2e-16
X_{206}	4027500	< 2.2e-16
X_{207}	23449000	< 2.2e-16
X_{208}	20822000	< 2.2e-16
X_{209}	22675000	< 2.2e-16
X_{210}	23123000	< 2.2e-16
X_{211}	22623000	< 2.2e-16
X_{212}	23230000	< 2.2e-16
X_{213}	22675000	< 2.2e-16
X_{214}	25045000	< 2.2e-16
X_{215}	23273000	< 2.2e-16
X_{216}	5982700	< 2.2e-16
X_{217}	6770300	< 2.2e-16
X_{218}	13155000	0.0001878
X_{219}	23815000	< 2.2e-16
X_{220}	13622000	0.04807
X_{221}	18634000	< 2.2e-16
X_{222}	15078000	0.0004614
X_{223}	20641000	< 2.2e-16
X_{224}	16323000	2.69E-16
X_{225}	11879000	< 2.2e-16
X_{226}	20590000	< 2.2e-16
X_{227}	10477000	< 2.2e-16
X_{228}	18203000	< 2.2e-16
X_{229}	13041000	3.12E-05
X_{230}	14582000	0.1019

续表：

变量	W	p-value
X_{231}	20299000	< 2.2e-16
X_{232}	16471000	< 2.2e-16
X_{301}	15120000	0.0002525
X_{302}	24111000	< 2.2e-16
X_{303}	21185000	< 2.2e-16
X_{304}	23815000	< 2.2e-16
X_{305}	14781000	0.01711
X_{306}	15942000	1.45E-11
X_{307}	22447000	< 2.2e-16
X_{308}	21933000	< 2.2e-16
X_{309}	21637000	< 2.2e-16
X_{310}	19128000	< 2.2e-16
X_{311}	9167400	< 2.2e-16
X_{312}	9324500	< 2.2e-16
X_{313}	23230000	< 2.2e-16
X_{314}	23046000	< 2.2e-16
X_{315}	24955000	< 2.2e-16
X_{316}	23123000	< 2.2e-16
X_{317}	23026000	< 2.2e-16
X_{318}	23011000	< 2.2e-16
X_{319}	16603000	< 2.2e-16
X_{320}	21507000	< 2.2e-16
X_{321}	6788300	< 2.2e-16
X_{322}	15191000	8.64E-05
X_{323}	22623000	< 2.2e-16
X_{324}	21579000	< 2.2e-16
X_{325}	23151000	< 2.2e-16

续表：

变量	W	p-value
X_{326}	14771000	0.01894
X_{327}	23269000	< 2.2e-16
X_{328}	20435000	< 2.2e-16
X_{329}	18852000	< 2.2e-16
X_{330}	9015400	< 2.2e-16
X_{331}	20432000	< 2.2e-16
X_{332}	14861000	0.007256
X_{333}	9003700	< 2.2e-16
X_{334}	8736300	< 2.2e-16
X_{335}	7180200	< 2.2e-16
X_{336}	4881400	< 2.2e-16
X_{337}	22321000	< 2.2e-16
X_{338}	21117000	< 2.2e-16
X_{339}	23750000	< 2.2e-16
X_{340}	20883000	< 2.2e-16
X_{341}	21863000	< 2.2e-16
X_{342}	6896000	< 2.2e-16
X_{343}	22113000	< 2.2e-16
X_{401}	17935000	< 2.2e-16
X_{402}	18008000	< 2.2e-16
X_{403}	18220000	< 2.2e-16
X_{404}	18098000	< 2.2e-16
X_{405}	15398000	2.53E-06
X_{406}	19615000	< 2.2e-16
X_{407}	19199000	< 2.2e-16
X_{408}	21584000	< 2.2e-16
X_{409}	21368000	< 2.2e-16

续表：

变量	W	p-value
X_{410}	18867000	< 2.2e-16
X_{411}	18863000	< 2.2e-16
X_{412}	18359000	< 2.2e-16
X_{413}	16140000	6.45E-14
X_{414}	18338000	< 2.2e-16
X_{415}	16146000	5.46E-14
X_{416}	18141000	< 2.2e-16
X_{417}	19868000	< 2.2e-16
X_{418}	18300000	< 2.2e-16
X_{419}	18703000	< 2.2e-16
X_{420}	18286000	< 2.2e-16
X_{421}	18630000	< 2.2e-16
X_{422}	22121000	< 2.2e-16
X_{423}	19179000	< 2.2e-16
X_{501}	13966000	0.4949
X_{502}	12398000	4.62E-11
X_{503}	13393000	0.004545
X_{504}	15191000	8.62E-05
X_{505}	17408000	< 2.2e-16
X_{506}	17313000	< 2.2e-16
X_{507}	19189000	< 2.2e-16
X_{508}	13408000	0.005378
X_{509}	18131000	< 2.2e-16
X_{601}	24515000	< 2.2e-16
X_{602}	24566000	< 2.2e-16
X_{603}	21292000	< 2.2e-16
X_{604}	21282000	< 2.2e-16

续表：

变量	W	p-value
X_{605}	16186000	1.71E-14
X_{606}	18592000	< 2.2e-16
X_{607}	24962000	< 2.2e-16
X_{608}	24514000	< 2.2e-16
X_{609}	24009000	< 2.2e-16
X_{610}	20041000	< 2.2e-16
X_{611}	24938000	< 2.2e-16
X_{612}	17374000	< 2.2e-16
X_{613}	24075000	< 2.2e-16
X_{614}	14769000	0.01938
X_{615}	16561000	< 2.2e-16
X_{616}	22425000	< 2.2e-16
X_{617}	17991000	< 2.2e-16
X_{618}	18084000	< 2.2e-16
X_{619}	15166000	0.0001261
X_{701}	15081000	0.0004448
X_{702}	16510000	< 2.2e-16
X_{703}	14321000	0.5129
X_{704}	17034000	< 2.2e-16
X_{705}	16256000	2.06E-15

