



长安大学

二〇一八届毕业论文

基于 Python 的 IPO 财务舞弊识别工具研究

学 院：经济与管理学院

专 业：会计学（文史）

姓 名：熊 璽 宇

学 号：201423030128

指导教师：郭 慧 婷

完成时间：2018 年 6 月

二〇一八年六月

摘 要

首次公开募股（Initial Public Offerings, IPO），又名股票市场启动，是公开上市集资的一种类型。通过证券交易所，公司首次将它的股票卖给一般公众，非上市公司通过这个过程会转化为上市公司。IPO 能够为公司带来重大效益的同时，对公司的财务状况有严格的要求，因而个别公司铤而走险，通过财务舞弊出具虚假财务报表以骗取上市机会。IPO 舞弊公司在上市后往往出现利润大幅下滑甚至严重亏损的情况，导致投资者蒙受损失，证券市场秩序被扰乱，因此对 IPO 公司的财务审查至关重要。

本文基于财务舞弊相关理论和主要的财务舞弊动机理论，选取了一批 IPO 舞弊公司和非舞弊公司作为样本，设计了一套与盈余管理关系密切的变量体系，通过逻辑回归分析建立了 IPO 财务舞弊识别模型。检验结果表明证明该模型准确率达到 91%，拟合度较佳，总体而言具有较强的 IPO 财务舞弊识别能力。

此外，本文以该模型为核心，设计了基于 Python 构建的 IPO 财务舞弊识别工具，该工具具有自动下载财务报表、模型变量数据接收、自动产生回归模型并使用模型进行 IPO 财务舞弊识别的功能。该工具能减少数据处理的工作量，在一定程度上提升处理效率，以期为市场参与者提供一套实用且易用的 IPO 财务舞弊识别工具。

关键词：IPO, 舞弊识别, 财务舞弊, 逻辑回归, Python

ABSTRACT

Initial Public Offerings (IPO) or stock market launch, is a type of public offering in which shares of a company are sold to institutional investors and usually also retail (individual) investors, transforming a company into a listed one. While benefitting a company, IPO put forward a high requirement to its financial status, that drives some companies to gain an opportunity to be listed by publishing falsified financial statements. These companies usually will produce performance falters or even heavy losses after being listed, causing investors' serious losses and market disorder. Therefore, it is essential to correctly audit companies which are planning an IPO.

This paper is based on relative theories of financial fraud and motivation of financial fraud and build a detection model of IPOs' financial frauds by logistic regression. The model is generated by samples of fraud and non-fraud IPOs and classified by some variables closely related to earnings management. The outcome of the experiment shows that the model is able to achieve a general accuracy of 91% with an ideal fitting degree. Overall, this model has a strong ability to detect IPOs' financial frauds.

Besides, revolving around the model, this paper raises a Python-based detection tool of IPOs' financial frauds, which can download financial statements, receive model variables and generate the regression model for calculating. This is designed to reduce manual operations and improve processing efficiency, providing a practicable and easy tool for market participants.

KEY WORDS: IPO, fraud detection, financial statement fraud, logistic regression, Python

目 录

1	绪论.....	1
1.1	研究背景.....	1
1.1.1	现实背景.....	1
1.1.2	问题产生.....	1
1.2	研究意义.....	3
1.2.1	理论意义.....	3
1.2.2	现实意义.....	4
1.3	文献综述.....	4
1.3.1	国外文献综述.....	4
1.3.2	国内文献综述.....	5
1.3.3	研究现状评述.....	6
1.4	研究内容与思路.....	6
1.5	研究方法.....	9
2	基本理论.....	10
2.1	财务舞弊相关理论.....	10
2.1.1	财务舞弊.....	10
2.1.2	盈余管理.....	10
2.1.3	会计信息失真.....	11
2.2	财务舞弊动机理论.....	11
2.2.1	舞弊冰山理论.....	11
2.2.2	舞弊三角理论.....	12
2.2.3	GONE 理论	12
2.2.4	企业舞弊风险因子理论.....	13
3	财务舞弊识别的逻辑回归分析模型.....	14
3.1	样本选择.....	14
3.2	模型变量体系设置.....	15
3.3	回归分析结果与模型性能评价.....	17
4	基于 Python 构建的 IPO 财务舞弊识别工具.....	21
4.1	工具简介.....	21
4.2	基于 Python 的财务报表下载模块的实现	21

4.3 基于 Python 的变量数值接收模块的实现	22
4.4 基于 Python 的舞弊性分析模块的实现	23
4.5 本工具的使用范例.....	24
结论与展望.....	26
致 谢.....	27
参考文献.....	28
附录.....	31
I. 逻辑回归模型训练集.....	31
II. 基于 Python 的财务报表下载模块源代码	31
III. 基于 Python 的财务报表数据提取模块	33
IV. 基于 Python 的财务报表舞弊性分析模块	34
V. 本文相关资源下载.....	35

1 绪论

1.1 研究背景

1.1.1 现实背景

首次公开募股（Initial Public Offerings，简称 IPO），又名股票市场启动，是公开上市集资的一种类型。通过证券交易所，公司首次将它的股票卖给一般公众，非上市公司通过这个过程会转化为上市公司。公司通常通过 IPO 募集资金，尽可能地将早期个人投资者的投资货币化和资本化，同时使公司的股份（股票）能够在交易所公开、自由地买卖。

IPO 对于公司具有诸多好处：使股本基数扩大化和多样化；更容易获得资本；提高公司曝光度、公司名声以及公司的公众形象；通过流动的股东权益分享，吸引和保留更好的管理层和员工；加速收购；创造多样的融资机会：普通股、可转换债券、低成本的银行贷款，等等。IPO 作为公司上市的必经环节，市场热情持续高涨，中国 IPO 数量和金额不断创造新高。2017 年 12 月 27 日，安永发布《安永全球 IPO 市场调研报告：2017 年回顾及 2018 年展望》报告称，预计 2017 全年 A 股市场共有 436 宗 IPO，共筹资 2,304 亿元人民币，分别较 2016 年同比增加 92% 和 53%^[1]。

同时，IPO 设置了较高的门槛。根据中国证监会《首次公开发行股票并上市管理办法》规定，首次公开发行股票并上市需要满足的主体（不包括在创业板上市的）要求，在财务方面就有以下几点：（1）最近 3 个会计年度净利润均为正数，且累计超过人民币 3000 万元，净利润以扣除非经常性损益前后较低者为计算依据；（2）最近 3 个会计年度经营活动产生的现金流量净额累计超过人民币 5000 万元；或者最近 3 个会计年度营业收入累计超过人民币 3 亿元；（3）发行前股本总额不少于 3000 万元；（4）最近一期末无形资产（扣除土地使用权、水面养殖和采矿权等后）占净资产的比例不高于 20%；（5）最近一期末不存在未弥补亏损^[2]。

1.1.2 问题产生

上市能够带来的诸多裨益使上市热情持续高涨，然而较高的门槛令许多公司望而却步，加之发审委往往抱持较为审慎的态度，因此新股发行审核的通过率并不高。2006 年至 2016 年，发审委审核新股共计 2145 家，审核通过 1834 家，审核未通过 311 家，市场整体通过率为 85.5%^[3]，但多数年份的通过率处在 50% 左

右的水平。在十八届三中全会明确提出“推进股票发行注册制改革”以来，新股审核的通过率连续 3 年保持在 90% 以上。不过自 2017 年 10 月 17 日新发审委上任后，其治下的 IPO 审核通过率又降低至 59.72%，新股发审通过率低已经成为“常态”。

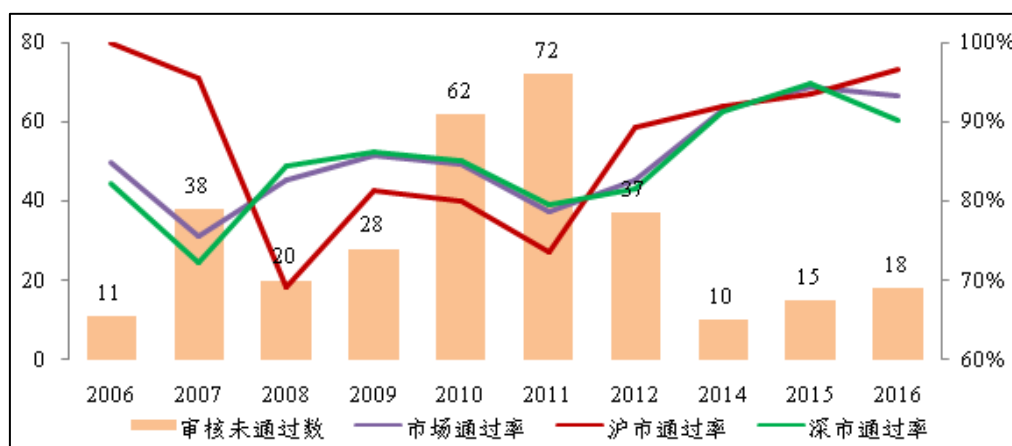


图 1-1: 2006-2016 年间中国大陆境内 IPO 审查通过率和通过数
数据来源: 知乎网

图 1-1 以柱状折线统计图形式描述了 2006 年至 2016 年的十年间大陆境内 IPO 审查的通过率和通过数。观察每年的审核未通过数可以发现，2006 年至 2008 年、2008 年至 2014 年间都形成了一个倒 U 形；相应地，同样时间段内的市场通过率呈现出 U 形，IPO 审查的松紧程度呈现出一定的周期性特征，结合 2017 年的审核情况，可以看出目前已进入第三个周期的下降阶段，IPO 审核再次收紧。

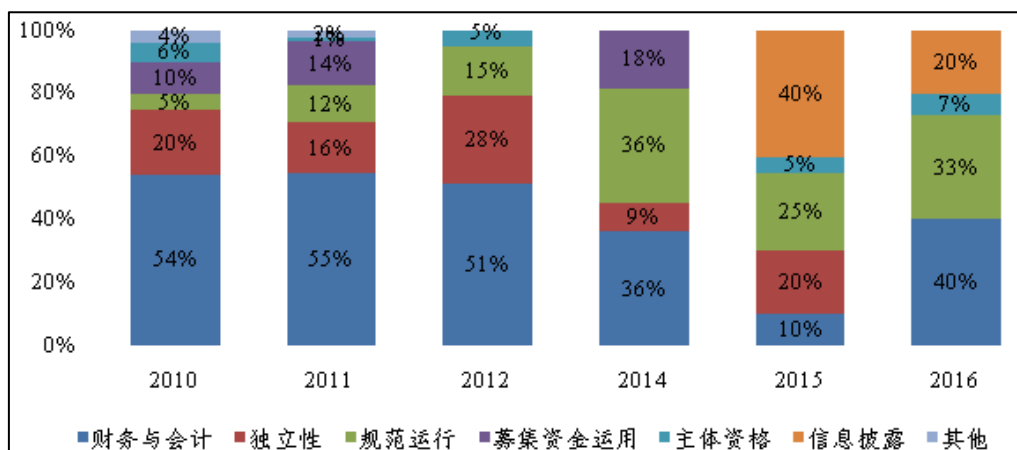


图 1-2: 2006-2016 年间中国大陆境内 IPO 审查不通过原因分类及占比
数据来源: 知乎网

由图 1-2 可见,“财务与会计”是导致 IPO 审查不通过的主要方面。财务报表是反映一个公司的财务信息、基本面信息的重要载体,因而成为了财务舞弊的主要作用对象,财务舞弊的结果也在财务报表中得到反映。为了争取上市机会、吸引市场投资,某些公司铤而走险,通过虚增资产、虚增收入、虚增利润、掩盖关联方交易等方式,出具虚假的财务报表,甚至某些券商和会计师事务所也参与到这个过程中。公司通过财务舞弊粉饰其盈利能力,对投资者产生严重误导,这些公司往往在成功上市后业绩急转直下,造成投资者的重大损失。近年来,万福生科、绿大地等多家公司的 IPO 财务舞弊行为被曝光, IPO 财务舞弊也随着舆论热议进入了市场参与者的视线。

