

An Empirical Study of the Relationship Between Financing Behavior and Credit Risk of  
Small and Micro Merchants in China

by

Chaoyang Li

A Dissertation Presented in Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Business Administration

Approved March 2021 by the  
Graduate Supervisory Committee:

Benjamin Shao, Co-Chair

Xianglin Li, Co-Chair

Xiaodan Dong

ARIZONA STATE UNIVERSITY

May 2021

国内小微商户融资行为与信用风险关系的实证研究

——基于 H 公司小微商户贷款业务的案例分析

厉朝阳

全球金融工商管理博士  
学位论文

研究生管理委员会  
于二零二一年三月批准：

邵保民， 联席主席  
李祥林， 联席主席  
董小丹

亚利桑那州立大学

二零二一年五月

## ABSTRACT

This dissertation studies the small and micro loan products of Company H, which provides financing services to domestic small and micro merchants in China, through the collection and analysis of the pre-loan and post-loan data of customers to reveal the financing channels used by such merchants over time and further explore the quantity and composition of these financing channels as well as their relationship with credit risk levels. The findings show that: (1) most of the financing channels of small and micro merchants in China are associated with non-traditional financial institutions other than traditional ones like banks; (2) in view of the diversification of these financing channels, PBOC Credit Reference from the People's Bank of China alone is not enough to evaluate multiplatform loan, and reports from other large big data companies should also be used; (3) the quantity and composition (such as banks, P2P, small loan companies, and consumer financial institutions) of financing channels used by small and micro merchants are significantly related to their credit risk levels; (4) both the PBOC Credit Reference and the multiplatform loan information provided by the third-party big data companies can help small and micro loan institutions to assess the credit risk levels of small and micro merchants more accurately; and (5) higher frequency of financing behavior leads to higher interest rate of loans and earlier credit default. Since a small and micro merchant is a business entity between individuals and enterprises, it has the

characteristics of small scale, large base, short life cycle, and high risk. It's a potential group that has been ignored and not been well served by traditional financial institutions. This dissertation aims to reveal the behavioral patterns of this borrowing group in applying for financing and identify the relationship with their credit risk, so as to help loan institutions like Company H, which are committed to serving the small and micro merchants, conduct better risk-based pricing.

## 摘要

本文以 H 公司（致力于为国内小微商户提供融资服务的贷款公司）的小微贷款产品为案例，通过收集和分析该产品放款客户的贷前贷后数据，对国内小微商户申请融资的渠道进行阐述和分析，并进一步探索一段时间内小微商户申请融资渠道的数量及组成成分，与其贷后信用风险水平之间的关系。

本研究表明：（1）国内小微商户的融资渠道很大部分来源于银行以外的非传统金融机构；（2）鉴于国内小微商户融资渠道的多样化，仅靠人民银行征信报告不足以评估其多头申请水平，可借助其他大型的第三方大数据公司；（3）小微商户申请融资渠道的数量和组成成分（如银行/P2P/小额贷款公司/消费金融机构），与其信用风险水平有显著关联；（4）结合人民银行征信报告和第三方大数据公司提供的多头申请信息，可帮助小微贷款机构更精准地评估小微商户的信用风险水平；（5）融资行为越频繁，融资渠道越下沉，越早发生较严重的信用违约，即信用风险水平越高。

本文研究对象——小微商户，是介于个人和企业之间的微型经营实体，具有单户规模小、总体基数大、生命周期短、风险水平高、信用评估困难等特征，是包括银行的众多正规传统金融机构，还无法深入渗透和服务的潜力客群。通过本文的探索，将揭示和刻画这类群体申请融资的主要渠道和客观现状，并寻求与信用风险水平关系的分析结论，从而帮助像 H 公司这样致力于服务小微客群的贷款机构，更精准地评估客户的资金需求程度和信用风险水平，给予更合理的风险定价。

## 目录

	页码
表格列表.....	vii
图表列表.....	ix
章节	
一、引言.....	1
1.1 小微商户定义.....	1
1.2 小微商户融资困境.....	1
1.3 研究的核心假设和意义 .....	2
二、文献综述.....	4
2.1 小微融资研究 .....	4
2.1.1 小微融资渠道.....	4
2.1.2 小微融资困境研究.....	5
2.1.3 解决小微融资困境对策研究 .....	6
2.2 小微企业信用风险研究.....	8
2.2.1 小微信用风险评价指标体系研究.....	8
2.2.2 小微信用评级特点.....	10
2.2.3 小微企信用评价方法研究.....	11
2.3 本文创新点 .....	12

章节	页码
三、理论基础和研究方法 .....	13
3.1 理论基础 .....	13
3.2 核心假设.....	21
3.3 数据来源.....	22
3.4 研究方法.....	22
3.5 计量模型和指标 .....	23
3.5.1 模型：逻辑回归(Logistic Regression).....	23
3.5.2 模型：决策树(Decision Tree).....	24
3.5.3 指标：IV 值(Information Value).....	25
3.5.4 指标：决策树特征重要性(Feature Importance).....	27
四、研究步骤简述 .....	29
4.1 贷前数据获取 .....	29
4.1.1 人行征信报告.....	29
4.1.2 同盾等第三方数据 .....	31
4.1.3 有效样本的筛选原则 .....	33
4.2 贷后数据的整合和目标定义.....	34
4.2.1 逻辑回归：好坏客户定义(Y=0 或 1) .....	35
4.2.2 决策树：不同违约期限结构下的客户分层 (A/B/C) .....	35

章节	页码
4.3 融资行为相关的 X 特征变量构造.....	36
4.4 核心假设验证和有效变量筛选.....	37
4.5 构建基于融资行为变量的子信用评分卡 .....	38
4.6 形成基于融资行为变量的风控规则.....	39
五、数据分析和结论.....	40
5.1 小微商户融资行为刻画.....	40
5.1.1 基于人行征信的融资行为刻画 .....	41
5.1.2 基于同盾多头的融资行为刻画.....	42
5.2 融资行为发生的频率.....	45
5.2.1 基于人行征信的融资行为频率统计 .....	45
5.2.2 基于同盾多头的融资行为频率统计 .....	50
5.3 融资渠道构成与分布.....	55
5.3.1 基于人行征信的融资渠道构成分析 .....	55
5.3.2 基于同盾多头的融资渠道构成分析 .....	62
5.4 融资行为频率与信用风险的关系 .....	65
5.4.1 基于人行征信的融资行为频率与信用风险的关系 .....	65
5.4.2 基于同盾多头的融资行为频率与信用风险的关系.....	79
5.4.3 “融资行为频率与信用风险关系”结论的稳定性分析 .....	89



章节	页码
5.5 融资渠道分布与信用风险的关系.....	97
5.5.1 基于人行征信的融资渠道分布与信用风险的关系.....	98
5.5.2 基于同盾多头的融资渠道分布与信用风险的关系.....	108
5.5.3 “融资渠道分布与信用风险关系”结论的稳定性分析.....	115
5.6 融资行为对风险预测的单变量分析.....	121
5.6.1 融资行为变量的 IV 值.....	121
5.6.2 融资行为变量的决策树特征重要性.....	125
5.7 基于融资行为的信用风险评分卡模型.....	127
5.7.1 评分模型训练.....	128
5.7.2 评分模型验证.....	135
5.8 融资行为变量对信用风险模型的提升效果.....	137
5.8.1 模型 1: 不含融资行为变量的评分卡模型效果.....	137
5.8.2 模型 2: 含融资行为变量的评分卡模型效果.....	142
5.9 融资行为与信用风险暴露期限的关系探索.....	147
5.9.1 决策树分析探索.....	148
5.9.2 生存曲线分析探索.....	149
六、总结陈述.....	150
参考文献.....	155

## 表格列表

表格	页码
1 : IV 值范围和对应预测力.....	26
2 : 人行征信查询记录字段 .....	30
3 : 同盾多头报告字段.....	32
4 : H 公司小微商户样本有效报告情况.....	34
5 : 评分卡示例.....	38
6 : 不同时间切片的融资行为频率 .....	45
7 : 同盾近 3 个月内申请平台总数分布 .....	51
8 : 机构“贷款审批”查询客户覆盖率 TOP50.....	56
9 : 机构“贷款审批”人均被查询次数 .....	60
10 : 不同融资渠道类型“贷款审批”查询情况.....	61
11 : 不同融资渠道类型“贷款审批”查询情况按年统计分析 .....	62
12 : 同盾不同融资渠道类型占比按年统计分析 .....	64
13 : 基于人行征信的融资行为频率相关指标变量.....	66
14 : 基于同盾多头的融资行为频率指标 .....	79
15 : 同盾主要的贷款融资类平台类型 .....	81
16 : 基于人行征信的融资渠道分布指标 .....	99
17 : 同盾多头融资渠道占比相关变量 .....	108

表格	页码
18 : 样本总体好坏客户分布.....	122
19 : IV 值较高的单变量及含义.....	123
20 : 训练集和验证集样本量及好坏样本.....	128
21 : WOE 分箱示例 (变量 var_P2P).....	129
22 : 逻辑回归最终入模变量 .....	130
23 : 逻辑回归最终入模变量含义.....	130
24 : “融资行为”子评分卡.....	132
25 : 模型 1 (不含融资行为变量) 评分卡.....	138
26 : 模型 2 (含融资行为变量) 评分卡.....	142

## 图表列表

图表	页码
1: sigmoid 函数 .....	24
2: 决策树 .....	25
3: 决策树特征重要性示例 .....	28
4: 特征变量构造方式 .....	36
5: 决策树式风控规则示例 .....	40
6: 人行征信查询记录示例 .....	42
7: 近 90 天贷款审批+个人查询 次数分布 .....	47
8: 近 180 天 贷款审批+个人查询 次数分布 .....	48
9: 近 90 天查询次数分布 .....	49
10: 近 180 天查询次数分 .....	50
11: 同盾近 3 个月申请 P2P 网贷平台数.....	52
12: 同盾近 3 个月申请小额贷款公司平台数.....	53
13: 同盾近 3 个月申请一般消费分期平台数.....	53
14: 同盾近 3 个月申请大型消费金融公司平台数 .....	54
15: 同盾近 3 个月申请银行消费金融公司平台数 .....	54
16: 机构“贷款审批”查询次数 TOP50 .....	58
17: 同盾近 3 个月申请融资渠道构成分析.....	63

图表	页码
18: 近 90 天 贷款审批+个人查询 次数分布和违约率.....	67
19: 近 180 天 贷款审批+个人查询 次数分布和违约率 .....	68
20: 近 360 天 贷款审批+个人查询 次数分布和违约率 .....	68
21: 近 2 年 贷款审批+个人查询 次数分布和违约率 .....	69
22: 近 90 天 个人查询 次数分布和违约率.....	70
23: 近 180 天 个人查询 次数分布和违约率.....	70
24: 近 360 天 个人查询 次数分布和违约率.....	71
25: 近 2 年 个人查询 次数分布和违约率 .....	71
26: 近 90 天 银行 贷款审批 查询次数分布和违约率 .....	72
27: 近 180 天 银行 贷款审批 查询次数分布和违约率.....	73
28: 近 360 天 银行 贷款审批 查询次数分布和违约率.....	73
29: 近 2 年 银行 贷款审批 查询次数分布和违约率 .....	74
30: 近 90 天 消费金融公司 贷款审批 查询次数分布和违约率 .....	75
31: 近 180 天 消费金融公司 贷款审批 查询次数分布和违约率 .....	75
32: 近 360 天 消费金融公司 贷款审批 查询次数分布和违约率 .....	76
33: 近 2 年 消费金融公司 贷款审批 查询次数分布和违约率.....	76
34: 近 90 天 小额贷款公司 贷款审批 查询次数分布和违约率 .....	77
35: 近 180 天 小额贷款公司 贷款审批 查询次数分布和违约率.....	78

36: 近 360 天 小额贷款公司 贷款审批 查询次数分布和违约率 .....	78
37: 近 2 年 小额贷款公司 贷款审批 查询次数分布和违约率 .....	79
38: 近 3 个月 同盾多头申请 平台数分布和违约率 .....	82
39: 近 3 个月 同盾多头申请 P2P 网贷 平台数分布和违约率.....	83
40: 近 3 个月 同盾多头申请小额贷款公司 平台数分布和违约率 .....	83
41: 近 3 个月 同盾多头申请一般消费分期平台 平台数分布和违约率 .....	84
42: 近 3 个月 同盾多头申请大型消费金融公司 平台数分布和违约率 .....	85
43: 近 3 个月 同盾多头申请银行消费金融公司 平台数分布和违约率 .....	86
44: 近 3 个月 同盾多头申请融资租赁 平台数分布和违约率.....	87
45: 近 3 个月 同盾多头申请银行小微贷款 平台数分布和违约率 .....	88
46: 近 3 个月 同盾多头申请互联网金融门户 平台数分布和违约率.....	89
47: H 公司小微商户样本中各省份的样本量及占比 .....	90
48: 四个省份内, 人行 近 90 天 贷款审批+个人查询 次数分布和违约率 .....	91
49: 四个省份内, 近 3 个月 同盾多头申请 平台数分布和违约率 .....	93
50: H 公司小微商户样本中各进件季度的样本量及占比.....	95
51: 2017-2019 各个年度, 近 90 天 贷款审批+个人查询 次数分布和违约率 .....	96
52: 基于人行征信的融资渠道分类 .....	98
53: 近 180 天 正规金融机构 贷款审批 查询次数占比分布和违约率.....	101

54: 近 360 天 正规金融机构 贷款审批 查询次数占比分布和违约率 .....	101
55: 近 2 年 正规金融机构 贷款审批 查询次数占比分布和违约率.....	102
56: 近 180 天 银行 贷款审批 查询次数占比分布和违约率 .....	103
57: 近 360 天 银行 贷款审批 查询次数占比分布和违约率 .....	103
58: 近 2 年 银行 贷款审批 查询次数占比分布和违约率 .....	104
59: 近 180 天 消费金融公司 贷款审批 查询次数占比分布和违约率.....	105
60: 近 360 天 消费金融公司 贷款审批 查询次数占比分布和违约率.....	105
61: 近 2 年 消费金融公司 贷款审批 查询次数占比分布和违约率 .....	106
62: 近 180 天 小额贷款公司 贷款审批 查询次数占比分布和违约率 .....	107
63: 近 360 天 小额贷款公司 贷款审批 查询次数占比分布和违约率 .....	107
64: 近 2 年 小额贷款公司 贷款审批 查询次数占比分布和违约率 .....	108
65: 近 3 个月 同盾多头申请 P2P 网贷 平台占比分布和违约率 .....	110
66: 近 3 个月 同盾多头申请 P2P 网贷 平台占比分布和违约率, 限定“近 3 个月申请平台 数>5” .....	110
67: 近 3 个月 同盾多头申请 小额贷款公司 平台占比分布和违约率 .....	111
68: 近 3 个月 同盾多头申请 一般消费分期平台 平台占比分布和违约率 .....	112
69: 近 3 个月 同盾多头申请 大型消费金融公司 平台占比分布和违约率 .....	113
70: 近 3 个月 同盾多头申请 银行消费金融公司 平台占比分布和违约率 .....	115

71: 四个省份内, 人行 近 2 年 正规金融机构 贷款审批 查询次数占比分布和违约率..	116
72: 2017-2019 各个年度内, 人行 近 2 年 正规金融机构 贷款审批 查询次数占比分布和 违约率.....	118
73: 决策树特征重要性结果 .....	126
74: 选定模型的 ROV 曲线和 AUC 值.....	131
75: 训练集(DEV)样本的评分结果分布和实际违约率.....	134
76: 训练样本(DEV)和验证样本(OOT)的 ROC 曲线.....	135
77: 验证集(OOT)样本的评分结果分布和实际违约率.....	136
78: 模型 1 的训练样本(DEV)和验证样本(OOT)的 ROC 曲线 .....	139
79: 模型 1 的训练集(DEV)样本的评分结果分布和实际违约率 .....	140
80: 模型 1 的验证集(OOT)样本的评分结果分布和实际违约率.....	141
81: 模型 2 的训练样本(DEV)和验证样本(OOT)的 ROC 曲线.....	145
82: 模型 2 的训练集(DEV)样本的评分结果分布和实际违约率 .....	145
83: 模型 2 的验证集(OOT)样本的评分结果分布和实际违约率 .....	146
84: var_P2P 与信用风险暴露期限关系决策树 .....	148
85: var_p2p 分组下的生存曲线.....	149



## 一、引言

### 1.1 小微商户定义

关于小微商户的定义，中国人民银行在对 2017 年底颁布的《中国人民银行关于规范支付创新业务的通知》的解读中，就明确指出，“小微商户指的是按法律、法规规定，可以免于工商登记的实体的商户”。更具体的，根据《银联卡收单机构小微商户风险防控指引》中对小微商户的划定标准，营业面积不超过 300 平米，单店从业人员不超过 10 人，单店月均营业收入不超过 100 万的商家，属于小微商户的范畴。中国有超 7000 万的小微商户，贡献了全国 80% 以上的就业，但在融资问题上，仍然面临重重障碍。

### 1.2 小微商户融资困境

小微商户客群，与个人相比，具有更频繁和大额的资金周转需求，与企业相比，又缺乏充足的抵押物和标准化的信息体系（包括财务报表、企业征信报告、税务数据等企业维度信息），这使得银行的个人消费贷款产品不足以满足小微商户客群的融资额度需求，但申请更大额的银行经营性贷款，又常常因无抵押物以及信用评估难度大，而被拒之门外。融资需求在传统金融机构得不到满足，许多小微商户往往只能选择利率更高但门槛更低的 P2P 公司和民间借贷公司，从而大大增加了融资成本，使得原本存活率就低的这类群体经营得更加艰难，信用违约风险不断增加，直至最终在高利率高负债下不堪重压无奈关门。

小微商户的融资难融资贵问题，是一个社会性难题，甚至是世界性难题。本研究中并不打算对这个广泛的命题探讨宏观长远的解决方案，“建立更完善的信用体系”、“降低融资

成本和门槛”、“大数据打破信息不对称”等等，都是当前热门的观点和方向。但对于一家贷款机构来说，最迫切和关注的难题是，如何基于贷前阶段掌握的信息和数据，从金融行为到风险属性，更精准刻画和深入了解小微商户这一群体，从而实现贷前风险的有效预测、贷中风险的主动管理，最终在保证自身财务收益的前提下，为更广大的小微商户缓解融资难问题，实现真正的“普惠金融”。

### 1.3 研究的核心假设和意义

本研究中的H公司，是一家致力于为小微商户提供专业融资服务的贷款机构，门店遍布全国十几个省份，服务客户规模达15000人左右，积累了丰富的小微信贷业务经验和评价客户信用资质的多维度数据。在与小微客群的长期密切接触中，可以深切感受到他们融资的迫切性和高难度。从最开始主要向银行申请低息的经营性贷款，到求助于高利率的P2P公司和小贷公司，直至最终一部分客户因超负债和资金链断裂走向违约和不良，这一系列的过程伴随着客户申请融资渠道的数量和结构的变化。

基于对这一系列变化的观察和总结，产生了本次研究中的核心假设：

“小微商户的近期融资行为，与其未来一段时间的信用风险水平，具有显著相关性”。

这个假设关系到以下几个子问题：

- 1) 小微商户融资行为如何刻画？如何获取数据并建立合理指标？
- 2) 小微商户申请融资的渠道（银行/P2P/小贷公司等）数量和分布情况？
- 3) 高风险的小微商户和低风险的小微商户，在申请融资行为上有什么区别，是否能找到具有区分力的特征指标？

4) 小微商户融资行为的相关特征指标，对预测客户短期和长期的违约率效果如何（用统计指标衡量）？

5) 如何将融资行为维度的特征变量，用于建立有效的风控模型和规则，最终提升小微商户的业务水平？

将融资行为和风险水平关联起来，是更具有前瞻性、及时性和解释力的风险预测方式，这是因为：

- a. 融资行为和风险水平之间是互为因果的关系，从逻辑上就很容易解释；
- b. 申请融资的行为刻画能更综合直接地反映客户的主观意愿和客观现状；
- c. 不需要等到客户出现实质性的逾期，就能更早监测到信用风险的暴露；
- d. 更动态地评估，客户在不同阶段的行为特征，所预示的风险水平变化；

## 二、文献综述

### 2.1 小微融资研究

#### 2.1.1 小微融资渠道

国内也有不少学者对小微企业融资渠道划分和选择影响因素进行了大量研究，比如田秀娟（2009）区分融资渠道为正规融资渠道和非正规融资渠道两种，并且提到中小企业的融资需求特征（无健全财务体系、缺乏可抵押和担保的有效资产、缺乏信用评级）强化了非正规金融的比较优势。梁琦等（2005）得出民营企业生命周期融资规律，即生命早期阶段的企业不得不依赖自有资金、商业信誉、民间借贷和亲友借款，这种依赖程度随企业发展阶段的演进而降低。

小微融资难融资贵的问题由来已久，在国内众多研究和报道中被广泛提到，某种程度上已经成为社会共识。在互联网金融快速发展的今天，除了银行等传统金融机构外，小微客群有了更多新型融资渠道。刘子琪（2019）提到，互联网金融环境下小微企业的主要融资渠道，包括：（1）大数据金融模式：例如基于阿里巴巴互联网平台的网贷模式，借助平台自身积累和数据模型进行放贷；（2）P2P网络贷款：尽管这种模式当前已经在监管引导下走向转型，但它代表了一种基于第三方信用评估的高利率低门槛贷款模式；（3）众筹模式：借助互联网平台，为需要贷款的投资项目筹集资金。这些新型融资渠道，解决了传统金融贷款模式下小微企业融资门槛高、金额小、低效手续繁琐等问题。

### 2.1.2 小微融资困境研究

Niklis（2014）根据信贷配给的概念，即尽管借款人愿意支付合同中的所有非价格条件和价格条件，但是其贷款需求仍然无法得到满足。其从信贷配给的角度对小微企业融资问题进行了探索，认为信贷配给来源于信息不对称，又分为逆向选择的影响和道德风险的影响两类。

Harris（2015）认为小微企业融资障碍主要来源于商业银行。此类研究认为银行为实现规模扩张，进行大量的并购活动，导致小规模企业的信贷数量下降；由于大银行与小银行在信贷标准、组织结构等方面存在差异，其借贷成本高于小银行，导致不愿意向小规模企业贷款。

Horata（2013）认为小微企业自身是导致其融资障碍的主因。这一类研究认为小微企业成立时间较短，缺少详尽的信贷记录，因而商业银行无法考察企业信用状况；另一方面中小企业经营不规范，财务信息混乱，较大的贷款风险导致小微企业贷款申请通过率低。

张璐昱（2018）从可提供融资服务的金融机构的角度来看，银行对小微企业普遍存在着“惜贷”行为；村镇银行、小额贷款公司等规模普遍较小，数量较少且发展并不完善，很难满足小微企业的融资需求；地方性商业银行或农村信用社等，往往受制于自身的资金规模，所以对小微企业的融资支持也非常有限；从我国信用担保机制来看也还不健全，缺少专门针对性配套服务小微企业提供融资服务的担保机构，这也在一定程度上制约了小微企业通过抵押和担保等有效手段在金融机构进行融资。

冯雪莲（2018）从主要提供小微企业融资的银行来看，银行作为理性的经济实体在续作贷款业务时必然会考虑其自身的盈利性和安全性，且侧重考虑其经济利益。所以在同等条件下银行发放贷款倾向会更偏好于大型或中型企业，再考虑到小微企业普遍信用评级较低或缺少信用评级，银行的“惜贷”行为也就不难理解。张苗苗（2019）认为小微企业资金需求往往体现为“短、小、急”，但目前商业银行大都缺乏创新，缺少适用于小微企业融资需求的产品。小微企业种类繁多，行业分布广泛，现有模式很难满足小微企业的融资需求。

王家辉（2019）提出由于小微企业或成立不久，或自身管理不完善，从而公开信息较少、信息披露不完全、财务报表不详细，加上社会上又缺乏针对小微企业的完善的信用评级体系，对小微企业进行信用评级很难，同时由于小微企业还未建立起自己的品牌，因此小微企业很难得到准确的信用评估。在这种情况下，只有小微企业自己知道自己的经营状况、风险状况、贷款动机和偿还能力等，但银行相对处在信息不对称劣势方，因此为了保险起见，为了减少道德风险的发生，银行就很容易从审慎经营角度出发选择减少对小微企业的贷款。另一方面，余青（2019）认为银行推出的融资产品，小微企业往往缺乏了解，不清楚产品的内容，不熟悉审批的过程，不知道要提交怎样的材料，而银行在不了解具体信息，没有提供更多咨询服务的情况下，也很难为小微企业提供贷款。

### 2.1.3 解决小微融资困境对策研究

陈凌白（2019）根据小微企业的实际情况，接受应收账款抵押或是用公司的商标权、知识产权做以抵押；在担保方式上，可以考虑小微企业联合进行担保，或是请地方政府在

政策上予以支持。对于一些更小规模的微型企业，在条件允许的情况下也可以考虑尝试引入自然人担保。而对信誉情况较好盈利情况明显的小微企业，可以逐步扩大无抵押无担保贷款的模式。

金融产品的开发，王文海（2019）针对于小微企业的特点，针对不同发展阶段、不同风险程度、不同行业类型的小微企业，所开发出的信贷产品要有所区别，从而满足不同类型企业的不同融资需求，同时还要考虑到所开发的产品在现实中的可操作性。比如，可以考虑小微企业生产产品的季节性或是所贷资金的用途等开发相应的融资产品；针对很多小型企业为大型企业提供中间产品或服务等特点，目前出现的供应链金融也不失为一种产品创新的非常有效的尝试。

陈隆（2018）认为由于信息披露不完全，在这种情况下，银行应更加重视对小微企业非财务信息的收集和利用。可以通过实地考察来调查收集和分析这些信息，包括产销情况、销售收入情况、雇主信息和不良贷款记录等，同时重点关注考察客户的还款来源和还款能力。同时，李素红（2018）提出各商业银行在小微企业贷款业务中应建立客户评级系统，引入自动评级体系，将收集来的各种信息转化为标准化的信用评级指标，由系统进行评分并自动审批，这样可以极大地提高审批处理效率。当然在授信体系的搭建方面，小微企业信息数据库的建立与完善肯定是离不开国家政策支持 and 引导的。

刘晓静（2018）提到在小微企业的贷款中银行还应加强贷后管理环节，保持对小微企业的持续动态监督并不断完善自身的信用评级体系，在风险早期可以预先警示，提高贷款安全性，不断降低小微贷款不良率。

韩曦莹（2018）提出银行要为小微企业提供全方位服务，帮助小微企业迅速成长。小微企业普遍缺乏高水平的财务人员和规范的财务制度，银行可以利用自身优势提供系统性的财务人员培训或进行针对性的财务指导，提升小微企业财务管理水平。同时银行利用自身信息资源优势，可以搭建信息平台，为小微企业答疑解惑，提供最新优惠政策解读，提供相应的咨询服务。看到小微企业经营管理中存在的风险问题，及时作出风险提示。

常磊（2019）面向小微企业的政策性银行是由政府财政出资，且不以盈利为目的，以小微企业为服务对象，为其提供融资支持或担保服务的金融机构。由于小微企业数量多且大部分分布在乡镇地区，因此针对于小微企业的政策性银行也应扎根在基层，只有这样，才能真正了解和掌握小微企业的经营情况和融资需求，切实为小微企业服务。王娜

（2018）在业务开展上，政策性银行除了专门为小微企业提供贷款服务外，还应该为当地的小微企业提供担保服务，弥补我国担保体系建设上的不足，缓解小微企业由于缺少抵押担保物而在银行很难获得贷款的尴尬。同时，可以拓展其业务范围，走基层下基层，为小微企业提供业务指导、风险防控、咨询服务和资金规划等金融服务。

## 2.2 小微企业信用风险研究

### 2.2.1 小微信用风险评价指标体系研究

“5C 评价原则”作为“权威评级指标体系”得到金融界普遍认可。它从品德（Character）、资本（Capital）、能力（Capacity）、担保（Collateral）、经营状况（Condition of Business）等五个方面，对借款人的信用风险（即还款能力和还款意愿）进行评价。雷晓敏等（2008）认为“企业家是决定企业信用最关键的因素”，因此基于“5C”



构造了（1）企业家的品质；（2）企业家的能力；（3）企业家的财产；（4）抵押担保品；（5）环境条件这五个模块的指标体系。

穆迪、标准普尔等企业信用评价指标体系，则纳入了大量的企业财务指标，例如流动比率、资产报酬率等。国内各大商业银行，例如中国工商银行（2005）的小企业法人客户信用风险评价指标体系就包含了资产负债率等指标；中国农业银行（2010）的小企业信用等级等指标体系包含了主营业务收入负债比等指标。

更全面的指标体系涵盖了宏观和微观、财务和运营、企业和人等更大范围。吾兰（2014）基于 1311 个某城商行贷款商户样本，通过分析商户基本情况及运营等方面因素，建立了商户贷款信用风险评价指标体系，这些指标来源于以下几个维度：（1）基本情况：学历、婚姻状况等；（2）保证联保：有无担保、保证人信息等；（3）偿债能力：流动比率、资产负债率等；（4）盈利能力：净利润、净资产报酬率等；（5）营运能力：应收账款周转率、经营稳定性等；（6）宏观环境：人均储蓄余额、经营区域定位等。同样的，石宝峰等（2017）利用某国有商业银行 2157 个商户小额贷款样本数据，从基本情况、保证联保、偿债能力、营运能力、宏观环境等六个方面海选指标，采用 Logistic 回归方法求出商户的信用评分，并在商业银行得到实践应用。孟斌（2014）则选用某全国性商业银行总行信贷数据库内京、津、沪、渝等地区的 14 个分行的 3847 笔商户的小额贷款借据，根据定性指标的打分标准对其数据进行打分，其指标维度同样来自于基本情况、保证联保、偿债能力、营运能力、宏观环境等六个方面。

### 2.2.2 小微信用评级特点

Xiong W(2013)从信用卡销售信息当中提取了有关小微企业销售规模和业务品类信息，希望以此类信息代替财务信息，进而解决小微企业财务信息难以取得和公开信息验证问题。PAN（2012）建立了包括盈利能力、资本构成、经营能力等在内的五维度中小企业破产预测指标体系。Shankar(2019)通过研究发现小微企业的生存与发展直接受行业发展情况、商业周期波动等因素影响。Johnny（2013）调用了美国社区银行的小微企业数据库，从中选取了部分小微企业的财务指标和非财务指标完成了对该部分小微企业的信用打分。GRAY等(2014)开创性的从企业信用信息共享程度角度建立了德国小微企业贷款违约状况指标体系。谭新莹（2018）指出，大部分小微企业都是劳动密集型企业，受规模限制其生产的产品很多都是中间产品，因此其市场中的生存状况很大程度上就取决于其上下游企业的发展。周良增（2018）指出，充分利用供应链和小微企业特点开发而成的供应链金融模式，成为解决小微企业融资难和商业银行增加新盈利渠道的“双赢”选择。供应链金融下的信用风险评价体系的构建不仅需要考察小微企业自身的风险因素，还应该对供应链的整体运营和竞争力上下游企业的合作状况、小微企业在核心企业供应链中所处地位等因素进行综合评价。夏晗（2019）指出，由于我国社会信用体系建设还处在起步阶段，信用环境差、违约成本低、导致恶意拖欠、信用缺失现象十分普遍，因此我国的企业信用评价需要更加注重对其偿债意愿的考量。

### 2.2.3 小微企业信用评价方法研究

李伟（2018）针对小微企业非财务指标的评分，认为影响小微企业信用评价的非财务指标为定性指标，指标之间并非平级关系且相互之间具有一定的模糊性，其优劣程度难以量化界定，因此采用模糊综合分析法对非财务指标进行评价，构建了一个二级三层非财务指标模糊综合评价模型，在求出企业信用综合评价值以后，找出相应的评语集，确定其信用等级从而实现定量评价。

钟琳（2018）融合“5C 要素分析法”和“5P 要素分析法”两种信用评价方式，结合 AHP 层次分析法和模糊综合评价决策的核心思想，建立了一套能够反映小微企业特性和互联网技术特性的小微企业综合信用评价模型，为互联网金融服务机构提供了一种专门针对小微企业的信用评价模型。

朱文超（2018）根据 AHP 所确定的矩阵以及在此基础上计算出的权重，运用模糊综合评价来进一步的完成模型的构建。

张建红（2018）基于神经网络算法，建立了多层感知器信用评价模型，并进行了实证检验。通过与传统模型比较，发现利用多层感知器模型能够对科技型小微企业进行较为准确的信用评价。神经网络模型具有较高的预测精度和较低的误分类代价，能够降低金融机构的信贷风险，同时缓解科技型小微企业与金融机构之间的信息不对称。

周超（2018）采用神经网络模型对小微企业进行信用风险评价，并证明该模型确实可行，能够应用与供应链中各小微企业的信用风险评价。

### 2.3 本文创新点

基于对小微融资渠道和信用风险相关的文献回顾，发现目前为止国内尚未把小微融资行为和风险水平关系进行关联的定量研究。对于小微融资渠道和选择的研究，主要集中在宏观定性的政策性分析，以及基于地方性样本的调研上。而对于小微信用风险的研究，则更关注全面指标体系的建设，特别是借鉴国外的成熟理论在国内加以验证，而极少有探究国内小微商户融资行为与风险之间的关联性问题的，更不用说基于真实样本定量地去分析和验证。

除了研究方向的创新外，本文使用的客群样本数据也与以往大量研究中主要来源于银行授信客群的情况显著不同，是相比之下更为下沉的客群。H 公司服务的小微商户客群以个体工商户和微型企业居多，无论是在经营规模上还是抗风险能力上整体差于银行客群，但却更能代表国内几千万小微商户，通过对这些客群的融资行为研究，也更能体现国内广大小微商户所面临的融资困境——在银行等正规传统金融机构很难获得足够授信，只能转而向新型融资渠道寻求解决途径。

综上，本文的创新点在于：聚焦于一个具体且符合逻辑的核心假设展开研究——“小微商户的融资行为，与其未来一段时间内的信用风险水平，具有显著的相关性”，这一切入点在国内现有的研究中尚处于空白，而本文将通过大量真实小微商户案例样本的定量分析来验证这一假设，并最终将结论用于实际业务上的风控优化，服务于更多尚未被正规传统金融机构准入的小微商户。

### 三、理论基础和研究方法

#### 3.1 理论基础

2020年5月8日，全国工商联等多方发布《2019-2020小微融资状况报告》。报告中将小型企业、微型企业和个体经营户这三类统称为“小微”，其中小微企业被称为“头部小微”，微型企业和个体经营户被称为“长尾小微”。针对小微的融资问题，报告明确指出：小微经营者融资难、融资贵和融资慢的问题还没有得到根本的缓解。“2019年，在头部小微经营者中，近三分之一（32.4%）使用过三种及以上的融资渠道。在长尾小微经营者中，也有将近三成（28.8%）企业使用过两种融资渠道，有18.7%企业使用过三种及以上的融资渠道。”