表 A4 财务指标成分分析

指标序号	初始部分			提取部分		
	载荷因子	比例%	累积比例%	载荷因子	比例%	累积比例%
1	6.099	8.591	8.591	6.099	8.591	8.591
2	5.596	7.882	16.473	5.596	7.882	16.473
3	3.455	4.866	21.338	3.455	4.866	21.338
4	2.944	4.146	25.484	2.944	4.146	25.484
5	2.590	3.648	29.132	2.590	3.648	29.132
6	2.395	3.374	32.505	2.395	3.374	32.505
7	2.350	3.310	35.815	2.350	3.310	35.815
8	2.055	2.894	38.709	2.055	2.894	38.709
9	1.999	2.815	41.525	1.999	2.815	41.525
10	1.934	2.725	44.249	1.934	2.725	44.249
11	1.670	2.352	46.602	1.670	2.352	46.602
12	1.551	2.184	48.786	1.551	2.184	48.786
13	1.388	1.955	50.741	1.388	1.955	50.741
14	1.343	1.891	52.632	1.343	1.891	52.632
15	1.243	1.751	54.383	1.243	1.751	54.383
16	1.210	1.704	56.087	1.210	1.704	56.087
17	1.175	1.655	57.742	1.175	1.655	57.742
18	1.128	1.589	59.331	1.128	1.589	59.331
19	1.106	1.558	60.889	1.106	1.558	60.889
20	1.082	1.524	62.413	1.082	1.524	62.413
21	1.066	1.502	63.915	1.066	1.502	63.915
22	1.038	1.462	65.377	1.038	1.462	65.377
23	1.018	1.434	66.811	1.018	1.434	66.811
24	1.001	1.409	68.220	1.001	1.409	68.220
25	.998	1.405	69.625			
26	.992	1.398	71.023			
27	.987	1.390	72.412			

续表：

指标序号	初始部分			提取部分		
	载荷因子	比例%	累积比例%	载荷因子	比例%	累积比例%
28	.975	1.373	73.786			
29	.969	1.365	75.151			
30	.959	1.351	76.502			
31	.955	1.345	77.846			
32	.938	1.321	79.167			
33	.913	1.285	80.453			
34	.876	1.233	81.686			
35	.866	1.219	82.905			
36	.854	1.203	84.108			
37	.832	1.172	85.280			
38	.778	1.096	86.376			
39	.744	1.048	87.424			
40	.685	.964	88.388			
41	.640	.901	89.289			
42	.611	.861	90.150			
43	.593	.835	90.986			
44	.552	.777	91.763			
45	.546	.769	92.531			
46	.515	.725	93.256			
47	.504	.710	93.966			
48	.461	.649	94.615			
49	.364	.513	95.128			
50	.354	.498	95.626			
51	.345	.486	96.111			
52	.313	.441	96.552			
53	.276	.388	96.941			
54	.254	.357	97.298			

续表：

指标序号	初始部分			提取部分		
	载荷因子	比例%	累积比例%	载荷因子	比例%	累积比例%
55	.248	.350	97.648			
56	.242	.341	97.988			
57	.210	.296	98.284			
58	.192	.270	98.554			
59	.186	.262	98.816			
60	.173	.244	99.059			
61	.107	.151	99.210			
62	.093	.131	99.342			
63	.077	.109	99.450			
64	.073	.103	99.554			
65	.068	.096	99.649			
66	.055	.077	99.726			
67	.049	.069	99.796			
68	.047	.066	99.862			
69	.038	.054	99.916			
70	.034	.047	99.963			
71	.026	.037	100.000			

表 A5 解释变量正态性检验表

变量	类别			
	标准类		非标准类	
	D	p-value	D	p-value
X_{101}	0.27991	< 2.2e-16	0.30177	< 2.2e-16
X_{102}	0.29945	< 2.2e-16	0.32773	< 2.2e-16
X_{103}	0.30698	< 2.2e-16	0.34307	< 2.2e-16
X_{104}	0.37491	< 2.2e-16	0.33444	< 2.2e-16
X_{105}	0.44133	< 2.2e-16	0.39146	< 2.2e-16
X_{106}	0.45307	< 2.2e-16	0.36564	< 2.2e-16
X_{107}	0.42651	< 2.2e-16	0.39568	< 2.2e-16
X_{108}	0.36976	< 2.2e-16	0.31813	< 2.2e-16
X_{109}	0.42541	< 2.2e-16	0.25527	< 2.2e-16
X_{110}	0.36133	< 2.2e-16	0.26568	< 2.2e-16
X_{111}	0.35998	< 2.2e-16	0.27796	< 2.2e-16
X_{112}	0.42675	< 2.2e-16	0.34175	< 2.2e-16
X_{113}	0.42758	< 2.2e-16	0.16567	< 2.2e-16
X_{114}	0.030898	< 2.2e-16	0.37845	< 2.2e-16
X_{115}	0.3904	< 2.2e-16	0.4112	< 2.2e-16
X_{116}	0.45011	< 2.2e-16	0.36819	< 2.2e-16
X_{117}	0.39647	< 2.2e-16	0.40773	< 2.2e-16
X_{118}	0.28279	< 2.2e-16	0.27368	< 2.2e-16
X_{119}	0.4911	< 2.2e-16	0.51071	< 2.2e-16
X_{120}	0.2813	< 2.2e-16	0.28074	< 2.2e-16
X_{121}	0.4899	< 2.2e-16	0.51781	< 2.2e-16
X_{122}	0.4191	< 2.2e-16	0.37913	< 2.2e-16
X_{123}	0.19058	< 2.2e-16	0.40899	< 2.2e-16
X_{124}	0.49039	< 2.2e-16	0.51741	< 2.2e-16
X_{125}	0.38555	< 2.2e-16	0.25498	< 2.2e-16
X_{126}	0.44264	< 2.2e-16	0.42154	< 2.2e-16