2017 年,于 7 年前成功登陆创业板并成为 A 股首家网络视频公司的乐视网(SZ:300104),深陷资金链危机的同时又被揭露涉嫌 IPO 财务造假,事态迅速发酵,多名发审委委员被采取强制措施,其中包括第一届创业板发审委委员谢忠平,其为亚太集团会计师事务所副主任会计;北京天圆会计师事务所副总经理孙小波,以及大华会计师事务所董事、执行合伙人韩建旻等,乐视面临重大退市危机^[4]。该事件使得 IPO 财务舞弊再次成为舆论中心,引起市场上广泛热议。

严格且准确的审计能够及时发现并及时拦截 IPO 财务舞弊公司,避免造成更多的损失。本文正是在这样的背景下,对 IPO 财务舞弊的识别工具进行了研究,以期为准确甄别 IPO 财务舞弊提供参考。

1.2 研究意义

1.2.1 理论意义

西方学术界对于财务舞弊的理论性研究较早、较丰富、较成熟,并且最先将

统计方法运用在审计过程中；国内学术界也在西方学术界的成果基础上，发展出一些具有特点的财务舞弊识别方法。但无论国内外学术界，以上方法都是针对上市公司提出的，较少涉及到 IPO 企业。本文参考了相关理论成果，设计了专门针对 IPO 财务舞弊的识别模型，具有鲜明的特色性。

此外，本文尝试性地基于 Python 语言构建了具有财务报表抓取、数据接收和统计分析识别等功能的轻量化工具，在财务自动化、信息化趋势越来越深入的今天，将数学工具程序化能够促使研究成果更快地转化为易用且实用的工具，或将起到抛砖引玉的作用。

1.2.2 现实意义

目前我国证券市场的 IPO 热潮方兴未艾，但如果放任各种公司都进入市场将带来极大的危害。在 IPO 审查阶段准确把关，对于维护证券市场秩序、保护投资者权益具有重要的意义。准确的 IPO 财务舞弊识别方法能够合理地评价被审计公司的审计风险水平，作为审计人员安排适当的审计程序的参考，既能辅助审计工作，也能对潜在的舞弊公司起到一定的震慑作用，减少舞弊公司“蒙混过关”的可能性。经过合理设计的 IPO 财务舞弊识别工具能方便地在审计工作中使用，这一定程度上能够提高审计工作的效率，减少手工处理数据的环节。

1.3 文献综述

1.3.1 国外文献综述

在对财务舞弊的特征的研究中，Persons（1995）发现舞弊公司较非舞弊公司具有更高的财务杠杆、更低的资本周转率、更高的流动资产比率（主要是存货和应收账款）^[5]。

一些研究论证了分析程序在财务舞弊识别中的有效程度和必要性。Coglitore 和 Berryman（1988）、Caldero 和 Green（1994）证明了简单的分析程序是识别会计舞弊的有效工具，能够揭示财务报告项目的显著变化和异常关系^[6]。Ameen 和 Strawser（1994）指出，审计人员在实务中运用最为广泛的定量分析程序是简单的模型，其往往将报告指与期望值相比较，并对研究结果提出建议；复杂的模型能够使审计人员与会计信息使用者同时考虑到公司财务状况及经营成果的各方面，以舞弊和未舞弊公司的过往的数据为基础，利用模型判别公司是否具备舞弊的动机和可能性^[7]。

许多研究都提出了具体的财务舞弊识别方法，通常都以数学统计模型为基础。

（1）Loebbecke、Eining 和 Willingham（1989）构造舞弊风险识别模型（L/W 模型）以判别公司的舞弊风险因素^[8]。（2）Beneish（1997）提出概率分析方法，该

方法由组合的财务变量产生利润操纵指数，进而判断公司操纵利润的可能性^[9]。

(3) Beneish (1999) 提出使用会计数据判别上市公司是否存在财务舞弊，通过 Probit 判别方法建立了会计舞弊的识别模型，准确率达 75%，且在实际中得到一些运用^[10]。(4) Spathis Doumpos 和 Zopounidis (2002) 以希腊的财务舞弊公司和非舞弊公司为样本，运用多指标分析和多变量统计技术对会计舞弊进行识别研究，研究表明该方法胜过传统的识别技术，对财务信息的研究有助于判别财务舞弊与否^[11]。(5) Dechow、Ge、Larson 等 (2010) 对 1982 年至 2005 年的 2190 份间美国证券交易委员会 (SEC) 会计与审计起诉公告 (AAERs) 进行了检验，通过应计质量、财务表现、非财务指标、表外活动和市场机制措施等五个维度，将舞弊公司的舞弊期间和非舞弊期间之间、舞弊公司和其他上市公司之间进行比较，发现舞弊公司通常有较多的应计项目、更高的 P/F (Price-to-fundamental) 比、更高的优先股股价，管理层具有自主定价权的资产占总资产比重也更高，并依此建立了 F-Score 模型^[12]。(6) Abbasi、Albrecht、Vance 等 (2012) 设计了一种科学的方法来建立称为 MetaFraud 的一种增强型财务舞弊识别的元学习框架，结果证明 MetaFraud 的每一个要素都能对总体的有效性产生贡献。追加的实验显示其有效性超过时下最先进的财务舞弊识别方法，对于事、投资者、审计公司和监管机构都有重要意义^[13]。

1.3.2 国内文献综述

关于拟上市公司财务舞弊的动机，邢婷婷 (2017) 指出我国上市公司管理舞弊存在显著的功利性特征，即相比西方发达国家通过舞弊达到少纳税、少付息的目的，我国上市公司管理舞弊通常是为了吸引投资、争取上市、抬高发行价、融资和再融资^[14]。

不少学者对于我国上市公司财务舞弊的方式和特征进行了分析和归纳。朱锦余、高善生 (2007) 基于 2000 至 2006 年间因舞弊性财务报告遭到证监会处罚公告的上市公司进行了统计分析，得出结论为我国上市公司舞弊性财务报告的主要类型是虚假披露、虚假利润表和虚假资产负债表，且多数公司同时存在两种及以上舞弊类型^[15]。黄世忠、黄京菁等 (2004) 分析公司财务报表舞弊行为特征和信号，总结了 Beneish (1999) 提出的应收账款周转指数、毛利率指数、资产质量指数、销售增长指数、应计总额资产指数等五项关键识别指标^[16]。

财务报表舞弊的量化统计识别也在向国外研究成果借鉴的同时不断完善。朱敏 (2005) 采用截止到 2005 年 2 月的 68 家存在财务报告舞弊的公司为样本，使用逻辑回归建立舞弊性财务报表的识别模型，发现资产负债率、销售毛利率、营运资金占总资金比等三项指标能够起到甄别作用^[17]。吴顺祥 (2005) 基于包含

29 份舞弊财务报告和 29 份非舞弊财务报告的样本，通过逻辑回归分析建立了能够预测财务报告舞弊的模型，发现资产负债率、销售毛利率、营运资金占总资产的比例可用于识别舞弊财务报告^[18]。陈国欣、吕占甲、何峰（2007）通过逻辑回归模型发现只需盈利能力、管理层持股比例、独立董事规模、审计意见四个变量就可以较好地识别预测上市公司财务报告舞弊，该模型的准确率达 95.1%^[19]。李清、任朝阳（2015）以 2003 年至 2012 年间我国沪深 A 股财务报告舞弊公司及其配对公司为样本，构建了准确率为 66.7% 的基于仿兰氏距离的案例推理模型，发现舞弊公司面对更大的盈利和偿债压力，并且频繁变动会计师事务所^[20]。

1.3.3 研究现状评述

纵观国外关于财务舞弊识别的研究成果，既有规范研究又有实证研究，首先提出通过数理统计方法来研究财务舞弊在财务数据上的特征，并建立了一系列数学模型。国外在会计舞弊研究方面已进入不断深化阶段，并积累了丰富的研究成果，且在实践中发挥了理论先行与实践指导作用（秦江萍，2007）^[21]。国内相关研究在对国外研究成果不断学习的同时，结合国内市场环境提出了一些不同点，并在此基础上取得了一些具有创新点的成果。

然而，上述国内外研究都以上市公司作为研究对象，对于 IPO 公司的财务舞弊识别研究仍然较少。作为将非上市公司转换为上市公司的关键一步，如果能在 IPO 审查时期就准确地识别出舞弊公司，就可以最大程度上防范经济损失并维护市场秩序。因此，本文以 IPO 财务舞弊识别作为切入点，试图提出一个可用的 IPO 财务舞弊识别工具。

1.4 研究内容与思路

第一章为绪论。本章首先介绍了选题的背景，并从理论意义和实践意义两方面说明了选题的意义；之后总结并评价了国内外相关研究现状；最后介绍了研究内容与思路，及以实证分析法为主的研究方法。

第二章为基本理论。本章解释了财务舞弊的定义、盈余管理和会计信息失真的概念，介绍了主要的财务舞弊动机理论和逻辑回归方法。

第三章介绍 IPO 舞弊性财务报表逻辑回归模型。本章基于逻辑回归方法建立了 IPO 财务舞弊的识别模型，介绍了样本的来源和选取方法、模型变量体系的设置方法以及最终产生的模型结果，最后对其显著性、拟合度、准确率进行了样本检验。

第四章为综合介绍了 Python 构建的 IPO 财务舞弊识别工具，分别说明了财务报表下载模块、财务数据提取模块和舞弊性分析验证模块的主要功能并解释其

实现方法。

最后为结论与展望，首先总结了回归模型的分析结果和准确性、拟合度等性能指标，之后指出了本文存在的两个难以克服的不足，最后提出了将来在财务领域有更多基于数学工具的程序涌现的愿景。