由此可见，对于本文研究的对象——小微商户（上文中的“长尾小微”）而言，多种融资渠道并存是非常普遍的现象，并且越来越多的小微商户，倾向于选择银行等传统金融机构以外的新型融资渠道，除去保证抵押类贷款、供应链金融、融资租赁等依托于资产或贸易的特定类型贷款，融资成本降序这些新型融资渠道主要包括：(1)P2P等网贷 (2)小额贷款公司 (3)消费金融公司 (4)互联网银行。这四类新型融资渠道，均有以下共同优点：(1)无抵押无担保 (2)申请门槛不高 (3)资料 and 手续简单 (4)审批放款快速便捷。正是具备这些优势，新型融资机构近年来如雨后春笋般涌现，贷款规模成指数型增长，一度成为实现“普惠”进而“惠普”的希望所在。但野蛮生长的背后一系列问题也纷至沓来——714高炮、暴力催收、违规过度收集借款人信息、套路贷校园贷等，引发政府金融监管机构的严厉整顿，导

致以 P2P 为代表的新型融资机构覆灭，这是后话。回到本文研究的 2017 年至 2019 年间，以及更早至 2015 年，P2P 一度成为普惠金融的急先锋。

小微商户作为这些新型融资渠道的主要受益群体之一，该如何合理配置自己的融资来源，他们的申请行为反映了何种现状。要回答这个问题，我们首先需要了解，各类新型融资渠道的起源以及分别有哪些优缺点，我们以下对此做简要介绍。

### 1) P2P 等网贷

P2P 是 peer-to-peer lending 的缩写，即个人对个人借款，又称点对点网络借款，是一种借助互联网收集客户信息，将分散于不同出借人的资金与有资金需求的借款人进行匹配，点对点地解决小额融资问题的平台渠道。P2P 的社会价值主要体现在满足个人资金需求、发展个人信用体系和提高社会闲散资金利用率等方面。P2P 小额贷款起源于诺奖得主穆罕默德·尤努斯教授的一次试验。1976 年，穆罕默德·尤努斯在一次乡间调查中把 27 美元借给了 42 位贫穷的村民用于经营周转，结果债务回收率高达 97%，那些所谓的穷人并不像当时传统信用体系认为的不可靠。自此，穆罕默德·尤努斯在商业银行体系内设立了一间分行，开始为穷人提供小额贷款的业务。2005 年 3 月，四个英国的年轻人——理查德·杜瓦、詹姆斯·亚历山大、萨拉·马修斯、大卫·尼克尔森创办了全球第一家 P2P 网贷平台 Zopa，这是 P2P 网贷模式的雏形。

在中国 P2P 则开始于 2007 年。从最初的不引人关注直至 2010 年随着陆金所、信而富、人人贷等头部平台冒出，开始进入飞速发展期。2012 年下半年至 2013 年是网贷平台

的爆发期，以每天 1-2 家上线的速度猛增。中国的 P2P 以投资方背景可分为民营系、银行系、上市公司系、国资系、风投系等几大派别，其中民营系数量最多（占 80%），质量也最参差不齐。毫无疑问当时的 P2P 平台的确缓解了中小微企业和个人的融资难题，作为银行业务的补充开辟出一条新型融资渠道。

P2P 网贷平台的兴盛，不能否认有部分原因是得益于瞄准了银行尚未触达的下沉客群，换句话说服务到了银行不做的客群——以小微企业和个体工商户为代表。反过来，P2P 网贷平台能受到小微借款人群的青睐，也因为以下两个最主要的优点：

#### 1. 有效解决融资难问题

在融资渠道较为单一过去，借款人如果难以从银行获得授信，就几乎只能选择费用高昂的民间借贷。P2P 网贷的出现，为借款人增加了新的融资渠道，利率费用介于两者之间，贷款成功率比银行高得多；

#### 2. 手续方便放款快速

由专门的的贷款顾问帮助收集借款人的申请材料（身份证明、人行征信报告、银行流水、经营实体材料等），材料提交后 P2P 贷款平台会对借款人资质进行核查，一般当天就能确认是否通过，T+1 日放款就能到账，审批到放款不超过 3 天。

尽管 P2P 能服务银行服务不了的客群，但银行仍然是优质借款人最优先考虑的融资渠道，P2P 只是急用钱时的周转方式，或是银行以外的补充选择。这是因为 P2P 相比银行有以下缺点：

### 1. 借款利率费用偏高

P2P 网贷的资金来源是平台上的众多个人投资者，平台承诺的收益率很多达到了 10% 的水平，再加上许多平台的风控能力不强导致坏账高企，要实现平台盈利基本要求年化 24% 以上的放款利率。即使有些平台号称 24% 以内的贷款利率，也绝大多数是巧立名目额外收费，比如砍头息、担保费、服务费等，最终的贷款利率甚至远超年化 36%。也正是如此，对于定价敏感的优质客户而言，除非紧急用款或需求金额较大，一般也仍然首选从银行贷款；

### 2. 个人信息泄露和暴力催收等隐患

P2P 是游离于监管所辖界限的存在，最直观的一点就是 P2P 贷款不上个人征信，P2P 机构也无权接入人行征信体系进行客户个人资信的审查。因此，绝大多数 P2P 网贷公司通过接入第三方数据公司来获取客户信息。正规的 P2P 平台在查询客户外部信息前，会让客户签署《征信查询授权书》，确保客户知情和同意。也有一些不正规平台不经客户授权查询个人信息，甚至非法购买客户资料用于营销。此外，暴力催收也是 P2P 平台常见的回款手段。人行征信报告除了提供客户资信明细外，还有一个作用在于震慑还款意愿不强的潜在违约客群。在人行征信上留下不良记录，会影响个人的信用卡使用、买房贷款、工作留学等方方面面。而在 P2P 平台贷款违约，则一般不影响人行征信记录，甚至没有多少贷款痕迹，这让一些老赖肆无忌惮。P2P 平台采取非常手段挽回损失可以理解，但有的平台偏激到暴力催收，骚扰威吓借款者身边人以达到震慑借款人的目的，就过于不择手段和涉黑性质了。



综上，P2P 在众多融资渠道中不受银行等持牌金融机构的青睐，但确实也解决了次级客群和资金需求旺盛的小微客群的融资难问题。换个角度，当一个客户的融资来源主要为 P2P 或在 P2P 平台申请行为频繁时，很可能是没有其他更好成本更低的选择了。在后面的分析中，我们将通过数据去检验这一推论。

## 2) 小额贷款公司

根据 2008 年发布的《关于小额贷款公司试点的指导意见》（以下简称“指导意见”），定义小额贷款公司为“由自然人、企业法人与其他社会组织投资设立，不吸收公众存款，经营小额贷款业务的有限责任公司或股份有限公司”。同时，指导意见还规定了小额贷款公司的资金来源，“小额贷款公司的主要资金来源为股东缴纳的资本金、捐赠资金，以及来自不超过两个银行业金融机构的融入资金。”

从指导意见中我们获取了以下信息：

1. 小额贷款公司受到银监会和央行的监督管理；
2. 小额贷款公司不得以任何形式吸收公众存款；
3. 小额贷款公司资金主要来源于自有资金、股东借款、银行融资、资产证券化；
4. 小额贷款公司贷款利率上限不能超过央行同期贷款基准利率的四倍。

小额贷款公司应坚持“小额、分散”的原则，鼓励面向农户和微型企业提供信贷服务，着力扩大客户数量和服务覆盖面。

相比较其他贷款而言，小额贷款有着显著的个性特征：

一是特定的目标群体。小额贷款瞄准的客户为中低收入贫困群体和微型企业。无论是制度主义的小额贷款还是福利主义的小额贷款在目标客户的选择上几乎没有差别，也即中低收入的贫困人口是他们共同的服务对象。但由于宗旨不同，制度主义小额贷款的目标客户更多地锁定为微型企业和贫困群体中收入相对较高者；

二是灵活的抵押担保方式。小额贷款原则上无需抵押品，或在抵押担保上采取灵活多样的形式。如果贷款必需物质担保，则多数目标客户由于贫困缺乏物质资本就无法参与。小额贷款常常也不需要传统担保，而是通过小组借贷，主要采用社会担保的方式为借款者提供担保；

三是简化的贷款流程。贷款申请程序和形式进行尽可能的简化，申请借款者无需提交各种书面材料，当然实际情况还是需要通过各种方式进行详细了解的。因此显著区别于多数正式金融机构申请方面和业务计划方面的复杂精确要求；

四是灵活的还款方式。小额贷款为减轻贫困人口的负担采取了灵活的还款方式，方便借款人根据自己的财务能力和现金流按时归还贷款；

五是利率差别较。小额贷款的不同类型机构所制定的利率差别较大。福利主义小额贷款由于对盈利性要求不高，贷款利率通常低于金融机构的商业贷款利率；制度主义小额贷款由于要考虑到机构的持续经营，故其贷款利率一般高于商业贷款利率。

与此同时，小额贷款也面临如下问题：一是放贷资金来源狭窄。小额贷款公司不吸收公众存款且只发放小额贷款，开展业务所需要的资金来源受《指导意见》规定制约；二是小额贷款公司贷款主要支持的对象是中小企业和“三农”，而这些放贷对象普遍为不符合金

融机构贷款抵押条件或在金融机构贷不到款的中小企业和“三农”项目。虽然小额贷款公司其中的一些贷款也有合格抵押品，但多数贷款实行的是一人或数人的信用担保。因此面临的贷款风险较大。

### 3) 消费金融公司

消费金融公司在西方很多国家以及日本、中国台湾等地区已经形成较为完善运作模式，与商业银行的个人贷款业务、信用卡业务在个人信贷领域中有着三分天下的重要地位。不同国家（地区）的消费金融公司有着各自的定义和特点。英国消费金融公司指专门为消费者提供服务的消费信贷商，如房屋信贷公司、汽车金融公司和消费者金融公司等；美国的消费金融公司是指以指向消费者提供房屋抵押贷款、信用卡业务等消费信用公司；中国的消费金融公司是指“经中国银行业监督管理委员会批准，在中华人民共和国境内设立的，不吸收公众存款，以小额、分散为原则，为中国境内居民个人提供以消费为目的的贷款的非银行金融机构”。贷款产品包括耐用消费品贷款和一般用途个人消费贷款等。

2010年中国银监会发放了首批消费金融公司试点的批筹，包括中银消费金融公司、北银消费金融公司、捷信消费金融公司和四川锦程消费金融公司等四家，使得消费金融公司陆续在上海、北京、天津和成都进行了区域性经营。2013年9月，批准了新增沈阳、南京、杭州等10个城市参与第二批试点。此外，合格的香港和澳门金融机构可在广东省试点设立消费金融公司。2015年6月10日，国务院常务会议提出决定放开消费金融市场的准入，将此前16个城市开展的消费金融公司试点扩大至全国。随着我国经济的发展，居民的消费需

求不断提高，消费金融公司的成立对我国来说是一项新的突破，它有着巨大的潜力和广阔的发展空间，将成为促进城镇化和消费升级的重要支撑。

消费金融公司有其自己的特点，与传统商业银行的贷款机构区别明显，具体如下：

### 1. 业务拓展区域广泛

国际上设立的消费金融公司，普遍具有远距离跨地区的商业网点和销售队伍，有广泛的有需求的客户基础支撑。只有这样，消费金融公司达到规模放款，才能降低平均每笔贷款成本，从而获得可持续发展。

### 2. 只放贷不吸收公众存款

国内外消费金融公司一个共同特点是不吸收存款，因此消费金融公司融资来源狭窄且成本不确定，消费金融公司的发展极易受此制约。当然相比较小额贷款公司来说，消费金融公司由于可以通过同业拆借、吸收股东存款、发行金融企业债券等渠道来获得资金，另外还可进行固定收益的投资，因此融资成本相对来说可以压到相对较低的水平。

### 3. 目标客户广泛化、年轻化

和传统的主要针对中高端收入人群的商业银行不同，消费金融公司目标客户人群更加的广泛化。作为商业银行金融机构的补充角色，消费金融公司主要目标客户群为中低端收入人群中财务状况较好者。这些人更加需要通过消费信贷来获得对教育等一般消费品的使用、对日常家电的更新换代等。而且这一批目标客户群还有一个特点就是平均年龄相对较低。年轻人有着更大的追逐快速迭代的电子产品的欲望，但其收入和信用卡额度又受到制

约，无法超前消费，转而选择消费金融公司获取贷款。另外这一批年轻人极其家庭对装修、教育等都存在大量的需求，受储蓄年限和储蓄水平的限制，也非常可能会成为消费金融公司的潜在客户。

#### 4. 贷款额度小、贷款利率相对高

消费金融公司的贷款用途主要是针对一般消费和耐用品消费，单笔贷款额度较小。尤其在我国，消费金融不包括房贷和车贷，且贷款额度不可超过借款人月收入的五倍，限制非常严格。由于贷款期限较短、额度较小等原因，消费金融公司发放贷款的利率相对较高，基准利率的四倍上限内可根据对不同借款人的不同评估进行浮动。

显而易见的是，消费金融公司也面临诸多困难，也有很多缺点，例如融资渠道窄导致资本不足；业务缺乏创新及地域限制导致供给渠道窄；个人征信体系不健全导致业务可操作性差以及法规制度不完善导致业务运行不规范等。

### 3.2 核心假设

本文的核心假设是：

小微商户的近期融资行为，与其未来一段时间的信用风险水平，具有显著相关性。

其中融资行为的刻画，又分为两个维度：融资行为发生的频率、融资渠道构成与分布。更具体的，我们将在本文研究中，首先对以下两个假设加以分析检验：

- 1) 小微商户融资发生的行为频率与信用风险呈正相关，即频率越高，风险越高；
- 2) 小微商户融资渠道构成与分布于信用风险显著相关，在正规金融机构的融资占比越高，风险越低；

验证上述两个假设后，我们将通过风控建模，评估小微商户融资行为变量对信用风险的违约率预测的贡献和提升效果，即：

含融资行为变量的评分卡模型，比不含融资行为变量的模型，预测信用风险违约率的准确度显著更高。

### 3.3 数据来源

研究的数据来源于 H 公司自 2017 至 2019 年 7 月的小微商户经营性融资贷款业务。每一个申请融资服务的客户，均会书面授权 H 公司获取其人民银行个人征信报告以及同盾等第三方大数据报告，提炼出一系列特征变量，用于风险评估和额度授信。而风控审核通过并放款的客户，均有每个月的还款记录作为其真实风险水平的判定依据。因此，本文研究所使用的贷前数据来源于小微商户在申请 H 公司贷款业务的时点授权 H 公司查询的当时信用记录，均可保证合法合规，而贷后数据则来源于 H 公司提供的每笔放款在每月的还款情况统计。

尽管信用风险和客户的基础属性（比如年龄、婚姻状况）、收入水平、负债水平、地理区域、宏观政策等多因素均有不同程度的相关性，但此次研究的目标是探索小微商户“申请融资行为”的风险关联性，因此对数据的处理和变量的衍生，也将聚焦于这一维度，而不会对其他维度变量做过多的展开和细致的分解。

### 3.4 研究方法

本论文将采取演绎研究方法，以申请 H 公司融资贷款的小微商户客群为研究对象，把他们最近一段时间内在多个不同融资渠道的申请行为（比如近三个月申请平台总数、近

三个月在银行/P2P/小贷公司申请机构数/占比)作为自变量  $X$ ，放款后一段时间内的逾期情况(即实际风险水平)作为因变量  $Y$ 。按照实际业务经验和风险逻辑，小微商户申请融资行为越频繁、高利率融资渠道越多，最终发生较严重逾期的概率越大，这是因为  $Z$ ：这类商户的资金饥渴度更大、融资成本更高、在银行等正规金融机构可获得的授信更少或无法满足需求，因此也就更容易发生资金链断裂和经营失败。

### 3.5 计量模型和指标

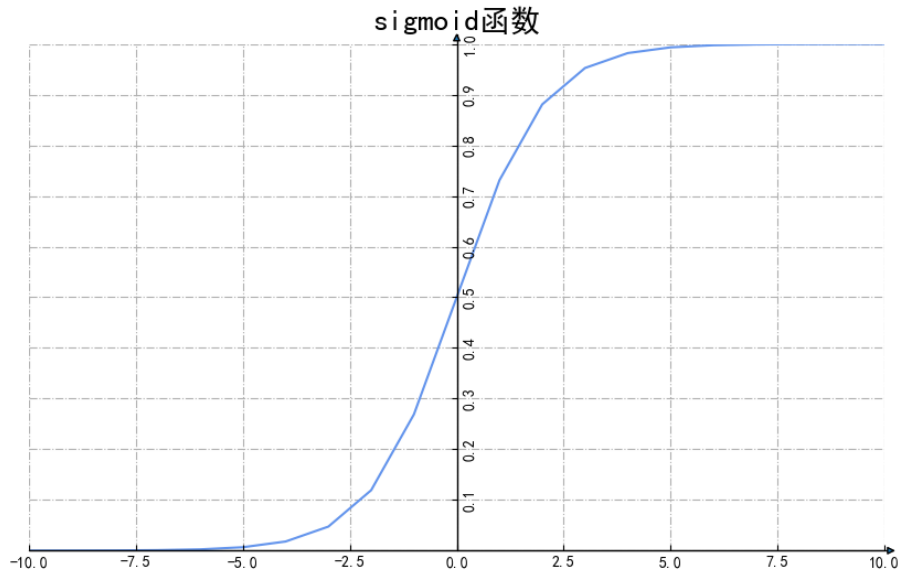
正如前文所述，本论文要探究的是“小微商户的融资行为对其信用风险水平的影响”，而在信用风险评级和决策分析领域中，最为广泛和常用的计量模型有“逻辑回归”和“决策树”，并且这两者都是解决“分类问题”的经典方法，在本次研究中将会采用。此外，为了衡量单个变量对于信用风险的预测力或影响，我们还会引入 IV 值(Information Value)作为重要指标。

#### 3.5.1 模型：逻辑回归(Logistic Regression)

我们在信用风险预测中，往往希望获得某个风险事件(比如某笔案件在放款后 6 个月内出现 30 天以上逾期)发生的概率值，如果用一个二进制因变量  $Y$  来定义这个事件的结果，那么  $Y=1$ ，反之风险事件不发生，则  $Y=0$ 。于是， $P(Y=1|X)$ 即为在自变量  $X$  条件下，风险事件  $Y$  发生概率，这是一个“是或否”的二分类问题。显然概率值  $P(Y=1|X)$ 是取值范围为  $0\sim 1$  的连续变量，而线性回归的输出值  $Y$  尽管连续，却并不能保证完全落在  $0\sim 1$  内，sigmoid 函数可以做一个很好的映射：

$$\text{sigmoid 函数: } S(z) = \frac{1}{1 + e^{-z}}$$

图 1: sigmoid 函数



与线性回归模型结合： $P(Y = 1|X) = S(ax + b) = \frac{1}{1+e^{-(ax+b)}}$

不难看出，逻辑回归本质上是从线性回归变化而来，因此它也具有线性回归的优点和缺点。优点包括：模型结构简单，可解释性强易理解，训练和实现速度快，输出的预测结果连续，可方便设定决策阈值，也可转化为评分卡形式。缺点包括：非线性问题需要转换，不适合处理大量特征以及缺失值过多变量。

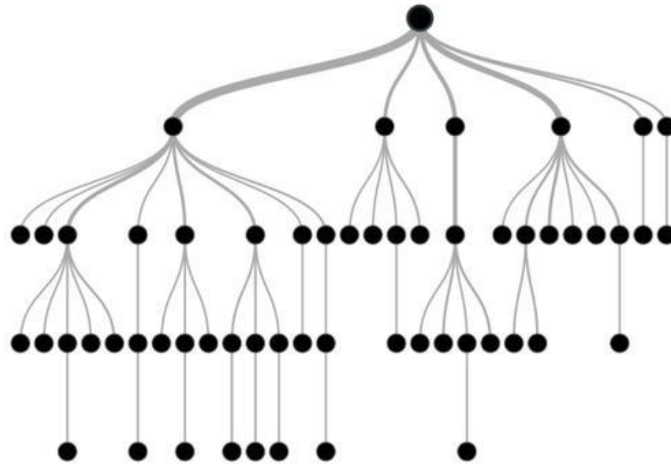
### 3.5.2 模型：决策树(Decision Tree)

不同于逻辑回归的线性关系属性，决策树更适用于应对非线性关系。决策树是一种非常直观的树形结构，其中每个内部节点表示一个属性上的测试，每个分支代表一个测试输出，每个叶节点代表一种类别。我们可以通过层层深入分解的方式，用筛选出的最优特征变量，归纳出一组分类规则（if-then），将整体中的每一个样本分配至树形末端的叶节点



上，不同的叶节点代表了明确的类，最终完成了决策树的分类功能。决策树的优点是：直观易理解、可以处理非线性关系、擅于挖掘局部具有明显特征的目标类别，缺点是：容易过拟合、缺乏全局性、分类结果断裂且不方便排序。

图 2：决策树



基于这两种模型的特性和优缺点，在风险分析领域中，常把逻辑回归用于构造信用评分卡，而把决策树用于风控规则制定。前者更注重全局把握，将整体按照预测的违约风险概率给予不同的评分结果和风险评级，关联不同的审核通过率和授信额度。后者更侧重具有明显特征的高风险剔除，从历史数据的学习中，当我们发现具有某一个或一系列特征的样本集合，违约率远远高于整体，那么以这些特征构成的风控规则，就能识别出这类高风险客户，并针对性地剔除或控制额度。

### 3.5.3 指标：IV 值(Information Value)

我们知道信用风险和多方面因素（年龄、行业、地区、负债、收入、历史信用等）有关，加上同一维度的变量衍生（如近 3/6/12 个月逾期次数）， $X$  变量通常高达几百个。

这么多的变量我们不会直接放入模型去做拟合训练，而是先从中挑选出一些作为入模变量。挑选会考虑变量的预测力、变量间的相关性、变量的业务可解释性等多方面因素，但其中最为重要和直观的衡量指标就是“变量的预测力”。IV 值就是衡量变量预测力的一个量化指标，顾名思义，Information Value 可理解为信息价值，IV 值越大则表示这个变量所蕴含的用于预测 Y 的信息价值越多，它的重要性也就越高，越应该加入到入模变量中。

将一个变量（无论是连续或离散）分成 n 组，整个变量的 IV 值是各组 IV 值的总和：

$IV = \sum_i^n IV_i$ ，而单个组的 IV 值计算如下：

$$IV_i = \left( \frac{g_i}{g_T} - \frac{b_i}{b_T} \right) \times \ln \left( \frac{g_i/g_T}{b_i/b_T} \right)$$

其中， $g_T$  表示样本所有好客户数（Y=0 的样本数）， $b_T$  表示样本所有坏客户数

（Y=1 的样本数）， $g_i$  表示变量第 i 组中的好客户数， $b_i$  表示变量第 i 组中的坏客户数。最

终，IV 越大，则表示好坏客户在该变量上的区分度越大，即该变量的预测力越强，IV 值

对应的预测力强弱衡量对照表如下：

表 1：IV 值范围和对应预测力

IV 值的范围	预测力
小于 0.02	无预测力
0.02-0.10	弱
0.10-0.30	中
大于 0.30	强

#### 3.5.4 指标：决策树特征重要性(Feature Importance)

构造决策树时，对于每个特征都会基于一定的目标准则（比如熵、基尼指数等）计算它的重要性，即用归一化指标来衡量每个特征对于因变量分类的贡献程度，这些重要性能够辅助进行特征选择。以常用的统计软件 **SAS** 为例，在设定目标变量、输入变量、目标准则后，运行决策树建模，除得到树模型外还会输出变量重要性，如下图：

图 3: 决策树特征重要性示例

变量名称	标签	拆分规则数	替代规则数	重要性	验证重要性	验证训练重要性的比率
var_p2p		2	1	1.0000	1.0000	1.0000
TD_3M_CNT		1	3	0.9418	0.9318	0.9894
TD_1M_CNT		0	3	0.8869	0.8836	0.9963
var_p2p_pct		0	3	0.4251	0.2635	0.6198
var_rzsl_pct		1	1	0.4047	0.2338	0.5776
var_rzsl		1	1	0.4001	0.2306	0.5764
var_ybxffq		1	1	0.2233	0.1684	0.7542
var_xedk_pct		0	1	0.2227	0.1684	0.7559
var_xedk		1	0	0.2227	0.1684	0.7559
var_yhxfjr_pct		1	0	0.2182	0.1036	0.4747
var_yhxfjr		0	1	0.2149	0.1020	0.4747
var_wsyh_pct		0	1	0.2148	0.1020	0.4747
var_dxxfjr_pct		0	1	0.2124	0.1646	0.7750
var_dxxfjr		1	0	0.2124	0.1646	0.7750
var_ybxffq_pct		0	1	0.1435	0.1039	0.7236

28

基于上图我们能直观了解到输入特征的重要性排序以及量化结果。

## 四、研究步骤简述

### 4.1 贷前数据获取

对 H 公司已放款的一万多笔小微商户经营贷案件，首先需要获取和整合案件相关的贷前和贷后数据。其中贷前数据，是以客户申贷环节作为节点，提取该节点时间附近的客户信息。由于本次研究的重点在于小微商户“融资行为”和“信用风险”之间的关系，因此我们更关注“融资行为”相关的贷前数据。在 H 公司已获取的内外部贷前数据中，与“融资行为”关联紧密的数据来源于人行征信报告的查询记录和同盾的多平台申请数据。

#### 4.1.1 人行征信报告

人行征信报告，是一份由中国人民银行记录的个人信用活动报告，里面真实的反馈个人的信用记录情况，是个人信用信息基础数据产品。其中的查询记录，包含了个人最近 2 年内本人查询和被机构查询的记录，数据结构如下：

表 2：人行征信查询记录字段

字段名称	字段解释	字段说明
查询日期	yyyy 年 mm 月 dd 日	2019 年 1 月 1 日
查询操作员	以 '/' 分割查询机构和操作员账号	中国工商银行北京分行/user
查询原因	本人查询 贷款审批 信用卡审批 担保资格审查 贷后管理 保前审查 保后管理 法人代表、负责人、高管等资信审查 特约商户实名 融资审批 客户准入资格 资信审查 公积金提取复 业务审批 异议核查 股指期货开户 .....	常见      其他

值得说明的是，当个人或关联企业向银行、消费金融公司等征信接入机构申请融资时，几乎必然会被查询个人征信报告，由此在上述查询记录中留下机构查询痕迹。而如果向 P2P 网贷等非征信接入机构申请融资时，往往会被要求提供本人查询的征信报告，由

此在上述查询记录中留下本人查询痕迹。因此，我们通过分析小微商户经营者在近期的查询记录，就能获取其最近的融资行为相关信息——申请融资日期、申请融资机构名称、申请融资的业务类别。

#### 4.1.2 同盾等第三方数据

近年来国内融资渠道呈现多样性，除传统的银行和正规的持牌机构外，还涌现了一大批“非正规军”，典型代表就是 P2P 网贷。这些平台数量繁多，2012 年开始野蛮增长，2015 年曾一度达到 3000 多家，直至平台爆雷、负责人跑路、暴力催收等恶性事件频发，2016 年 P2P 迎来“监管元年”，截至 2019 年 10 月，正常运营平台数据已经下降至 500 多家。P2P 网贷尽管利息高，且濒临消亡边缘，但门槛低，它曾经解决了以小微商户为代表的未被银行青睐的庞大下沉群体的资金需要，规模不容小觑。与银行一样，P2P 网贷也有风控需要；与银行不同，P2P 网贷无法接入人行征信。于是，像同盾、百融等这样的第三方数据公司应运而生，从某个角度来看，第三方数据服务本身就是一个生于灰色地带边缘的产业。

同盾是第三方数据行业的头部公司，数据显示，截至 2018 年底，同盾累计客户数量为 2421 个，完成了对互金行业 90% 以上的覆盖。也正是这样的覆盖率，使得同盾的“多平台申请”数据成为该公司的核心产品之一。当一个客户向某个互金公司申请融资时，如果该公司使用了同盾的产品，那么每次调用同盾接口，就会在同盾产生一次查询记录。记录越多，就说明该客户申请的频次越高，融资需求越旺盛。同盾的多平台申请数据结构如下：

表 3: 同盾多头报告字段

字段名称	字段解释	字段说明
规则名称	7天/1个月/3个月内申请人在多个平台申请借款	
平台总个数	platform_count:10	
平台详情	行业名称	<b>P2P 网贷</b> 小额贷款公司 一般消费分期平台 融资租赁 银行消费金融公司 大型消费金融公司 财产保险 银行小微贷款 互联网金融门户 银行个人业务 网上银行 综合类电商平台 第三方服务商 大数据金融 第三方支付 直销银行 .....
	子行业平台个数	count:2

值得说明的是，同盾的多平台申请数据，不仅包含申请的平台总数，还将平台细分到



了 20 多个子行业，记录了子行业内的申请平台个数。这使得我们通过分析近期内客户的申请平台数和行业结构，能刻画出客户的申请融资行为——申请借款频次、申请平台类型、申请数量构成明细和平台偏好。

#### 4.1.3 有效样本的筛选原则

在对 H 公司历史客户的贷前数据整理中，我们常常能碰到这样的两种情况：1)部分授信客户一人对应着多份人行征信报告，或是多次同盾等第三方数据查询结果;2)部分授信客户已无法找到任何人行征信报告，或同盾等第三方数据缺失。这两种情况的客户，均不能直接加入分析样本，我们在筛选有效样本筛选时会遵循以下原则：

- 1) 每笔授信只能对应一条样本记录，即一份贷前数据和一个贷后标签。比如当一笔授信对应多份人行征信报告时，我们只会选取其中一份报告反映该笔授信中客户的贷前情况。
- 2) 每笔授信的贷前数据，应对应审批时点客户的情况。比如银行机构对授信客户在存续期间常常会发起多次信用数据查询，时点可能是贷前审批时，也可能是贷后管理时。而对本文分析而言，关注的是客户在贷前的融资行为数据，是否与其贷后表现出来的信用风险水平显著相关，因此只能取贷前审批时点的客户人行征信报告和同盾等第三方数据。
- 3) 每笔授信的贷前数据有效期为 30 天。前一条原则中提到，我们要取的是贷前审批时点客户的数据，实务中征信拉取时间与审批时间一致的情况不多，基于约定俗成的行业内有效征信的定义，我们会以 30 天作为分界点。即早于 30 天的征信

报告无效，需退回客户重新拉取最新记录，而拉取时间距离审批时间在 30 天以内的，才是有效征信，同盾等第三方数据的有效性也是这样界定。

- 4) 当有多份有效贷前数据时，以距离审批时点最近的记录为准。第一条原则中，我们强调了每笔授信只有一条记录，因此当有多条有效贷前数据时，我们也只会选取最能反映审批时点客户情况的记录，即时效性最强的。

选取分析样本时，要求有效的人行征信报告和有效的第三方数据，两者缺一不可。于是 H 公司 2017 年至 2019 年 7 月的样本，根据是否留存有效的人行征信报告和有效的三方数据，分为以下四种情况。

表 4: H 公司小微商户样本有效报告情况

		有效的人行征信报告		总计
		有	无	
有效的同盾三方数据	有	10081	5223	15304
	无	120	0	120
总计		10201	5223	15424

如上表，本次研究的样本为 10081 笔，即有效的人行征信和有效的同盾三方，两个数据源均无缺失的案件。

#### 4.2 贷后数据的整合和目标定义

对于每一笔放款，H 公司都记录了放款编号、放款对象、放款日期、放款金额、放款期数、当前逾期状态、当前结清状态、当前本金/利息余额、每一期计划和实际还款明细等贷中贷后数据。这使得我们可以清楚了解客户的实际违约情况，不仅仅是当前也包括历

史。

#### 4.2.1 逻辑回归：好坏客户定义(Y=0 或 1)

在定义目标好坏客户(Y=0 或 1)时，我们采用 6 期（即 6 个月）作为好客户的表现期，30 天作为坏客户逾期天数界限，即如果客户已结清或还款满 6 期,且未发生逾期，则为好客户 Y=0，反之，出现过 30 天及以上逾期，那么该客户为坏客户(Y=1)。这样定义的原因如下：

- 1) 6 个月的表现期，基本能将潜在信用违约风险客户暴露，也基本能判断客户是否是具备足够还款能力的好客户，并且能尽可能将近期的放款案件纳入样本；
- 2) 30 天的逾期天数，基本排除了因还款习惯差或突发事件无法及时还款的情况，尽管大多数金融机构定义“不良”的分界线为 90 天，但实际上逾期 30 天后，再回收的可能性一般不足 20%，且需要投入较多人力去电话催收、实地催收、律师函催收，这样的客户也属于我们业务中认为的坏客户范畴；

除好客户和坏客户外，还有两类客户存在：一类客户出现过 1-29 天逾期，另一类客户尚未结清且表现期不足 6 个月，这两类客户排除在分析样本之外。

#### 4.2.2 决策树：不同违约期限结构下的客户分层 (A/B/C)

决策树模型和逻辑回归模型一样都能解决“是否会发生违约”这样的二分类问题，在本文的研究中也会用决策树来制定风控规则。除此之外我们还可以通过决策树方式，更细化研究“短期内违约/中长期内违约/未违约”这三类客户在融资行为上的表现，换句话说，具有什么样融资行为特征的客户，会更早发生违约。因此我们按如下期限结构将客户贷后表

现分为信用风险递增的三类：

A 客户：已结清或还款满 6 期,且未发生逾期；

B 客户：(3,+ ) 期内发生了 30 天及以上逾期；

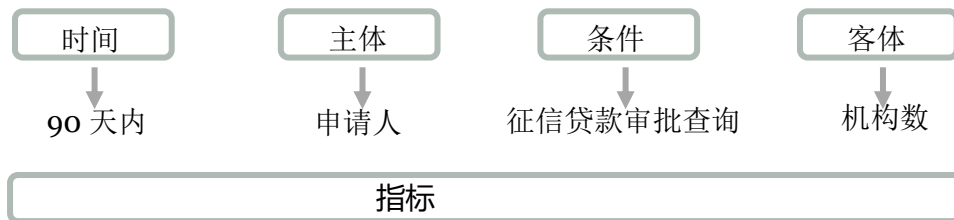
C 客户：(0,3] 期内发生了 30 天及以上逾期；

不难发现，A 客户的定义与逻辑回归中好客户（ $Y=0$ ）的定义相同，而 B、C 客户则均属于逻辑回归中坏客户（ $Y=1$ ）的定义，只是在发生违约的期限上有所不同。除 A、B、C 三类客户外，同样还有两类客户存在：一类客户出现过 1-29 天逾期较轻微逾期，另一类客户尚未结清且表现期不足 6 个月，这两类客户也均排除在分析样本之外。本文中我们将进一步探索是否能从融资行为变量来区分“短期内违约/中长期内违约/未违约”这三类风险递减的客户。

#### 4.3 融资行为相关的 X 特征变量构造

融资行为相关的指标变量，可按照“时间+主体+客体+条件”的方式衍生。以人行征信报告的查询记录为例，原始的数据是以“日期+查询原因+查询机构”的方式陈列，但显然这不能直接用来作为刻画客户融资行为的指标，于是按上述方式转化为“90 天内申请人贷款审批查询机构数”类似的指标：

图 4：特征变量构造方式



需要说明的是，尽管我们已经知道放款客户截止当前时点为止的许多信息，但仍然要回溯到客户申请 H 公司贷款时间点的信息更新状态，上例中的“90 天内”指的也是该客户（即申请人）在 H 公司申贷日期前的 90 天内。本次研究中，主体不变，即为申请人，而时间、条件、客体这三个元素均有多种选择，且可以任何组合，从而构造出数以百计的融资行为相关 X 特征变量。

#### 4.4 核心假设验证和有效变量筛选

在验证核心假设——“小微商户的融资行为与信用风险有显著关系”前，我们首先要确定合适的样本范围。前文在 X 和 Y 变量的定义过程中，我们已经对放款案件样本的筛选明确了如下要求：

- a) 已结清或有至少 6 个月表现期的好客户，或发生 30 天及以上逾期的坏客户；
- b) 保留了完整且在 H 公司贷款申请日附近的人行征信报告或同盾多头报告；

筛选完成后，我们得到了用于分析的样本数据，并对每个样本产生了融资行为相关的 X 变量集合以及信用风险相关的二元变量 Y，其中 Y=1 代表信用违约的风险客户，Y=0 代表非信用违约的好客户。

如何确定某个 X 变量对于 Y 的取值是否有显著相关性，我们采取前面提到过的 IV 值（Information Value）和决策树特征重要性（Feature Importance）作为指标，IV 值或特征重要性越大，说明 X 变量对 Y 的取值影响越大。将 IV 值  $\geq 0.02$  作为逻辑回归模型中筛选有效 X 变量的条件，IV 值  $< 0.02$  则说明该变量对预测 Y 基本无效。特征重要性排

序则主要用于决策树模型中。

#### 4.5 构建基于融资行为变量的子信用评分卡

融资行为变量中，与信用风险预测有显著效力的指标，在前述的 IV 值  $\geq 0.02$  的筛选中可被找出。接下来我们希望探索如何将这些有效变量真正应用于风控的实际业务中，以达到整体优化的目标，构建进件阶段的信用评分卡是其中一个尝试的方向。

评分卡是综合了客户多方面的信息，运用统计模型，最终形成一个单一的分值。一个有效的评分卡，可以用来根据评分结果，预测客户未来一段时间内的风险水平，并进行排序，从而辅助信贷决策。这里的统计模型就是逻辑回归模型，是一种广义的线性回归模型。

表 5：评分卡示例

维度	变量	区间	评分
基准分			650
基础信息	客户年龄	(-, 22 ]	-10
		( 22, 28 ]	-6
		( 28, 40 ]	+3
		( 40, 50 ]	-4
		( 50, + )	-8
		未知	0
	客户性别	('女性')	+15
		('男性')	-9
		未知	0
		征信报告	最近一年征信查询次数
( 2, 5 ]	+18		
( 5, 10 ]	-10		
( 10, + ]	-30		

---

.....