续表：

变量	类别			
	标准类		非标准类	
	D	p-value	D	p-value
X_{127}	0.4605	< 2.2e-16	0.48491	< 2.2e-16
X_{128}	0.22502	< 2.2e-16	0.32685	< 2.2e-16
X_{129}	0.35461	< 2.2e-16	0.39285	< 2.2e-16
X_{130}	0.49072	< 2.2e-16	0.50913	< 2.2e-16
X_{131}	0.21914	< 2.2e-16	0.31487	< 2.2e-16
X_{132}	0.22083	< 2.2e-16	0.31925	< 2.2e-16
X_{201}	0.39161	< 2.2e-16	0.49387	< 2.2e-16
X_{202}	0.34485	< 2.2e-16	0.40542	< 2.2e-16
X_{203}	0.30652	< 2.2e-16	0.48233	< 2.2e-16
X_{204}	0.30299	< 2.2e-16	0.44844	< 2.2e-16
X_{205}	0.45972	< 2.2e-16	0.42845	< 2.2e-16
X_{206}	0.33258	< 2.2e-16	0.47439	< 2.2e-16
X_{207}	0.38489	< 2.2e-16	0.46936	< 2.2e-16
X_{208}	0.39593	< 2.2e-16	0.49161	< 2.2e-16
X_{209}	0.14691	< 2.2e-16	0.36857	< 2.2e-16
X_{210}	0.17232	< 2.2e-16	0.37098	< 2.2e-16
X_{211}	0.39398	< 2.2e-16	0.471	< 2.2e-16
X_{212}	0.34819	< 2.2e-16	0.38478	< 2.2e-16
X_{213}	0.14691	< 2.2e-16	0.36857	< 2.2e-16
X_{214}	0.26096	< 2.2e-16	0.41572	< 2.2e-16
X_{215}	0.44087	< 2.2e-16	0.4196	< 2.2e-16
X_{216}	0.4094	< 2.2e-16	0.25259	< 2.2e-16
X_{217}	0.45582	< 2.2e-16	0.43318	< 2.2e-16
X_{218}	0.42386	< 2.2e-16	0.43283	< 2.2e-16
X_{219}	0.42484	< 2.2e-16	0.19099	< 2.2e-16

续表：

变量	类别			
	标准类		非标准类	
	D	p-value	D	p-value
X_{220}	0.29812	< 2.2e-16	0.43057	< 2.2e-16
X_{221}	0.41307	< 2.2e-16	0.46528	< 2.2e-16
X_{222}	0.44039	< 2.2e-16	0.27706	< 2.2e-16
X_{223}	0.3755	< 2.2e-16	0.41558	< 2.2e-16
X_{224}	0.03758	< 2.2e-16	0.037814	0.0022
X_{225}	0.037591	< 2.2e-16	0.037403	0.002607
X_{226}	0.019253	< 2.2e-16	0.37212	< 2.2e-16
X_{227}	0.14521	< 2.2e-16	0.32943	< 2.2e-16
X_{228}	0.12367	< 2.2e-16	0.30103	< 2.2e-16
X_{229}	0.17952	< 2.2e-16	0.21221	< 2.2e-16
X_{230}	0.16452	< 2.2e-16	0.18452	< 2.2e-16
X_{231}	0.19805	< 2.2e-16	0.40273	< 2.2e-16
X_{232}	0.44202	< 2.2e-16	0.37404	< 2.2e-16
X_{301}	0.38383	< 2.2e-16	0.32959	< 2.2e-16
X_{302}	0.38074	< 2.2e-16	0.27997	< 2.2e-16
X_{303}	0.44689	< 2.2e-16	0.33952	< 2.2e-16
X_{304}	0.38452	< 2.2e-16	0.27789	< 2.2e-16
X_{305}	0.41865	< 2.2e-16	0.36394	< 2.2e-16
X_{306}	0.46488	< 2.2e-16	0.34717	< 2.2e-16
X_{307}	0.39872	< 2.2e-16	0.26197	< 2.2e-16
X_{308}	0.42546	< 2.2e-16	0.25628	< 2.2e-16
X_{309}	0.39233	< 2.2e-16	0.4936	< 2.2e-16
X_{310}	0.070415	< 2.2e-16	0.098936	< 2.2e-16
X_{311}	0.070415	< 2.2e-16	0.098936	< 2.2e-16
X_{312}	0.30666	< 2.2e-16	0.47744	< 2.2e-16

续表：

变量	类别			
	标准类		非标准类	
	D	p-value	D	p-value
X_{313}	0.34819	< 2.2e-16	0.38478	< 2.2e-16
X_{314}	0.24893	< 2.2e-16	0.40136	< 2.2e-16
X_{315}	0.29405	< 2.2e-16	0.35562	< 2.2e-16
X_{316}	0.17232	< 2.2e-16	0.37098	< 2.2e-16
X_{317}	0.3982	< 2.2e-16	0.24837	< 2.2e-16
X_{318}	0.1813	< 2.2e-16	0.37143	< 2.2e-16
X_{319}	0.39642	< 2.2e-16	0.33153	< 2.2e-16
X_{320}	0.45583	< 2.2e-16	0.43318	< 2.2e-16
X_{321}	0.45583	< 2.2e-16	0.43318	< 2.2e-16
X_{322}	0.42357	< 2.2e-16	0.22628	< 2.2e-16
X_{323}	0.39398	< 2.2e-16	0.471	< 2.2e-16
X_{324}	0.39476	< 2.2e-16	0.49388	< 2.2e-16
X_{325}	0.35905	< 2.2e-16	0.29939	< 2.2e-16
X_{326}	0.41291	< 2.2e-16	0.34897	< 2.2e-16
X_{327}	0.35396	< 2.2e-16	0.30753	< 2.2e-16
X_{328}	0.46347	< 2.2e-16	0.51409	< 2.2e-16
X_{329}	0.2759	< 2.2e-16	0.51478	< 2.2e-16
X_{330}	0.49418	< 2.2e-16	0.51575	< 2.2e-16
X_{331}	0.46356	< 2.2e-16	0.51409	< 2.2e-16
X_{332}	0.4715	< 2.2e-16	0.51413	< 2.2e-16
X_{333}	0.49495	< 2.2e-16	0.51625	< 2.2e-16
X_{334}	0.4893	< 2.2e-16	0.51309	< 2.2e-16
X_{335}	0.48582	< 2.2e-16	0.51408	< 2.2e-16
X_{336}	0.49164	< 2.2e-16	0.51414	< 2.2e-16
X_{337}	0.45607	< 2.2e-16	0.51415	< 2.2e-16