本文的组织结构如图 1-3 所示（见下页）：

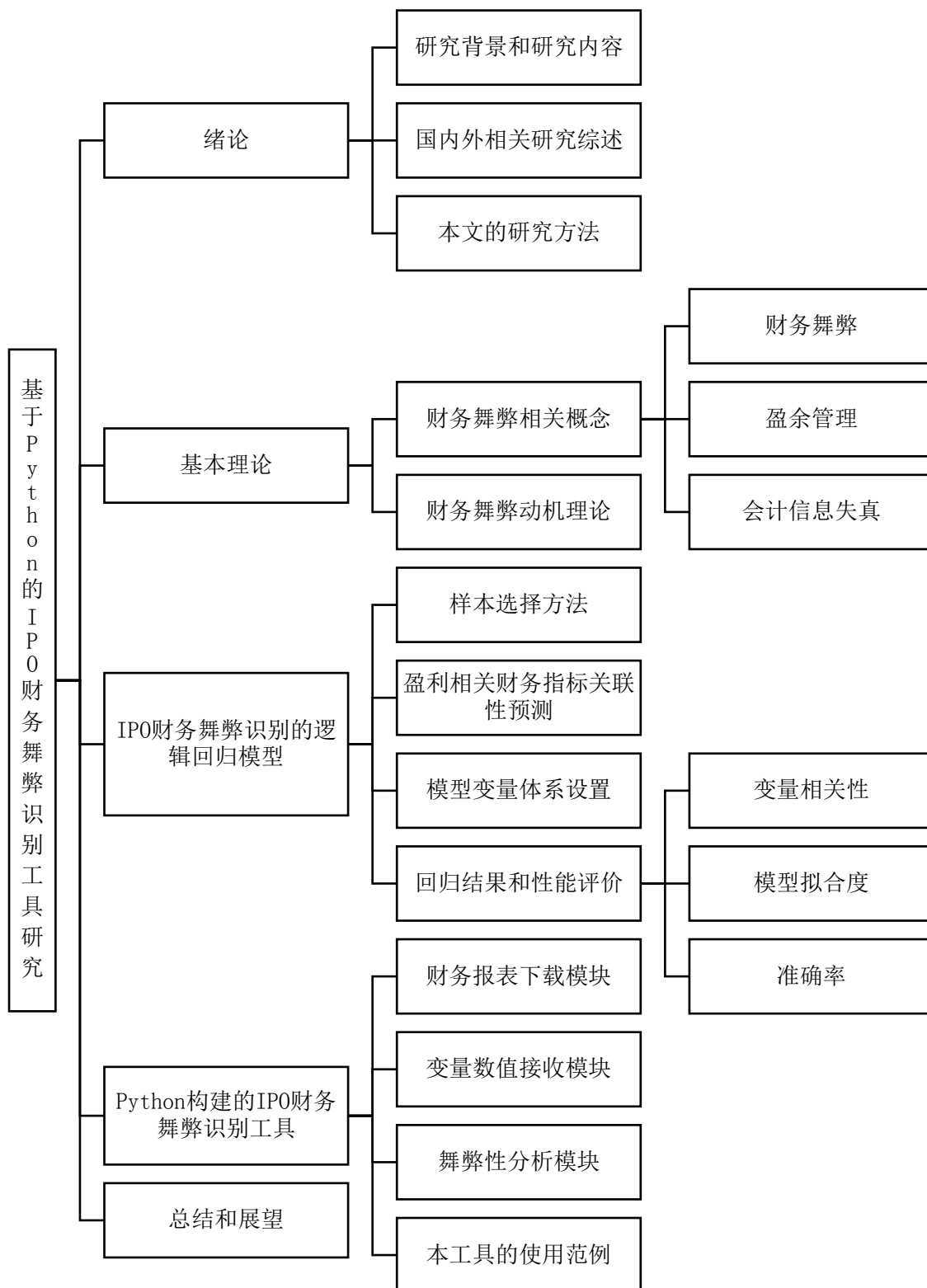


图 1-3：本文组织结构框架

1.5 研究方法

(1) 统计分析法。本文选取了 17 家 IPO 财务舞弊公司（包括 4 家被严重怀疑的）和 20 家非 IPO 财务舞弊公司作为样本，按照设置的变量公式计算各样本的变量值，通过逻辑回归产生能够依据设置的变量的值判定 IPO 公司的舞弊概率的模型。

(2) 工具模型的使用。本文以统计学方法为基础，通过一些设置的变量建立起 IPO 财务舞弊识别的逻辑回归模型，该模型具有根据 IPO 公司的财务指标和数据计算其舞弊性的能力。本文还以 Python 语言将该模型封装，实现了具有财务报表下载、模型数值接收和舞弊性分析等三项功能的计算机程序。

2 基本理论

本章对财务舞弊的概念、盈余管理的概念、会计信息失真的概念、主要的财务舞弊动机理论、逻辑回归模型进行了相关概述，为本文建立 IPO 财务舞弊识别的逻辑回归模型进行了铺垫。

2.1 财务舞弊相关理论

2.1.1 财务舞弊

审计准则公告 99 号（Statement on Auditing Standards No.99，简称 SAS99）将财务舞弊划分为挪用资产（偷盗或占用公司资产并设法在财务报表中掩饰）和财务报表舞弊（通过失真财务报表欺骗使用者，骗取资本市场认可，试图影响投资者的决策），指出财务舞弊主要的三种形式：伪造、操纵或篡改会计记录和凭证文件；蓄意忽略或错误披露重要事项、交易及其他重要信息；滥用会计政策、会计估计造成金额、分类等产生错误。^[22]

我国《独立审计具体准则第 8 号——错误与舞弊》中对舞弊定义为“导致会计报表产生不实反映的故意行为”^[23]。《中国注册会计师审计准则第 1141 号——财务报表审计中对舞弊的考虑》则认为财务舞弊是指被审计单位的管理层等有意识、有目的的使用欺骗、造假等手段获取不当利益的行为^[24]。

2.1.2 盈余管理

Scott 在其所著的《财务会计理论》一书中认为，盈余管理是会计政策的选择具有经济后果的一种具体表现。他认为，只要企业的管理人员有选择不同会计政策的自由，他们必定会选择使其效用最大化或使企业的市场价值最大化的会计政策，这就是所谓的盈余管理。Schipper（1989）认为盈余管理是为了获得某种私人利益（而并非仅仅为了中立地处理经营活动），对外部财务报告进行有目的地干预。

被普遍认可的是 Healy 和 Wahlen（1999）对盈余管理所作出的解释：当管理者在编制财务报告和构建经济交易时，运用判断改变财务报告，从而误导一些利益相关者对公司根本经济收益的理解，或者影响根据报告中会计数据形成的契约结果，盈余管理就产生了。公司主要可以通过变更会计政策、应计项目管理、改变交易时间、创造特殊交易和资产评估。与财务舞弊违反一般公认会计准则不同，盈余管理虽然本质上并不能真实反映公司的财务信息，却并不违反一般公认会计准则^[25]。应计项目操控是通过会计政策选择、会计估计变更等会计方法（例

如少提坏帐费用、推迟摊销费用等等)来管理盈余。由于应计项目有回转的特性,企业调整应计项目而管理盈余的能力是有限的,会计弹性制约着管理盈余的程度^[26]。

2.1.3 会计信息失真

会计信息失真,是指会计信息的输出与输入不一致产生的信息虚假,即财务报告反映的情况与原始凭证(包括依据原始凭证所作的进一步处理)不符。吴联生(1998)将会计信息失真分为三类:违法性会计信息失真、规范性会计信息失真、技术性会计信息失真^[27]。钟黄聪(2011)认为会计信息失真是会计主体的经济属性和人脑所反映出的经济属性的不一致,包括企业内各部门的信息过滤和曲解、会计部门的信息过滤和曲解、会计信息使用者的信息过滤和曲解等三种类型^[28]。会计信息失真同时包含了财务舞弊和盈余管理,是客观的结果,既可能是违法的,也可能是合法的。

2.2 财务舞弊动机理论

2.2.1 舞弊冰山理论

舞弊冰山理论建立在奥地利精神学家西格蒙德·弗洛伊德的冰山理论基础之上,把舞弊比喻为海平面上的一座冰山,露在海平面上的只是冰山的一角,更庞大的危险部分隐藏在海平面以下^[29]。从结构和行为方面考察舞弊,海平面上的是结构部分,海平面下的是行为部分。舞弊结构的内容实际上是组织内部管理方面的,这是客观存在且容易鉴别的。而舞弊行为的内容则是更主观化、更个性化、更容易被刻意掩饰起来的。示意图如图 2-1 所示:

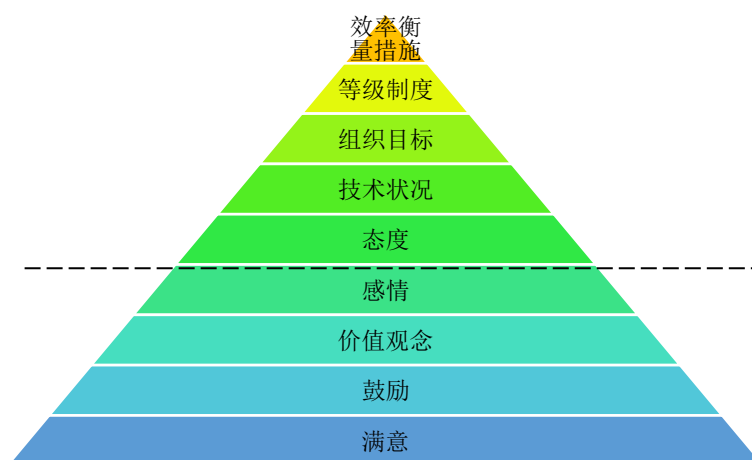


图 2-1：舞弊冰山理论示意图

2.2.2 舞弊三角理论

该理论由 Albrecht（2014）提出，他认为，企业舞弊的产生是由压力、机会和自我合理化三要素组成，缺少了上述任何一项要素都不可能真正形成企业舞弊^[30]。压力可能是经营或财务上的困境以及对资本的急切需求。机会可能是宽松的或松懈的控制以及信息不对称。而自我合理化则可能是“我只是向公司借而不是偷”、“我们只是为了暂时渡过困难时期”、“我的出发点是为了一个很好的愿望”等。舞弊三角中的三个因素是两两相互作用的。舞弊三角理论可用图 2-2 表示：

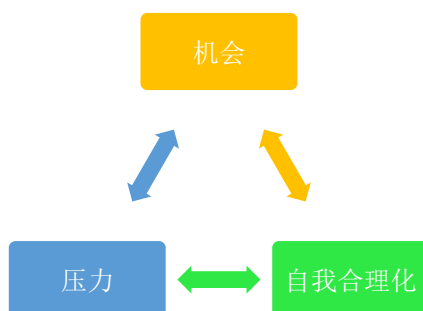


图 2-2：舞弊三角理论示意图

2.2.3 GONE 理论

GONE 理论认为企业会计舞弊由 G、O、N、E 四个因子组成，它们相互作用、密不可分、同等重要，共同决定了企业舞弊风险的程度。GONE 由 4 个英语单词的开头字母组成，其中：G 为 Greed，指贪婪；O 为 Opportunity，指机会；N 为 Need，指需要；E 为 Exposure，指暴露。GONE 理论可表述为图 2-3 所示：