---

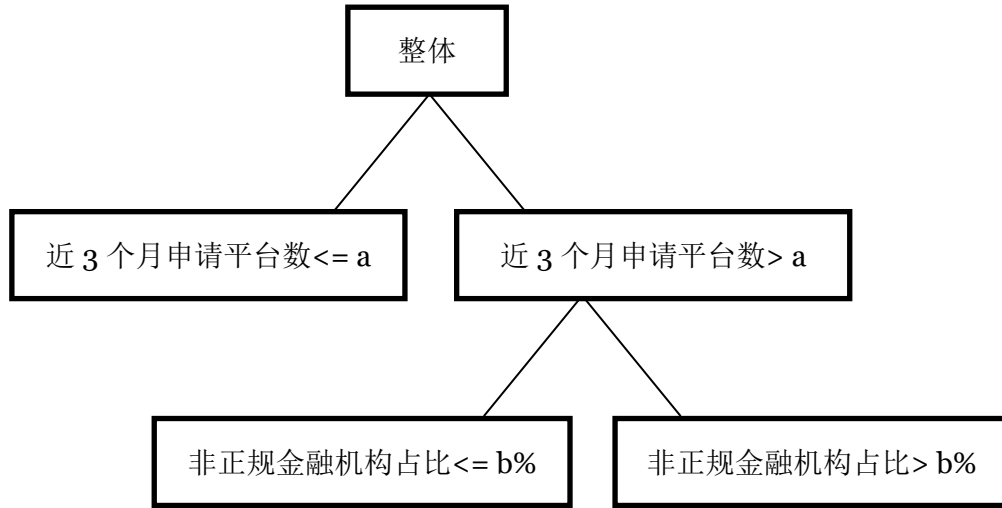
我们在构造逻辑回归模型和评分卡时，除了关注变量的预测效力，还需要注意变量间的相关性，尤其是本次研究中涉及的变量都是融资行为这一维度相关的变量，因此更要注意入选变量的相关性检验。

#### 4.6 形成基于融资行为变量的风控规则

风控实务的另一尝试，是将这些有效变量用于制定新的风控规则，以决策树的形式，将信用违约率高的枝节（即一小簇客群）细分出来并在贷前阶段就加以剔除。例如，按照风控逻辑和经验，近期融资行为频繁，且主要向非正规金融机构融资的客群，往往是资金需求旺盛且在银行等正规金融机构授信不足的客群。除此以外，这样的融资行为也会导致这群人的融资成本越来越高，并接近超负债，在正规金融机构获得授信会更加困难，出现融资渠道的进一步下沉。这是一个恶性循环，如果数据验证显示这个推论成立，那么就能形成一条类似这样结构的风控规则——“if 近 3 个月申请平台数 > a, 且非正规金融机构占

比> b%, then 拒绝”。

图 5：决策树式风控规则示例



## 五、数据分析和结论

### 5.1 小微商户融资行为刻画

小微商户融资行为如何刻画？如何获取数据并建立合理指标？这是我们进行数据分析前首先要解决的问题。基于本次研究所能获取的贷前数据，我们将从以下两个方向衍生融资行为变量并进行后续的数据分析：

#### 1) 融资行为发生的频率

频率是单位时间内完成某个事件的次数，在本文对融资行为的刻画中，指的就是一段时间内（例如近 1 个月/近 3 个月/近 12 个月）申请融资的次数或机构数。单位时间内融资行为次数或关联机构数越多，意味着对资金的需求越急迫。

#### 2) 融资渠道构成与分布

样本中的小微商户主要向何种机构申请融资，向银行等正规传统金融机构申请得



更多，还是更偏重于 P2P、小额贷款、消费金融公司等新型机构？通过对融资渠道的构成分布进行数据分析，我们能基本回答这一问题。

以下将按数据源的不同，分为“基于人行征信”和“基于同盾多头”两个部分来统计分析小微商户融资行为发生的频率和融资渠道构成与分布。回顾前文 4.1.3 中我们筛选有效样本的原则，查得有效的人行征信报告的 H 公司小微客户共 10201 个，查得有效的同盾三方数据的 H 公司小微客户共 15304 个，它们将分别作为“基于人行征信”和“基于同盾多头”刻画小微商户融资行为的分析样本来源。

#### 5.1.1 基于人行征信的融资行为刻画

人行征信报告的查询次数，是反映客户近期融资行为频率的最有效指标之一。无论是房贷、车贷还是经营性贷款，客户只要有向接入征信系统的正规金融机构申请融资的行为，就会被该金融机构查询人行征信来评估其借款资质，即在人行征信报告中留下一条查询记录。

关于什么是“查询记录”，中国人民银行征信中心官网的官方解释是：“查询记录”反映您的信用报告被查询的历史记录，显示何机构或何人在何时以何种理由查询过您的信用报

告。展示内容包括查询日期、查询操作员和查询原因。

那么什么样的查询记录最能客观反映客户的融资行为，主要有以下两种：

- 1) 查询原因为“贷款审批”的记录；
- 2) 查询原因为“本人查询”的记录；

查询记录 这部分包含您的信用报告最近 2 年内被查询的记录。			
编号	查询日期	查询操作员	查询原因
1	2010 年 5 月 5 日	中国工商银行北京分行/user	贷后管理
2	2009 年 4 月 23 日	中国征信中心北京分中心/user	本人查询
3	2008 年 12 月 10 日	中国农业银行北京分行/user	贷款审批
4	2008 年 12 月 2 日	中国农业银行北京分行/user	贷款审批

此外，2010 年您通过互联网进行了 3 次查询。

图 6：人行征信查询记录示例

前者很好理解，以上图人行征信报告样本的中第 3 条记录为例：2008 年 12 月 10 日，中国农业银行北京分行以“贷款审批”为由查询了该客户的人行征信，说明该客户当时向农行发起了贷款申请。而后者需结合实际场景理解，我们很少会主动去查询自己的人行征信报告，一般只有当被要求提供时才会进行。而最普遍的提供人行征信的场景就是申请非银行的贷款时，例如尚未接入征信的 P2P 网贷平台、典当行、民间借贷等。这也是零售信贷行业中的一个普遍性惯例。

### 5.1.2 基于同盾多头的融资行为刻画

不同于人行征信报告的查询记录以明细的形式呈现，同盾多头报告则直接形成了多头申请相关的统计变量，且我们可获取的 H 公司同盾多头报告均是以 3 个月为统计周期，只能得到小微商户在申请 H 公司贷款时间节点近 3 个月的多头申请变量，例如“近 3 个月

申请平台数”。尽管时间跨度有局限性，但在查询机构的分类上同盾报告则更为细化，至少可分为以下 28 种子行业（不全是金融信贷机构）：

- 1) P2P 网贷
- 2) 小额贷款公司
- 3) 一般消费分期平台
- 4) 大型消费金融公司
- 5) 银行消费金融公司
- 6) 互联网金融门户
- 7) 信用卡中心
- 8) 财产保险
- 9) 融资租赁
- 10) 第三方服务商
- 11) 银行小微贷款
- 12) 网上银行
- 13) 理财机构
- 14) 银行个人业务
- 15) 综合类电商平台
- 16) 大数据金融
- 17) 房地产金融

- 18) 直销银行
- 19) 资产转让
- 20) 第三方支付
- 21) O2O
- 22) 厂商汽车金融
- 23) 汽车租赁
- 24) 担保
- 25) 设备租赁
- 26) 信息中介
- 27) 交通工具租赁
- 28) 房屋租赁

同盾之所以有如此丰富的多头申请信息，是因为它作为第三方风控数据的头部供应商，自 2003 年成立以来数据合作的机构几乎覆盖了所有头部的消费金融公司、互联网金融平台、小额贷款公司、P2P 公司，甚至是越来越多的银行。同盾合作的信贷机构每调用一次同盾的数据接口用于查询客户的三方信用报告，就会在同盾产生一条多头申请记录。同盾将每个客户在不同信贷机构的多头申请记录（包括查询时间、查询机构名称、查询机构类别）整合加工后就形成了一系列“同盾多头申请指标”，这也是我们在下文中刻画客户融资行为的重要数据来源。但需要注意的是，与官方的人行征信报告相比，同盾的多头变

量用于计算融资行为的确切次数，会有一定程度的夸大和不准确，这是由于查询同盾报告的机构不一定是用于贷款审批，但总体而言用于衡量融资行为的频繁程度和渠道构成大致分布，仍然是三方大数据公司中最具行业公认效果的。

## 5.2 融资行为发生的频率

以下我们分别基于 1)人行征信的查询记录 2)同盾多头申请报告 这两个数据源来统计分析 H 公司小微商户客群的融资行为频率，即一段时间内向信贷机构申请贷款融资的次数，并阐述得到的主要结论。

### 5.2.1 基于人行征信的融资行为频率统计

我们以 H 公司小微客户中保留有效人行征信报告的 10201 个客户为样本，统计这 10201 个客户的查询记录中，近 90 天/近 180 天查询原因为“贷款审批”、“本人查询”的查询次数和机构数，可以得到以下结论：

结论 1：近 90 天/近 180 天/近 360 天/近 2 年的融资行为频率平均水平统计如下：

表 6：不同时间切片的融资行为频率

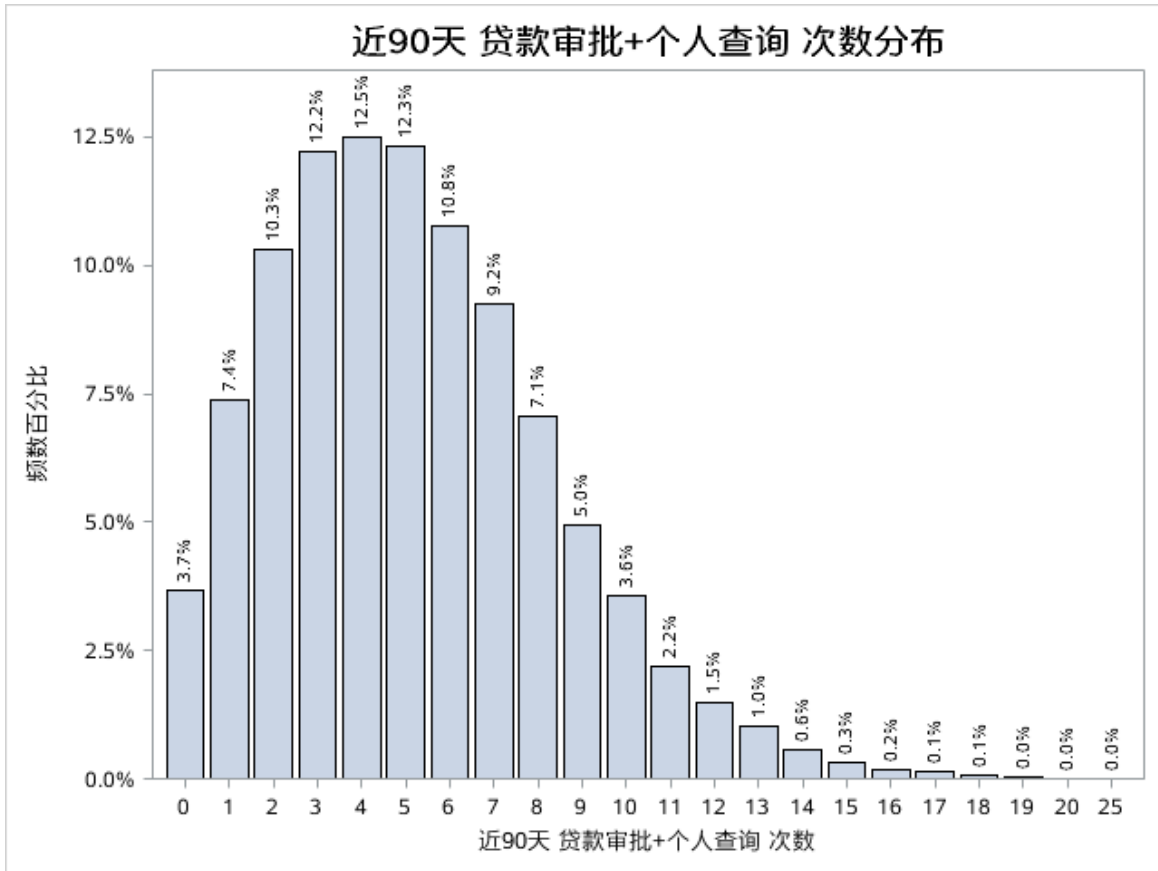
统计时间	贷款审批查询		个人查询	贷款相关
	机构数	次数	次数	总次数
近 90 天	2.59	2.80	2.38	5.18
近 180 天	4.49	5.25	4.47	9.72
近 360 天	7.39	9.45	8.08	17.53
近 2 年	11.23	16.10	13.17	29.26

说明：以第一行数据为例，这 10210 个客户在申请 H 公司贷款时间节点的近 90 天

内，人均被 2.59 家接入征信的金融机构因“贷款审批”而查询过，且查询次数为人均 2.80 次。同时，“个人查询”为原因的人均次数为 2.38 次，意味着这批客户近 90 天内向未接入征信的融资渠道（P2P 网贷平台、典当行、民间借贷等）申请贷款的人均次数也为 2-3 次。贷款相关总次数=贷款审批查询次数+个人查询次数，约为人均 5.18 次。从上表结果来看，H 公司小微商户客群这样的融资行为频率可以说是偏高的，并且不仅仅是向正规金融机构（接入人行征信的）频繁申请，在非正规金融机构（未接入人行征信的）的融资频率甚至接近向正规金融机构的融资频率。这表明 H 公司小微商户客群的资金饥渴度很高。

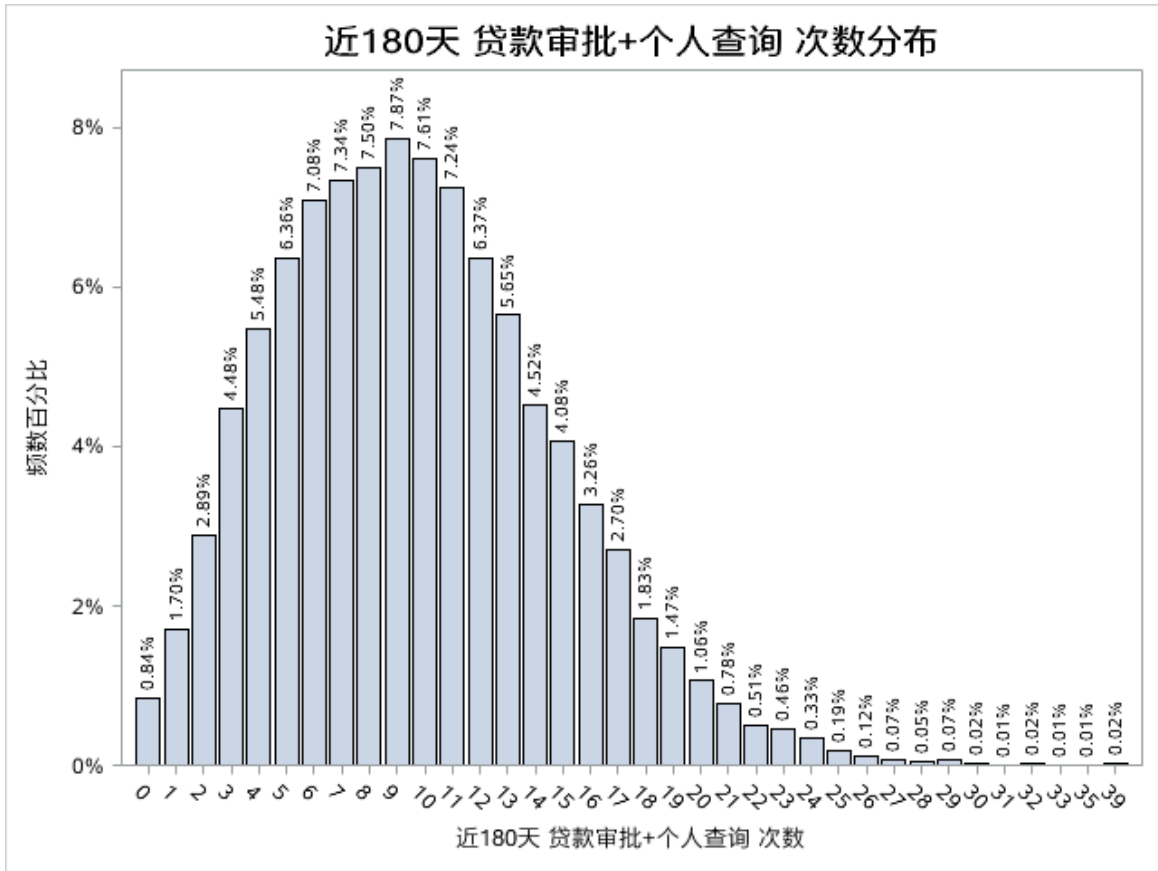
结论 2：仅 3.7% 的客户近 90 天内几乎无融资行为，接近 80% 的客户近 90 天内有 3 次或 3 次以上贷款融资申请行为，10% 的客户近 90 天内有 10 次或 10 以上的贷款融资申请行为。

图 7: 近 90 天贷款审批+个人查询 次数分布



结论 3: 仅 0.84% 的客户近 180 天内几乎无融资行为, 接近 80% 的客户近 180 天内有 5 次或 5 次以上贷款融资申请行为, 17% 的客户近 180 天内有 15 次或 15 以上的贷款融资申请行为。

图 8: 近 180 天 贷款审批+个人查询 次数分布



结论 4: 与贷款审批查询次数的分布相比, 个人查询次数的分布更集中, 这或许是由于正规融资渠道更多样化, 客户可选择的机构更多。此外, 贷款审批次数为 0 的客户比例高于个人查询次数为 0 的客户比例, 这说明正规融资渠道的覆盖率高于非正规融资渠道。



图 9: 近 90 天查询次数分布

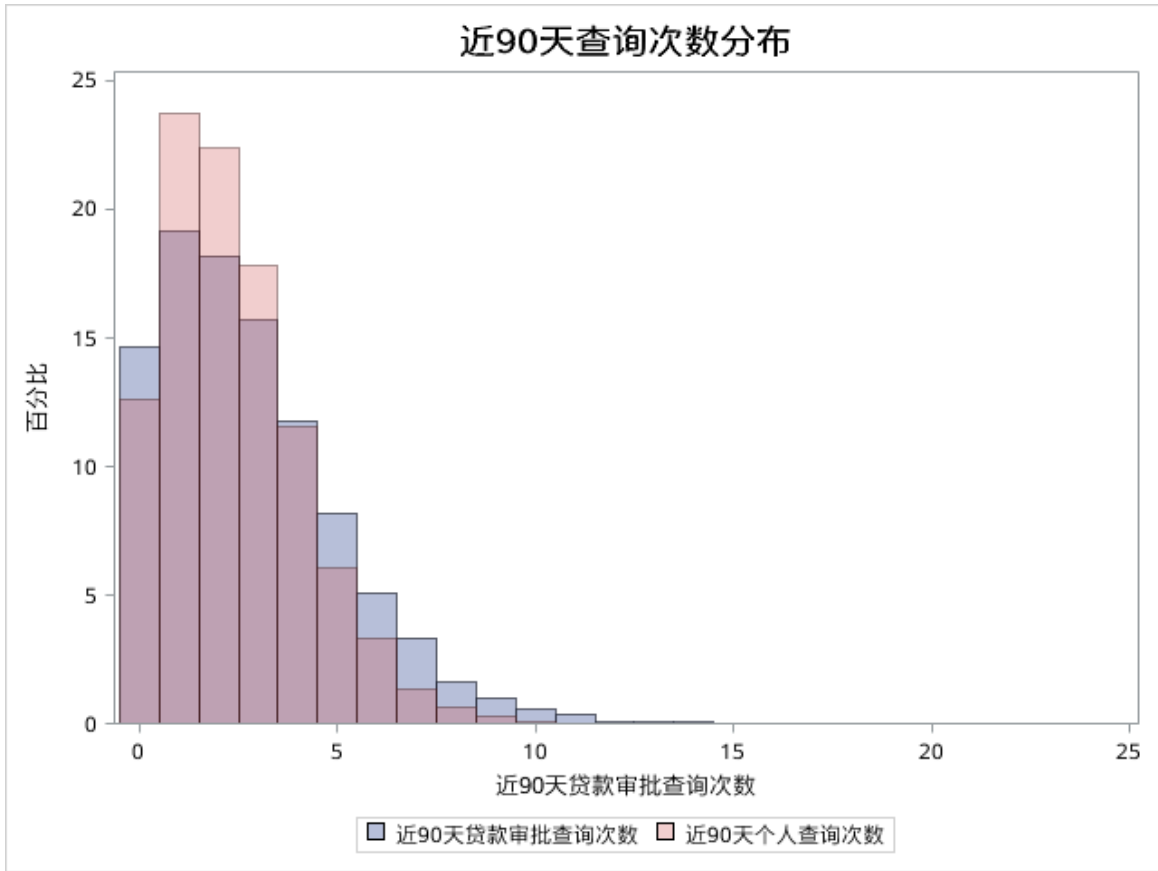
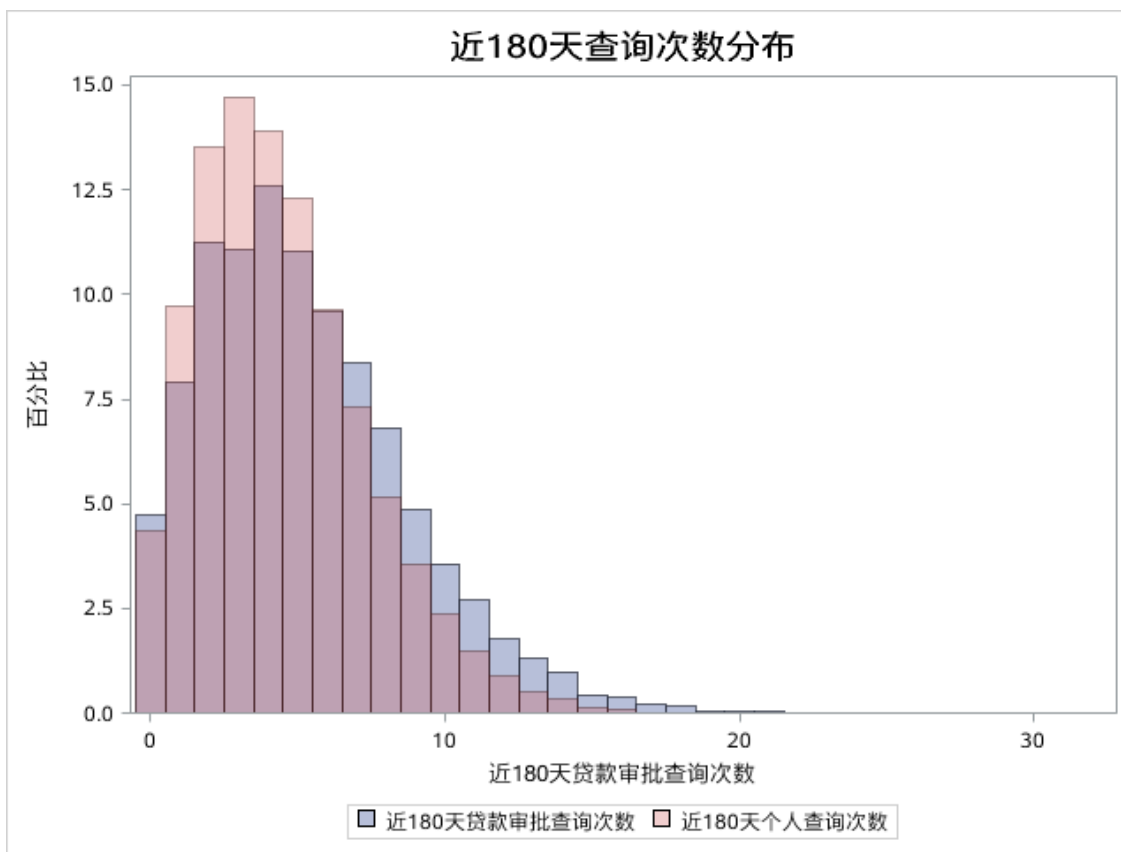


图 10: 近 180 天查询次数分



### 5.2.2 基于同质多头的融资行为频率统计

我们以 H 公司小微客户中查得有效同盾三方数据的 15304 个客户为样本，统计分析这 15304 个客户的同质多头申请类指标，可以得到以下结论：

结论 1：近 3 个月内申请平台总数>1 的客户占比高达 95%以上，即至少有 2 家机构近 3 个月内向同盾查询过客户的三方信用信息，近 3 个月内申请平台总数>5 的客户占比高达 60%以上，近 3 个月内申请平台总数>10 的客户占比大约为 25%，这说明 H 公司的小微商户客群与各类信用机构的日常接触极为频繁。

表 7: 同盾近 3 个月内申请平台总数分布

近 3 个月内申请平台总数	频数	百分比	累积频数	累积百分比
0-1	685	4.48%	685	4.48%
2	1045	6.83%	1730	11.30%
3	1396	9.12%	3126	20.43%
4	1440	9.41%	4566	29.84%
5	1418	9.27%	5984	39.10%
6	1292	8.44%	7276	47.54%
7	1228	8.02%	8504	55.57%
8	1126	7.36%	9630	62.92%
9	1008	6.59%	10638	69.51%
10	857	5.60%	11495	75.11%
11	676	4.42%	12171	79.53%
12	599	3.91%	12770	83.44%
13	447	2.92%	13217	86.36%
14	414	2.71%	13631	89.07%
15	337	2.20%	13968	91.27%
16	253	1.65%	14221	92.92%
17	196	1.28%	14417	94.20%
18	151	0.99%	14568	95.19%
19	141	0.92%	14709	96.11%
20	104	0.68%	14813	96.79%
20+	491	3.21%	15304	100.00%