续表：

变量	类别			
	标准类		非标准类	
	D	p-value	D	p-value
X_{338}	0.36339	< 2.2e-16	0.32572	< 2.2e-16
X_{339}	0.31745	< 2.2e-16	0.35524	< 2.2e-16
X_{340}	0.21764	< 2.2e-16	0.33391	< 2.2e-16
X_{341}	0.4895	< 2.2e-16	0.4179	< 2.2e-16
X_{342}	0.48725	< 2.2e-16	0.37667	< 2.2e-16
X_{343}	0.47515	< 2.2e-16	0.2712	< 2.2e-16
X_{401}	0.45676	< 2.2e-16	0.43325	< 2.2e-16
X_{402}	0.42193	< 2.2e-16	0.47687	< 2.2e-16
X_{403}	0.40777	< 2.2e-16	0.47062	< 2.2e-16
X_{404}	0.47469	< 2.2e-16	0.43624	< 2.2e-16
X_{405}	0.43689	< 2.2e-16	0.40515	< 2.2e-16
X_{406}	0.44599	< 2.2e-16	0.42422	< 2.2e-16
X_{407}	0.401	< 2.2e-16	0.44186	< 2.2e-16
X_{408}	0.2749	< 2.2e-16	0.41496	< 2.2e-16
X_{409}	0.40147	< 2.2e-16	0.43891	< 2.2e-16
X_{410}	0.38587	< 2.2e-16	0.40296	< 2.2e-16
X_{411}	0.38569	< 2.2e-16	0.40292	< 2.2e-16
X_{412}	0.50221	< 2.2e-16	0.47312	< 2.2e-16
X_{413}	0.49338	< 2.2e-16	0.45247	< 2.2e-16
X_{414}	0.50221	< 2.2e-16	0.47306	< 2.2e-16
X_{415}	0.49324	< 2.2e-16	0.45247	< 2.2e-16
X_{416}	0.46472	< 2.2e-16	0.38996	< 2.2e-16
X_{417}	0.43532	< 2.2e-16	0.44508	< 2.2e-16
X_{418}	0.47019	< 2.2e-16	0.477	< 2.2e-16
X_{419}	0.47962	< 2.2e-16	0.40865	< 2.2e-16

续表：

变量	类别			
	标准类		非标准类	
	D	p-value	D	p-value
X_{420}	0.43276	< 2.2e-16	0.47477	< 2.2e-16
X_{421}	0.46113	< 2.2e-16	0.41702	< 2.2e-16
X_{422}	0.40441	< 2.2e-16	0.4458	< 2.2e-16
X_{423}	0.41442	< 2.2e-16	0.34659	< 2.2e-16
X_{501}	0.48197	< 2.2e-16	0.44119	< 2.2e-16
X_{502}	0.24604	< 2.2e-16	0.46167	< 2.2e-16
X_{503}	0.49477	< 2.2e-16	0.49758	< 2.2e-16
X_{504}	0.49667	< 2.2e-16	0.47421	< 2.2e-16
X_{505}	0.13833	< 2.2e-16	0.25037	< 2.2e-16
X_{506}	0.44264	< 2.2e-16	0.44534	< 2.2e-16
X_{507}	0.1231	< 2.2e-16	0.3007	< 2.2e-16
X_{508}	0.47706	< 2.2e-16	0.43402	< 2.2e-16
X_{509}	0.45327	< 2.2e-16	0.45322	< 2.2e-16
X_{601}	0.17026	< 2.2e-16	0.21978	< 2.2e-16
X_{602}	0.16575	< 2.2e-16	0.22001	< 2.2e-16
X_{603}	0.28469	< 2.2e-16	0.37998	< 2.2e-16
X_{604}	0.28473	< 2.2e-16	0.38	< 2.2e-16
X_{605}	0.16665	< 2.2e-16	0.16823	< 2.2e-16
X_{606}	0.18322	< 2.2e-16	0.1914	< 2.2e-16
X_{607}	0.23148	< 2.2e-16	0.084777	< 2.2e-16
X_{608}	0.29084	< 2.2e-16	0.22198	< 2.2e-16
X_{609}	0.16676	< 2.2e-16	0.13577	< 2.2e-16
X_{610}	0.21625	< 2.2e-16	0.19166	< 2.2e-16
X_{611}	0.22197	< 2.2e-16	0.081966	< 2.2e-16
X_{612}	0.23663	< 2.2e-16	0.22527	< 2.2e-16

续表：

变量	类别			
	标准类		非标准类	
	D	p-value	D	p-value
X_{613}	0.28656	< 2.2e-16	0.21218	< 2.2e-16
X_{614}	0.1559	< 2.2e-16	0.18948	< 2.2e-16
X_{615}	0.1581	< 2.2e-16	0.18043	< 2.2e-16
X_{616}	0.17558	< 2.2e-16	0.24352	< 2.2e-16
X_{617}	0.41312	< 2.2e-16	0.4662	< 2.2e-16
X_{618}	0.41237	< 2.2e-16	0.46616	< 2.2e-16
X_{619}	0.43013	< 2.2e-16	0.40846	< 2.2e-16
X_{701}	0.35363	< 2.2e-16	0.2712	< 2.2e-16
X_{702}	0.35188	< 2.2e-16	0.25671	< 2.2e-16
X_{703}	0.43129	< 2.2e-16	0.50906	< 2.2e-16
X_{704}	0.48029	< 2.2e-16	0.51406	< 2.2e-16
X_{705}	0.42387	< 2.2e-16	0.2518	< 2.2e-16