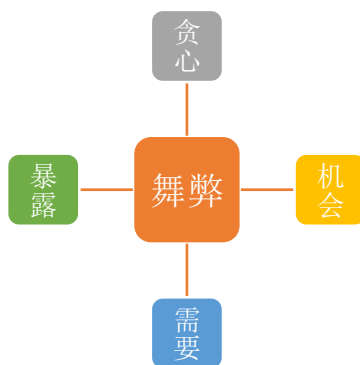


图 2-3：GONE 理论示意图

上述 4 个因子实质上表明了舞弊产生的 4 个条件，即舞弊者有贪婪之心且又十分需要钱财、自尊时，只要有机会，并认为事后不会被发现，他就一定会进行舞弊，导致 “You can consider your money gone”（被欺骗者的钱、物、权益等离

他而去)。因此,产生了一种很巧妙的说法,即“在贪婪、机会、需要和暴露四因子共同作用的特定环境中,会滋生舞弊,促使被欺骗者的钱、物、权益等离他而去”。

2.2.4 企业舞弊风险因子理论

企业舞弊风险因子理论旨在 GONE 理论的基础上发展形成的,由 Singleton、Bologna 和 Lindquist (2006) 提出,是比较完善的关于形成企业舞弊的风险因子的学说^[31]。它把舞弊风险因子分为个别风险因子与一般风险因子:个别风险因子指因人而异、且在组织控制范围之外的因素,包括道德品质与动机;一般风险因子是指由组织或实体来控制的因素,包括舞弊的机会、舞弊被发现的概率以及舞弊被发现后舞弊者受罚的性质和程度。当一般风险因子与个别风险因子结合在一起,并且被舞弊者认为有利时,舞弊就会发生。舞弊风险因子结构如图 2-4 所示:

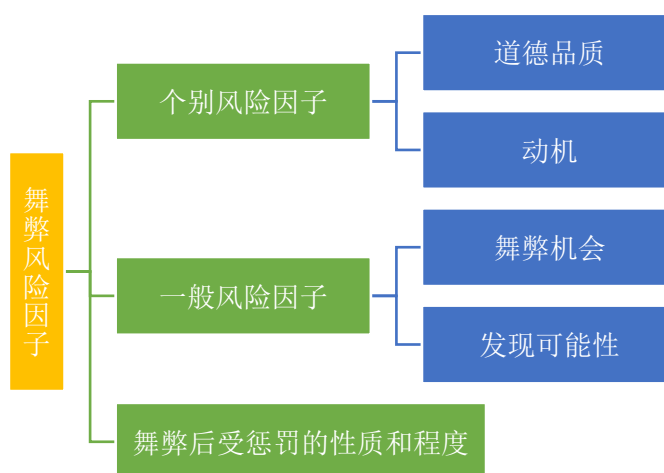


图 7：舞弊风险因子示意图

本文将基于财务舞弊的概念,对舞弊的 IPO 公司进行舞弊行为的分类统计,结合盈余管理的概念和主要的财务舞弊动机理论分配相关的财务指标作为模型变量,通过二元逻辑回归方法建立起 IPO 财务舞弊识别模型。

3 财务舞弊识别的逻辑回归分析模型

3.1 样本选择

本文将样本划分为两组：舞弊组和非舞弊组。舞弊组选取了能够在主流搜索引擎中检索出 IPO 财务舞弊资讯，且中国证监会发布了相应行政处罚公告的企业共 13 家。出于谨慎考虑，另有 4 家被强烈质疑存在 IPO 财务舞弊行为、尚在调查中的公司也加入了舞弊组样本中。表 3-1 列示了上述公司的主要信息以及其舞弊方式的分类统计：

表 3-1：舞弊组样本情况统计（顺序不分先后）

序号	企业简称	股票代码	利润操纵	虚增收入	其他虚假记载	重大遗漏
1	新荷花	审查中止	√			
2	嘉寓股份	300117	√		√	√
3	迪威视讯	300167	√			
4	新大地	审查中止	√		√	
5	胜景山河	许可撤销				√
6	华锐风电	601558	√	√		
7	登云股份	002715			√	√
8	振隆特产	审查中止	√	√	√	
9	欣泰电气	已退市			√	
10	绿大地	002200		√	√	
11	海联讯	300277		√	√	
12	天能科技	审查中止	√	√		
13	万福生科	300268	√	√		√
14 ¹	乐视网	300104	?	?	?	?
15 ¹	鼎捷软件	300378	?	?	?	?
16 ¹	快意电梯	002774	?	?	?	?
17 ¹	天地数码	300743	?	?	?	?

1：该公司被强烈质疑存在 IPO 财务舞弊行为，但证监会尚未做出行政决定。
2：“?” 代表相应的舞弊方式是否存在暂时无法证实。

非舞弊组的样本则通过在中国证监会“监管信息公开目录”(<http://www.csrc.gov.cn/pub/zjhpublic>) 上按照发布时间升序排列 IPO 公司发布的招股说明书，选取深沪两市中在 2011 至 2012 年之间通过 IPO 审查并成功上市、并且至今未因 IPO 而受任何证监会行政处罚的公司。鉴于舞弊组样本容量有限，为了使模型能

更加准确地对样本进行拟合，非舞弊组的样本容量允许在最低等于舞弊组样本容量的基础上适当扩大，但不应超过舞弊组样本容量的 1.5 倍，以适当减少舞弊组样本数偏低所带来的潜在不利影响。

3.2 模型变量体系设置

结合 IPO 对财务方面的要求、第二章介绍的财务舞弊动机理论及表 3-1 的统计结果，可以发现编造虚假的盈利能力是实施 IPO 舞弊的主要目的。要设置合理的模型变量体系，变量能够有效反映公司的盈利能力是必要前提。本文首先对一些能够影响盈利能力的财务指标进行关联性预测，作为设置模型变量体系的参考依据。各项财务指标与利润之间的预测关联性如表 3-2 所示：

表 3-2：各项财务指标与盈利能力的预测关联性

序号	财务指标	与利润间关联性预测
1	应收账款周转率	+
2	毛利率	+
3	净利率	+
4	营业收入	+
5	营业利润	+
6	折旧与摊销	-
7	坏账准备	-

注：“+”代表正相关关系，“-”代表负相关关系。

表 3-2 的各项预测解释如下：

- （1）应收账款周转率的关联性预测为正相关，应收账款周转率越高，在不变的经营周期内将有更多应收账款被收回并被确认为当期收入。
- （2）毛利率的关联性预测为正相关，毛利率越高，说明公司的增殖能力越高。
- （3）净利率的关联性预测为正相关，净利率越高，代表公司获得利润的能力越高。净利率越接近毛利率，则企业控制期间费用的能力越强。
- （4）营业收入的关联性预测为正相关，在单位成本不变的情况下，营业收入越高，公司最终可获得的利润就越多。
- （5）营业利润的关联性预测为正相关，在税费占比不发生显著变化的情况下，营业利润越多，公司最终可获得的净利润就越多。
- （6）折旧与摊销的关联性预测为负相关，当期发生的折旧与摊销越多，当期应计费用相应增加，从而导致了当期利润减少。

(7)坏账准备的关联性预测为负相关,计提的坏账准备将被转为当期费用。计提的坏账准备越多,当期费用越多,导致当期利润减少。

上述财务数据或指标,在不同的行业、不同的规模中通常存在较大的差异,不能作为直接的判断依据。但是将某些数据组合并按照一定公式计算之后,可以获得为以百分比为单位的新的变量,这些变量通常反应公司某些财务方面的水平值,具有更强的可比性。本文选取一些基于上述财务数据或指标组合计算而得到的、单位为百分比的变量,作为模型的变量集使用。

此外,一些非财务指标数据也可以作为评价舞弊风险的参考。IPO 舞弊公司的保荐机构和会计师事务所所有可能存在屡次包庇的行为,因此,若保荐机构或会计师事务所曾经涉及舞弊事件,则某 IPO 公司的舞弊性应当受到质疑^[32]。

基于上述预测,逻辑回归模型的变量集解释及计算公式列示为表 3-3:

表 3-3: 逻辑回归模型初始变量

序号	变量代码	变量解释	计算公式
1	id1	应收账款周转率变动水平	$= \frac{\text{本年应收账款周转率}}{\text{上年应收账款周转率}}$
2	id2	毛利率变动水平	$= \frac{\text{本年毛利率}}{\text{上年毛利率}}$
3*	id3	资产质量指数变动水平	$= \frac{\text{本年资产质量指数}}{\text{上年资产质量指数}}$
4	id4	营业收入变动水平	$= \frac{\text{本年营业收入}}{\text{上年营业收入}}$
5	id5	净资产回报率变动水平	$= \frac{\text{本年净利润} \div \text{本年所有者权益}}{\text{上年净利润} \div \text{上年所有者权益}}$
6	fraudspns	保荐机构是否曾涉及舞弊	是, 值为 1; 否则为 0
7**	isfraud	是否为舞弊公司	是, 值为 1; 否则为 0
*资产质量指数 = 1 - (流动资产 + 固定资产净值) ÷ 资产总额			
**该变量是样本分类变量。			

表 3-3 中的各项变量解释如下:

(1) 应收账款周转率变动水平是某期应收账款周转率与上一期的比值, 异常低的变动水平可能意味着存在虚增应收账款的行为。

(2) 毛利率变动水平是某期毛利率与上一期的比值。异常高的变动水平可能意味着存在费用操纵的行为。

(3) 资产质量变动水平是某期资产质量指数与上一期的比值。过低的值意味着应收项目在资产中的比例过高, 须警惕是否存在虚构应收账款的可能。

(4) 营业收入变动水平是某期营业水平与上一期的比值。异常增高的水平可能意味着存在关联方交易或虚构收入。

(5) 净资产回报率变动水平是某期净资产回报率与上一期的比值。净资产回报率代表了每一单位权益的收益，异常的水平可能意味着虚增利润或关联方交易等行为。

(6) 保荐机构是否曾涉及舞弊的判断依据是舞弊样本组的保荐机构清单。如果某公司的保荐机构出现在清单中，则被认为曾涉及舞弊。

回归模型将根据样本集中的以上变量生成。生成回归模型的原始样本数据请参见附录 I，本章节不予列示。

3.3 回归分析结果与模型性能评价

逻辑回归 (Logistic regression 或 logit regression)，即逻辑模型 (Logit model，也译作评定模型、分类评定模型)，是离散选择法模型之一，属于多重变量分析范畴，是社会学、生物统计学、临床、数量心理学、计量经济学、市场营销等统计实证分析的常用方法。逻辑回归的分布公式可以表达为 $P(Y = 1|X = x) = \frac{e^{x'\beta}}{1+e^{x'\beta}}$ 。

回归分析生成的逻辑回归模型为：

$$P = \frac{e^{4.726 \times id1 - 4.302 \times id5 - 6.478 \times fraudspns + 1.578}}{1 + e^{4.726 \times id1 - 4.302 \times id5 - 6.478 \times fraudspns + 1.578}}$$