说明：上表中无法区分近 3 个月内申请平台总数为 0 或者 1，这是由于触发同盾多头报告的条件是“近 3 个月内申请平台总数 $\geq 2$ ”。

结论 2：P2P 网贷、小额贷款公司、一般消费分期平台、大型消费金融公司、银行消费金融公司这五类最常申请融资渠道的“近 3 个月申请平台数”客户数分布如下图。

图 11：同盾近 3 个月申请 P2P 网贷平台数

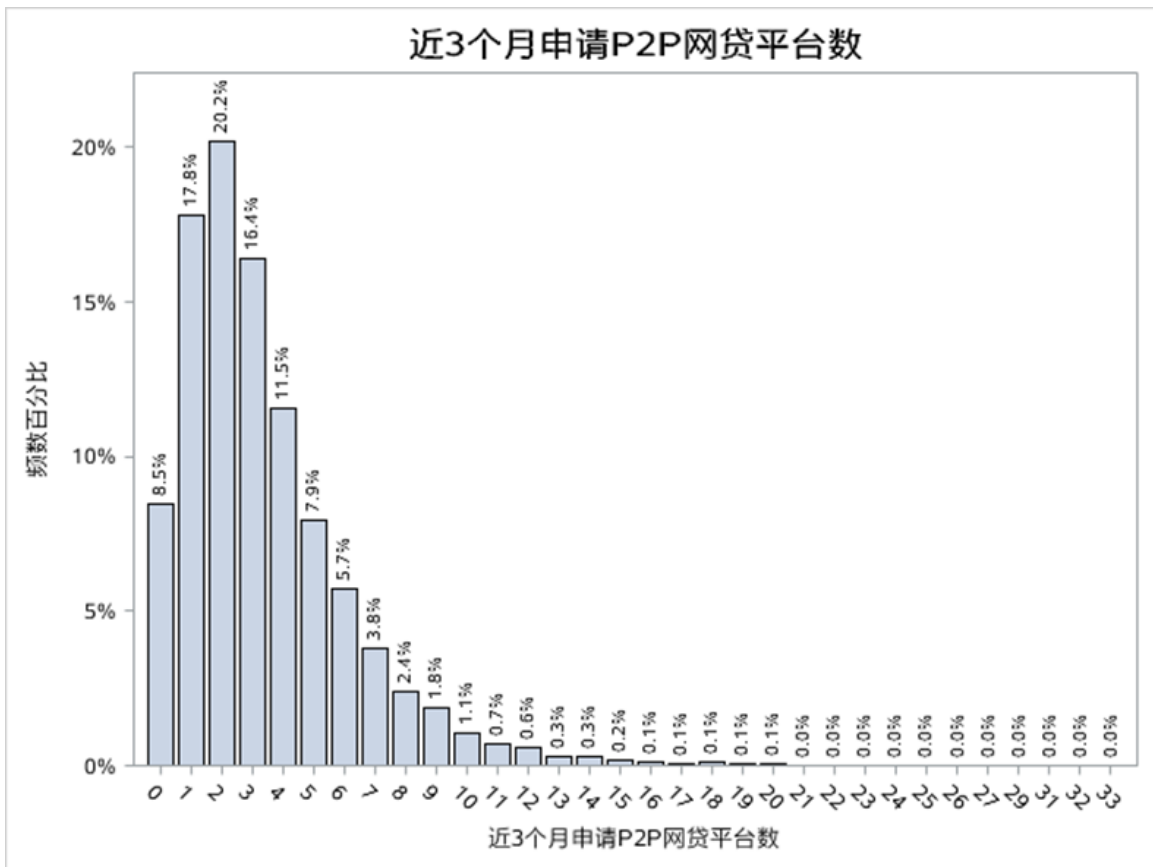


图 12: 同盾近 3 个月申请小额贷款公司平台数

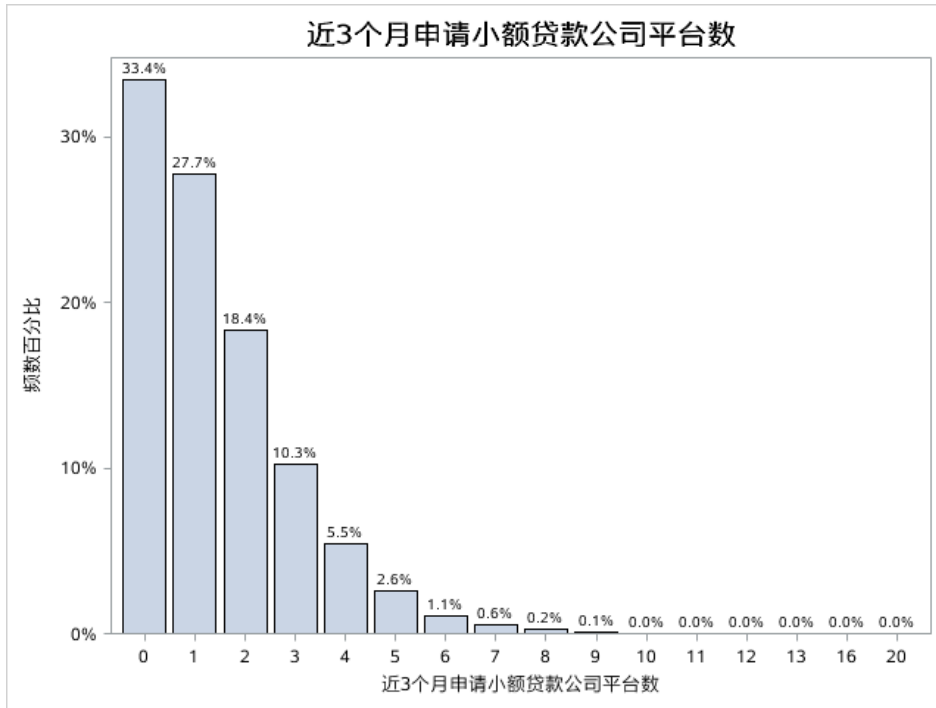


图 13: 同盾近 3 个月申请一般消费分期平台数

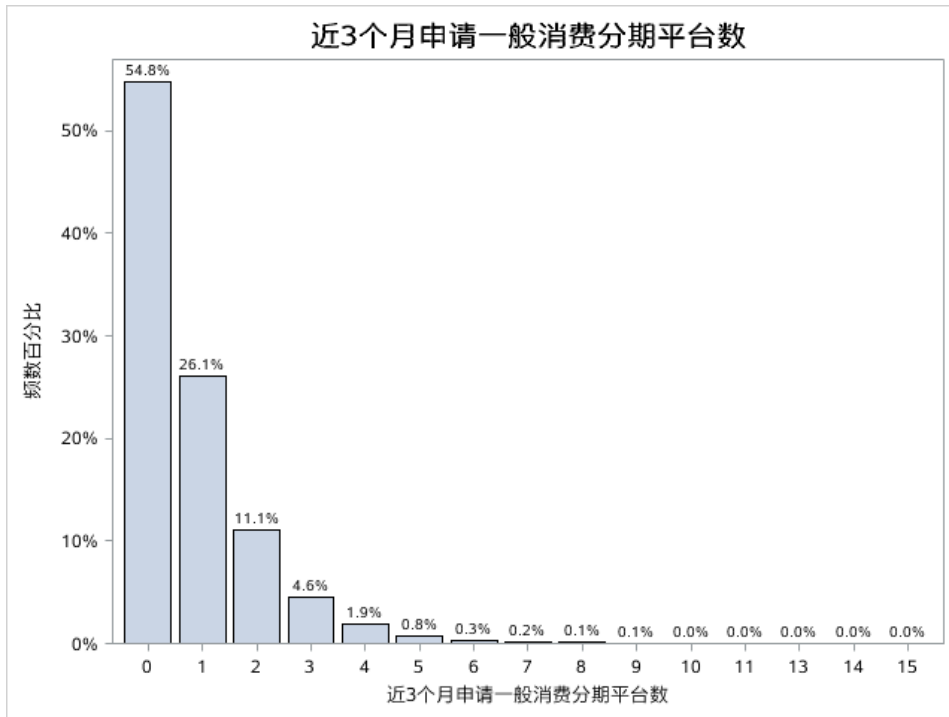


图 14: 同盾近 3 个月申请大型消费金融公司平台数

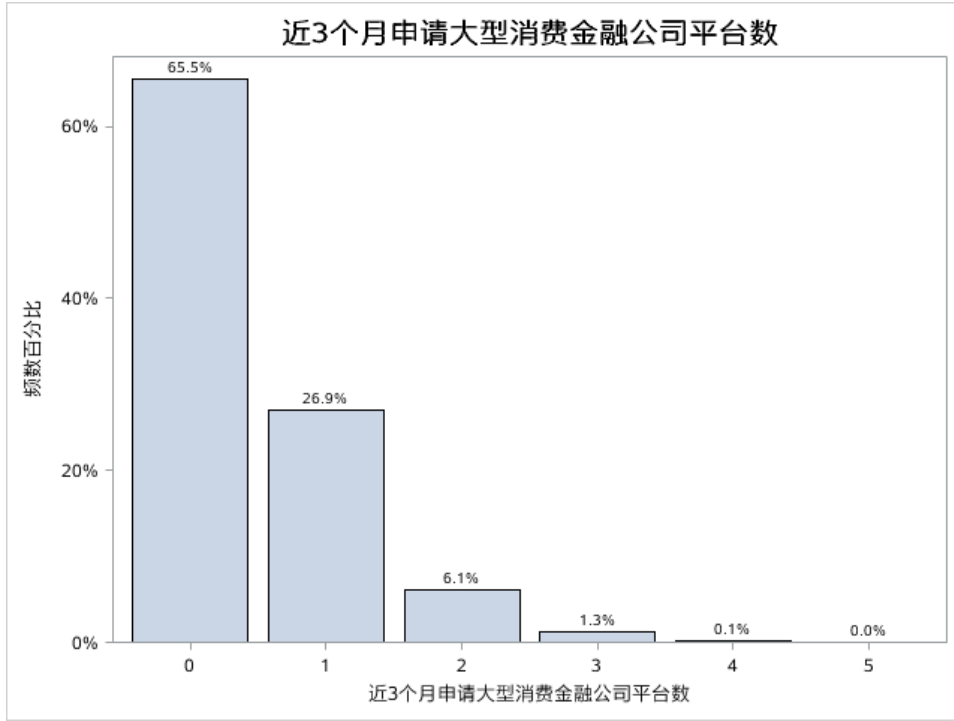
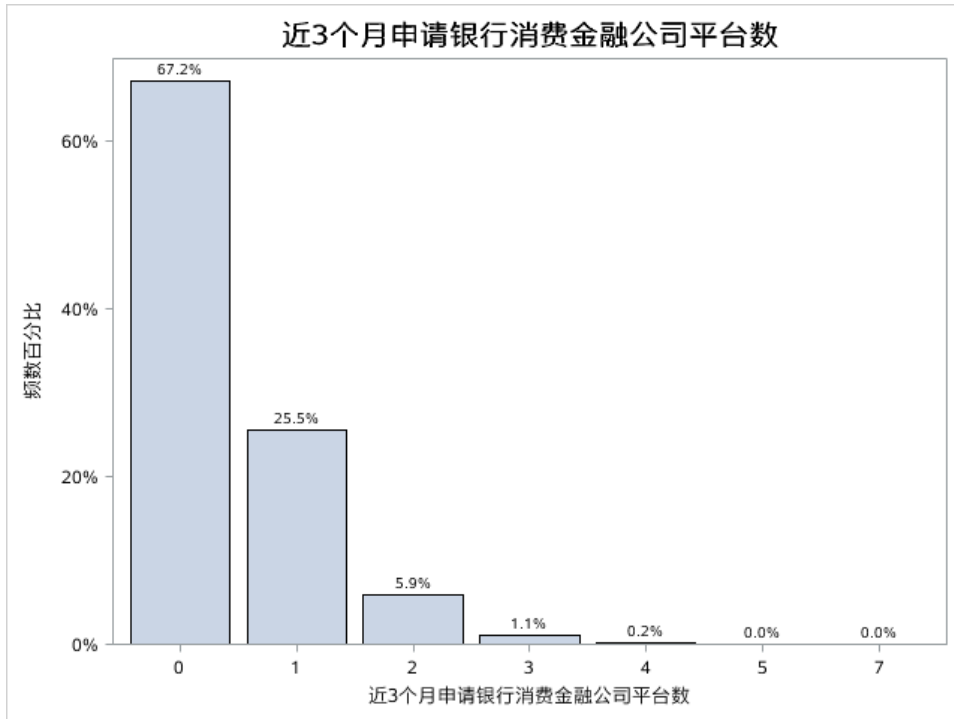


图 15: 同盾近 3 个月申请银行消费金融公司平台数



### 5.3 融资渠道构成与分布

除了关注融资行为频率外，哪些机构是小微商户最常申请或最偏好的融资渠道，是另一个非常值得研究的问题。正如我们在 3.1 理论基础中的阐述，当前国内的融资渠道已日趋多样化，不仅仅是传统银行，诸如消费金融公司、小额贷款公司、P2P 网贷平台等新型融资渠道不断涌现。新老共存的现状下，小微商户更倾向于哪种融资渠道和机构，新老占比是否在悄悄发生此消彼长的变化，都是我们接下来要分析的问题。

#### 5.3.1 基于人行征信的融资渠道构成分析

我们以 H 公司小微客户中保留有效人行征信报告的 10201 个客户为样本，统计这 10201 个客户的查询记录中，近 2 年查询原因为“贷款审批”的查询机构名称和类型，可以得到以下结论：

**结论 1：机构“贷款审批”查询客户覆盖率 TOP50。**以下 50 个机构是 H 公司小微客群样本中，查询原因为“贷款审批”的客户覆盖率最高的 TOP50。其中“平安银行”、“兴业消费金融股份有限公司”、“深圳平安普惠小额贷款有限公司”分别是银行、消费金融公司、小额贷款公司这三类融资渠道中贷款客户占有率最高的机构。平安集团旗下的“平安银行”和“深圳平安普惠小额贷款有限公司”贷款客户覆盖率高达 50.17%和 32.97%，在小微商户客群中的优势明显。

表 8：机构“贷款审批”查询客户覆盖率 TOP50

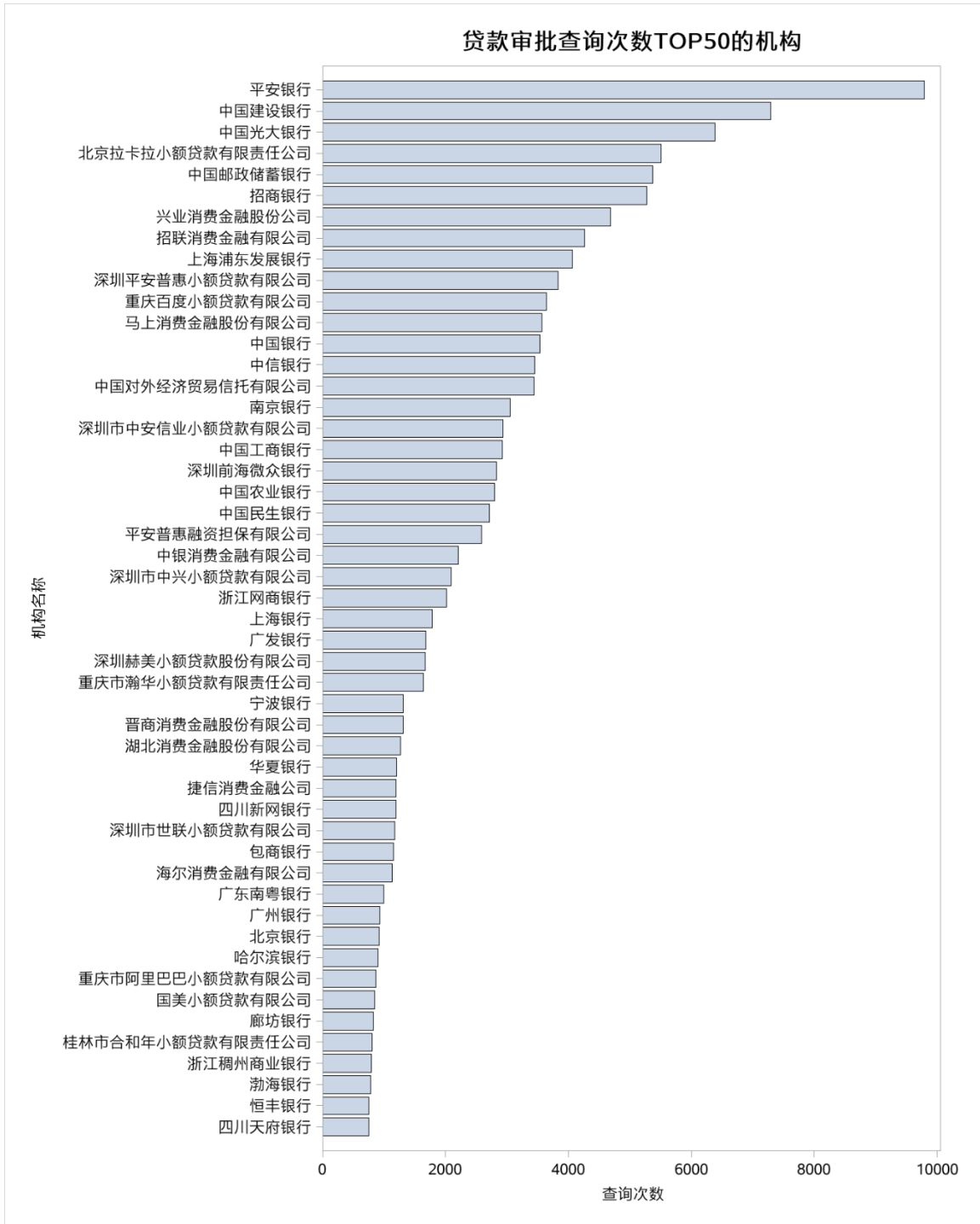
排名	贷款审批查询机构	被查询客户数	被查询客户覆盖率
1	平安银行	5118	50.17%
2	中国光大银行	4061	39.81%
3	兴业消费金融股份公司	3816	37.41%
4	深圳平安普惠小额贷款有限公司	3363	32.97%
5	中国建设银行	3323	32.58%
6	招联消费金融有限公司	3257	31.93%
7	中国邮政储蓄银行	3197	31.34%
8	中国对外经济贸易信托有限公司	2768	27.14%
9	深圳前海微众银行	2755	27.01%
10	马上消费金融股份有限公司	2703	26.50%
11	北京拉卡拉小额贷款有限责任公司	2553	25.03%
12	重庆百度小额贷款有限公司	2537	24.87%
13	南京银行	2412	23.65%
14	上海浦东发展银行	2320	22.74%
15	深圳市中安信业小额贷款有限公司	2290	22.45%
16	招商银行	2287	22.42%
17	中国银行	1978	19.39%
18	平安普惠融资担保有限公司	1886	18.49%
19	中信银行	1841	18.05%
20	中银消费金融有限公司	1816	17.80%
21	中国民生银行	1762	17.27%
22	中国农业银行	1718	16.84%
23	中国工商银行	1582	15.51%
24	深圳赫美小额贷款股份有限公司	1428	14.00%
25	深圳市中兴小额贷款有限公司	1415	13.87%
26	重庆市瀚华小额贷款有限责任公司	1385	13.58%
27	上海银行	1384	13.57%
28	浙江网商银行	1319	12.93%



29	晋商消费金融股份有限公司	1104	10.82%
30	广发银行	1098	10.76%
31	湖北消费金融股份有限公司	1080	10.59%
32	四川新网银行	1032	10.12%
33	海尔消费金融有限公司	988	9.69%
34	深圳市世联小额贷款有限公司	975	9.56%
35	包商银行	942	9.23%
36	捷信消费金融公司	933	9.15%
37	宁波银行	847	8.30%
38	华夏银行	817	8.01%
39	广东南粤银行	800	7.84%
40	重庆市阿里巴巴小额贷款有限公司	793	7.77%
41	北京银行	778	7.63%
42	桂林市合和年小额贷款有限责任公司	756	7.41%
43	廊坊银行	755	7.40%
44	哈尔滨银行	753	7.38%
45	恒丰银行	716	7.02%
46	四川天府银行	695	6.81%
47	国美小额贷款有限公司	689	6.75%
48	苏宁消费金融有限公司	684	6.71%
49	渤海银行	654	6.41%
50	广州银行	630	6.18%

机构“贷款审批”查询次数 TOP50。以下 50 个机构是 H 公司小微客群样本中，查询原因为“贷款审批”的查询次数最高的 TOP50，其中 TOP3 均为银行，而“中国建设银行”、“招联消费金融股份有限公司”、“北京拉卡拉小额贷款有限责任公司”分别是银行、消费金融公司、小额贷款公司这三类融资渠道中贷款审批查询次数最多的机构。

图 16: 机构“贷款审批”查询次数 TOP50



值得关注的是“北京拉卡拉小额贷款有限公司”，能打败众多银行在小微商户申请次数最多的 TOP5 机构中占领一席之地，与其跟小微商户经营活动的紧密联系是分不开的。“拉卡拉”是国内首批获得央行颁发牌照的第三方支付公司，专注于整合信息科技，服务线下实体，从支付切入，全维度为中小微商户的经营赋能。正是由于“拉卡拉”服务的群体是众多小微商户，才更了解小微商户的经营受季节和客源的影响会常出现资金周转困难，且银行常因为小微商户经营规模小、抗风险能力差而将其拒之门外。于是以“拉卡拉”在帮助小微商户收单的同时，提供线上线下的无抵押信贷服务，既增加了商户对“拉卡拉”支付的黏性，也迅速扩张了自己的金融版图。

机构“贷款审批”查询人均次数排名 TOP20。小微商户的融资需求是频繁且长期存在的，信贷机构是否能将新客转化为老客，体现了它们产品和服务的竞争力。因此“贷款审批”人均申请次数多的机构，老客黏性应该更强，复借率（向机构首次借款后结清再借）也应该更高。10201 个有效征信的 H 公司小微商户样本结果显示，以下 20 家机构在客户近 2 年的“贷款审批”查询记录中，查询总次数大于 1000 且人均出现次数最高。招商银行的“闪电贷”、中国建设银行的“快贷”、平安银行的“新一贷”都是无抵押信贷的“明星产品”，拥有良好的新客流量和老客基础是它们成功的原因和结果。

表 9：机构“贷款审批”人均被查询次数

排名	机构名称	贷款审批查 询次数	被查询人数	人均被查询 次数
1	招商银行	5271	2287	2.30477
2	中国建设银行	7293	3323	2.19470
3	北京拉卡拉小额贷款有限责任公司	5507	2553	2.15707
4	平安银行	9786	5118	1.91208
5	中信银行	3456	1841	1.87724
6	中国工商银行	2927	1582	1.85019
7	中国银行	3531	1978	1.78514
8	上海浦东发展银行	4058	2320	1.74914
9	中国邮政储蓄银行	5372	3197	1.68033
10	中国农业银行	2795	1718	1.62689
11	中国光大银行	6382	4061	1.57153
12	宁波银行	1316	847	1.55372
13	中国民生银行	2714	1762	1.54030
14	广发银行	1682	1098	1.53188
15	浙江网商银行	2015	1319	1.52767
16	深圳市中兴小额贷款有限公司	2090	1415	1.47703
17	华夏银行	1202	817	1.47124
18	重庆百度小额贷款有限公司	3642	2537	1.43555
19	平安普惠融资担保有限公司	2584	1886	1.37010
20	马上消费金融股份有限公司	3565	2703	1.31890

融资渠道客户覆盖率统计。将融资渠道分为银行、消费金融公司、小额贷款公司、其他这四种类型（P2P 无人行征信查询权限，因此不在统计之列），并基于前述 10201 个有

有效征信的 H 公司小微客户，统计近 2 年查询原因为“贷款审批”的查询记录，发现客户申请贷款融资的偏好程度（无论从查询次数还是被查询客户覆盖度）降序排列为：1.银行 2.小额贷款公司 3.消费金融公司 4.其他。银行依旧在信贷领域有绝对优势，接近 99%的覆盖率说明，当有融资需求时，几乎所有的小微商户都会寻求正规低利率的银行授信。但小额贷款公司和消费金融公司的覆盖率也令人吃惊，须知这两个融资渠道是近 5 年才快速发展起来的，如今已经达到 80%左右的小微商户客群覆盖率。

表 10：不同融资渠道类型“贷款审批”查询情况

排名	贷款审批查询机构类型	查询记录数	查询记录数占比	被查询客户数	被查询客户覆盖率
1	银行	99897	60.8%	10077	98.78%
2	小额贷款公司	34004	20.7%	8429	82.63%
3	消费金融公司	21973	13.4%	7924	77.68%
4	其他	8340	5.1%	5141	50.40%
总计		164214	100.0%	10201	100.00%

融资渠道受客户偏好程度随时间变化。按查询年份和融资渠道类型细分，统计前述 10210 个客户的 164214 条近 2 年“贷款审批”的查询记录，可以发现银行虽然仍然为 H 公司小微客户最偏好的融资渠道，但其偏好程度（查询记录数占比）在逐年下降，从 2015 年的 77.1%下降至 2019 年的 52.6%，而小额贷款公司和消费金融公司的偏好程度（查询记录数占比）在逐年上升。

表 11：不同融资渠道类型“贷款审批”查询情况按年统计分析

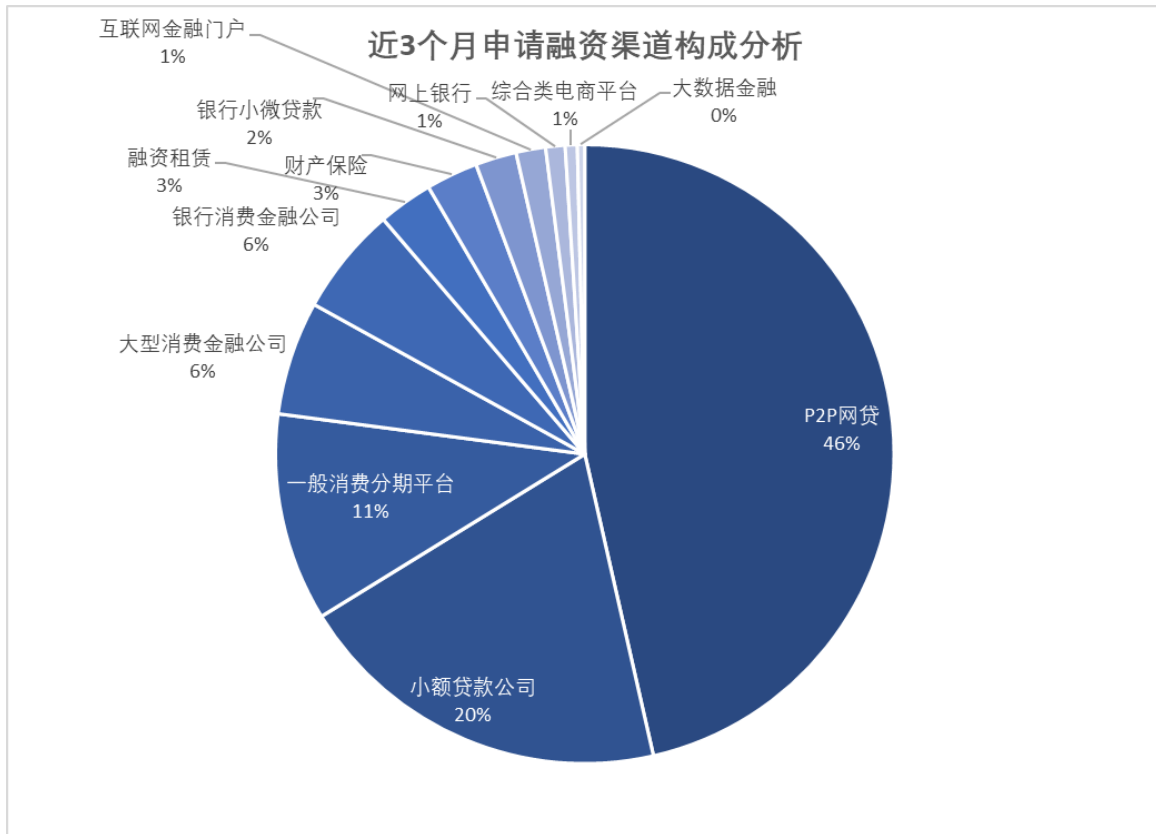
查询年份	融资渠道类型				总计
	银行	小额贷款公司	消费金融公司	其他	
2015	77.1%	9.6%	5.8%	7.5%	100.0%
2016	65.5%	18.6%	9.9%	6.0%	100.0%
2017	60.6%	20.8%	13.9%	4.7%	100.0%
2018	59.6%	21.7%	14.2%	4.5%	100.0%
2019	52.6%	20.9%	15.8%	10.7%	100.0%

### 5.3.2 基于同盾多头的融资渠道构成分析

我们以 H 公司小微客户中保留有效人行征信报告的 15304 个客户为样本，筛选出贷款融资相关的 12 个子行业（P2P 网贷、小额贷款公司、一般消费分期平台、大型消费金融公司、银行消费金融公司、融资租赁、银行小微贷款、互联网金融门户、网上银行、综合类电商平台、大数据金融、直销银行），统计这 15304 个客户在同盾多头报告中的近 3 个月申请 12 类融资渠道平台数及占比，得到以下结论：

结论 1：P2P 网贷平台在同盾近 3 个月申请平台数量中占据 46% 的绝对优势份额占比，即这 15304 个客户申请的平台中 46% 是 P2P 网贷平台。其次是小额贷款公司在近 3 个月查询平台数中占 20% 位居第二。接下来占比较高的融资渠道分别为一般分期消费平台（11%）、大型消费金融公司（6%）、银行消费金融公司（6%）。

图 17: 同盾近 3 个月申请融资渠道构成分析



结论 2: 将这 12 类融资渠道细分年份 (2017-2019 年) 分析构成占比变化, 可以看到 P2P 网贷的申请平台数占比在逐年下降, 2019 年甚至降至只有 2017 年的一半。这样的趋势变化主要与近几年推进的“P2P 专项整治行动”有关。头部和经营较好的 P2P 网贷平台选择申请金融牌照、或转型金融科技公司、或良性退出等出路, 而经营不佳或社会影响恶劣的 P2P 平台则不少已发生暴雷、跑路、被经济侦查等。2020 年 3 月银保监会宣布了一项重大的消息, 三年中 P2P 数量减少了 90%。目前来看 P2P 网贷平台退出历史舞台只是时间问题。

表 12：同盾不同融资渠道类型占比按年统计分析

融资渠道类型	2017 年	2018 年	2019 年
P2P 网贷	53.5%	47.0%	26.3%
小额贷款公司	15.7%	20.8%	26.2%
一般消费分期平台	11.3%	10.8%	10.0%
大型消费金融公司	10.2%	3.7%	3.9%
银行消费金融公司	4.7%	6.1%	6.4%
融资租赁	0.2%	1.9%	13.0%
财产保险	1.8%	2.9%	4.2%
银行小微贷款	0.6%	2.7%	3.9%
互联网金融门户	1.8%	1.4%	1.4%
网上银行	0.1%	1.5%	1.5%
综合类电商平台	0.0%	0.8%	1.8%
大数据金融	0.0%	0.4%	1.4%
总计	100.0%	100.0%	100.0%

小额贷款公司类申请平台数的占比则从 2017 年的 15.7% 迅速上升至到 2019 年的 26.2%，2019 年已几乎和 P2P 网贷类平台占比持平。2017 年末开始的网贷行业整顿给行业带来的一大趋势就是“持牌化”，而“网络小贷牌照”更是最受追捧的香饽饽，不受地域限制且性价比高的优势吸引了众多亟需转型的 P2P 平台和意图进入小额信贷行业的新玩家。仅我们熟知的互联网巨头阿里巴巴旗下的蚂蚁金服就拥有 4 张网络小贷牌照。

银行消费金融公司、银行小微贷款、网上银行等银行类申请平台数的占比近几年也均有显著上升，事实上无论是国有银行、股份制银行还是各地城商行甚至农信社都在积极响应国家号召扶持小微企业和个体工商户，推出了各种小微特色产品，尤其是可随借随还的



循环贷产品，更是一次授信，额度三年有效，期间有资金周转需要可随时提款，资金宽裕后可随时还款，不仅最大程度降低不必要的资金占用和资金成本，更是免除繁琐的一年一次申请手续，深受小微客群喜爱。

#### 5.4 融资行为频率与信用风险的关系

融资行为频率与客户的资金饥渴程度紧密相关，而小微商户正是一群经常需要资金周转用于备货采购等经营活动的客群，因此在前面对于融资行为频率的统计分析中我们可以看到 H 公司提供的小微商户样本中，近期申请银行、消费金融公司、小额贷款公司、P2P 网点平台等各种新老融资渠道贷款的客户比例相当高，人均申请平台数也居高不下。站在风险管理的角度，融资行为频繁并不是件好事，这一方面说明客户的负债率可能不断增加，另一方面客户的经营是否遇到较大问题也是值得担忧的。为了证实“融资行为频率高，信用风险很可能水平”这一推理，我们接下来基于“人行征信”和“同盾多头”两方面数据分析验证。

##### 5.4.1 基于人行征信的融资行为频率与信用风险的关系

首先我们基于人行征信的查询记录中查询原因为“贷款审批”或“个人查询”的记录，衍生了以下 24 个和融资行为频率相关的指标变量：

表 13: 基于人行征信的融资行为频率相关指标变量

Obs	变量名	变量含义
1	QUERY_CNT_3M	近 90 天_贷款审批+个人查询_查询次数
2	QUERY_CNT_6M	近 180 天_贷款审批+个人查询_查询次数
3	QUERY_CNT_12M	近 360 天_贷款审批+个人查询_查询次数
4	QUERY_CNT_24M	近 2 年_贷款审批+个人查询_查询次数
5	QUERY_L_CNT_3M	近 90 天_贷款审批_查询次数
6	QUERY_L_CNT_6M	近 180 天_贷款审批_查询次数
7	QUERY_L_CNT_12M	近 360 天_贷款审批_查询次数
8	QUERY_L_CNT_24M	近 2 年_贷款审批_查询次数
9	QUERY_OWN_CNT_3M	近 90 天_个人查询_查询次数
10	QUERY_OWN_CNT_6M	近 180 天_个人查询_查询次数
11	QUERY_OWN_CNT_12M	近 360 天_个人查询_查询次数
12	QUERY_OWN_CNT_24M	近 2 年_个人查询_查询次数
13	QUERY_BANK_CNT_3M	近 90 天_银行_贷款审批_查询次数
14	QUERY_BANK_CNT_6M	近 180 天_银行_贷款审批_查询次数
15	QUERY_BANK_CNT_12M	近 360 天_银行_贷款审批_查询次数
16	QUERY_BANK_CNT_24M	近 2 年_银行_贷款审批_查询次数
17	QUERY_XF_CNT_3M	近 90 天_消费金融公司_贷款审批_查询次数
18	QUERY_XF_CNT_6M	近 180 天_消费金融公司_贷款审批_查询次数
19	QUERY_XF_CNT_12M	近 360 天_消费金融公司_贷款审批_查询次数
20	QUERY_XF_CNT_24M	近 2 年_消费金融公司_贷款审批_查询次数
21	QUERY_XD_CNT_3M	近 90 天_小额贷款公司_贷款审批_查询次数
22	QUERY_XD_CNT_6M	近 180 天_小额贷款公司_贷款审批_查询次数
23	QUERY_XD_CNT_12M	近 360 天_小额贷款公司_贷款审批_查询次数

基于 10201 个获取有效人行征信报告的 H 公司小微商户样本，将这 24 个变量取值对应的违约率曲线画出，我们可以发现以下结论：

尽管违约率有波动，但从大趋势上看，融资行为频率越高，实际违约率水平越高，即信用风险也越高。如下图中，近 90 天/180 天/360 天/2 年这四个时间切片下，查询原因为“贷款审批”和“个人查询”的次数（QUERY\_CNT\_3M/ QUERY\_CNT\_6M/ QUERY\_CNT\_12M/ QUERY\_CNT\_24M）均与信用风险正相关。