表 A6 旋转后成分矩阵

变量	成分																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
X101	-0.016	-0.005	0.378	0.001	0.005	-0.006	0.041	-0.104	0.016	0.849	-0.006	-0.042	0.009	0.065	-0.018	-0.004	0.009	-0.057	-0.051	-0.003	0.006	0.009	0.006	-0.022
X104	0.602	0.009	-0.050	0.003	-0.003	-0.001	-0.036	0.035	-0.154	-0.005	0.046	0.662	-0.018	-0.109	0.019	-0.110	-0.015	0.050	0.004	0.018	-0.003	-0.015	-0.011	0.014
X108	0.545	-0.025	-0.039	0.003	-0.002	0.002	-0.032	0.011	0.000	-0.003	0.062	0.734	-0.022	-0.091	0.023	-0.115	0.010	0.038	0.022	0.029	-0.005	-0.007	-0.007	0.005
X109	0.879	0.018	-0.027	-0.001	-0.001	-0.008	-0.022	0.012	-0.253	0.027	0.009	0.185	0.006	-0.004	0.017	-0.036	-0.032	-0.021	-0.042	0.019	-0.003	-0.028	-0.022	0.014
X110	0.010	0.048	-0.009	0.001	0.002	0.002	0.002	-0.052	0.876	0.017	-0.017	-0.234	0.011	0.048	-0.004	0.053	-0.009	-0.016	0.009	-0.004	0.002	-0.001	-0.005	0.003
X111	0.024	0.001	-0.028	0.004	-0.001	0.005	-0.018	-0.033	0.909	-0.002	0.033	0.182	-0.008	0.032	0.011	-0.020	0.001	0.009	0.010	0.023	-0.002	-0.006	-0.005	0.002
X114	0.014	-0.009	-0.101	-0.060	0.005	-0.011	-0.781	0.067	0.038	-0.139	0.016	0.056	-0.114	-0.012	-0.008	0.040	0.003	0.325	0.030	-0.006	0.006	0.049	0.056	-0.011
X115	-0.008	-0.032	0.023	-0.013	-0.010	-0.120	-0.044	-0.034	-0.036	-0.049	-0.084	-0.013	-0.123	0.098	0.043	-0.060	-0.071	0.646	0.029	0.033	0.039	-0.035	0.039	0.049
X116	-0.001	-0.003	-0.028	0.006	-0.006	-0.018	0.019	0.016	0.016	-0.014	0.049	0.037	-0.868	0.002	-0.052	-0.013	-0.006	0.099	0.009	-0.014	-0.016	0.003	-0.042	0.009
X121	-0.010	-0.001	-0.028	0.048	0.003	0.039	0.045	0.010	-0.020	0.082	-0.023	0.016	0.137	0.076	-0.024	0.105	0.034	0.196	0.015	-0.107	-0.019	-0.566	-0.031	-0.052
X122	-0.009	-0.001	0.090	-0.004	0.002	-0.011	0.022	-0.022	0.015	-0.041	-0.019	0.019	-0.068	0.157	-0.018	-0.125	0.009	-0.369	0.039	-0.019	0.044	-0.099	0.104	0.054
X127	-0.041	0.007	0.010	0.014	-0.003	-0.005	0.026	-0.005	-0.011	-0.058	-0.165	0.703	-0.003	-0.113	-0.045	0.088	0.008	0.008	0.004	-0.037	0.012	0.028	0.042	-0.035
X128	-0.009	0.051	0.783	0.000	0.026	0.041	0.123	0.018	-0.026	0.291	0.024	-0.021	0.004	0.074	0.273	0.089	0.032	-0.036	-0.035	-0.021	0.007	-0.007	0.030	-0.021
X129	-0.010	0.019	0.363	0.001	0.017	-0.004	0.029	-0.082	-0.003	0.840	-0.007	-0.040	-0.004	0.024	-0.024	-0.038	0.012	-0.026	-0.008	-0.011	0.008	0.016	0.009	-0.010
X131	-0.001	0.062	0.873	0.018	0.001	0.067	0.075	0.006	0.003	0.277	0.034	0.005	0.026	0.022	0.006	0.105	0.037	-0.009	-0.018	-0.016	0.003	-0.014	0.034	-0.016
X132	-0.006	0.066	0.891	0.018	-0.002	0.043	0.054	0.016	-0.004	0.240	0.030	-0.001	0.030	0.108	-0.006	0.082	0.037	-0.034	-0.012	-0.016	0.002	-0.021	0.021	-0.013
X202	-0.023	0.017	0.109	-0.546	-0.005	-0.009	-0.054	-0.133	-0.035	-0.091	-0.011	-0.009	-0.008	0.052	0.027	-0.002	0.001	-0.166	0.022	0.136	0.052	0.207	-0.015	-0.038
X205	-0.002	-0.010	-0.066	-0.784	-0.003	-0.055	-0.010	0.016	0.009	0.037	-0.037	-0.017	0.006	-0.024	-0.053	-0.023	-0.022	0.073	-0.007	0.116	-0.047	-0.080	-0.020	0.014

续表：

变量	成分																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
X207	-0.003	0.009	0.017	0.927	0.000	-0.004	0.000	-0.033	-0.012	-0.018	-0.014	-0.002	0.006	0.008	0.049	-0.006	-0.002	-0.047	0.012	0.015	0.009	0.049	-0.003	-0.011
X210	0.013	0.080	0.249	0.005	0.042	0.082	0.579	0.037	0.008	-0.065	0.042	-0.017	0.000	0.052	0.689	0.049	0.048	-0.032	-0.005	-0.014	0.012	0.038	0.010	-0.008
X211	-0.010	-0.001	-0.033	-0.045	0.014	0.003	0.166	0.007	-0.021	0.039	0.019	0.014	0.023	0.012	0.516	0.178	0.010	0.149	-0.009	0.034	-0.006	-0.079	0.066	-0.012
X214	0.009	0.030	0.053	0.006	-0.011	0.896	0.005	-0.001	-0.007	-0.006	0.021	0.001	0.228	0.038	0.008	0.013	0.017	0.014	0.006	0.013	0.003	-0.042	0.002	0.007
X215	-0.003	0.002	-0.008	-0.003	0.004	0.015	-0.001	0.025	-0.025	0.063	0.046	-0.007	-0.014	-0.005	0.044	0.027	0.693	-0.062	0.232	0.017	-0.009	-0.066	0.011	0.008
X216	-0.034	-0.047	-0.079	-0.009	-0.007	-0.023	-0.048	-0.007	-0.013	0.040	-0.002	0.028	-0.018	-0.031	0.027	0.069	-0.782	-0.051	0.224	0.014	-0.023	-0.040	-0.022	-0.013
X218	0.054	0.046	0.049	0.010	0.000	0.010	0.065	-0.020	0.045	-0.102	-0.035	-0.073	0.009	-0.041	-0.100	-0.240	0.283	0.187	-0.613	-0.047	0.008	0.099	-0.017	-0.015
X219	0.006	0.013	0.048	0.017	0.003	0.014	-0.035	0.008	0.012	0.002	0.047	0.058	0.054	0.052	0.028	0.260	0.373	-0.150	0.027	-0.009	-0.016	-0.045	-0.006	-0.012
X221	0.000	0.012	0.020	0.854	-0.002	-0.009	0.009	-0.009	-0.001	-0.018	-0.014	-0.007	0.012	-0.002	0.032	-0.013	-0.001	-0.003	-0.001	0.421	-0.007	0.002	-0.013	0.000
X224	-0.017	0.056	0.008	-0.001	0.015	0.012	0.006	-0.086	0.189	0.223	0.025	-0.069	0.026	0.803	0.040	0.089	0.067	-0.020	-0.044	0.045	-0.004	0.081	0.025	-0.049
X227	0.005	-0.039	-0.182	0.013	-0.029	-0.022	-0.217	0.094	0.098	-0.109	-0.076	0.175	-0.170	-0.230	0.028	-0.152	-0.051	0.558	0.080	0.004	0.019	-0.003	0.019	0.044
X229	-0.051	0.006	0.047	-0.013	0.007	0.006	0.051	0.126	-0.088	-0.100	-0.020	-0.185	-0.022	0.758	-0.002	-0.020	0.007	-0.139	-0.006	0.007	0.007	-0.038	0.007	0.004
X231	0.000	0.054	0.680	0.005	0.017	-0.012	-0.002	0.021	-0.007	-0.069	-0.001	-0.006	0.019	-0.102	-0.069	-0.136	-0.012	-0.059	0.045	0.015	-0.018	-0.048	-0.061	0.079
X304	0.936	0.067	0.023	0.004	0.000	0.005	0.003	0.015	-0.035	-0.012	-0.010	-0.002	0.010	-0.033	-0.004	0.076	0.011	-0.009	0.020	-0.007	0.000	0.008	0.003	-0.007
X306	0.745	0.001	-0.029	-0.004	-0.001	-0.006	-0.026	0.024	-0.490	0.028	0.023	0.284	-0.002	0.007	0.024	-0.067	-0.024	-0.009	-0.050	0.025	-0.004	-0.023	-0.020	0.013
X309	0.011	0.008	0.050	0.163	-0.006	-0.013	-0.082	-0.016	0.023	-0.043	-0.008	-0.038	0.022	-0.014	0.867	-0.139	0.008	-0.057	-0.013	-0.055	-0.003	0.037	-0.056	0.001
X310	-0.014	0.110	0.465	-0.078	0.010	0.023	0.042	-0.462	-0.044	-0.007	0.024	-0.031	0.062	-0.023	0.054	0.189	0.146	-0.163	0.038	0.107	0.046	0.107	0.015	-0.098
X313	-0.002	0.033	0.013	0.017	0.026	-0.014	0.532	0.024	0.025	-0.007	0.304	0.041	0.355	-0.005	-0.067	-0.053	-0.006	0.023	0.023	-0.022	0.000	-0.010	-0.044	0.016