其中P为被检测的IPO公司实施财务舞弊的概率。该模型的临界点为0.5，即当 $P \geq 0.5$ 时，被测试IPO公司被判定为舞弊；若 $P < 0.5$ ，则被判定为非舞弊。

为了对模型的可用性做出合理的评价、了解该模型的性能表现，有必要对其进行一系列的检验，包括模型变量显著性、模型拟合度、模型正确率等。下文将基于建立模型所使用的样本集数据，分别对以上内容进行检验。

表3-4：模型变量显著性检验结果

方程中的变量							
		B	S.E.	Wals	df	Sig.	Exp (B)
步骤 1 ^a	id1	6.836	3.956	2.986	1	0.084	930.392
	id2	3.003	2.848	1.112	1	0.292	20.144
	id3	0.688	1.562	0.194	1	0.660	1.990
	id4	6.514	5.319	1.500	1	0.221	674.297
	id5	-10.542	6.191	2.899	1	0.089	0.000
	fraudspns(1)	-9.221	3.877	5.657	1	0.017	0.000
	常量	-6.026	6.540	0.849	1	0.357	0.002
步骤 2 ^a	id1	7.028	3.985	3.111	1	0.078	1127.824
	id2	2.614	2.630	0.988	1	0.320	13.652
	id4	5.802	4.830	1.443	1	0.230	330.822
	id5	-10.102	6.017	2.819	1	0.093	0.000
	fraudspns(1)	-8.910	3.668	5.901	1	0.015	0.000
	常量	-4.678	5.439	0.740	1	0.390	0.009
步骤 3 ^a	id1	6.045	3.370	3.217	1	0.073	421.987
	id4	4.439	4.054	1.199	1	0.274	84.688
	id5	-7.233	4.453	2.638	1	0.104	0.001
	fraudspns(1)	-8.062	3.148	6.560	1	0.010	0.000
	常量	-2.300	4.518	0.259	1	0.611	0.100
步骤 4 ^a	id1	4.726	2.656	3.165	1	0.075	112.851
	id5	-4.302	2.898	2.204	1	0.138	0.014
	fraudspns(1)	-6.478	2.222	8.497	1	0.004	0.002
	常量	1.578	3.084	0.262	1	0.609	4.848
a. 在步骤 1 中输入的变量: id1, id2, id3, id4, id5, fraudspns.							
不在方程中的变量							
			得分	df		Sig.	
步骤 2 ^a	变量	id3	0.176	1		0.675	
	总统计量		0.176	1		0.675	
步骤 3 ^b	变量	id2	1.103	1		0.294	
		id3	0.000	1		0.985	
	总统计量		1.263	2		0.532	
步骤 4 ^c	变量	id2	0.638	1		0.425	
		id3	0.000	1		0.983	
		id4	0.791	1		0.374	
	总统计量		1.615	3		0.656	
a. 在步骤 2 中移去的变量: id3。							
b. 在步骤 3 中移去的变量: id2。							
c. 在步骤 4 中移去的变量: id4。							

表3-4显示了模型变量显著性的检验结果,为了淘汰不相关变量并适当克服变量间可能的线性关系,本文选择了“向后:Wald”的方法,在每次回归迭代中抛弃一个P值最大的变量。考虑到总体样本量较少的问题,本模型适当对变量的P值筛选标准做出了适当放松:变量的 $P \leq 0.1$ 则进入, $P > 0.15$ 则退出。由表格可见,在第二至第四次迭代中,id3(资产质量指数变动水平)、id2(毛利率变动水平)、id4(营业收入变动水平)被先后移除,最终被保留的是id1(应收账款周转率变动指数)、id5(净资产回报率变动水平)和fraudspns(保荐机构是否曾涉舞弊)三个变量以及常数项1.578。表中各项因子的释义列示为表3-5:

表 3-5: 各回归参数释义

参数名	参数释义	与显著性的相关性
B	解释系数	—
S.E.	标准误	—
Wals	Wald 卡方检验	+
df	自由度	×
Sig.	显著性	+
Exp (B)	优势比	—
“×”代表无相关性,“+”代表正相关性;“—”代表负相关性。		

回归结果表明应收账款周转率变动水平(id1)在10%的水平上显著,净资产回报率(id5)在15%的水平上显著,保荐机构是否曾涉舞弊(fraudspns)在5%的水平上具有统计学意义。这反映出操纵应收账款、操纵净利润很可能为IPO舞弊公司的常用手段,并且保荐机构的诚信度也左右着舞弊的可能性。

表3-6: 模型拟合程度

步骤	-2对数似然值	Cox & Snell R ²	Nagelkerke R ²
1	16.006 ^a	0.612	0.818
2	16.157 ^a	0.611	0.816
3	17.170 ^a	0.600	0.801
4	18.718 ^b	0.583	0.779
a. 因为参数估计的更改范围小于0.001,所以估计在迭代次数8处终止。			
b. 因为参数估计的更改范围小于0.001,所以估计在迭代次数7处终止。			

表 3-6 的三项数值均显示了模型的拟合程度。-2 对数似然值的值越小,表示模型的拟合度越高,本模型的值为 18.718;Cox&Snell R²和 Nagelkerke R²的值越大,则模型的拟合度越高,本模型的值分别为 0.583 和 0.779,以上结果均表

明本模型的拟合效果较佳。与剔除变量前（步骤 1）的结果相比较，本模型仅牺牲了较小的拟合度，而得到了更加精简的模型。

表3-7：样本正确率检验结果

已观测			已预测		
			isfraud		百分比校正
			0	1	
步骤 1	isfraud	0	18	2	90.0
		1	1	16	94.1
	总计百分比				91.9
步骤 2	isfraud	0	18	2	90.0
		1	1	16	94.1
	总计百分比				91.9
步骤 3	isfraud	0	18	2	90.0
		1	2	15	88.2
	总计百分比				89.2
步骤 4	isfraud	0	18	2	90.0
		1	1	16	94.1
	总计百分比				91.9
a. 切割值为0.500。					

表 3-7 显示了原始样本代入模型中的检验结果。37 家样本公司中包含 17 家舞弊公司（isfraud=1）和 20 家非舞弊公司（isfraud=0）。其中，非舞弊公司有 2 家被误判为舞弊公司，准确率为 90.0%，舞弊公司有 1 家被误判为非舞弊公司，准确率为 94.1%，总体正确率为 91.9%。可以看到，与未剔除任何变量的结果（步骤 1）相比较，本模型在保持了同样的准确率的情况下，做到了进一步的精简。检验结果表明本模型的回归方法能较为准确地判定 IPO 公司的财务舞弊可能性，未来经过样本补充之后，准确率能得到更加稳健的保证。需注意的是，尽管本工具表现出了良好的识别能力，但不能代替一般审计程序，仅能作为审计风险的参考依据，并据此适当调整审计程序。

第四章将以该逻辑回归模型为核心，设计一个基于 Python 语言的简单的 IPO 财务舞弊识别工具。

4 基于 Python 构建的 IPO 财务舞弊识别工具

4.1 工具简介

前文中已经建立 IPO 财务舞弊识别的逻辑回归模型，但添加样本和生成模型的过程较为繁琐，且在样本产生变化后模型必须重新生成，这不利于使用者快速上手并重复使用，但是本文提出的整个 IPO 财务舞弊识别的过程可以通过数学语言和计算机语言简洁地描述。因此，本文提出了基于 Python 构建的 IPO 财务舞弊识别工具（以下简称“本工具”）。

Python 作为一种动态语言，包含丰富的库资源，可以简洁轻量地设计出各种程序，而不用投入大量精力考虑语法、类型等底层的工作，因此非常利于非计算机专业人士学习，并开发一些轻量的、自动化的工具来提升工作效率，本工具也是基于这样的理念而被提出。本工具兼具体积小、运行快、可重写的特点，没有使用门槛，应当能够为审计人员带来便利。

本工具主要由三个模块组成，分别为财务报表批量下载模块、变量数值接收模块和舞弊性识别模块。本工具的工作流程见图 4-1。本工具的下载地址参见本文附录 V，运行本工具所必要的组件的下载地址也一同给出。

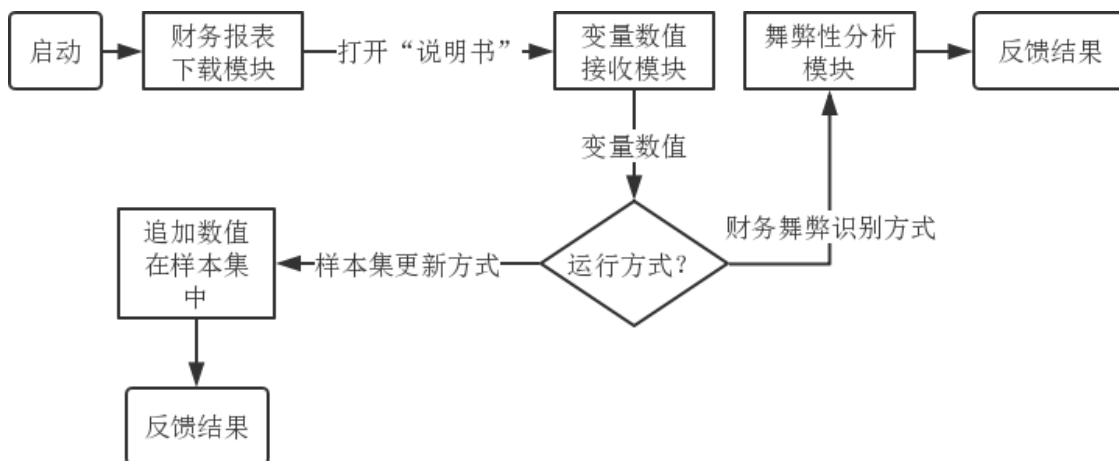


图 4-1：基于 Python 的 IPO 财务舞弊识别工具流程图

4.2 基于 Python 的财务报表下载模块的实现

(1) 确定下载源。IPO 公司以首次公开发行股票招股说明书（以下简称“说明书”）的形式披露财务内容。中国证监会“证券期货监督管理信息公开目录”

（<http://www.csrc.gov.cn/pub/zjhpublic>，下文简称“目录”）自 2011 年 7 月起集中收录并披露“说明书”，同时，证监会作为国务院直属事业单位必然长期存在。

因此，“目录”兼具权威性和持久性，本工具即以“目录”作为财务报表的下载源。

(2) 发起检索。“目录”支持关键字查找，结果页中列出了所有标题中含关键字的页面链接，因此可以构造一个搜索请求并获取结果页。以“公司名称+首次公开发行股票招股说明书(或创业板首发招股说明书)”作为关键字即可实现。