图 18: 近 90 天 贷款审批+个人查询 次数分布和违约率

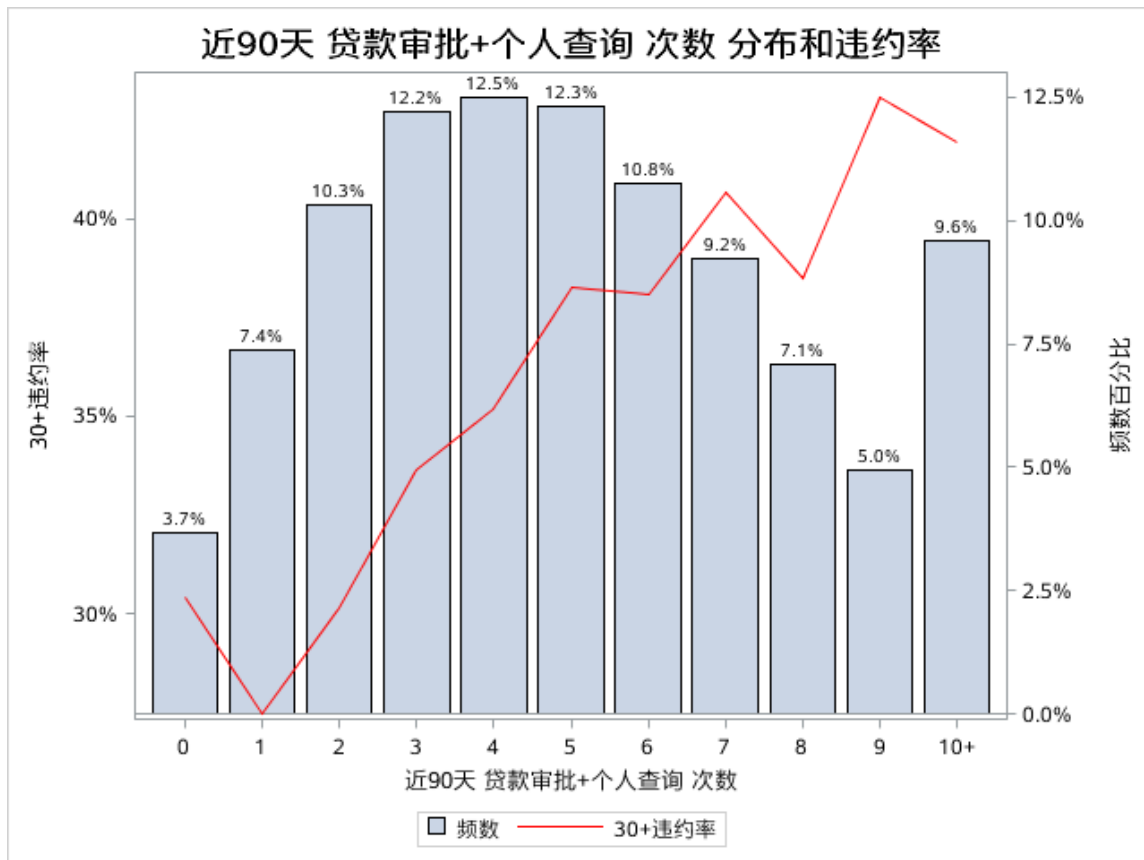


图 19: 近 180 天 贷款审批+个人查询 次数分布和违约率

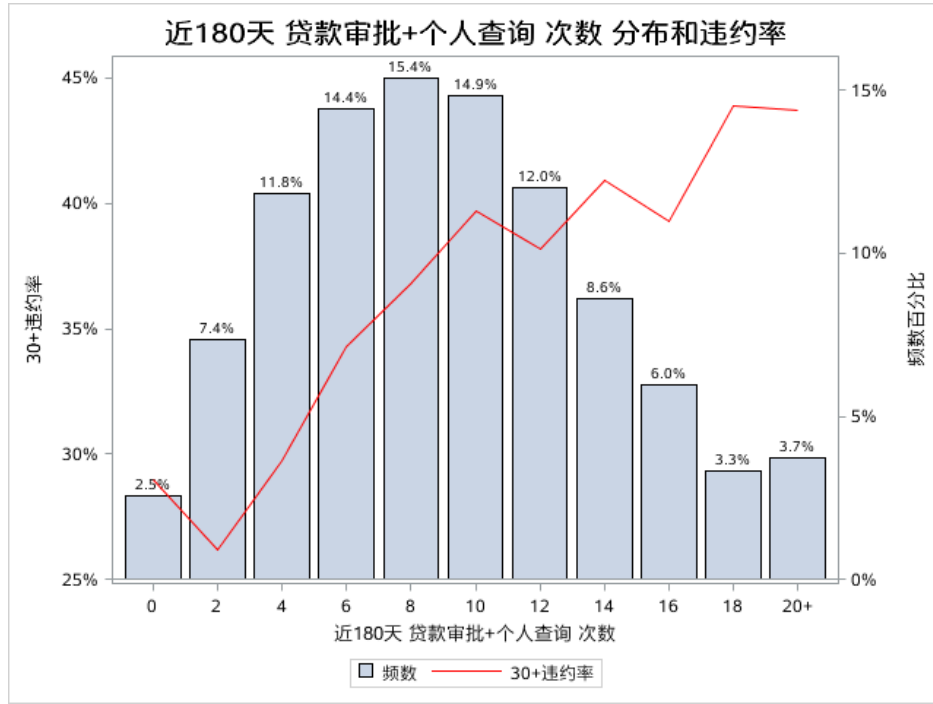


图 20: 近 360 天 贷款审批+个人查询 次数分布和违约率

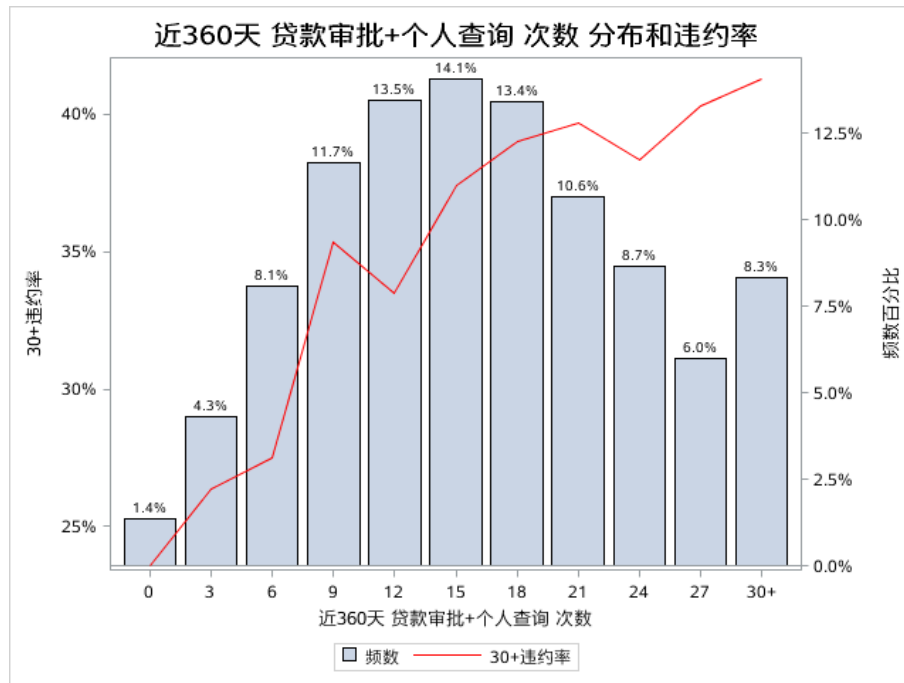
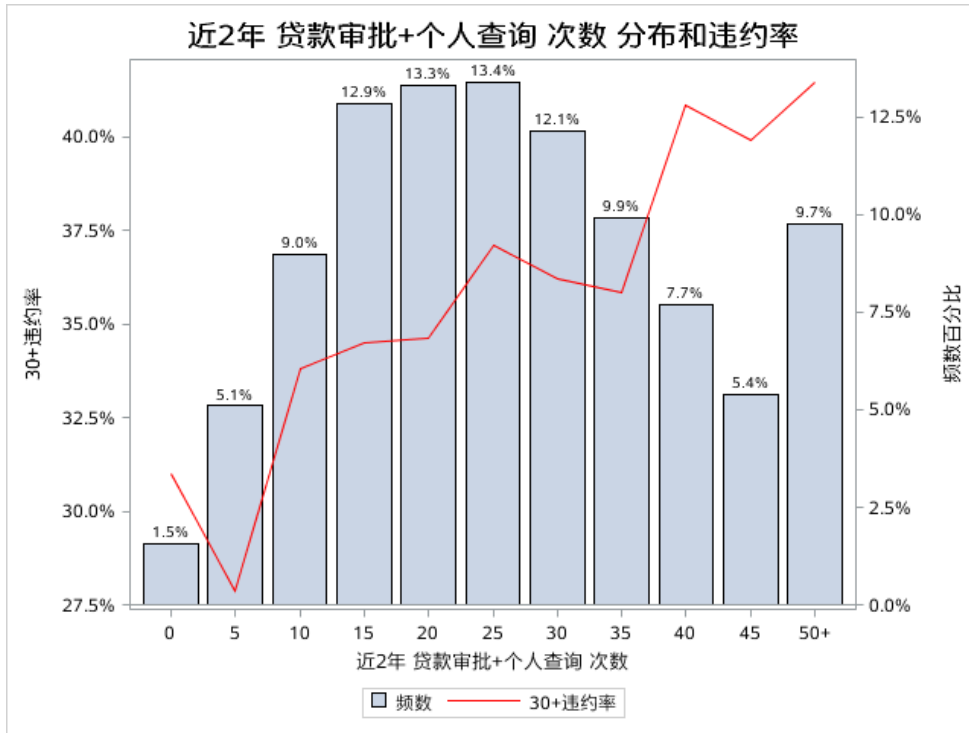


图 21: 近 2 年 贷款审批+个人查询 次数分布和违约率



个人查询次数对信用风险的影响非常显著。查询因为“个人查询”很可能与客户在未接入征信的非正规金融机构的贷款申请行为有关，相比向银行贷款这种行为应该更具风险特征。在下图中，我们可以看到近 90 天/180 天/360 天/2 年这四个时间切片下，个人查询次数（QUERY\_OWN\_CNT\_3M/ QUERY\_OWN\_CNT\_6M/ QUERY\_OWN\_CNT\_12M/ QUERY\_OWN\_CNT\_24M）与违约率的正相关关系表现得更加显著，甚至极少波动。

图 22: 近 90 天 个人查询 次数分布和违约率

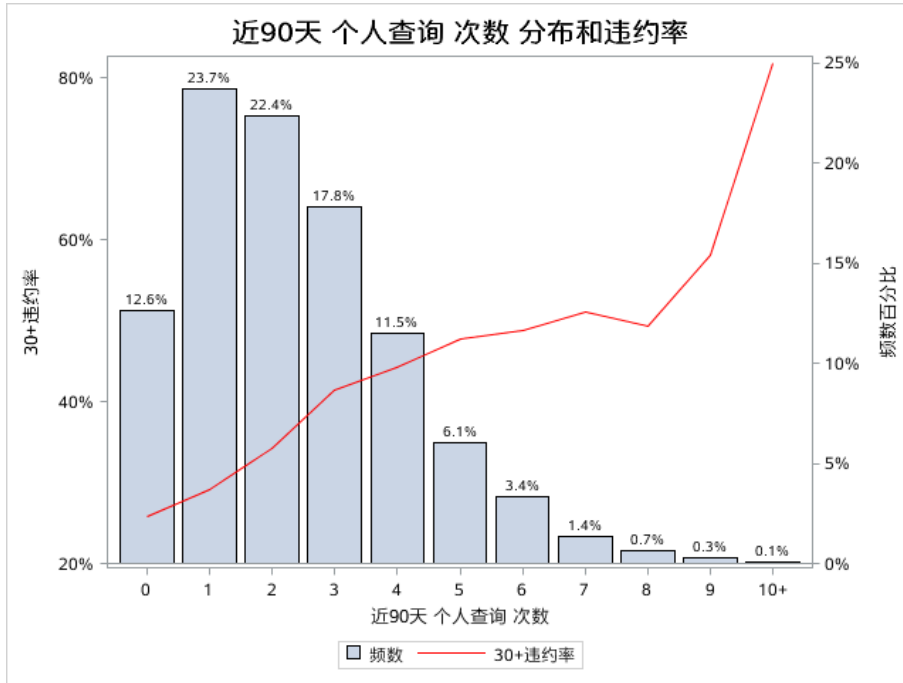


图 23: 近 180 天 个人查询 次数分布和违约率

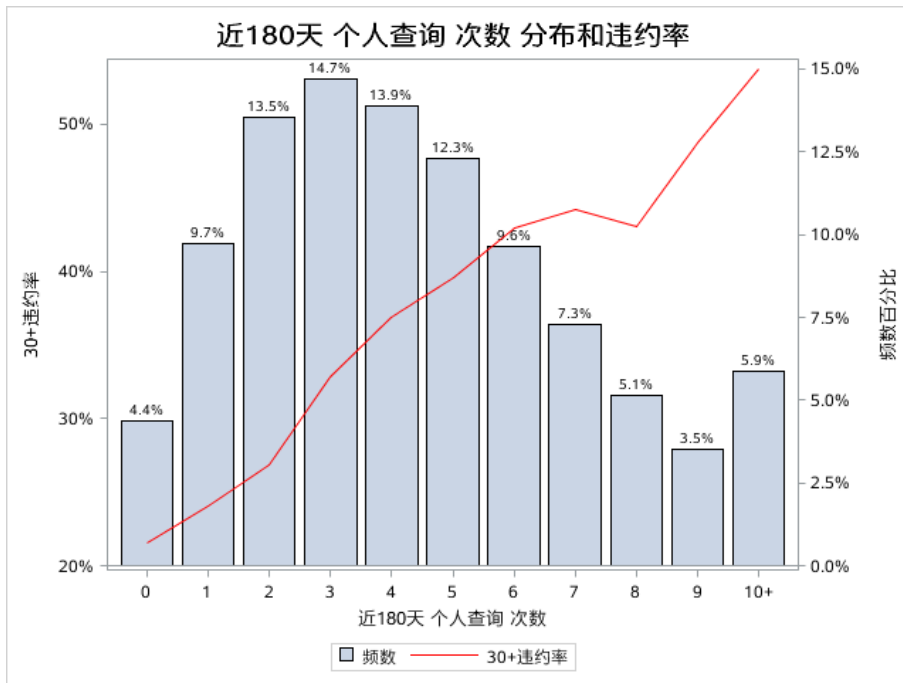


图 24: 近 360 天 个人查询 次数分布和违约率

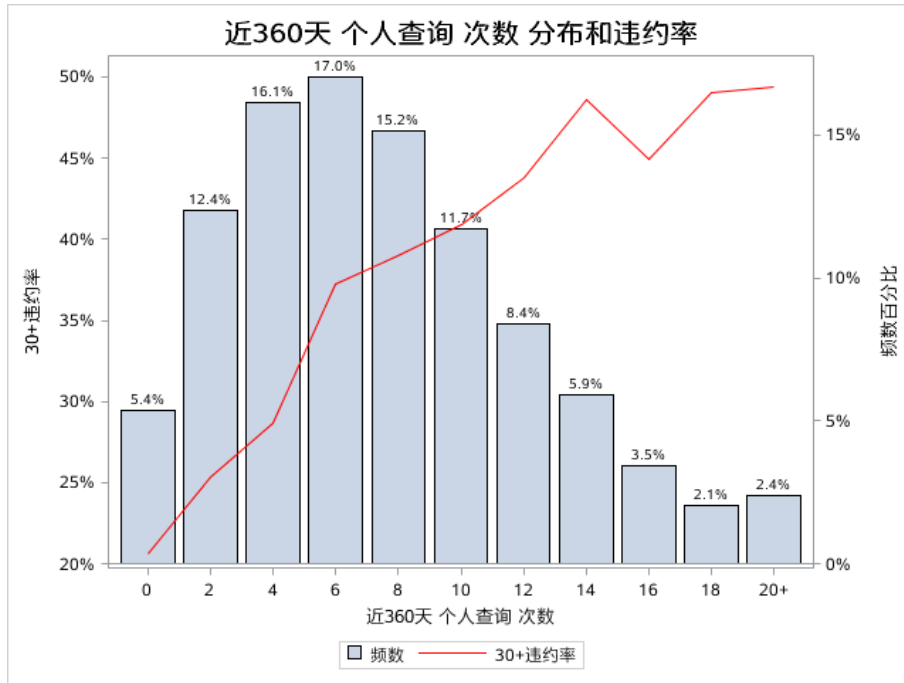
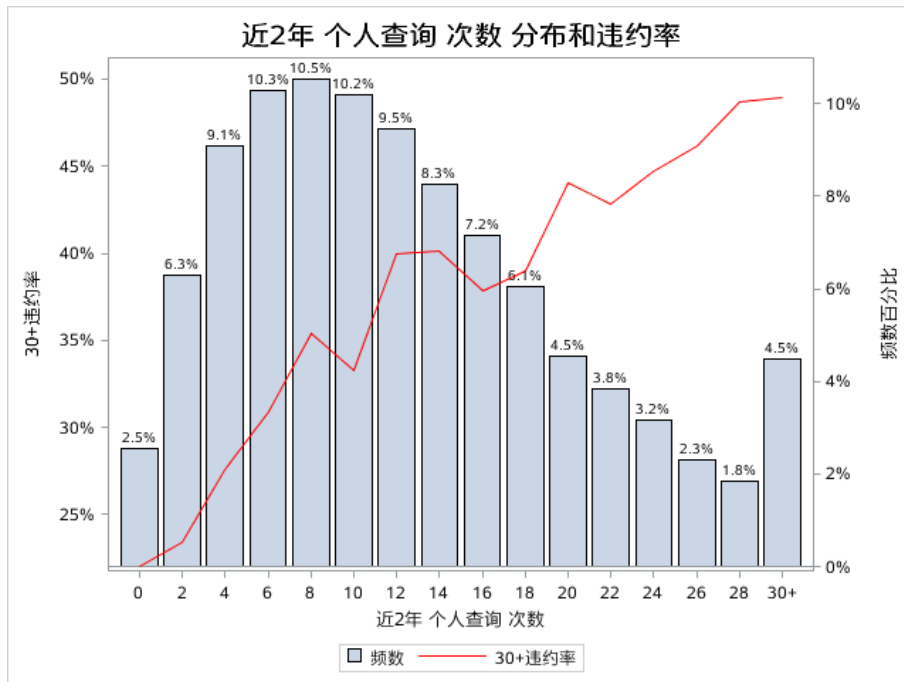


图 25: 近 2 年 个人查询 次数分布和违约率



在银行机构的融资行为频率越高，违约率反而越低，即信用风险更低。如下图我们对不同时间切片统计了银行“贷款审批”查询次数（QUERY\_BANK\_CNT\_3M/ QUERY\_BANK\_CNT\_6M/ QUERY\_BANK\_CNT\_12M/ QUERY\_BANK\_CNT\_24M）以及对应的违约率，均可发现尽管有波动，但融资行为频率和违约率还是大致呈负相关。这一现象可以解释为，银行是融资渠道中最为优质门槛也最高的机构类型，当一个小微商户经常向银行申请贷款，很可能他本身就是银行的目标客群，并早已在银行机构有多次贷款记录，这样的客户通常有良好的资质和信用记录。

图 26：近 90 天 银行 贷款审批 查询次数分布和违约率

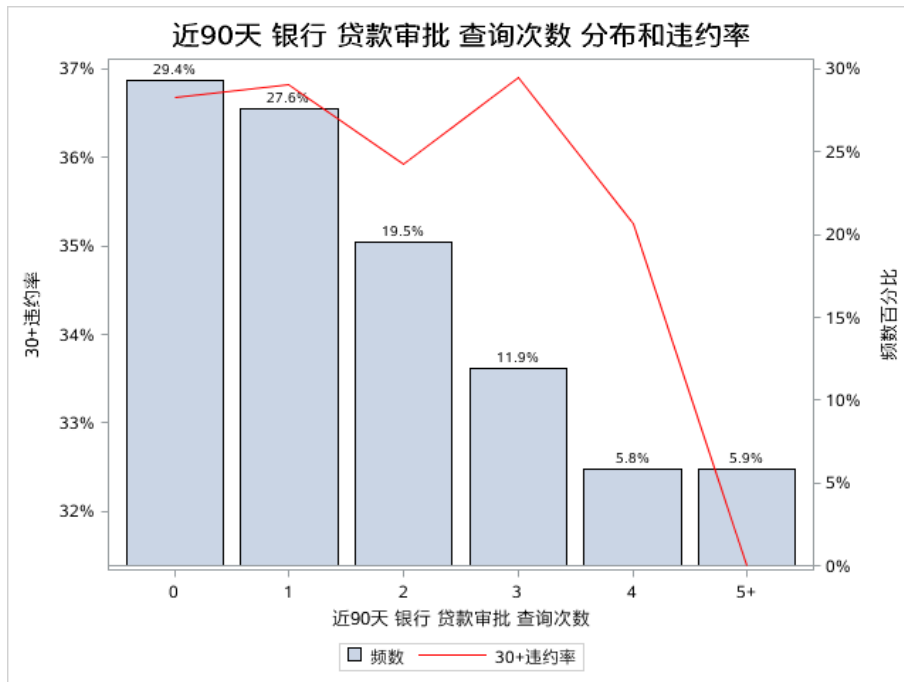




图 27: 近 180 天 银行 贷款审批 查询次数分布和违约率

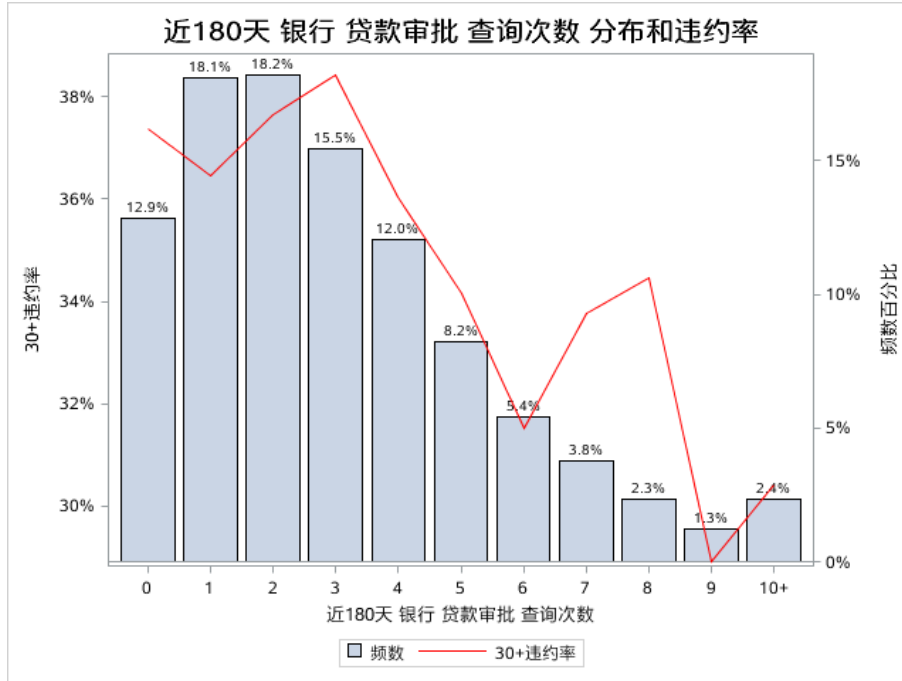


图 28: 近 360 天 银行 贷款审批 查询次数分布和违约率

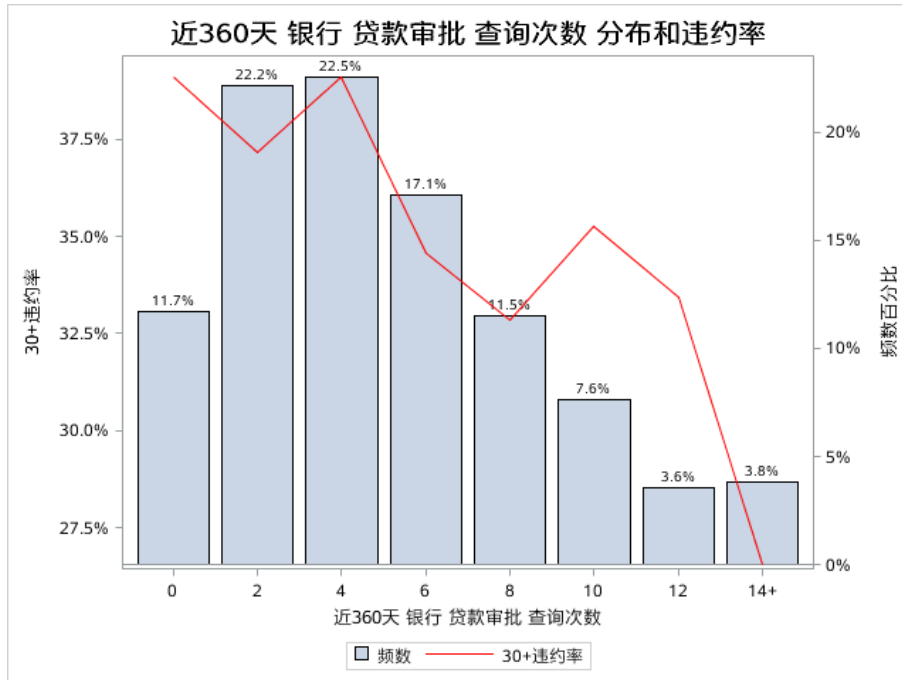
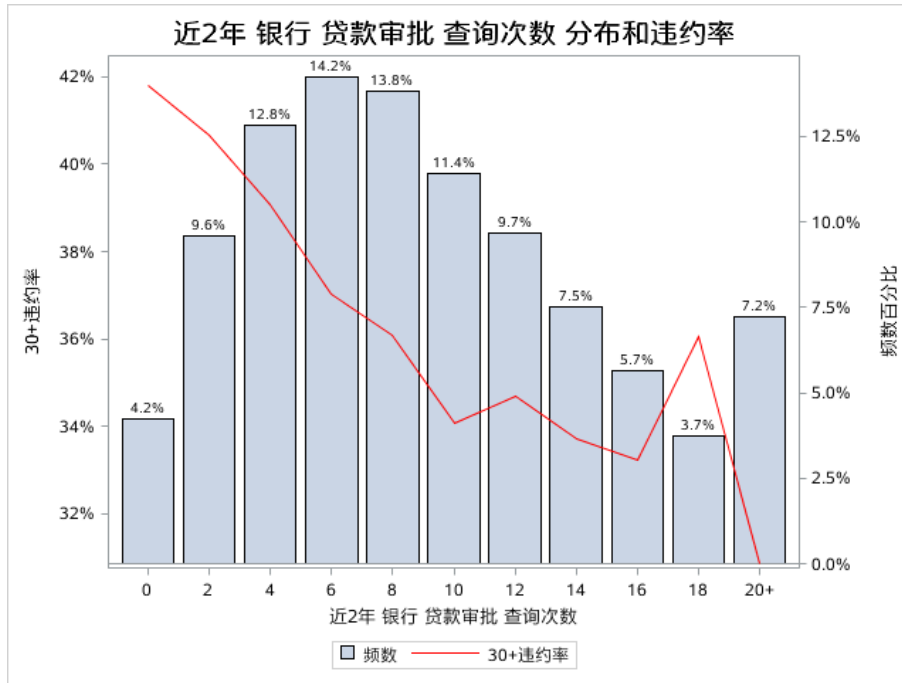


图 29: 近 2 年 银行 贷款审批 查询次数分布和违约率



在消费金融公司的融资行为频率越高，违约率也越高，即信用风险更高。如下图可见，无论哪种时间切片下，消费金融公司“贷款审批”查询次数（ $QUERY\_XF\_CNT\_3M/$   
 $QUERY\_XF\_CNT\_6M/$   $QUERY\_XF\_CNT\_12M/$   $QUERY\_XF\_CNT\_24M$ ）与违约率是非常显著的正相关。

图 30: 近 90 天 消费金融公司 贷款审批 查询次数分布和违约率

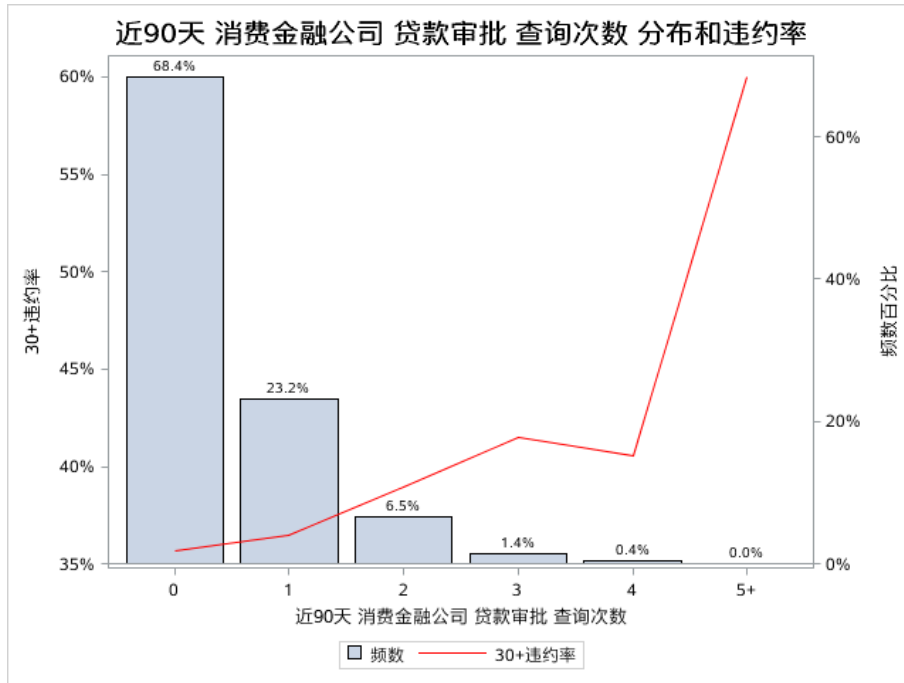


图 31: 近 180 天 消费金融公司 贷款审批 查询次数分布和违约率

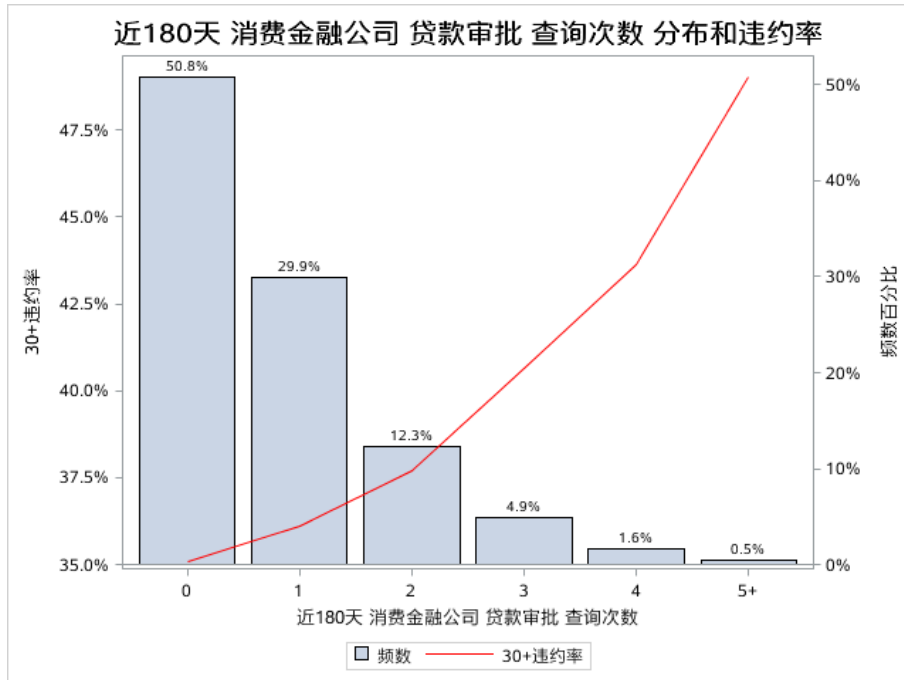


图 32: 近 360 天 消费金融公司 贷款审批 查询次数分布和违约率

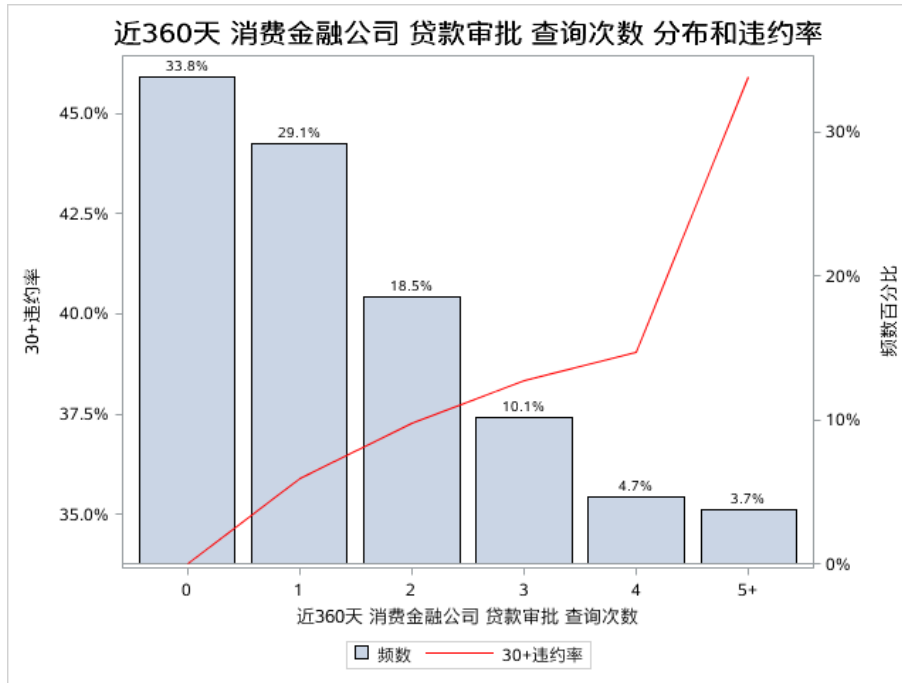
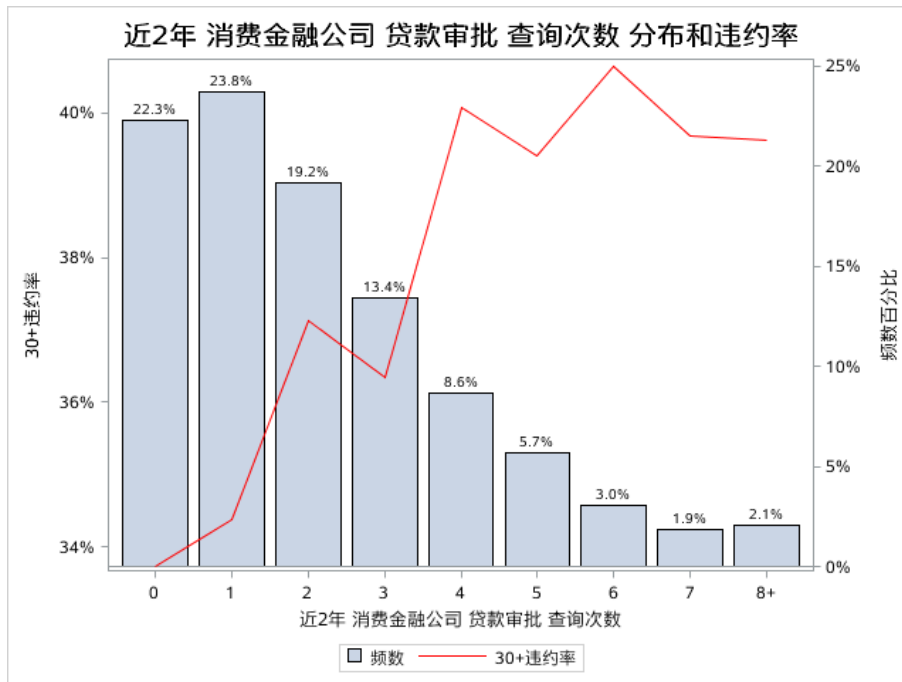


图 33: 近 2 年 消费金融公司 贷款审批 查询次数分布和违约率



短期内在小额贷款公司的融资行为频率越高，违约率也越高，即信用风险更高，但长期来看这一关系则很难成立。如下图可见，近 90 天的时间切片下，小额贷款公司“贷款审批”查询次数（QUERY\_XD\_CNT\_3M）与违约率是比较显著的正相关，但在 180 天/360 天/2 年的时间切片下（QUERY\_XD\_CNT\_6M/ QUERY\_XD\_CNT\_12M/ QUERY\_XD\_CNT\_24M），则并不能发现这两者之间的单调关系。