续表：

变量	成分																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
X314	0.008	0.033	0.058	0.010	0.006	0.926	0.039	0.004	-0.010	-0.008	0.021	-0.005	-0.031	0.037	0.004	0.028	-0.001	0.054	0.008	0.007	0.006	-0.042	0.003	0.008
X315	0.016	0.050	0.065	0.021	0.016	0.113	0.221	0.023	0.022	-0.017	0.296	0.015	0.855	0.015	-0.002	0.011	0.050	-0.054	0.016	-0.009	0.026	-0.014	0.013	-0.004
X322	0.007	0.017	0.055	0.000	0.003	0.012	0.013	-0.011	0.001	-0.084	0.008	-0.018	-0.013	-0.025	-0.015	-0.169	-0.016	-0.004	0.012	-0.006	0.000	0.020	0.132	0.545
X325	0.852	0.062	0.040	0.010	0.006	0.027	0.034	-0.004	0.227	-0.024	0.012	-0.136	0.001	-0.031	-0.016	0.158	0.037	-0.005	0.044	-0.023	0.003	0.034	0.044	-0.016
X327	0.856	0.069	0.049	0.012	0.008	0.041	0.038	0.001	0.158	-0.028	0.030	-0.149	0.002	-0.032	-0.012	0.199	0.039	-0.016	0.010	-0.025	0.003	0.031	0.053	-0.019
X328	0.001	0.001	0.007	0.018	0.000	0.006	0.025	0.003	-0.003	-0.004	-0.033	0.012	-0.014	0.012	-0.011	0.033	0.005	-0.005	0.002	-0.026	0.761	0.010	0.000	0.001
X338	0.019	0.021	0.030	0.009	0.026	0.046	0.089	0.007	0.002	-0.006	0.927	-0.055	0.023	0.001	0.007	0.045	0.021	-0.026	0.001	-0.005	0.029	0.008	0.017	-0.003
X339	0.032	0.034	0.047	0.012	0.006	0.055	0.069	0.004	0.007	0.001	0.887	-0.074	0.149	0.002	0.029	0.037	0.066	-0.090	-0.001	-0.003	0.037	0.004	0.014	-0.006
X340	0.010	0.029	0.077	0.004	0.063	0.110	0.836	0.034	0.003	-0.025	0.055	0.025	-0.058	0.036	0.215	0.141	0.014	0.043	-0.003	0.021	0.022	0.027	0.078	0.000
X342	0.018	0.026	0.054	-0.002	0.003	0.025	0.047	0.004	0.014	-0.081	0.012	-0.069	-0.029	-0.006	-0.026	-0.418	-0.050	0.016	0.012	-0.012	0.010	-0.031	0.279	-0.551
X406	0.019	0.011	0.008	0.033	0.039	0.672	0.066	0.014	0.021	0.012	0.054	-0.001	-0.056	-0.051	0.012	0.030	0.030	-0.184	-0.008	-0.023	-0.006	0.059	-0.010	-0.003
X407	-0.001	0.013	0.007	0.001	0.878	-0.002	0.023	0.005	0.000	0.009	0.012	0.001	0.017	0.007	0.003	-0.002	0.001	-0.014	-0.004	-0.001	0.000	0.003	0.007	0.001
X409	-0.004	0.020	0.008	0.002	0.910	0.006	0.033	-0.010	-0.004	0.039	0.014	-0.001	0.021	0.020	0.005	0.002	0.003	-0.014	-0.013	-0.001	0.000	0.030	0.003	-0.001
X419	0.004	0.006	-0.012	-0.001	0.003	-0.003	-0.012	0.016	-0.001	0.014	-0.005	0.000	0.014	-0.013	0.016	-0.024	-0.010	-0.060	-0.004	0.013	-0.006	-0.005	0.675	0.059
X420	0.000	-0.046	-0.045	-0.029	0.013	0.007	0.004	0.323	0.017	0.145	-0.014	-0.032	-0.020	-0.182	0.055	0.264	0.049	-0.117	0.003	0.167	0.045	0.058	0.124	-0.126
X423	0.721	0.046	-0.048	-0.003	-0.002	-0.003	-0.024	0.043	0.074	0.003	0.009	0.135	0.003	0.063	0.018	-0.106	-0.022	0.001	0.043	0.014	0.001	-0.043	-0.034	0.029
X502	-0.006	-0.010	-0.003	0.021	-0.001	-0.005	0.007	-0.039	0.014	-0.010	-0.008	-0.015	0.004	0.035	-0.030	-0.010	-0.007	0.056	-0.009	0.941	-0.013	-0.010	-0.014	0.004
X505	0.068	0.049	0.062	0.020	0.003	0.010	0.008	0.854	-0.140	-0.151	0.008	0.036	0.001	-0.205	-0.002	0.006	0.024	0.011	0.022	-0.036	0.010	0.023	-0.003	0.015