(3) 从检索结果中查找目标页面。“目录”的搜索结果按照时间降序排列，通常第一项搜索结果通常为所求页面（如果搜索结果为空则说明不存在）。本模块使用 BeautifulSoup 实现对搜索结果的寻获。BeautifulSoup 是一个能够从 HTML 或 XML 格式的文档中提取数据的 Python 库，能够生成结构清晰的网页剖析树，分解网页结构以提取需要的信息。本模块将结果页通过 BeautifulSoup 打开，使用其标签搜索功能按序查找所有标题符合（2）中格式的超链接并保存为列表。

(4) 从目标页面中下载“说明书”。本模块首先判断（3）中的超链接列表是否为空列表，如果为空则查找失败并退出，否则打开列表中的第一个超链接，同样按（3）的方法分析和提取超链接，下载文件名同（2）中格式的 PDF 格式的“说明书”并保存在本模块所在的目录下。

本模块的工作流程图如图 4-2 所示。本模块的源代码请参见附录 II。

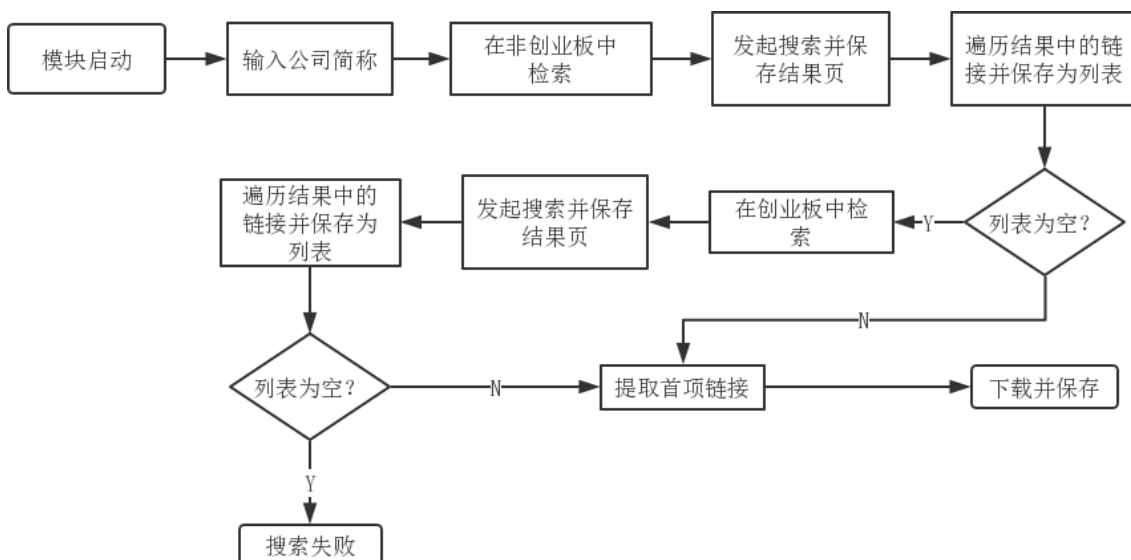


图 4-2：基于 Python 的财务报表下载模块工作流程

4.3 基于 Python 的变量数值接收模块的实现

(1) 本模块启动时首先询问工作方式。本模块可以以两种方式运行：财务舞弊识别方式和样本集更新方式。

(2) 本模块以数组形式预先定义了一组待接收的财务数据或指标，并通过循环要求使用者逐一提供（由于各公司发布的财务报表格式不一，故无法实现自动提取）。

(3) 使用者在每次循环中应当按照“数值 1+空格+数值 2”的格式提供相应的数值。本模块以空格为界，将两个以字符串格式传入的数值分别截取并转换为浮点数保存。当遗漏第二个数值或空格时，本模块会因为发生数组下标越界的异常而终止运行，防止非法格式的数据被接受。如果没有提供完所有预设变量的数值，将返回（1）继续。

(4) 全部财务指标数值接受完毕后，将根据预设的公式（见表 3-2）计算模型中各个变量的值并保存。

(5) 本模型如果运行在财务舞弊识别方式下，本模块将呼叫舞弊性分析模块，并传递上述变量的值以进行舞弊性分析；如果运行在样本集更新方式下，仅将这组值作为一个新样本追加在样本集文件中保存，不会启动舞弊性分析模块。财务舞弊识别方式下，输入的变量值不会被更新在样本集中。

本模块的工作流程图如图 4-3 所示。本模块的源代码请参见附录 III。

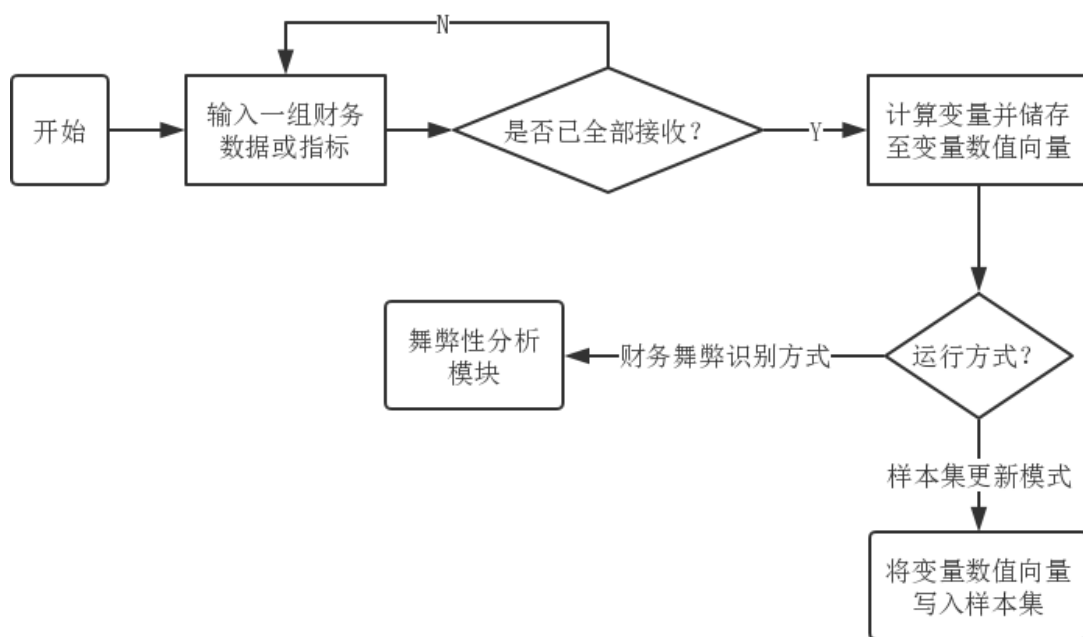


图 4-3：基于 Python 的变量数值接收模块工作流程

4.4 基于 Python 的舞弊性分析模块的实现

(1) 本模块从变量数值接收模块接受并保存各项变量数值，按序保存为一个向量（变量数值向量）。

(2) 本模块调用 R 语言解释器，执行一系列操作：打开样本集文件并通过

glm()函数连接至逻辑回归函数，逻辑回归函数根据样本集生成 IPO 公司财务舞弊识别的逻辑回归模型。

(3) 本模块读取模型内容，将各变量的系数按序保存为一个向量，再将保存的模型变量向量与该向量进行向量点积运算，得到财务舞弊概率 (P)。若 $P \geq 0.5$ 则判定存在财务舞弊，否则不存在，并将该结果反馈给使用者。

图 4-4 解释了本模块的工作流程。本模块的源代码请参考附录 IV。

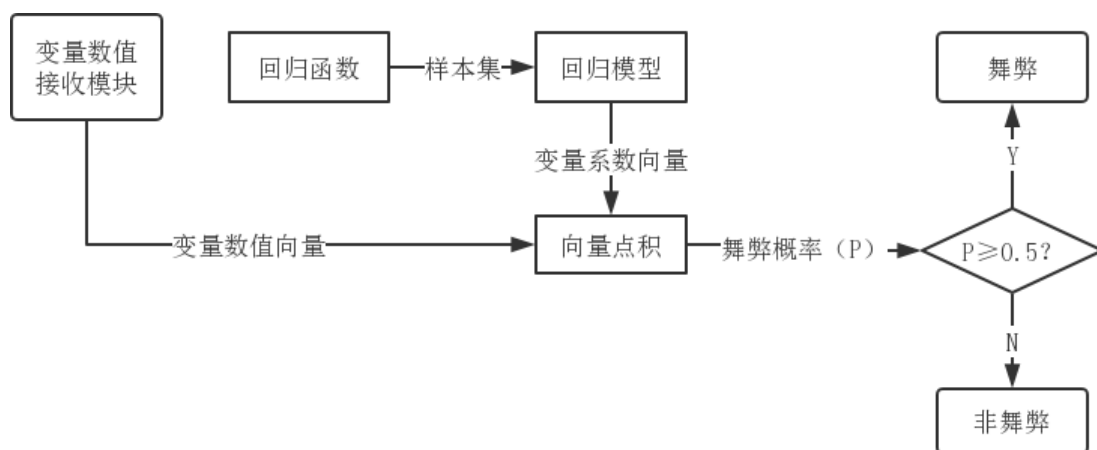


图 4-4: 基于 Python 的舞弊性分析模块工作流程

4.5 本工具的使用范例

本节选择嘉寓股份 (300117) 为例，展示本工具完整的使用过程。

(1) 启动本工具。命令行界面首先要求提供公司简称，本例中输入“嘉寓”二字后按回车提交，显示“已经找到说明书，正在下载”字样。稍等片刻后提示下载完成 (图 4-5)。此时在本工具的目录中可以看到被保存的说明书 (图 4-6)。

```

请输入待查找的公司名称（尽可能完整），按回车键确认。
输入“exit”以退出。
嘉寓
已经找到说明书，正在下载。
下载完成。

Process finished with exit code 0
    
```

图 4-5: 成功查找并下载说明书

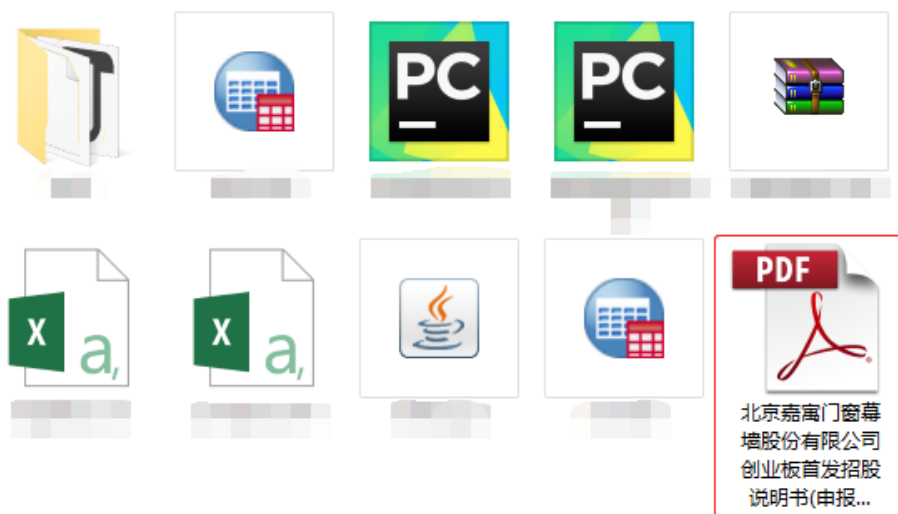


图 4-6: 保存的说明书

(2) 说明书下载完成后即进入变量数据接收模块，选择模式之后，命令行以文字提醒输入各项财务数据和指标的值。

```
选择模式：输入“1”进入财务舞弊识别方式，输入“2”进入样本集更新方式。
↓
请输入最近一年和上一年的【应收账款周转率】，两个数据之间以单个空格分隔：
2.99 3.05
请输入最近一年和上一年的【营业收入】，两个数据之间以单个空格分隔：
```