图 34：近 90 天 小额贷款公司 贷款审批 查询次数分布和违约率

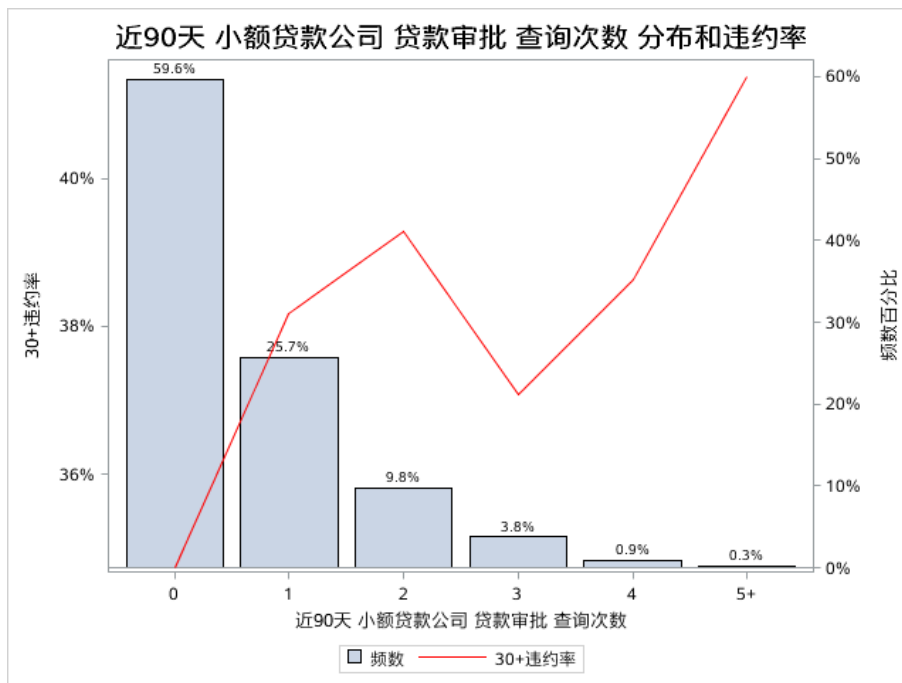


图 35: 近 180 天 小额贷款公司 贷款审批 查询次数分布和违约率

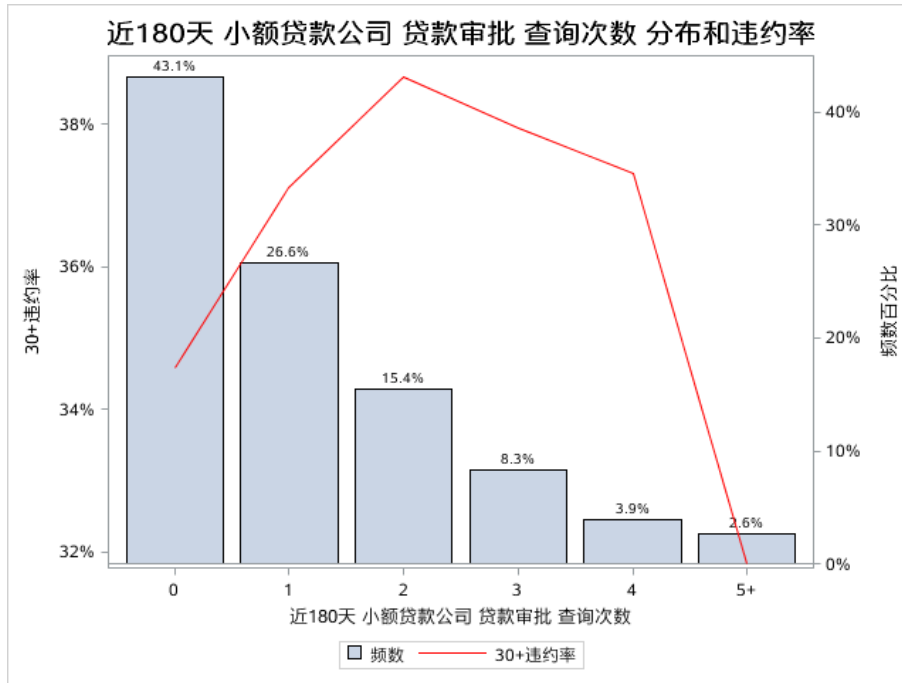


图 36: 近 360 天 小额贷款公司 贷款审批 查询次数分布和违约率

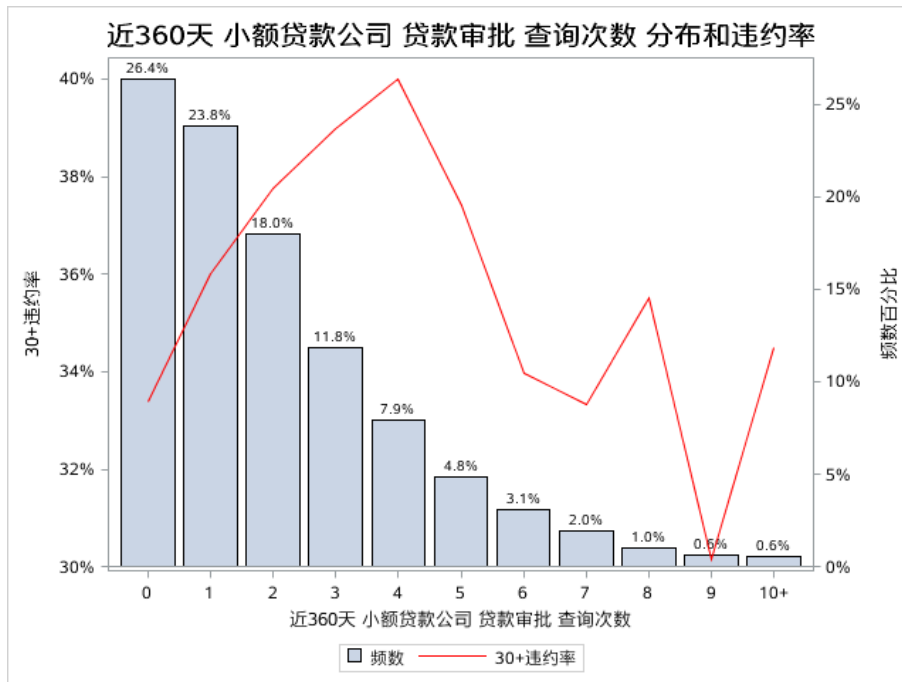
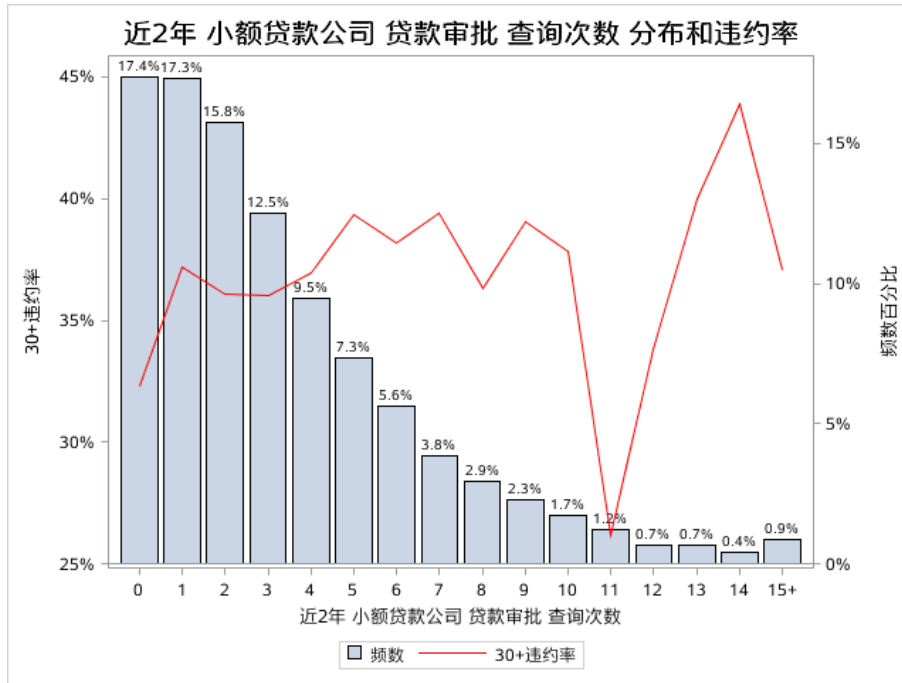


图 37: 近 2 年 小额贷款公司 贷款审批 查询次数分布和违约率



#### 5.4.2 基于同盾多头的融资行为频率与信用风险的关系

正如前文提到的，同盾本身已经基于客户被不同类型查询的记录提取出了一系列多头相关指标，比如近 3 个月查询平台数，以及在 28 个不同子行业的细分查询平台数，变量列表如下：

表 14: 基于同盾多头的融资行为频率指标

Obs	变量名	变量含义
1	TD_3M_CNT	近 3 个月申请平台数
2	var_p2p	近 3 个月申请_P2P 网贷_平台数
3	var_xedk	近 3 个月申请_小额贷款公司_平台数
4	var_ybxffq	近 3 个月申请_一般消费分期平台_平台数
5	var_dxxfjr	近 3 个月申请_大型消费金融公司_平台数
6	var_yhxfjr	近 3 个月申请_银行消费金融公司_平台数

7	var_hlwjr	近 3 个月申请_互联网金融门户_平台数
8	var_xyk	近 3 个月申请_信用卡中心_平台数
9	var_ccbx	近 3 个月申请_财产保险_平台数
10	var_rzzl	近 3 个月申请_融资租赁_平台数
11	var_dsffws	近 3 个月申请_第三方服务商_平台数
12	var_yhxwdk	近 3 个月申请_银行小微贷款_平台数
13	var_wsyh	近 3 个月申请_网上银行_平台数
14	var_lcjg	近 3 个月申请_理财机构_平台数
15	var_yhgryw	近 3 个月申请_银行个人业务_平台数
16	var_zhldspt	近 3 个月申请_综合类电商平台_平台数
17	var_dsjjr	近 3 个月申请_大数据金融_平台数
18	var_fdcjr	近 3 个月申请_房地产金融_平台数
19	var_zxyh	近 3 个月申请_直销银行_平台数
20	var_zczt	近 3 个月申请_资产转让_平台数
21	var_dsfcz	近 3 个月申请_第三方支付_平台数
22	var_o2o	近 3 个月申请_O2O_平台数
23	var_csqcjr	近 3 个月申请_厂商汽车金_平台数融
24	var_qczl	近 3 个月申请_汽车租赁_平台数
25	var_db	近 3 个月申请_担保_平台数
26	var_sbzl	近 3 个月申请_设备租赁_平台数
27	var_xxzj	近 3 个月申请_信息中介_平台数
28	var_jtgjzl	近 3 个月申请_交通工具租赁_平台数
29	var_fwzl	近 3 个月申请_房屋租赁_平台数

上表中细分的子行业高达 28 个，其中属于贷款融资类平台，且查得率较高或者覆盖面较大的子行业如下表，人均申请平台数>0.1，共有 8 家：



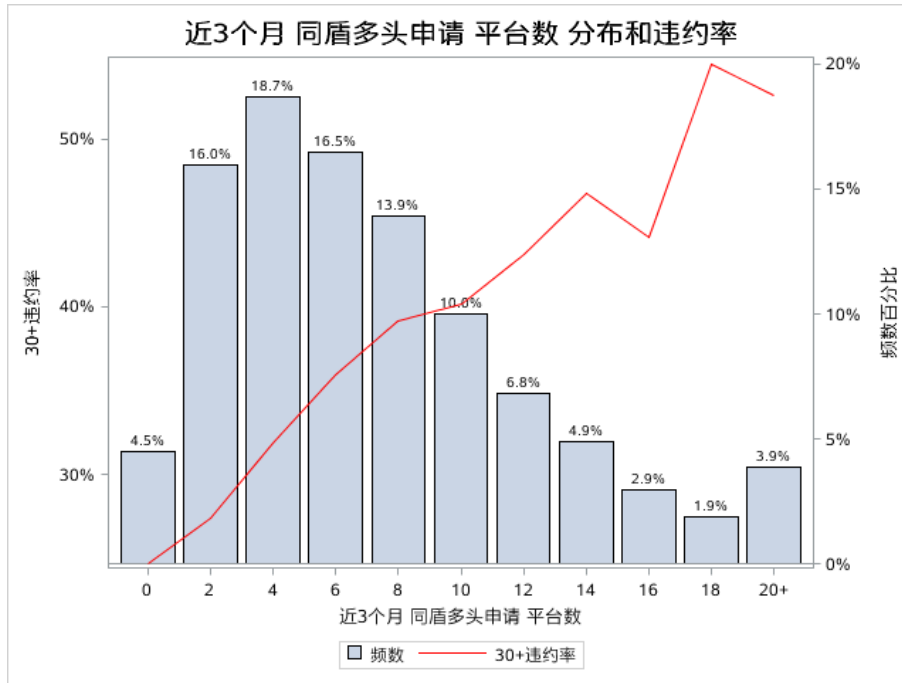
表 15: 同盾主要的贷款融资类平台类型

Obs	变量名	子行业	人均申请平台数
1	var_p2p	P2P 网贷	3.401
2	var_xedk	小额贷款公司	1.448
3	var_ybxffq	一般消费分期平台	0.796
4	var_dxxfjr	大型消费金融公司	0.436
5	var_yhxfjr	银行消费金融公司	0.416
6	var_rzzl	融资租赁	0.209
7	var_yhxwdk	银行小微贷款	0.157
8	var_hlwjr	互联网金融门户	0.113

基于 15304 个查得有效同盾三方数据报告的 H 公司小微商户样本，通过分析同盾整体多头申请频率和细分以上 8 个子行业的申请频率，将这相应变量取值对应的违约率曲线画出，我们可以发现以下结论：

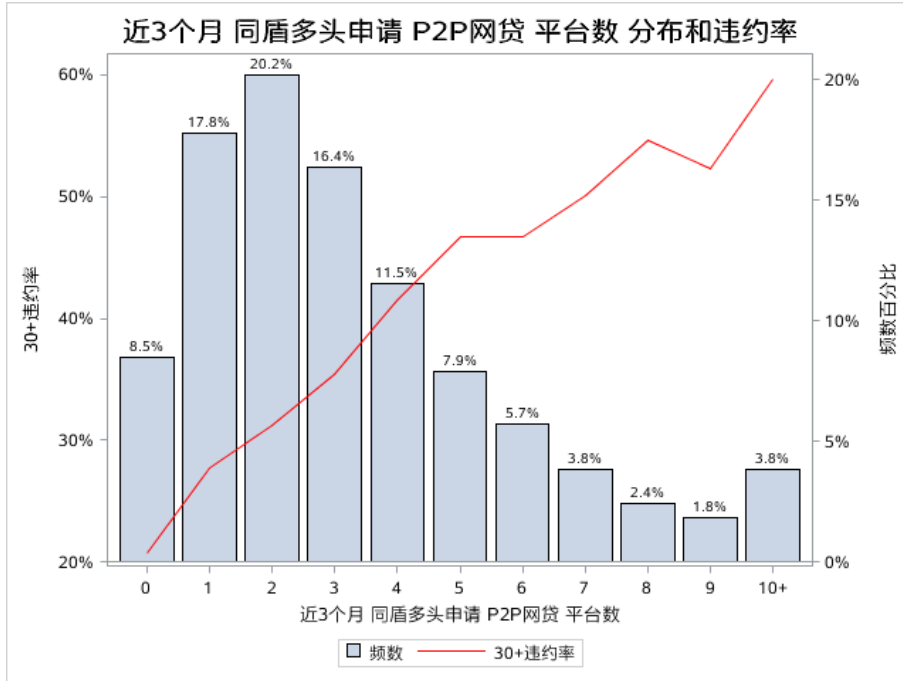
- 1) 尽管违约率有波动，但从大趋势上看，同盾多头报告中近 3 个月申请平台数（TD\_3M\_CNT）越多，实际违约率水平越高，即信用风险也越高。如下图中，违约率曲线呈现显著的上升趋势，仅在多头申请平台数>15 时有些波动。

图 38: 近 3 个月 同盾多头申请 平台数分布和违约率



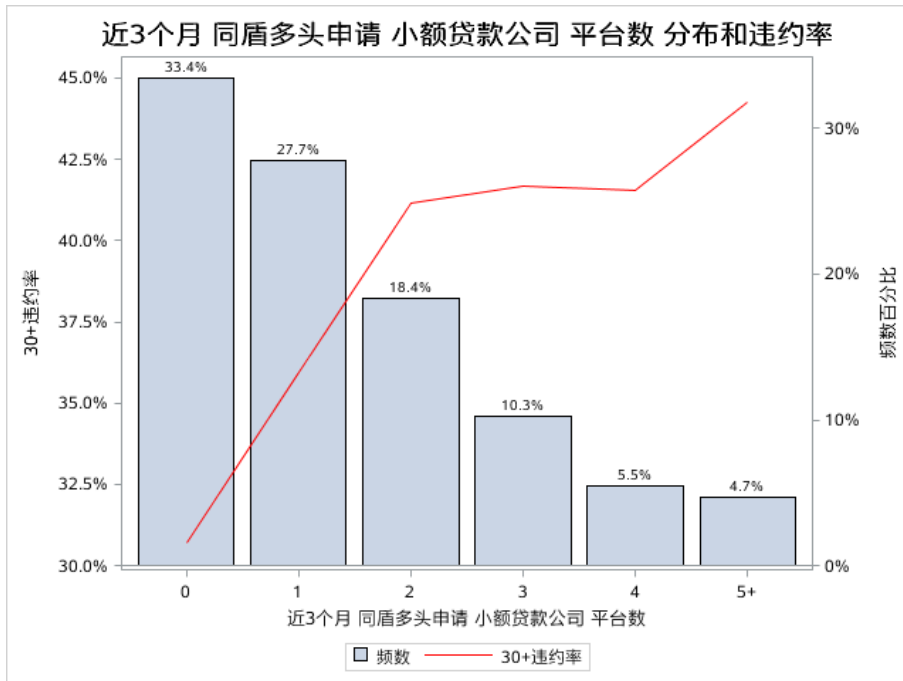
2) 在 P2P 网贷平台的融资行为频率，与违约率显著正相关。从下图可见，近 3 个月申请 P2P 网贷平台数 (var\_p2p) 越多，违约率基本越高，即信用风险也越高。另外，观察左侧第一纵坐标可以发现，var\_p2p 这一指标对违约率的区分程度 (20%-60%左右) 很大，比整体多头指标 TD\_3M\_CNT 对违约率的区分程度 (25%-55%左右) 更大。

图 39: 近 3 个月 同盾多头申请 P2P 网贷 平台数分布和违约率



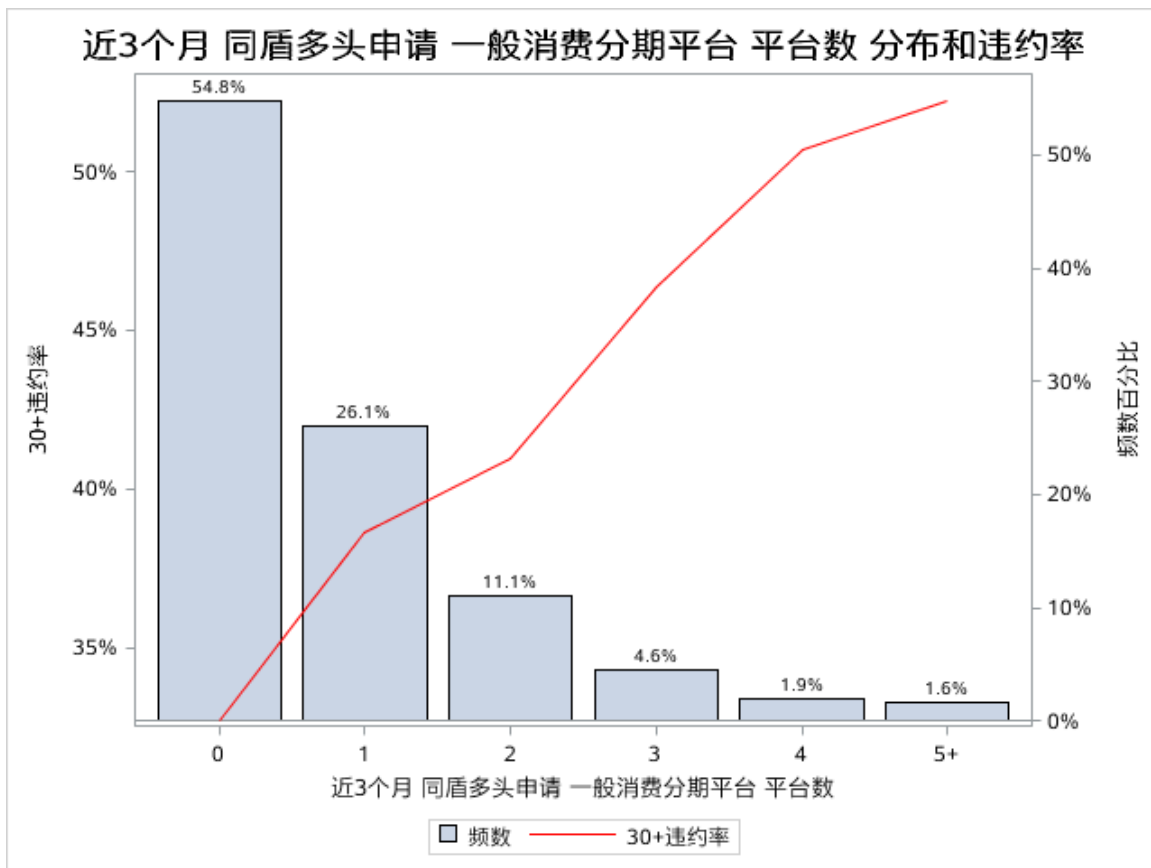
3) 在小额贷款公司平台的融资行为频率 (var\_xedk)，与违约率显著正相关。

图 40: 近 3 个月 同盾多头申请小额贷款公司 平台数分布和违约率



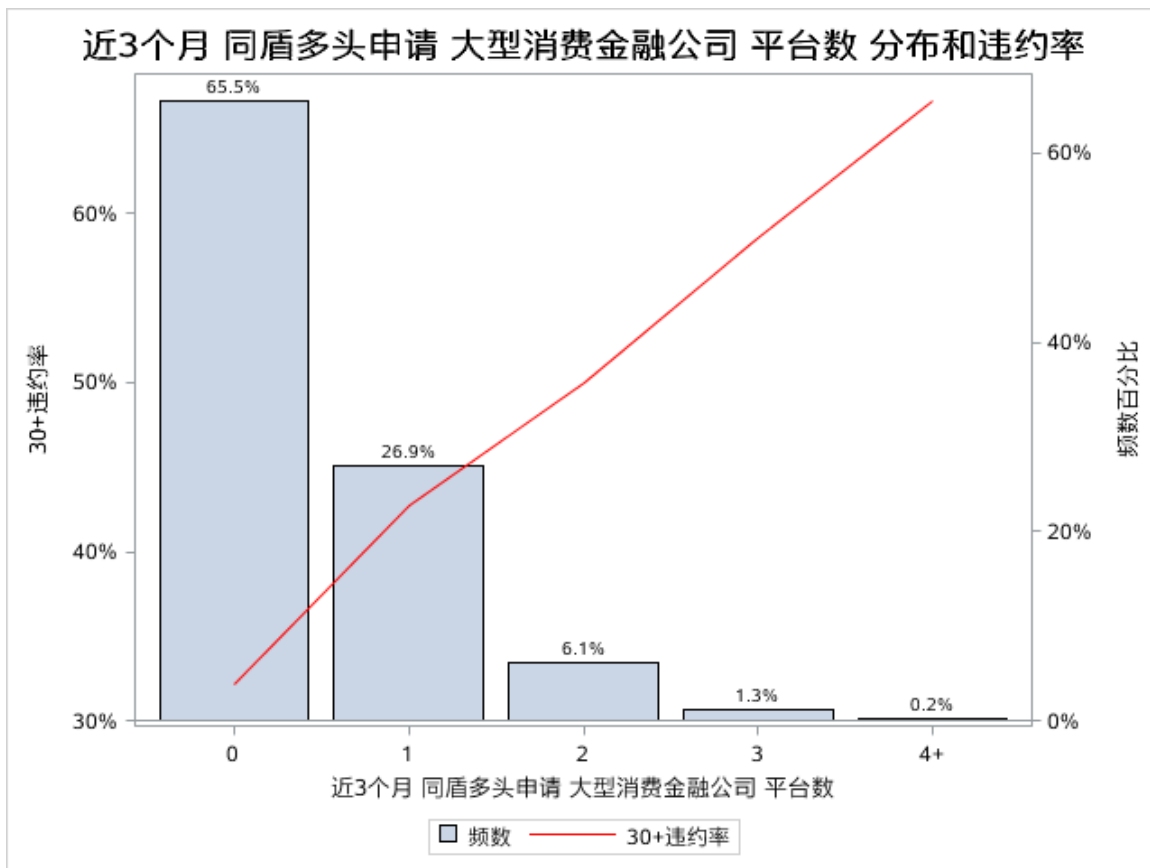
4) 在一般消费分期平台的融资行为频率 (var\_ybxffq) , 与违约率显著正相关。

图 41: 近 3 个月 同盾多头申请一般消费分期平台 平台数分布和违约率



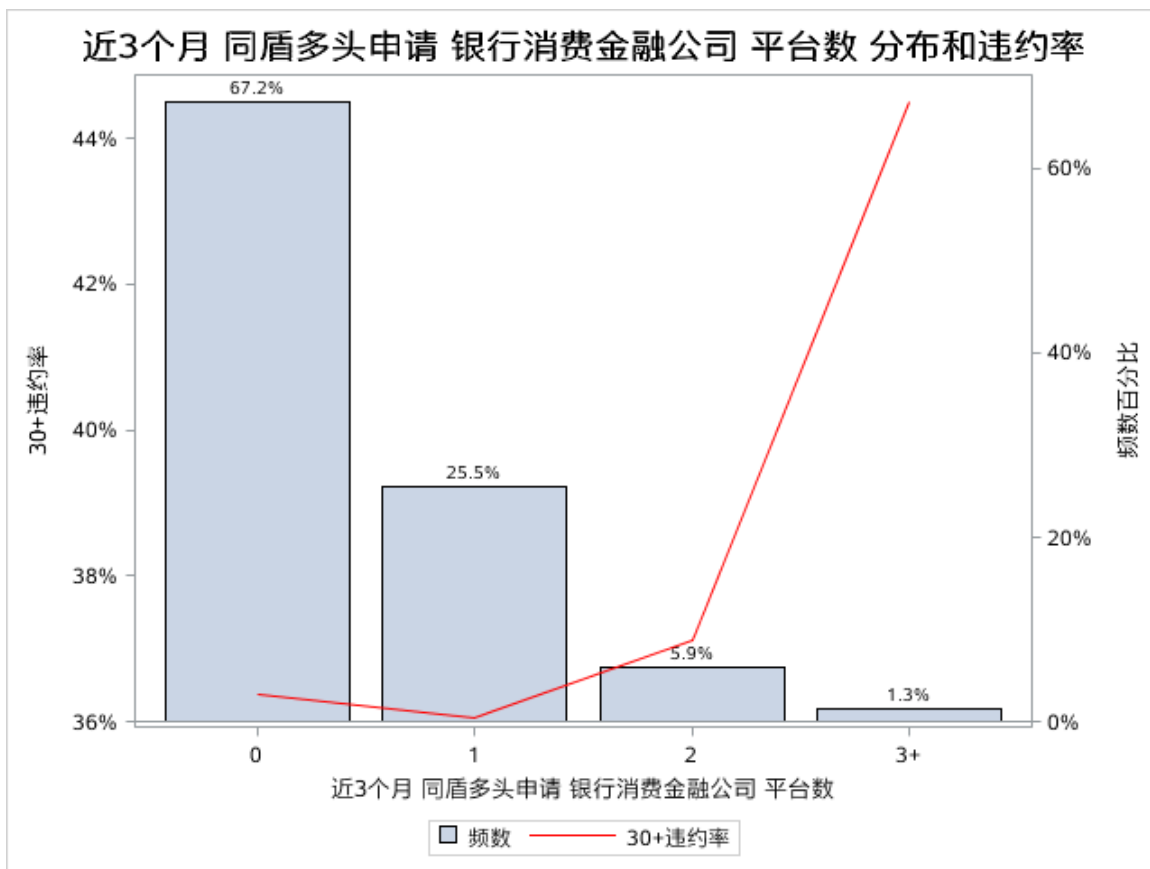
5) 在大型消费金融公司的融资行为频率 (var\_dxxfjr)，与违约率显著正相关。

图 42: 近 3 个月 同盾多头申请大型消费金融公司 平台数分布和违约率



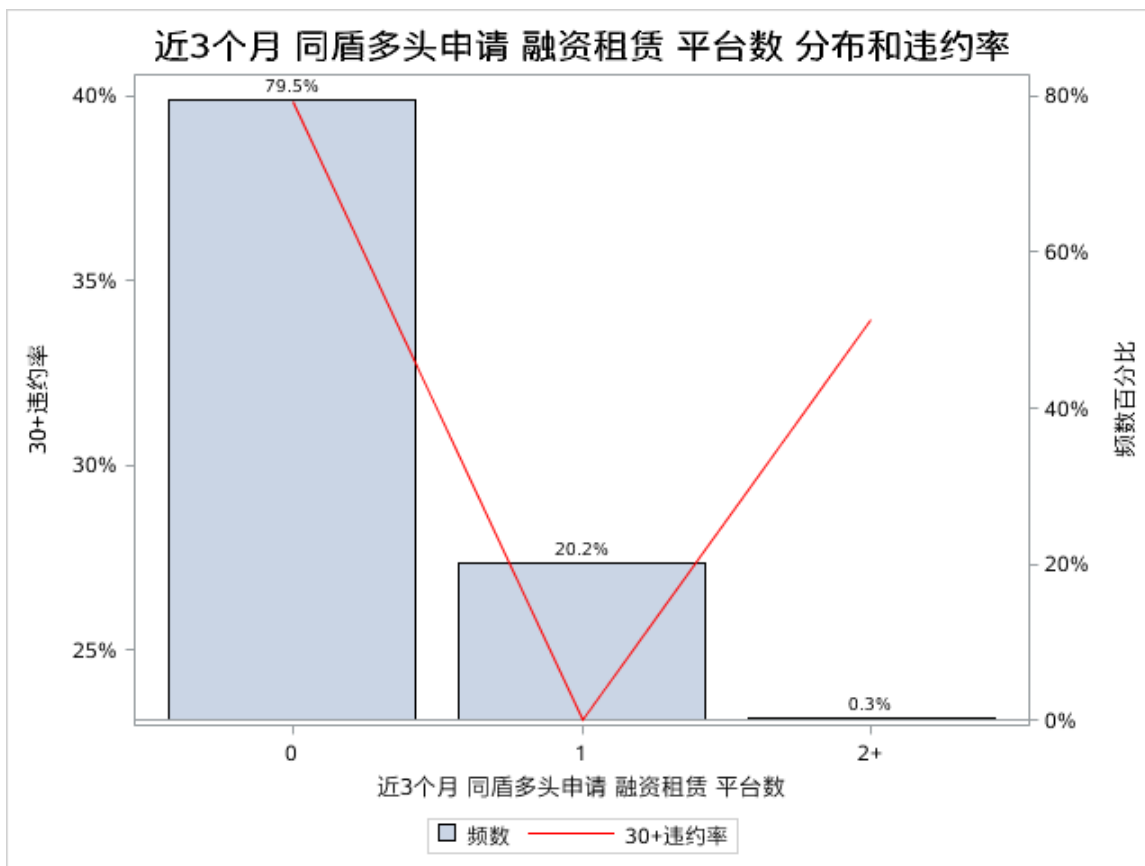
6) 在银行消费金融公司的融资行为频率 (var\_yhxfjr)，较高时违约率有一些上升，但对违约率区分度很小 (35%-45%左右)。

图 43: 近 3 个月 同盾多头申请银行消费金融公司 平台数分布和违约率



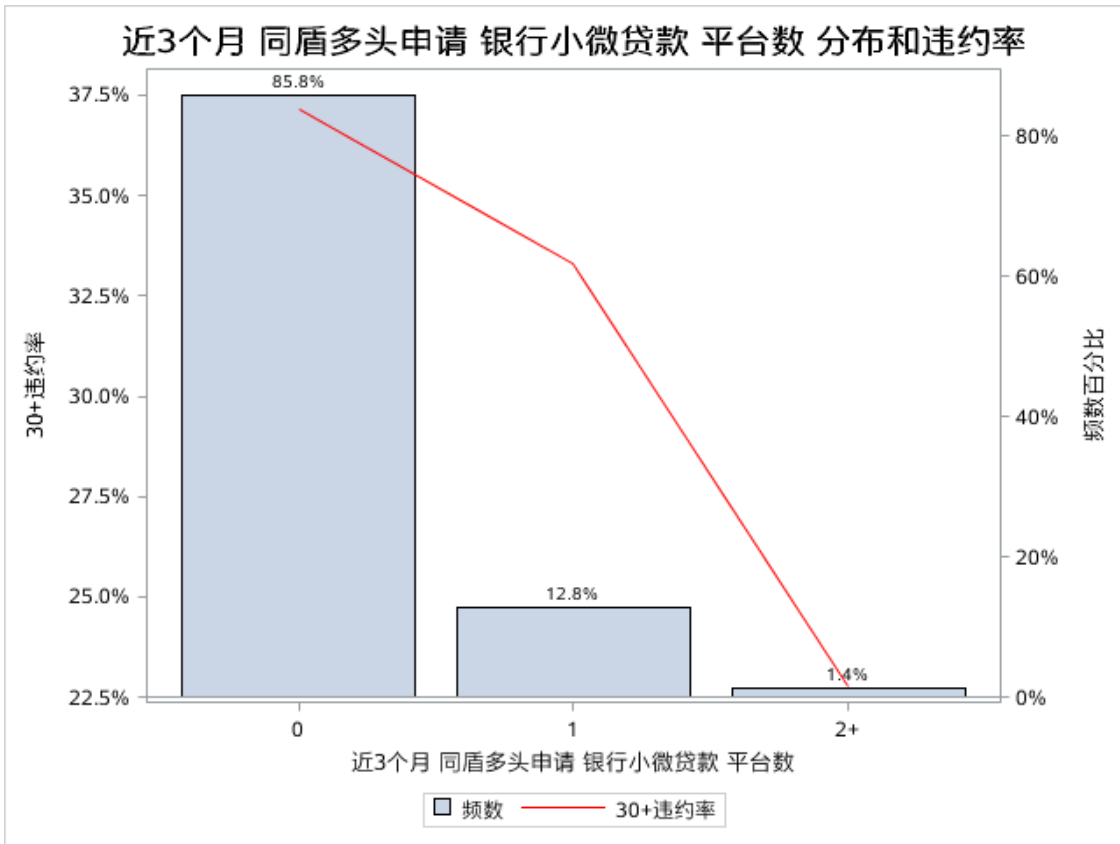
7) 在融资租赁机构的融资行为频率 (var\_rzsl), 0 和 1+ 的信用风险不同。

图 44: 近 3 个月 同盾多头申请融资租赁 平台数分布和违约率



8) 在银行小微贷款机构的融资行为频率 (var\_yhxwdk),与违约率呈负相关,即申请的银行小微贷款机构平台数越多,客户的信用风险反而较低。这一现象与 5.4.1 中基于人行征信的第 3 个结论一致:能在银行获得小微贷款授信的本身已经属于这一群体中的优质客户。