续表：

变量	成分																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
X506	-0.015	-0.022	-0.028	0.009	-0.010	0.003	-0.009	0.161	0.039	0.022	0.007	0.083	0.041	0.255	-0.028	0.012	-0.005	0.098	0.051	-0.031	-0.019	0.413	-0.066	-0.014
X507	0.003	0.087	0.050	0.014	0.001	0.018	0.016	0.877	-0.008	-0.068	0.015	-0.039	0.027	0.292	0.010	0.044	0.049	-0.001	0.004	-0.019	0.007	0.091	-0.004	-0.004
X601	0.151	0.435	0.365	0.021	0.053	0.132	0.174	0.013	0.075	-0.145	0.104	-0.078	0.023	0.143	0.072	0.504	0.112	0.029	0.048	-0.057	0.012	0.032	0.104	-0.102
X604	0.051	0.613	-0.102	0.008	-0.011	0.003	-0.046	0.531	0.078	-0.010	0.019	0.057	0.005	0.181	0.000	-0.056	-0.002	0.062	-0.006	-0.013	-0.002	0.016	-0.036	0.018
X606	0.226	0.401	0.128	0.025	-0.023	-0.033	-0.002	0.033	0.051	-0.128	-0.045	-0.023	0.030	-0.054	-0.059	0.279	-0.033	0.036	0.059	-0.081	0.005	-0.104	-0.039	-0.070
X607	0.068	0.893	0.062	0.011	-0.021	0.019	0.038	-0.018	0.045	-0.020	0.022	-0.011	0.055	0.014	-0.017	0.115	0.020	-0.089	-0.021	-0.003	0.000	-0.049	-0.013	-0.020
X610	0.037	0.803	0.103	-0.005	0.001	-0.002	0.000	0.035	-0.081	-0.066	-0.013	-0.006	-0.001	-0.116	-0.001	-0.047	0.003	-0.045	0.057	0.015	0.000	-0.017	-0.014	0.052
X612	-0.017	0.631	-0.028	-0.007	0.066	0.019	0.009	-0.037	0.000	0.277	0.015	-0.006	-0.017	0.095	0.029	-0.122	0.032	0.025	0.099	0.018	0.003	0.075	0.035	0.020
X613	0.038	0.930	0.078	-0.001	0.014	0.048	0.039	0.023	0.030	-0.018	0.045	0.015	0.000	0.023	0.054	0.097	0.022	0.005	-0.001	0.005	0.002	0.010	0.023	-0.010
X615	0.090	0.181	0.041	0.008	-0.011	0.013	0.010	0.004	0.020	-0.112	-0.009	-0.042	0.011	-0.055	-0.080	-0.068	0.166	0.128	0.761	-0.031	0.004	0.069	-0.017	-0.004
X616	0.128	0.259	0.274	0.019	0.091	0.219	0.279	-0.004	0.078	-0.122	0.202	-0.047	-0.049	0.137	0.070	0.543	0.097	0.068	0.005	-0.038	0.022	0.057	0.202	-0.089
X617	0.009	0.007	0.025	0.003	0.874	0.034	0.017	0.006	0.005	-0.025	0.008	-0.005	-0.017	-0.009	0.017	0.034	0.008	0.002	0.007	0.002	0.002	-0.011	0.001	-0.002
X702	0.502	0.046	0.017	0.001	-0.002	0.000	0.045	-0.010	0.107	-0.027	-0.032	0.020	0.005	-0.024	-0.044	-0.134	0.104	0.089	0.484	-0.033	0.006	0.053	-0.009	-0.014
X703	-0.001	-0.004	0.005	0.007	-0.001	0.004	0.012	-0.008	-0.003	-0.013	-0.084	0.004	-0.041	0.011	-0.008	0.025	0.005	-0.011	0.002	-0.017	-0.734	0.022	0.014	-0.009
X705	0.006	-0.002	-0.021	0.002	-0.003	0.005	0.011	0.014	0.004	0.025	-0.008	-0.025	-0.006	-0.004	-0.005	0.015	0.004	0.021	-0.005	0.001	0.012	-0.013	0.064	0.587

提取方法：主成分分析法

旋转方法：凯撒正态化分析法

a.旋转在 16 次迭代后已收敛

表 A7 贝叶斯优化迭代结果

迭代次数	指标得分	参数						
		colsample_bytree	max_depth	min_child_samples	min_child_weight	reg_alpha	reg_lambda	subsample
1	0.9375	0.501	12.05	25.58	12.27	0.8718	0.1032	0.9364
2	0.9394	0.4184	12.41	10.27	11.32	0.6129	0.6693	0.9451
3	0.9434	0.4159	7.921	15.93	5.788	0.2578	0.7495	0.5079
4	0.9409	0.9455	7.488	15.06	12.84	0.8043	0.3507	0.7879
5	0.9438	0.6492	12.19	11.46	3.529	0.9118	0.3783	0.5038
6	0.9433	0.6919	5.049	10.96	3.052	0.5816	0.3699	0.6027
7	0.9437	0.7272	5.016	10.04	3.049	0.8995	0.959	0.6328
8	0.9428	0.7649	5.55	29.89	3.038	0.619	0.8349	0.9302
9	0.9406	0.8977	5.034	10.27	3.179	0.6197	0.4795	0.8378
10	0.9422	0.5658	14.86	10.01	3.204	0.3483	0.8593	0.6098
11	0.943	0.4637	5.082	29.51	3.1	0.9999	0.5807	0.7465
12	0.9439	0.8917	14.93	29.6	3.04	0.9604	0.4268	0.7811
13	0.9435	0.6634	14.78	29.87	3.005	0.1435	0.8506	0.7621
14	0.9435	0.9588	5.269	29.9	3.051	0.3906	0.2815	0.7872
15	0.9459	0.9192	14.52	29.13	3.032	0.4589	0.5739	0.5954
16	0.9436	0.9687	15	29.66	3.013	0.5883	0.2848	0.9
17	0.9429	0.7069	14.94	29.46	3.043	0.8154	0.9834	0.7757
18	0.9419	0.7306	14.7	29.85	3.019	0.3431	0.7437	0.6557