图 4-7: 输入财务数据或指标

(3) 所有需要提供的內容都输入完毕后，本工具将自动开始计算舞弊性。本例中嘉寓股份被认定为舞弊，并将该结果在命令行中反馈，如图 4-8 所示。事实上，嘉寓股份确系 IPO 财务舞弊企业（见表 3-1）。

```
该公司被认定为实施了IPO财务舞弊。
Process finished with exit code 0
```

图 4-8: 识别结果反馈

本工具以第三章建立的 IPO 财务舞弊识别的逻辑回归模型为基础，功能向前延伸至财务报表的搜寻，向后延伸至自动完成模型生成和概率计算。本工具初始设想使变量数值接收模块能够自动完成从财务报表中提取财务数据或指标的工作，然而各个公司所发布的说明书格式以及使用的货币单位不尽相同，给程序规则设置带来极大的困难，不得不转而要求手动输入，这是本工具的一大遗憾。

结论与展望

本文基于 17 家 IPO 舞弊公司和 20 家非 IPO 舞弊公司的样本数据，以应收账款周转率变动水平、毛利率变动水平、资产质量指数变动水平、营业收入变动水平、净资产回报率变动水平等五项变量为维度，建立了 IPO 财务舞弊识别的逻辑回归模型。通过分析变量的显著性水平，该模型反映出 IPO 舞弊公司可能主要通过操纵应收账款、操纵净利润来扭曲财务报表的真实性。该模型能够较好地对数据进行拟合，以样本数据进行检验的结果表明，该模型的平均识别准确率为 91%，其中被舞弊公司误判为非舞弊公司的仅 1 家，漏判率为 5.9%。对模型的各项检验结果都表明该模型具有识别 IPO 财务舞弊的较好潜能。

本文在建立了 IPO 财务舞弊识别的逻辑回归模型的同时，也以该模型为基础，通过 Python 构建了具有财务报表下载、财务数据接收和舞弊性分析功能的 IPO 财务舞弊识别工具。该工具能有效减少使用者为了识别 IPO 财务舞弊而收集数据和进行计算的时间，并能够成为衡量审计风险的参考内容。同时，该工具具有可拓展性，第一是可以随时补充和更新样本库，以得到更加准确的逻辑回归模型；第二是可以随时根据需要，对该工具进行功能上的修改或添加。以 Python 构建的该工具具有体积小、运行快、代码易读的优点，非常利于财务和审计人员对其进行使用和维护。

但是，本文也存在一些明显的不足之处。第一，由于 IPO 财务舞弊可供研究的样本极少，该模型的准确率受到一定的制约，这需要在将来通过补充更多被曝光的无比企业的数据来继续完善和提升该模型的性能表现；第二，各个企业发布的“说明书”格式参差不齐，导致工具无法实现正确解析和提取其中的财务数据，只能转通过人工输入代替，对工具的效率造成了比较明显的影响，这需要将来的证券市场“说明书”的格式进行规范统一，如此方能实现本文中工具的完全自动化。

当今社会正处于信息时代和电子时代，通过计算机存储数据能够对数据进行快速地传递、修改和计算，同时，巨量的数据中蕴藏了丰富的隐藏信息，可以通过合理的方式得到挖掘。会计电算化实现了会计工作方式的大变革，可以预见，未来在会计领域，计算机技术和数理统计理论将在审计领域得到应用，使财务和审计工作者能更加准确和理性地分析、判断公司财务状况，指导他们合理安排审计程序，提高工作效率。

致 谢

写毕业论文的过程中，我总是会回想曾经发生的点点滴滴，又偷偷地设想一下未来的光景。当时在台湾交换的我错过了大三的学年论文，所以这篇毕业论文，是我本科生涯的第一篇，也是最后一篇论文。初次撰写论文，免不了有很多生疏的地方，但是我尽力修缮每一个注意到的细节，希望能让它臻于完善，好给这段时间的忙碌画上一个完美的句号。

首先要感谢的是我的导师郭慧婷老师，从了解到我考研的志向开始，她就一直无条件地支持我，论文写作期间她也尽心尽力地指导。她对我的激励和对我的严格要求，都让我对自己的工作充满了干劲，让我感受到全身心投入的快乐与充实。这篇本身粗糙的论文，如果不经她的打磨，绝不能达到这样的状态。

然后要感谢我的家人和朋友们，他们总是支持着我，无论我有怎样的异想天开，他们绝对不会泼冷水，总是鼓励我去实现自己的设想。我很庆幸能在大学有一群完美的伙伴，他们能和我讨论各种话题，随时陪我胡吃海塞，我从他们身上学习到如何做一个友善、大方和体贴的人，如果没有大家的陪伴，我的大学四年将要暗淡不少。在大学之前认识的挚友们，虽然平时不能见面，彼此之间的感情却从来没有降温，心中许许多多的困惑和烦恼都会和他们倾诉，他们帮助我塑造起基本人格，如果没有他们，我也许不会有现在的样子。

最后想在这里缅怀身在天国的爱犬 Google。她是我去年养的一只黄色拉布拉多，乖巧、温驯，会在我写作业的时候安静地在一旁打盹，会在我开门的时候第一时间冲到面前，会在感到害怕的时候爬上我的肩头。本以为可以和她度过十几年的美好时光，却没料到犬瘟带走了她的生命。对于我来说，这是一生都无法弥补的缺憾。我曾经看过《我在雨中等你》这本书，书里说，狗在死后会投胎成人。我想如果她重生成人的话，一定是一个美丽、善良又幸福的人吧。

临近毕业，心中难免感到萧索。对于我来说，会计生涯已经告一段落，也许很多年都不会再拾起，虽然自己也难免感到可惜，但是过去的这段生活并不会因此被遗忘。我衷心祝愿各位同学和老师将来在会计领域继续发光发热，也相信自己一定能在崭新的领域找到属于自己的位置。我们的每分每秒，都不该被辜负。

参考文献

- [1] 中国证券报. 安永发布《安永全球IPO市场调研报告: 2017年回顾及2018年展望》[EB/OL]. <http://finance.sina.com.cn/roll/2017-12-27/doc-ifyqchnr6317362.shtml>, 2017-12-27.
- [2] 中国证券监督管理委员会. 中国证券监督管理委员会令第32号[EB/OL]. http://www.gov.cn/gongbao/content/2007/content_678934.htm, 2006-05-17.
- [3] 朱宝琛. IPO审核两条路径并行: 从严监管与平稳推进[EB/OL]. <http://finance.china.com.cn/stock/zqyw/20170113/4067746.shtml>, 2017-01-13.
- [4] 财经网. 因乐视IPO财务造假发审委多人被抓 最终名单超过10人[EB/OL]. <http://stock.eastmoney.com/news/1699,20171031797215473.html>, 2017-10-31.
- [5] Persons O. Using Financial Statement Data to Identify Factors Associated with Fraudulent Financing Reporting[J]. Journal of Applied Business Research, 1995, 11: 38-46.
- [6] Calderon T. G and B. P. Green. Signaling Fraud by Using Analytical Procedures[J]. The Ohio CPA Journal, 1994, 53(April): 27-38.
- [7] Ameen E. C and J. R. Strawser. Investigating the Use of Analytical Procedures: An Update and Extension[J]. Auditing, 1994, 13: 68-79.
- [8] Loebbecke J. K, M. M. Eining and J. J. Willingham. Auditors' Experience with Material Irregularities: Frequency, Nature and Detectability[J]. Auditing: A Journal of Practice and Theory, 1989, 9: 1-28.
- [9] Beneish, M. D. 1997. Detecting GAAP violation: implications for assessing earnings management among firms with extreme financial performance[J]. Journal of Accounting and Public Policy, 1997, 16: 271-309.
- [10] Beneish, M. D. Incentives and penalties related to earnings overstatements that violate GAAP[J]. Accounting Review, 1999.
- [11] Ch. Spathis, M. Doumpos and C. Zopounidis. Detecting falsified financial statements: a comparative study using multicriteria analysis and multivariate statistical techniques[J]. The European Accounting Review, 2002, 11: 509-535.
- [12] Patricia M. Dechow, Weili Ge, Chad R. Larson, et al. Predicting Material Accounting Misstatements[J]. Contemporary Accounting Research, forthcoming; AAA 2008 Financial Accounting and Reporting Section (FARS) Paper.
- [13] Ahmed Abbasi, Conan Albrecht, Anthony Vance, and James Hansen. MetaFraud: A Meta-Learning Framework for Detecting Financial Fraud[J]. MIS Quarterly, 2012, 36(4): 1293-1327.
- [14] 邢婷婷. 我国上市公司管理舞弊的特征和审计对策[J]. 中国国际财经, 2017,

15: 138.

- [15] 朱锦余, 高善生. 上市公司舞弊性财务报告及其防范与监管——基于中国证券监督管理委员会处罚公告的分析[J]. 会计研究, 2007, 11: 17-23.
- [16] 朱敏. 上市公司财务报告舞弊的识别方法及模型研究[D]. 成都: 四川大学, 2005.
- [17] 吴顺祥, 陈险峰. 试析内部审计质量的影响因素及控制措施[J]. 技术经济与管理研究, 2005, 06: 75-76.
- [18] 陈国欣, 吕占甲, 何峰. 财务报告舞弊识别的实证研究——基于中国上市公司经验数据[J]. 审计研究, 2007, 3: 88-93.
- [19] 李清, 任朝阳. 基于案例推理的财务报告舞弊识别研究[J]. 财经理论与实践, 2015, 36(03): 84-89.
- [20] 黄世忠, 黄京菁. 财务报表舞弊行为特征及预警信号综述[J]. 财会通讯, 2004, 12: 4-9.
- [21] 秦江萍. 上市公司会计舞弊:国外相关研究综述与启示[J]. 会计研究, 2005, 6: 69-74.
- [22] Wikipedia. Statement on Auditing Standards No. 99: Consideration of Fraud[EB/OL].
https://en.wikipedia.org/wiki/Statement_on_Auditing_Standards_No._99:_Consideration_of_Fraud, 2018-2-22.
- [23] 中国注册会计师协会. 独立审计具体准则第8号——错误与舞弊[EB/OL]. <http://www.chinaacc.com/new/63/66/1996/12/ad598290111622169911818.htm>, 1996-12-26.
- [24] 财政部. 中国注册会计师审计准则第1141号——财务报表审计中对舞弊的考虑[EB/OL]. <http://www.chinaacc.com/new/63/66/2006/2/ma5129451402422600220142-0.htm>, 2006-02-15.
- [25] MBA智库百科. 盈余管理[EB/OL]. <http://wiki.mbalib.com/wiki/%E7%9B%88%E4%BD%99%E7%AE%A1%E7%90%86>, 2016-08-19.
- [26] 李彬, 张俊瑞, 郭慧婷. 会计弹性与真实活动操控的盈余管理关系研究[J]. 管理评论, 2009, 06: 99-107.
- [27] 吴联生. 会计信息失真的类型划分及其治理[J]. 浙江财税与会计, 1998, 04: 14-15.
- [28] 钟黄聪. 会计信息失真的重新定义与分类[J]. 财会月刊, 2011, 02: 8-10.
- [29] Wikipedia. Id, Ego and Super-ego[EB/OL]. https://en.wikipedia.org/wiki/Id,_ego_and_super-ego#cite_note-Carlson-7, 2018-05-14.
- [30] W. Steve Albrecht. Iconic Fraud Triangle Endures[J]. Fraud Magazine, 2014, July/August: 1-7.