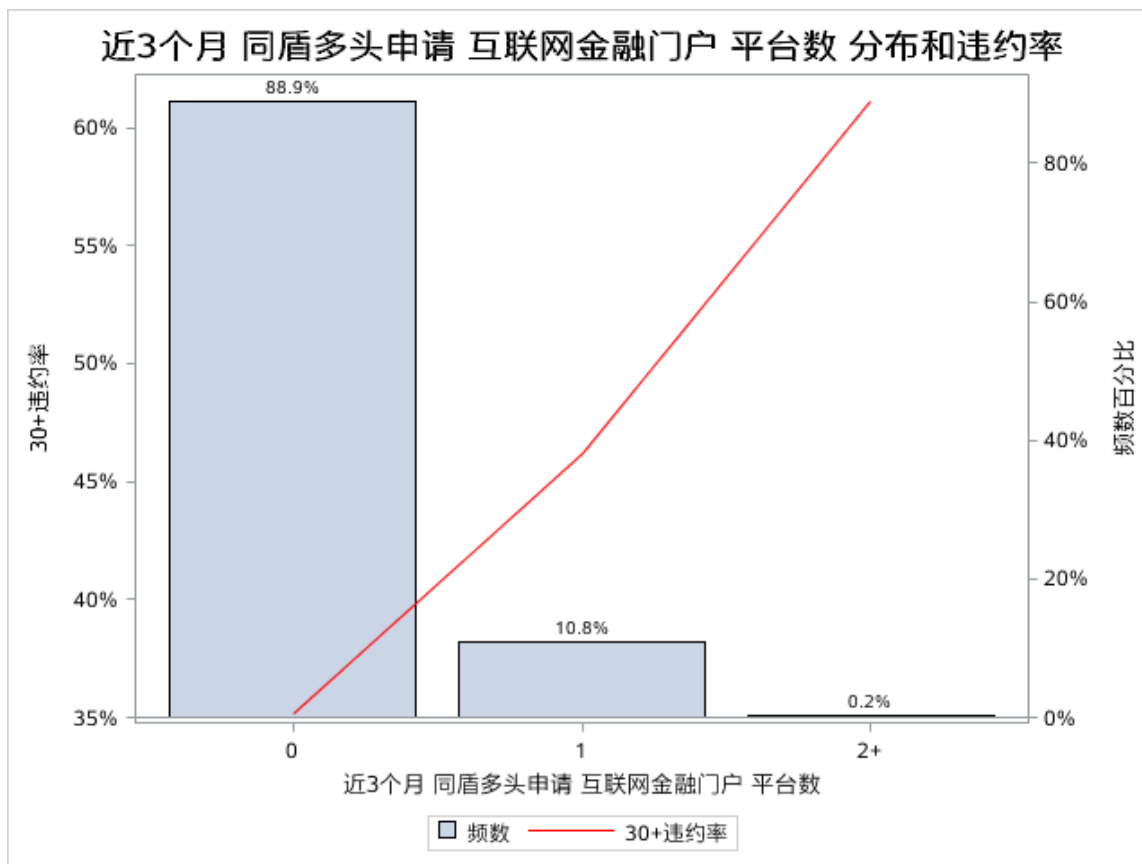
图 45: 近 3 个月 同盾多头申请银行小微贷款 平台数分布和违约率





9) 在互联网金融门户的融资行为频率（var\_hlwjr），与违约率显著正相关。

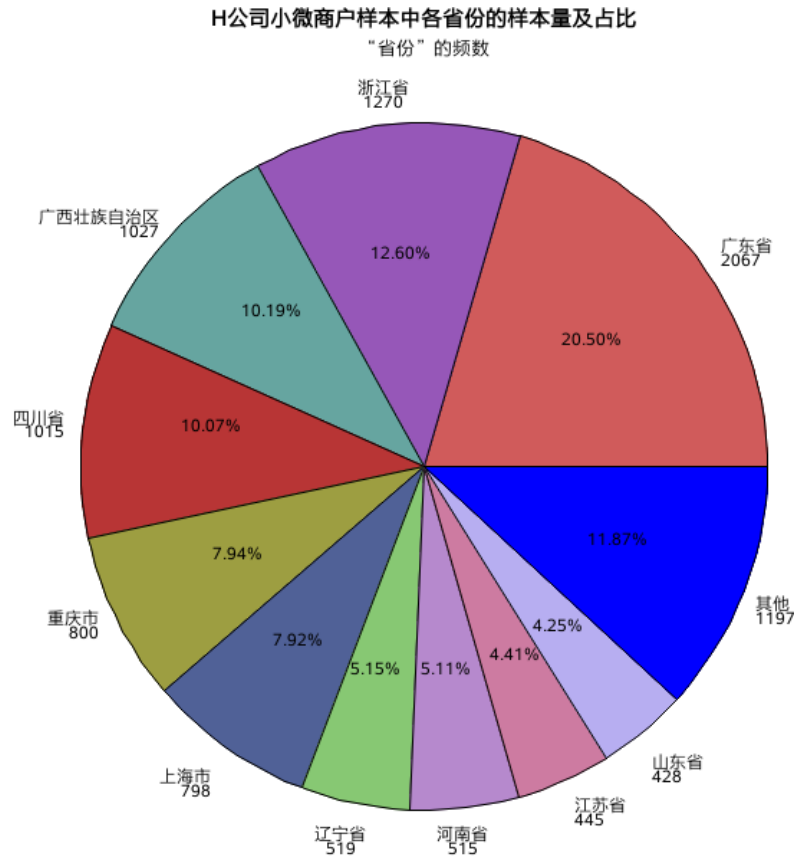
图 46：近 3 个月 同盾多头申请互联网金融门户 平台数分布和违约率



#### 5.4.3 “融资行为频率与信用风险关系”结论的稳定性分析

H 公司是一家全国范围内设有网点且展业的公司，此次研究的小微商户样本，按进件省份的样本占比情况如下：

图 47: H 公司小微商户样本中各省份的样本量及占比



为验证“融资行为频率与信用风险呈正相关”这一结论的稳定性，尽可能排除地区因素对结论的影响，我们对占比超 10% 的四个主要省份——广东省、浙江省、广西壮族自治区、四川省，分别统计同省份内“人行近 90 天 贷款审批+个人查询 次数”以及“近 3 个月 同盾多头申请 平台数”这两个融资行为频率指标与违约率之间关系（如下图），同样发现两者基本呈正相关。尽管不同地区的融资行为活跃程度有差异，例如广东省征信近 90 天贷前查询为 10+ 的占比高达 10.9%，而广西仅为 5.1%，分布图也明显左偏，但在各自省内仍然是查询次数越多，风险相对越高。

图 48: 四个省份内, 人行近 90 天 贷款审批+个人查询 次数分布和违约率

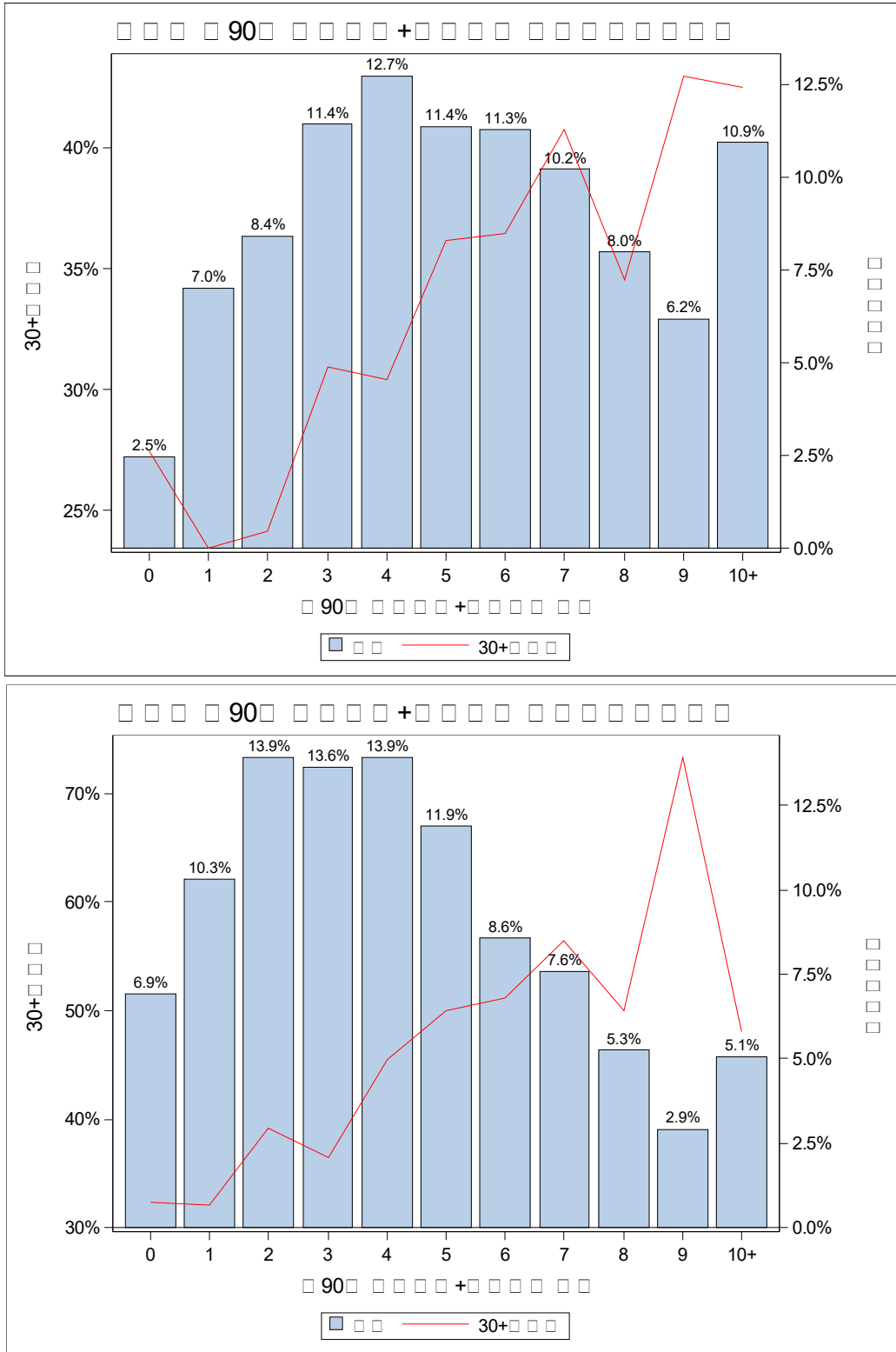


图 48 (续)：四个省份内，人行近 90 天 贷款审批+个人查询 次数分布和违约率

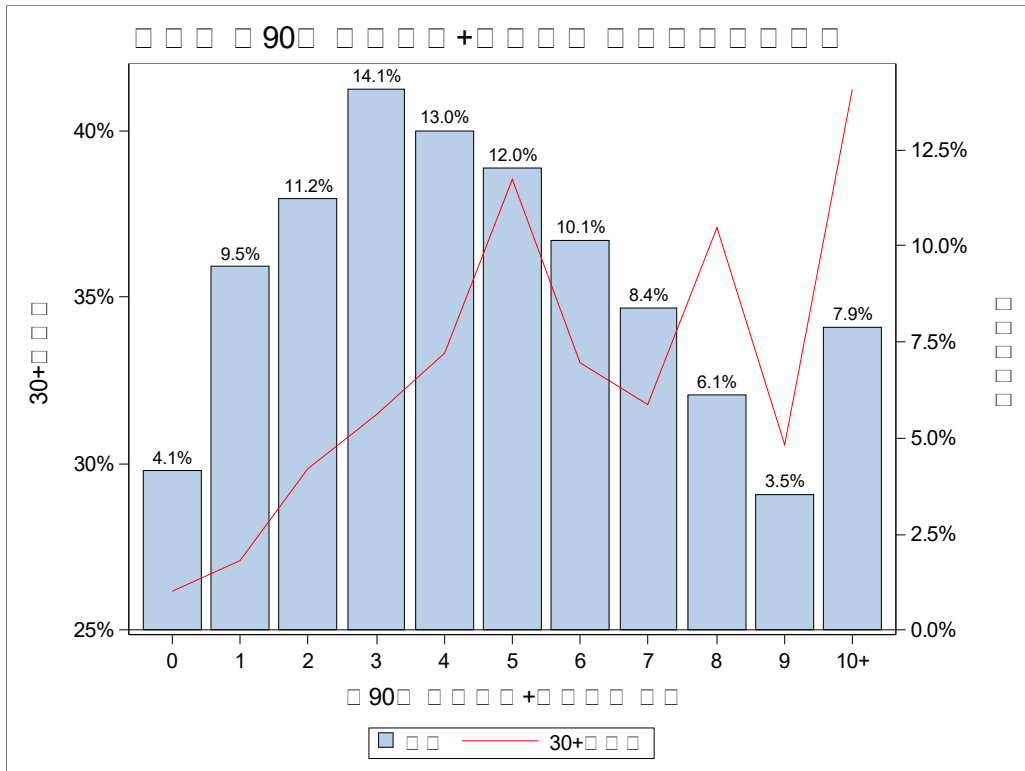
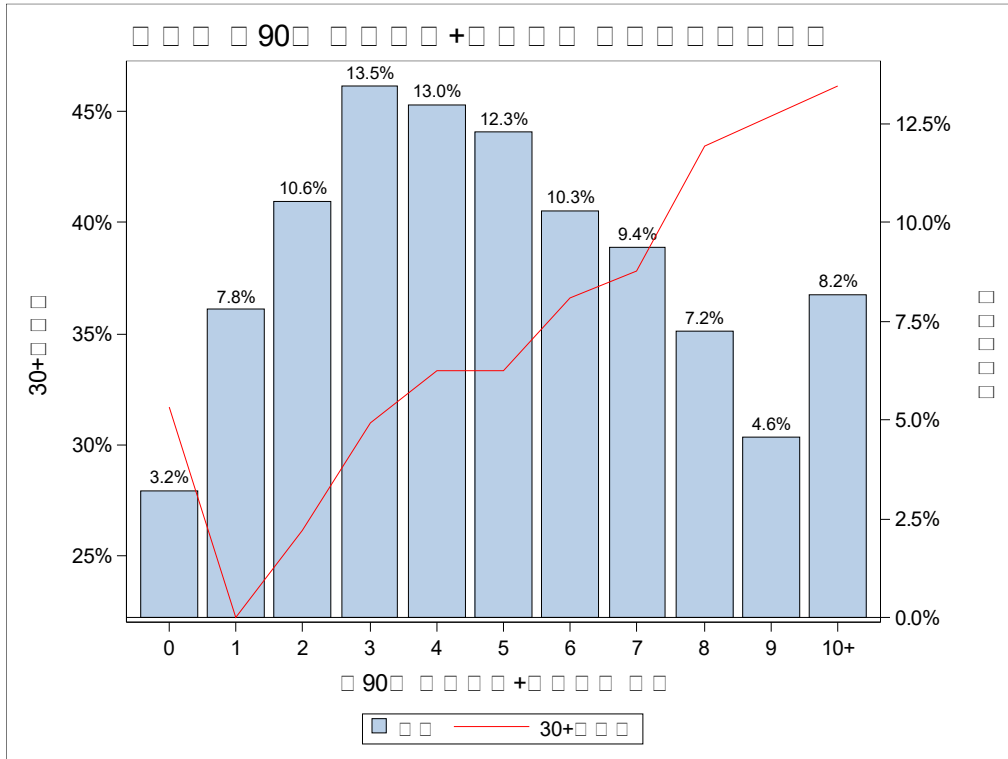


图 49: 四个省份内, 近 3 个月 同盾多头申请 平台数分布和违约率

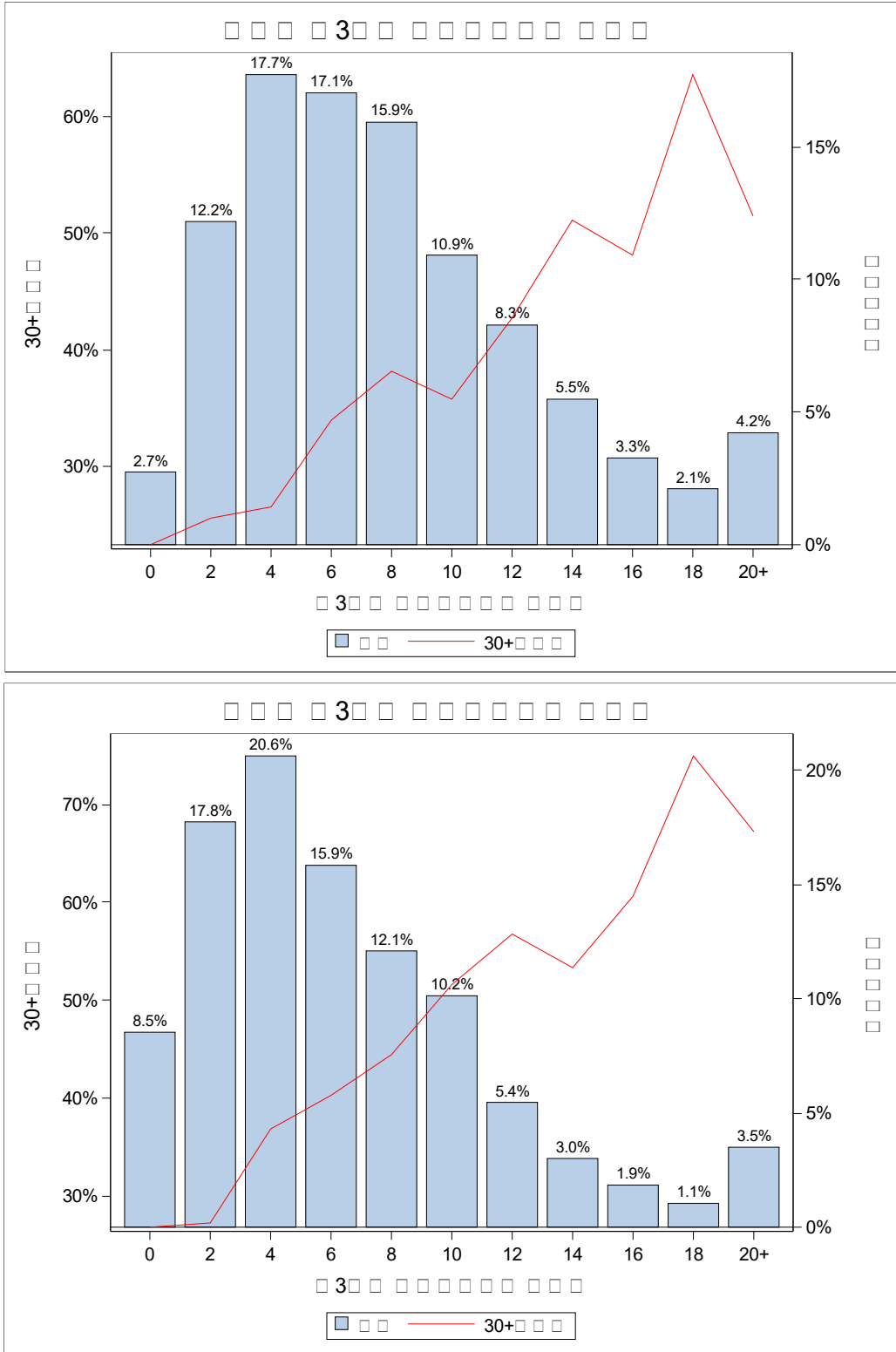
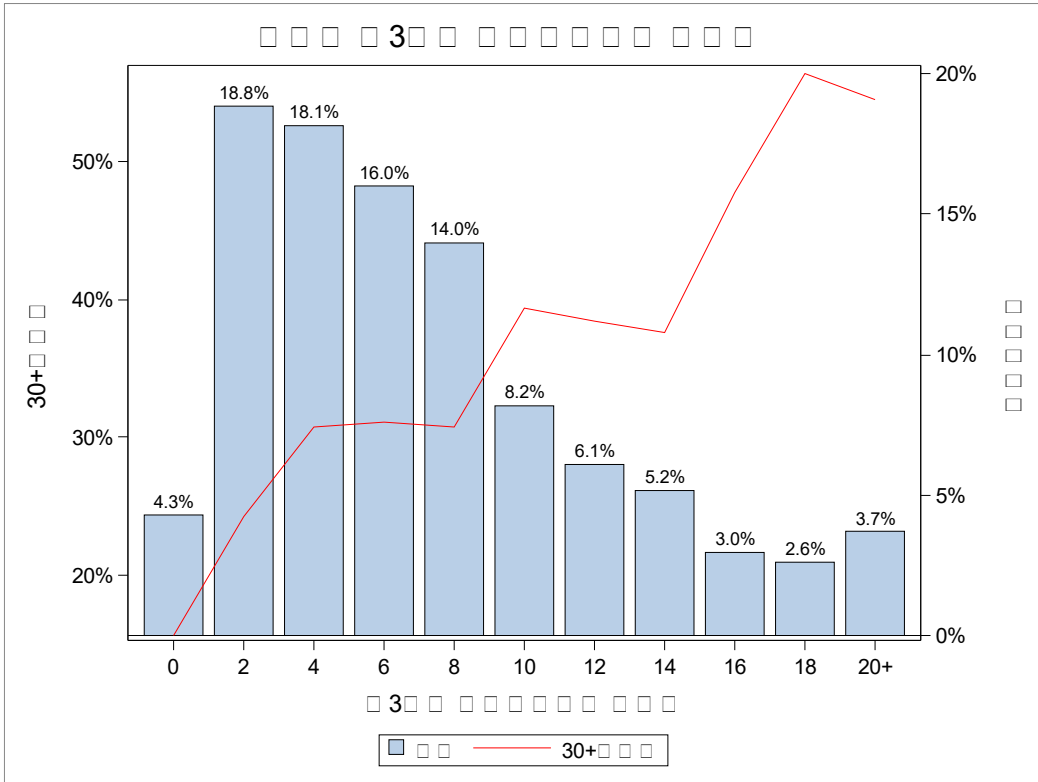
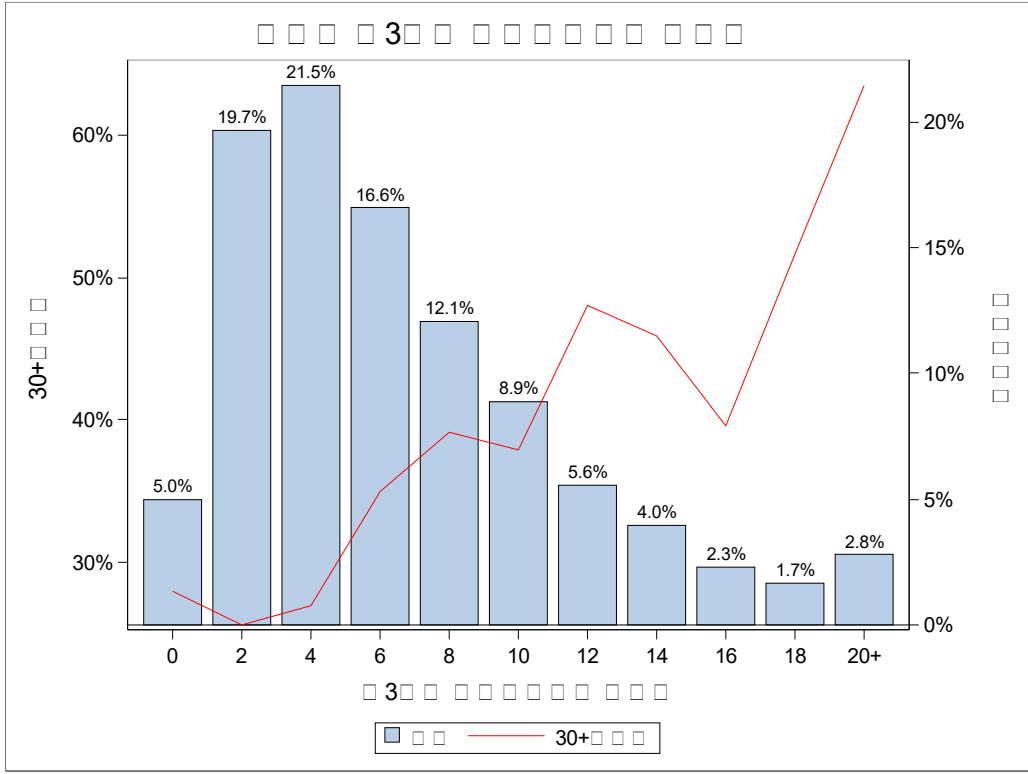
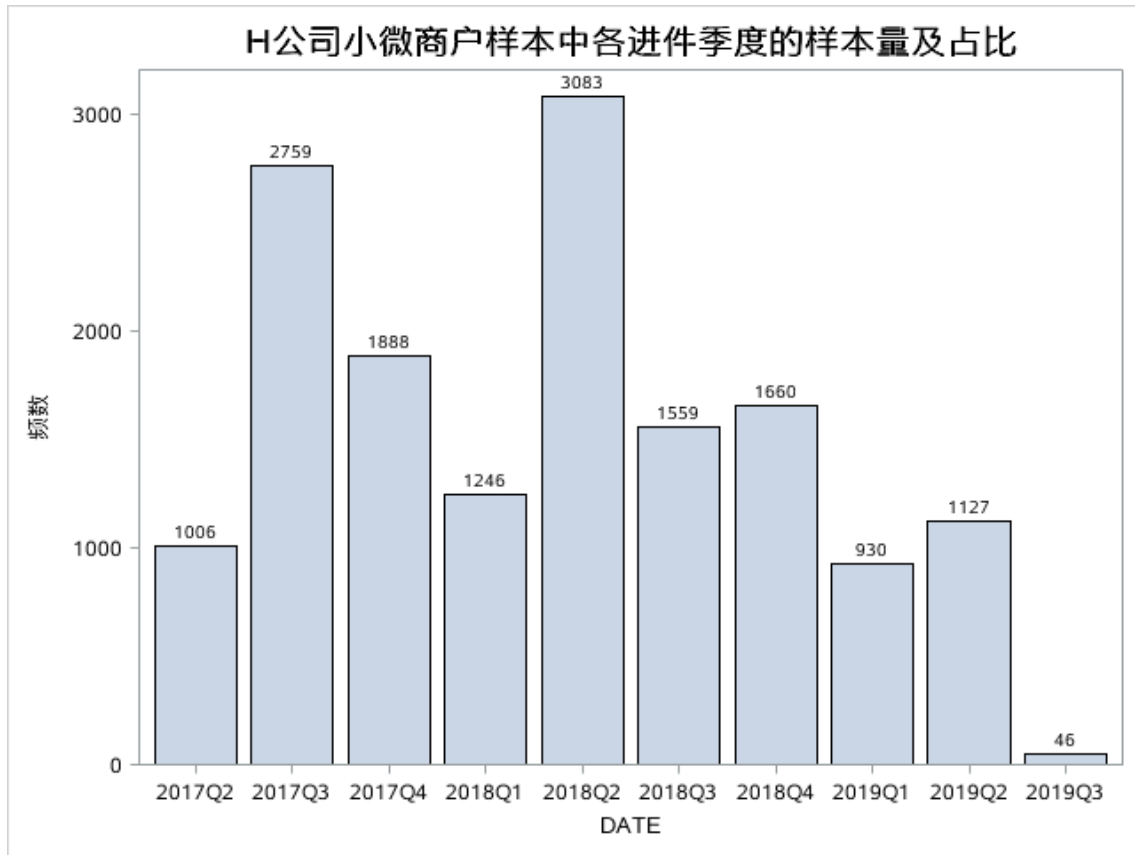


图 49（续）：四个省份内，近 3 个月 同盾多头申请 平台数分布和违约率



宏观经济、展业情况、风控政策等都会随时间变化，这些因素也都会对不同阶段放款客户的贷后表现产生影响。此次研究的小微商户样本中，按进件时间（按季度）的样本分布情况如下：

图 50: H 公司小微商户样本中各进件季度的样本量及占比



同样的，为尽可能排除时间因素对“融资行为频率与信用风险呈正相关”结论的影响，我们对每个放款年份（2017-2019 年）分别统计该年度内“人行近 90 天 贷款审批+个人查询次数”与违约率之间关系（如下图），可以发现 2017-2018 年的 30+ 违约率与查询次数基本成正相关走势，而 2019 年的 30+ 违约率曲线在首尾出现较大波动，中间段保持较好的正相关走势，或与 2019 年总体样本量较少有关。

图 51: 2017-2019 各个年度, 近 90 天 贷款审批+个人查询 次数分布和违约率

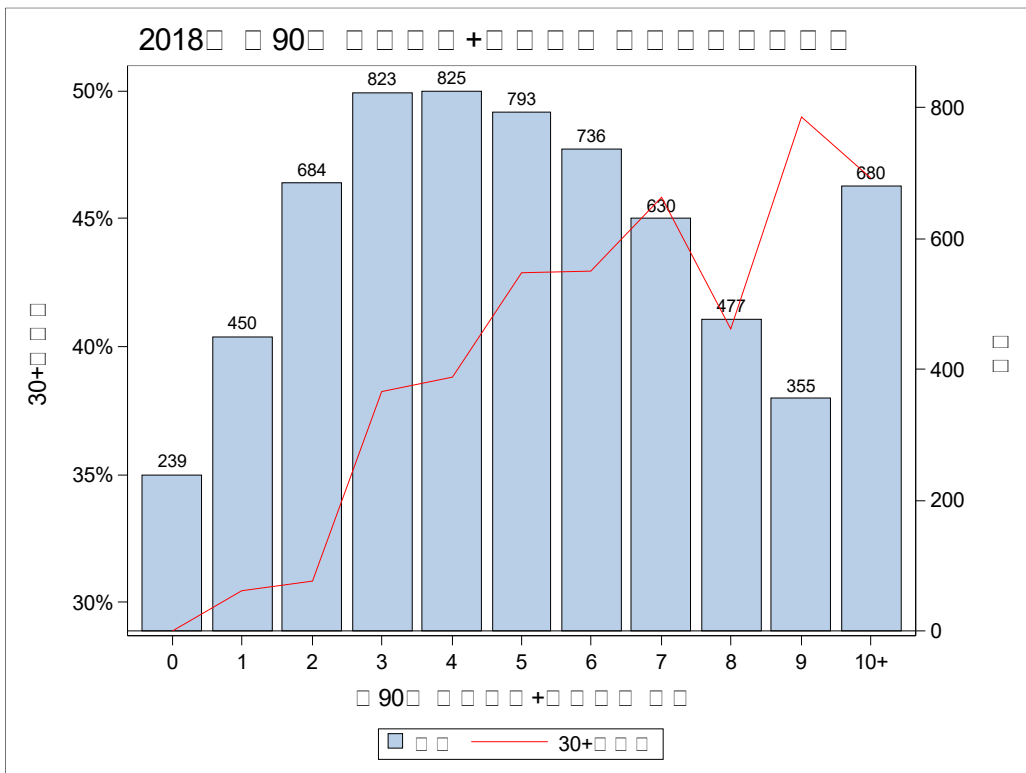
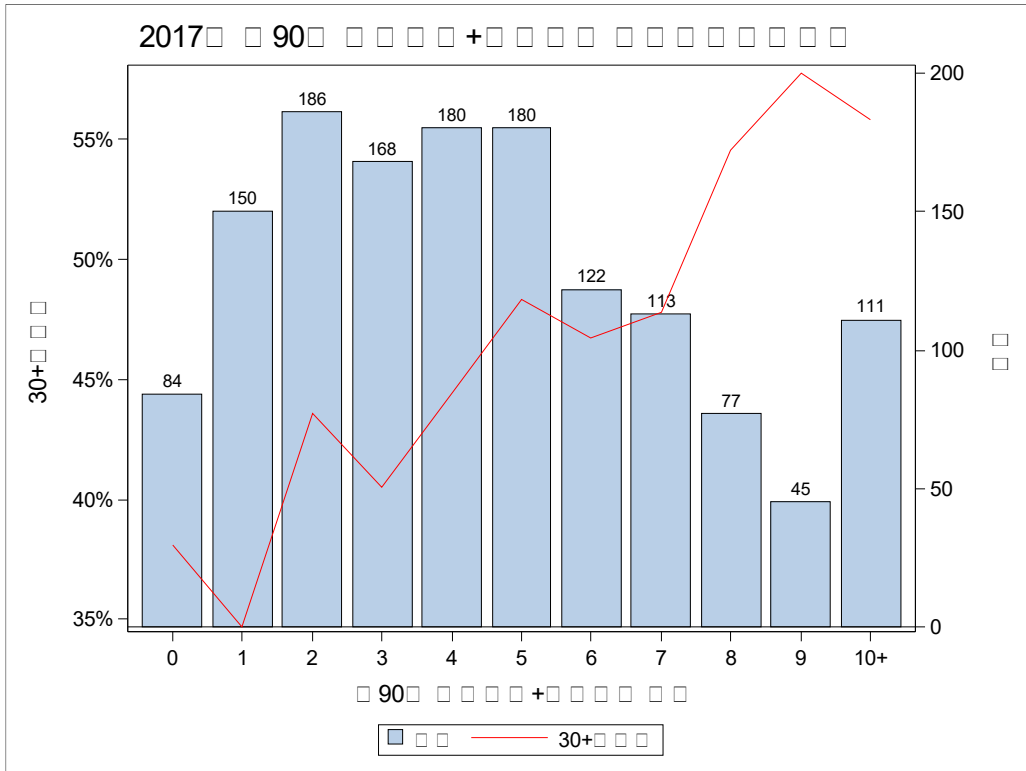
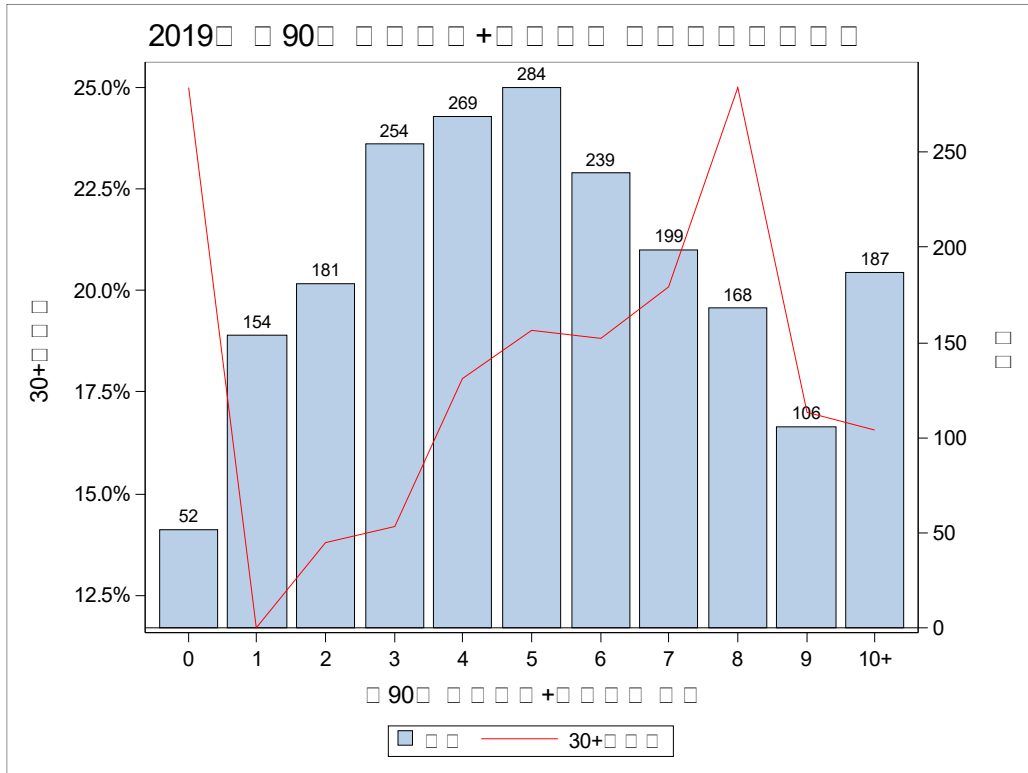