续表：

迭代次数	指标得分	参数						
		colsample_bytree	max_depth	min_child_samples	min_child_weight	reg_alpha	reg_lambda	subsample
19	0.9426	0.9176	5.264	29.47	3.012	0.9698	0.4252	0.5661
20	0.9438	0.9581	14.74	10.23	3.004	0.5116	0.232	0.8786
21	0.9425	0.7357	14.71	10.02	3.037	0.4831	0.3832	0.531
22	0.9437	0.9814	14.79	29.81	3.087	0.9782	0.4929	0.5513
23	0.9425	0.6987	14.93	28.72	3.049	0.9096	0.879	0.5568
24	0.9432	0.7381	5.249	29.9	3.076	0.1542	0.9703	0.9418
25	0.9431	0.5105	14.91	29.51	3.007	0.2461	0.4478	0.5668
26	0.9438	0.5599	14.81	10.14	3.03	0.7929	0.9785	0.9419
27	0.943	0.6008	14.61	10.89	3.002	0.327	0.934	0.6261
28	0.9443	0.8728	14.52	10.27	3.143	0.1019	0.2721	0.7179
29	0.9439	0.761	14.97	11.19	3.004	0.9068	0.2467	0.6829
30	0.9427	0.9854	14.99	12.48	3.063	0.1865	0.2044	0.6772
31	0.9427	0.9766	5.816	10.3	3.108	0.1295	0.8399	0.5171
32	0.9425	0.4512	14.78	29.41	3.014	0.6936	0.1508	0.9162
33	0.9432	0.9018	14.57	10.8	3.038	0.7406	0.2806	0.5193
34	0.9445	0.9658	14.99	10.93	3.028	0.9064	0.9299	0.5732
35	0.9434	0.5324	14.99	10.5	3.032	0.583	0.2024	0.9549
36	0.9409	0.9819	5.233	10.29	3.008	0.4554	0.2991	0.7784

续表：

迭代次数	指标得分	参数						
		colsample_bytree	max_depth	min_child_samples	min_child_weight	reg_alpha	reg_lambda	subsample
37	0.9441	0.9624	14.82	29.71	3.059	0.1927	0.325	0.6099
38	0.9445	0.7634	14.92	29.01	3.122	0.5251	0.8497	0.8526
39	0.9438	0.6893	14.65	29.93	3.011	0.7145	0.6957	0.5525
40	0.9436	0.8584	14.8	28.42	3.024	0.1635	0.1107	0.6339
41	0.9443	0.7446	14.74	29.3	3.006	0.5111	0.8096	0.7555
42	0.9443	0.9404	14.57	29.92	3.199	0.725	0.8702	0.5332
43	0.9433	0.7342	14.73	29.92	3.019	0.8139	0.7398	0.9359
44	0.9419	0.4257	14.98	29.88	3.169	0.5451	0.9204	0.7303
45	0.9436	0.9452	14.5	29.9	3.058	0.949	0.9274	0.8124
46	0.9431	0.9333	14.8	28.27	3.001	0.7269	0.8186	0.6874
47	0.9428	0.9891	14.39	10.18	3.062	0.8876	0.9575	0.5413
48	0.943	0.5085	14.7	29.07	3.002	0.9976	0.2107	0.9237
49	0.944	0.663	14.91	29.21	3.003	0.2504	0.6333	0.9571
50	0.9445	0.9999	14.95	26.38	3.094	0.1943	0.9862	0.8521
51	0.9425	0.9376	14.88	29	3.043	0.1284	0.1135	0.9917
52	0.9432	0.8385	14.83	11.53	3.023	0.5483	0.9734	0.9382
53	0.9443	0.4253	14.62	29.28	3.01	0.3097	0.9276	0.9628
54	0.9433	0.7357	14.91	29.89	3.014	0.1973	0.3202	0.6084

续表：

迭代次数	指标得分	参数						
		colsample_bytree	max_depth	min_child_samples	min_child_weight	reg_alpha	reg_lambda	subsample
55	0.9427	0.8488	14.66	29.25	3.168	0.105	0.9649	0.5049
56	0.9432	0.8865	14.91	28.04	3.012	0.457	0.9612	0.5427
57	0.9439	0.6865	14.63	11.56	3.042	0.3633	0.8918	0.5012
58	0.9451	0.922	14.92	29.45	3.028	0.1155	0.5806	0.9846
59	0.9429	0.7753	14.74	28.54	3.087	0.2451	0.9775	0.5517
60	0.9429	0.7021	14.93	10.18	3.045	0.906	0.9733	0.9537
61	0.9421	0.6488	14.48	28.87	3.074	0.1413	0.951	0.9582
62	0.9442	0.9363	14.97	28.66	3.004	0.5327	0.7754	0.5339
63	0.9429	0.4954	14.87	29.62	3.005	0.9298	0.7628	0.5573
64	0.9424	0.9968	14.97	28.89	3.234	0.1628	0.2252	0.7621
65	0.9443	0.9816	14.96	29.84	3.238	0.1542	0.1291	0.5499
66	0.9447	0.5943	14.64	29.66	3.043	0.272	0.1201	0.6349
67	0.9419	0.9888	14.76	29.94	3.023	0.7451	0.1597	0.9512
68	0.9418	0.7947	14.94	11.32	3.051	0.1985	0.1447	0.7546
69	0.943	0.7734	14.66	29.24	3.027	0.2363	0.6914	0.6232
70	0.9438	0.6941	14.66	30	3.141	0.8673	0.8808	0.6367
71	0.9435	0.9183	14.73	29.85	3.028	0.719	0.14	0.5573
72	0.9434	0.4892	14.44	29.71	3.011	0.4528	0.8346	0.6996

续表：

迭代次数	指标得分	参数						
		colsample_bytree	max_depth	min_child_samples	min_child_weight	reg_alpha	reg_lambda	subsample
73	0.9437	0.7495	14.75	29.75	3.155	0.8299	0.9771	0.663
74	0.943	0.5053	14.75	29.64	3.005	0.9796	0.2693	0.6152
75	0.9442	0.5327	15	28.81	3.122	0.1389	0.8768	0.514