- [31] Singleton, T. W., Singleton, A., Bologna, J. & Lindquist, R. Fraud auditing and forensic accounting, 3rd edition[M]. NJ: John Wiley & Sons, Inc., 2006.
- [32] 新浪财经. IPO乱象系列之券商[EB/OL]. <http://finance.sina.com.cn/focus/ipolxqs/index.shtml>, 2010-8-29.

附录

I. 逻辑回归模型训练集

id1	id2	id3	id4	id5	isfraud	fraudspns
0.980328	1.07431	0.733689	1.228744	0.675211	1	1
0.886179	0.980595	0.342184	1.173005	0.76033	1	0
1.035211	0.990594	1.424998	1.352932	1.063666	1	0
1.725057	0.915127	0.793597	1.408295	0.722293	1	0
0.859555	1.132191	1.265931	1.300013	1.158907	1	1
1.048223	1.051402	1.560949	2.668117	1.092255	1	0
1.187097	0.828884	2.19869	1.164466	0.8979	1	0
0.913043	1.024855	0.769964	1.25767	0.893057	1	0
0.98722	1.119361	0.690322	1.207571	0.969215	1	0
0.852605	0.938589	1.522338	1.178348	0.492357	1	1
1.532909	0.665135	0.788725	1.283504	0.494142	1	0
1.355237	0.961442	4.026703	2.601819	1.04132	1	0
0.830409	1.266556	0.882196	1.323345	1.18936	1	1
1.175781	1.47373	0.646739	1.390097	0.910046	0	0
0.944954	1.062217	1.053788	1.326394	0.563581	0	0
1.246507	1.442069	5.439534	1.288847	1.316058	0	0
1.370683	1.232354	2.220678	1.476069	1.78828	0	0
0.508348	0.768618	0.837077	1.271636	0.990822	0	0
0.944149	0.812817	1.632351	1.415664	0.874296	0	0
1.28436	0.922422	0.569211	1.771289	1.149949	0	0
0.978698	1.026438	0.877176	1.380088	1.245996	0	0
1.066019	0.606898	1.052002	1.688317	0.948519	0	0
1.123028	1.506879	1.993322	1.286574	1.571946	0	0
0.793558	1.08444	0.7722	1.572078	1.037564	0	0
1.081197	0.774235	2.017403	1.300008	0.788363	0	0
0.906836	0.4947	0.538957	1.221352	0.517166	0	0
1.324365	1.646874	0.828275	1.785182	1.178571	0	0
1.276451	0.993195	0.947514	1.566829	1.544171	0	1
1.746238	1.029235	1.136075	1.270856	1.042398	0	0
1.426009	0.755926	1.106702	1.979855	1.121589	1	0
1.184834	1.693883	0.776321	1.333308	1.44823	1	1
1.41958	1.169926	1.293804	1.681281	1.701137	0	0
1.699591	0.707928	2.15841	1.141356	0.807873	0	0
1.117464	1.561034	1.253127	1.098024	1.701258	0	0
0.693256	1.523168	0.432824	1.742615	0.89048	0	0
1.004559	1.16654	1.470031	1.087438	0.890945	1	0
0.829305	0.706913	0.950677	0.901744	0.619224	1	0

II. 基于 Python 的财务报表下载模块源代码

```

from bs4 import BeautifulSoup
import re
import sys
import urllib.request
import urllib.parse

def getSearchURL_1(keyword):
    return
    str("http://www.csrc.gov.cn/wcm/govsearch/simp_gov_list.jsp?sword="+urllib.parse.quote(str(keyword+" 首次公开发行股票招股说明书"))+"&searchColumn=all&searchYear=all&pubwebsite=/zjhpublic/&SType=1&page=1")
def getSearchURL_2(keyword):
    return
    str("http://www.csrc.gov.cn/wcm/govsearch/simp_gov_list.jsp?sword="+urllib.parse.quote(str(keyword+" 创业板首发招股说明书"))+"&searchColumn=all&searchYear=all&pubwebsite=/zjhpublic/&SType=1&page=1")

print("请输入待查找的公司名称（尽可能完整），按回车键确认。\\n 输入“exit”以退出。")
keyword = str(input())
if keyword == 'exit':
    sys.exit()

donecnt = 0

_soup = BeautifulSoup(urllib.request.urlopen(getSearchURL_1(keyword)),
    "html.parser")
linklist = []
for x in _soup.find_all('a', href=re.compile('pub')):
    link = x.get('href')
    if link:
        linklist.append('http://www.csrc.gov.cn'+link[0:link.find('?')])
for _link in linklist:
    if donecnt == 0:
        original = urllib.request.urlopen(_link)
        soup = BeautifulSoup(original, "html.parser")
        for x in soup.find_all('a', string=re.compile('首次公开发行股票招股说明书')):
            dlink = x.get('href')

```

```

        string = x.get_text()
        downloadlink = str((_link[0:_link.rfind('/')])+dlink[1:])
        print("已经找到说明书, 正在下载。")
        urllib.request.urlretrieve(downloadlink, string)
        print("下载完成。")
        donecnt = donecnt + 1

if donecnt == 0:
    _soup = BeautifulSoup(urllib.request.urlopen(getSearchURL_2(keyword)),
    "html.parser")
    linklist = []
    for x in _soup.find_all('a', href=re.compile('pub')):
        link = x.get('href')
        if link:
            linklist.append('http://www.csrc.gov.cn' + link[0:link.find('?')])
    for _link in linklist:
        if donecnt == 0:
            original = urllib.request.urlopen(_link)
            soup = BeautifulSoup(original, "html.parser")
            for x in soup.find_all('a', string=re.compile('创业板首发招股说明书')):
                dlink = x.get('href')
                string = x.get_text()
                downloadlink = str((_link[0:_link.rfind('/')]) + dlink[1:])
                print("已经找到说明书, 正在下载。")
                urllib.request.urlretrieve(downloadlink, string)
                print("下载完成。")
                donecnt = donecnt + 1

if donecnt == 0:
    print("没有找到说明书。")

```

III. 基于 Python 的财务报表数据提取模块

```

import csv

print("选择模式: 输入“1”进入财务舞弊识别方式, 输入“2”进入样本集更新方式。")
modeid = int(input())

indexName = ['应收账款周转率', '营业收入', '营业利润', '流动资产', '固定资产', '资产总计', '净利润', '所有者权益']

```

```
indexList = ['RTR', 'OI', 'OR', 'CA', 'FA', 'TA', 'NR', 'SE']

if modeid == 1:
    loopend = 0
    while loopend == 0:
        csvfile = open("logistics.csv", "a+", newline='')
        writer = csv.writer(csvfile)

        for i, value in enumerate(indexList):
            print("请输入最近一年和上一年的【"+indexName[i]+"】，两个数据之间以
            单个空格分隔：")
            _temp = str(input())
            locals()[value] = [float(_temp[0:_temp.find(" ")]),
            float(_temp[_temp.find(" ")+1:])]
            id1 = RTR[0] / RTR[1]
            id2 = OR[0] / OI[0] / OR[1] * OI[1]
            id3 = (1 - (CA[0] + FA[0]) / TA[0]) / (1 - (CA[1] + FA[1]) / TA[1])
            id4 = OI[0] / OI[1]
            id5 = NR[0] / SE[0] / NR[1] * SE[1]

            print("保荐机构是否曾涉及舞弊？0 代表无，1 代表是。")
            fraudspns = int(input())

            writer.writerow([id1, id2, id3, id4, id5, fraudspns])
            print("如果训练集录入没有结束，请输入 0 继续，输入 1 结束。")
            loopend = int(input())

        csvfile.close()
    else:
        if modeid == 2:
            for i, value in enumerate(indexList):
                print("请输入最近一年和上一年的【"+indexName[i]+"】，两个数据之间以
                单个空格分隔：")
                _temp = str(input())
                globals()[value] = [float(_temp[0:_temp.find(" ")]),
                float(_temp[_temp.find(" ")+1:])]
                id1 = RTR[0] / RTR[1]
                id2 = OR[0] / OI[0] / OR[1] * OI[1]
                id3 = (1 - (CA[0] + FA[0]) / TA[0]) / (1 - (CA[1] + FA[1]) / TA[1])
                id4 = OI[0] / OI[1]
                id5 = NR[0] / SE[0] / NR[1] * SE[1]

                print("保荐机构是否曾涉及舞弊？0 代表无，1 代表是。")
                fraudspns = int(input())
                valvar = [int(1), id1, id2, id3, id4, id5, fraudspns]
```

IV. 基于 Python 的财务报表舞弊性分析模块

```
import math
import os

cmd = "cmd R BATCH LogisticGen.R"
os.system(cmd)

coefvar = []
coefreader = csv.reader("coef.csv")
column = [row[2] for row in reader]
coefvar.append(float(column))

expn = 0
i = 0
while i <= len(valvar) + 1:
    expn += valvar[i] * coefvar[i]
prob = math.exp(expn) / (1 + math.exp(expn))

if prob >= 0.5:
    print("该公司被认定为实施了 IPO 财务舞弊。")
else:
    print("该公司被认定为未实施 IPO 财务舞弊。")

os.remove(r'coef.csv')
'''
LogisticGen.R 源代码:
#!/path/to/Rscript
samples = read.csv("samples.csv")
gen = glm(formula = isfraud~., data = samples, family = "binomial")
for (i in 1:7) write.csv(coef(gen)[i], "coef.csv")
'''
```

V. 本文相关资源下载

(1) IPO 财务舞弊识别工具: 链接: <https://pan.baidu.com/s/1PpsmBJChlGQ S38TqKfz8HA>, 密码: 6qkm。

(2) Python 3.6.5 (64-bit): <https://www.python.org/ftp/python/3.6.5/python-3.6.5-amd64.exe>。

(3) R: <https://cran.r-project.org/mirrors.html>。