图 51（续）：2017-2019 各个年度，近 90 天 贷款审批+个人查询 次数分布和违约率



总体而言，排除地区和时间因素，小微商户“融资行为频率与信用风险呈正相关”这一结论仍然成立，无论是沿海或是内陆省份，2017年 2018年还是 2019年，即具有较好的稳定性。

### 5.5 融资渠道分布与信用风险的关系

融资渠道分布受小微商户能准入的机构类型影响，同时关系到小微商户的融资成本。通常来说，在非紧急和大量资金需求的情况下，我们通常都会选择利率低的融资渠道（比如银行），甚至在同一融资渠道选择利率更低的产品（抵押类>信用类，经营类>消费类）。但对于资金需求极为旺盛，但在银行等低息融资渠道较难获取足够授信的小微商户而言，选择融资渠道时常常会不得不考虑银行以外的机构，甚至如 P2P 网贷、民间借贷等高息渠道。最差的情况，是小微商户已经处于资金链断裂边缘，需要用高息但容易获取

的贷款，以偿还即将到期的低息贷款，借新还旧并不断提高自己的融资成本，直至最终无法偿还而陷于破产和信用不良。因此，接下来我们将基于“人行征信”和“同盾多头”两方面数据来探索融资渠道分布与信用风险的关系。

### 5.5.1 基于人行征信的融资渠道分布与信用风险的关系

首先我们将融资渠道分布用以下 12 个衍生变量来刻画，旨在将正规融资渠道（接入征信的，以“贷款审批”方式获取征信）和非正规融资渠道（未接入征信，以“个人查询”方式获取征信）区分开，或将银行、消费金融公司、小额贷款公司等不同融资渠道类型区分开。

图 52：基于人行征信的融资渠道分类

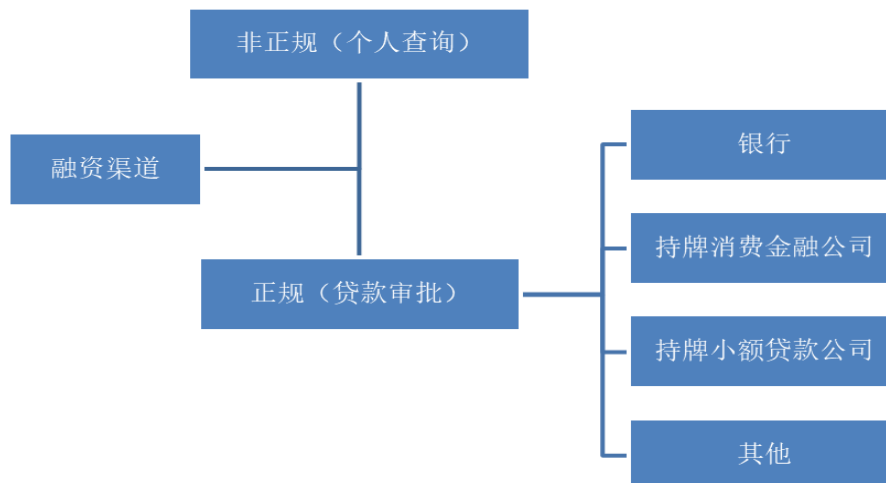


表 16：基于人行征信的融资渠道分布指标

Obs	变量名	变量含义
1	QUERY_L_R_6M	分子：近 180 天_贷款审批_查询次数 分母：近 180 天_贷款审批+个人查询_查询次数
2	QUERY_L_R_12M	分子：近 360 天_贷款审批_查询次数 分母：近 360 天_贷款审批+个人查询_查询次数
3	QUERY_L_R_24M	分子：近 2 年_贷款审批_查询次数 分母：近 2 年_贷款审批+个人查询_查询次数
4	QUERY_BANK_R_6M	分子：近 180 天_银行_贷款审批_查询次数 分母：近 180 天_贷款审批_查询次数
5	QUERY_BANK_R_12M	分子：近 360 天_银行_贷款审批_查询次数 分母：近 360 天_贷款审批_查询次数
6	QUERY_BANK_R_24M	分子：近 2 年_银行_贷款审批_查询次数 分母：近 2 年_贷款审批_查询次数
7	QUERY_XF_R_6M	分子：近 180 天_消费金融公司_贷款审批_查询次数 分母：近 180 天_贷款审批_查询次数
8	QUERY_XF_R_12M	分子：近 360 天_消费金融公司_贷款审批_查询次数 分母：近 360 天_贷款审批_查询次数
9	QUERY_XF_R_24M	分子：近 2 年_消费金融公司_贷款审批_查询次数 分母：近 2 年_贷款审批_查询次数
10	QUERY_XD_R_6M	分子：近 180 天_小额贷款公司_贷款审批_查询次数 分母：近 180 天_贷款审批_查询次数
11	QUERY_XD_R_12M	分子：近 360 天_小额贷款公司_贷款审批_查询次数 分母：近 360 天_贷款审批_查询次数

---

12	QUERY_XD_R_24M	分子：近 2 年_小额贷款公司_贷款审批_查询次数 分母：近 2 年_贷款审批_查询次数
----	----------------	---

---

从上表中可以看到，12 个衍生变量按时间切片分为近 180 天/360 天/2 年三种，机构类型分上关注 1) 接入人行征信的可进行“贷款审批”查询的正规金融机构，例如用 QUERY\_L\_R\_6M 来表示此类机构的近 180 天贷款申请查询次数，占整体近 180 天“贷款审批+个人查询次数”的比例，而非正规金融机构的近 180 天贷款审批查询次数占比，则为  $1 - \text{QUERY\_L\_R\_6M}$  可计算得到，因此不再额外衍生；2) 细分银行、消费金融公司、小额贷款公司的查询次数占比，例如 QUERY\_BANK\_R\_6M 代表近 180 天银行机构贷款审批查询次数，占近 180 天“贷款审批”的比例。

基于 10201 个获取有效人行征信报告的 H 公司小微商户样本，将这 12 个变量取值对应的违约率曲线画出，我们可以发现以下结论：

- 1) 能通过“贷款审批”原因查询的正规金融机构的查询次数占比越高，则违约率基本越低，尤其在近 2 年的时间切片下，正规金融机构的查询次数占比，与违约率呈显著的负相关性。这说明，客户的融资申请渠道中，持牌机构等正规渠道占比越高，其信用风险水平越低；反之，在非正规金融机构的查询占比越高，则信用风险水平越高。究其原因，一方面是客户的整体融资成本较低且资金需求通过正规融资渠道就能满足，尚未出现资金过度饥渴的情况，另一方面则是客户本身资质较好，是正规金融机构的目标客群，也更容易获得授信。

图 53: 近 180 天 正规金融机构 贷款审批 查询次数占比分布和违约率

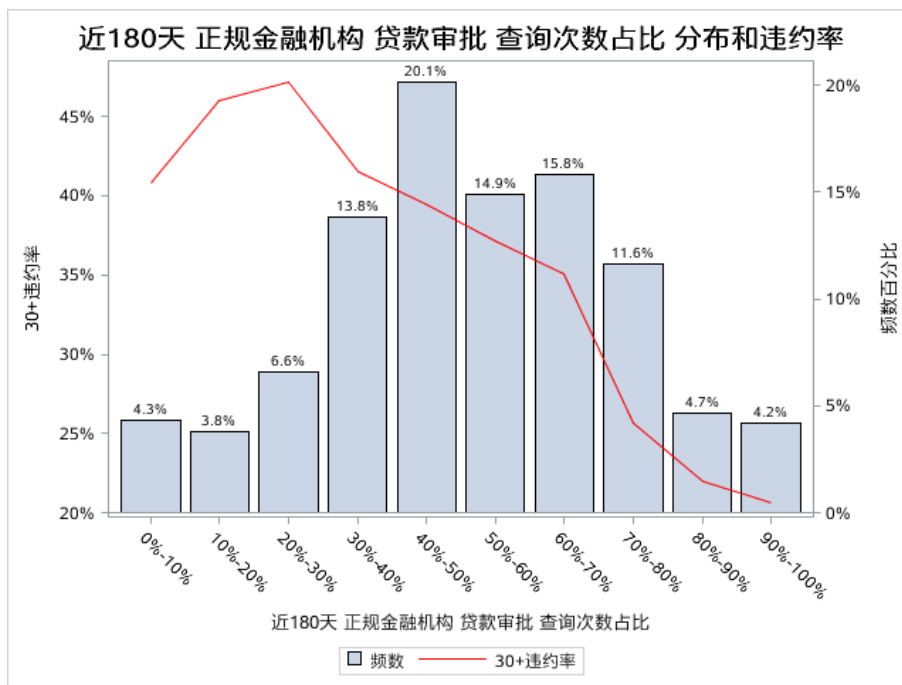


图 54: 近 360 天 正规金融机构 贷款审批 查询次数占比分布和违约率

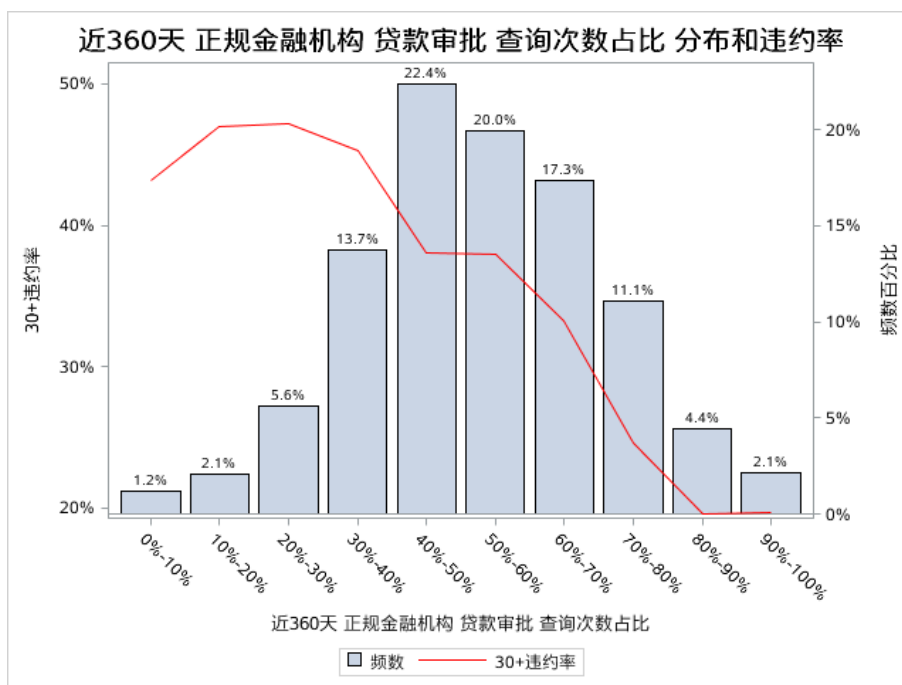
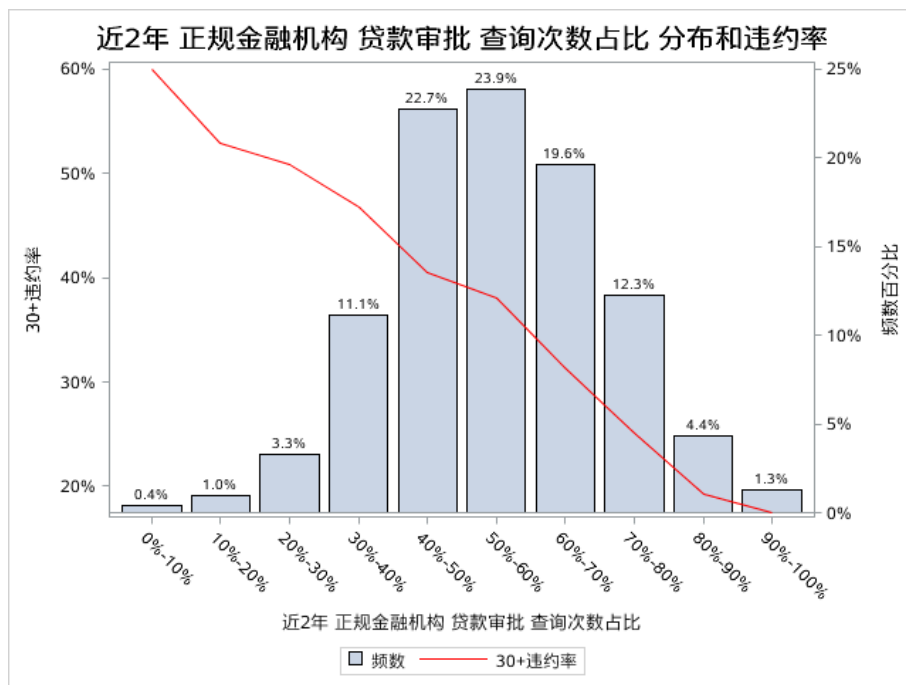


图 55: 近 2 年 正规金融机构 贷款审批 查询次数占比分布和违约率



2) 尽管违约率曲线有波动，但整体来看，银行的贷款审批查询次数占正规金融机构贷款审批查询次数的占比较高时，违约率较低，信用风险也较低。

图 56: 近 180 天 银行 贷款审批 查询次数占比分布和违约率

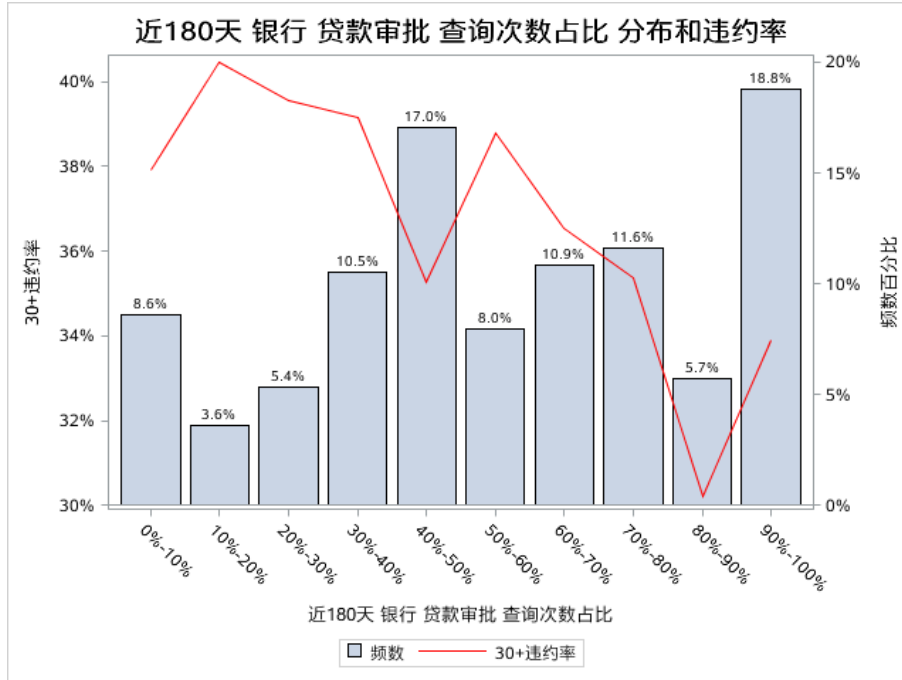


图 57: 近 360 天 银行 贷款审批 查询次数占比分布和违约率

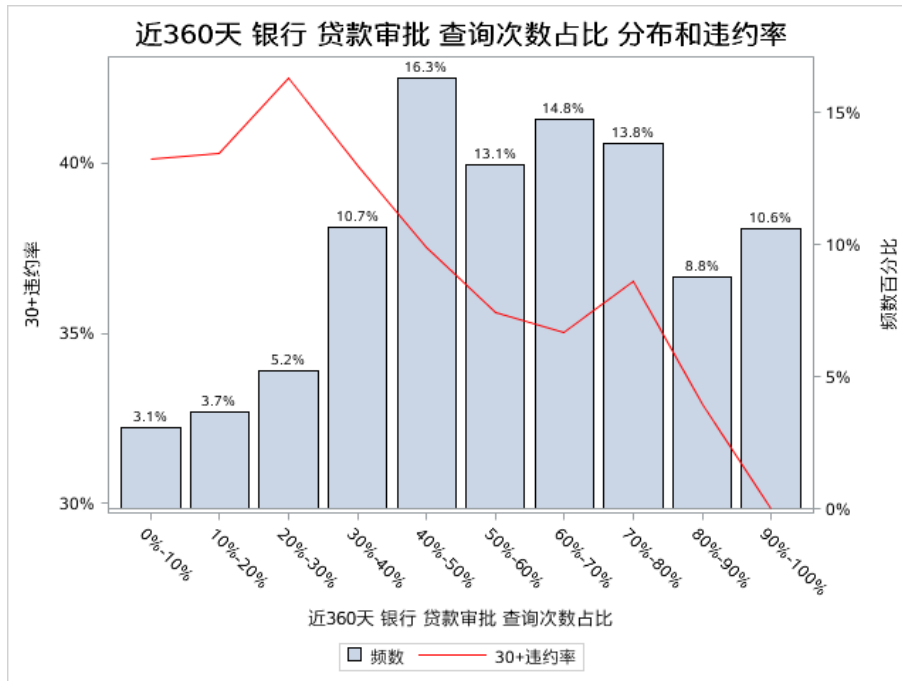
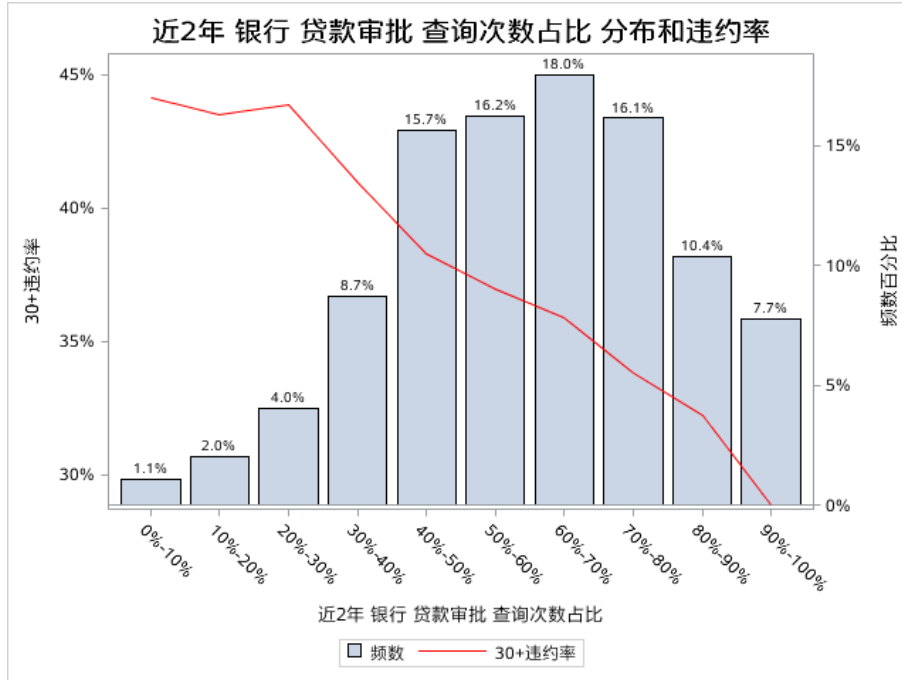


图 58: 近 2 年 银行 贷款审批 查询次数占比分布和违约率



- 3) 同是持牌正规金融机构，消费金融公司查询次数占比，与违约率大致呈正相关，说明相对于银行而言，消费金融公司是“下沉”渠道，当融资渠道从银行向消费金融公司转移时，客户的信用风险开始有所恶化。



图 59: 近 180 天 消费金融公司 贷款审批 查询次数占比分布和违约率

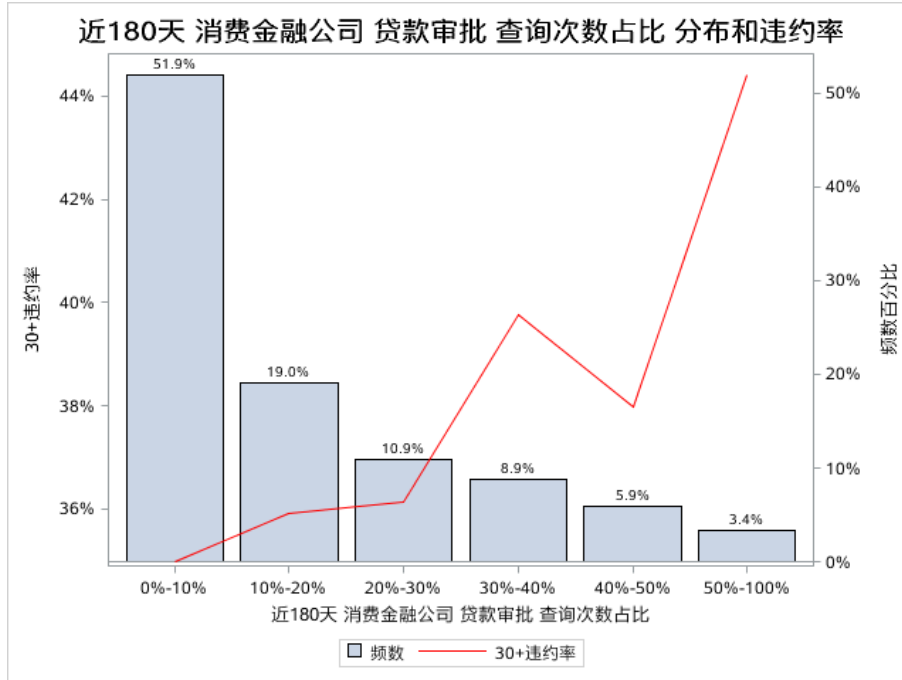


图 60: 近 360 天 消费金融公司 贷款审批 查询次数占比分布和违约率

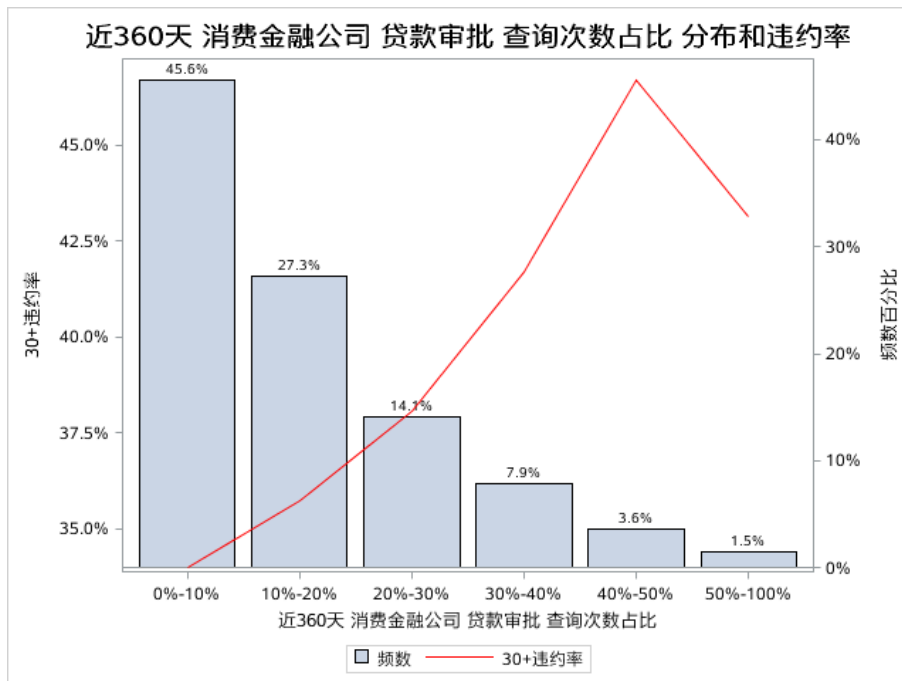
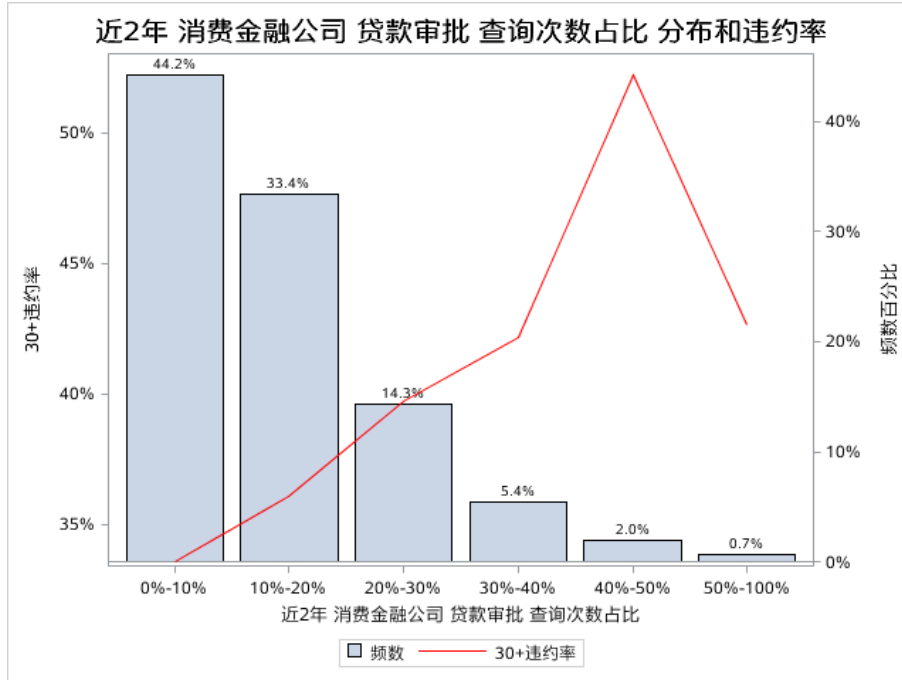


图 61: 近 2 年 消费金融公司 贷款审批 查询次数占比分布和违约率



- 4) 小额贷款公司查询次数占比,与违约率的关系呈倒U型曲线,占比在40%左右时,违约率曲线达到最高。

图 62: 近 180 天 小额贷款公司 贷款审批 查询次数占比分布和违约率

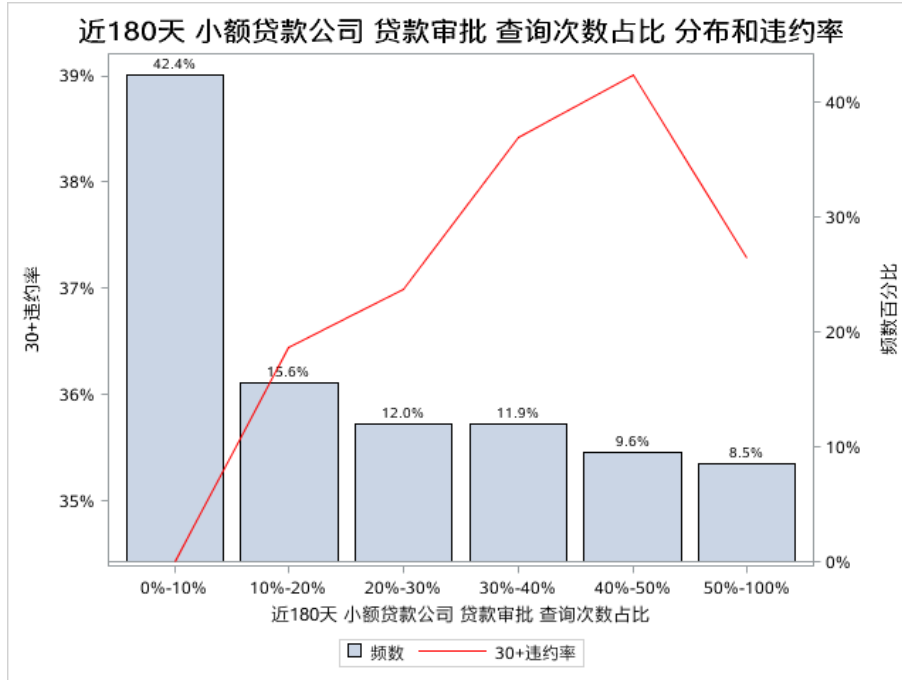


图 63: 近 360 天 小额贷款公司 贷款审批 查询次数占比分布和违约率

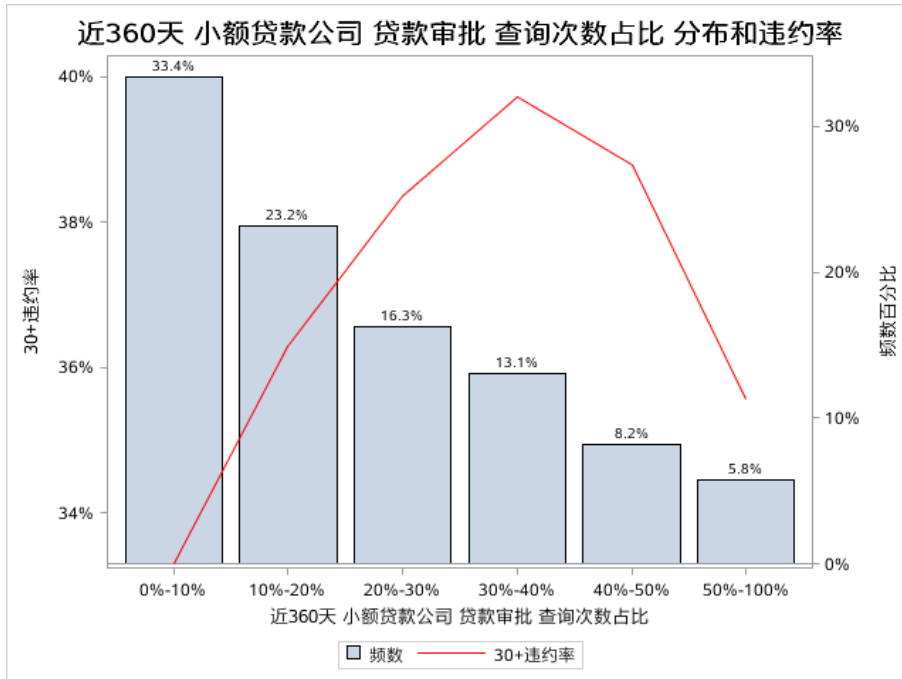
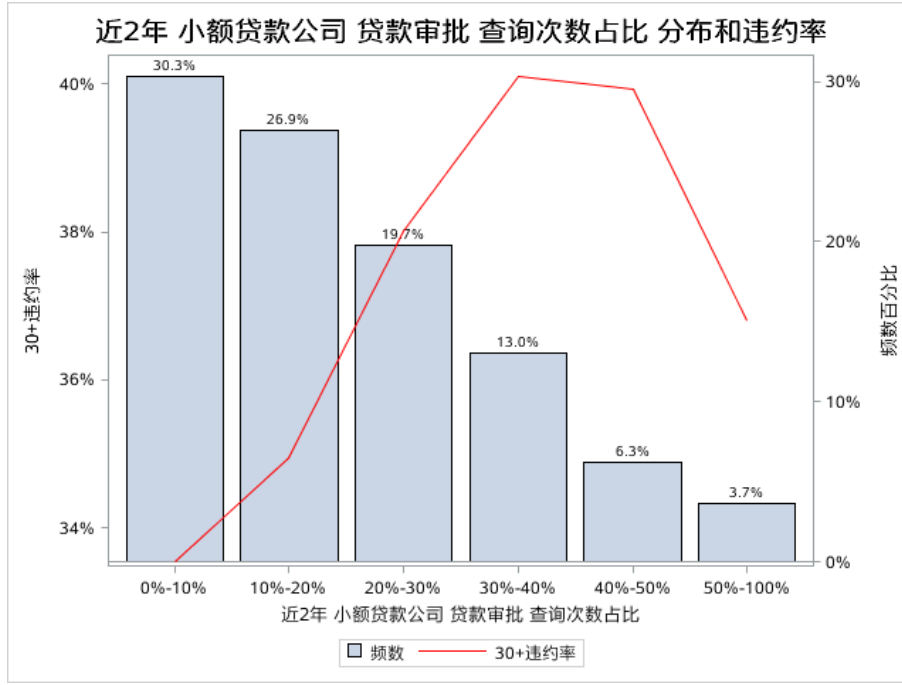


图 64：近 2 年 小额贷款公司 贷款审批 查询次数占比分布和违约率



### 5.5.2 基于同盾多头的融资渠道分布与信用风险的关系

同盾多头申请平台可细分为高达 28 个子行业，但其中占比较高（整体大于 5%）且与贷款融资相关的子行业主要为以下五个：1）P2P 网贷 2）小额贷款公司 3）一般消费分期平台 4）大型消费金融公司 5）银行消费金融公司

表 17：同盾多头融资渠道占比相关变量

Obs	变量名	变量含义
1	var_p2p_r	近 3 个月申请 P2P 网贷平台数/近 3 个月申请平台数
2	var_xedk_r	近 3 个月申请小额贷款公司数/近 3 个月申请平台数
3	var_ybxffq_r	近 3 个月申请一般消费分期平台数/近 3 个月申请平台数
4	var_dxxfjr_r	近 3 个月申请大型消费金融公司数/近 3 个月申请平台数

以上五个变量表示五类融资渠道在总申请平台中的占比，基于 15304 个获取有效人行征信报告的 H 公司小微商户样本，将这 5 个变量取值对应的违约率曲线画出，我们可以发现以下结论：

- 1) P2P 网贷平台占比 (var\_p2p\_r) <60% 时，占比越高违约率会越高，但如果超过 60%，则违约率曲线有较大波动。一个合理的猜测是，当 P2P 网贷平台占比高于 60% 时，整体申请平台数 (分母) 也很可能偏小，比如一共只申请了 3 家平台，其中 2 家为 P2P 网贷平台。根据 5.4 中融资行为频率和信用风险的关系，申请平台数少，违约率越低。因此融资行为频率和 P2P 网贷平台占比两者对违约率往相反方向影响，造成了波动剧烈的情况。未验证这一猜想，我们再次假定“近 3 个月申请平台数 > 5”，得到以下第二张表，可以看到违约率曲线的确不再剧烈波动，符合猜想的预期。

图 65: 近 3 个月 同盾多头申请 P2P 网贷 平台占比分布和违约率

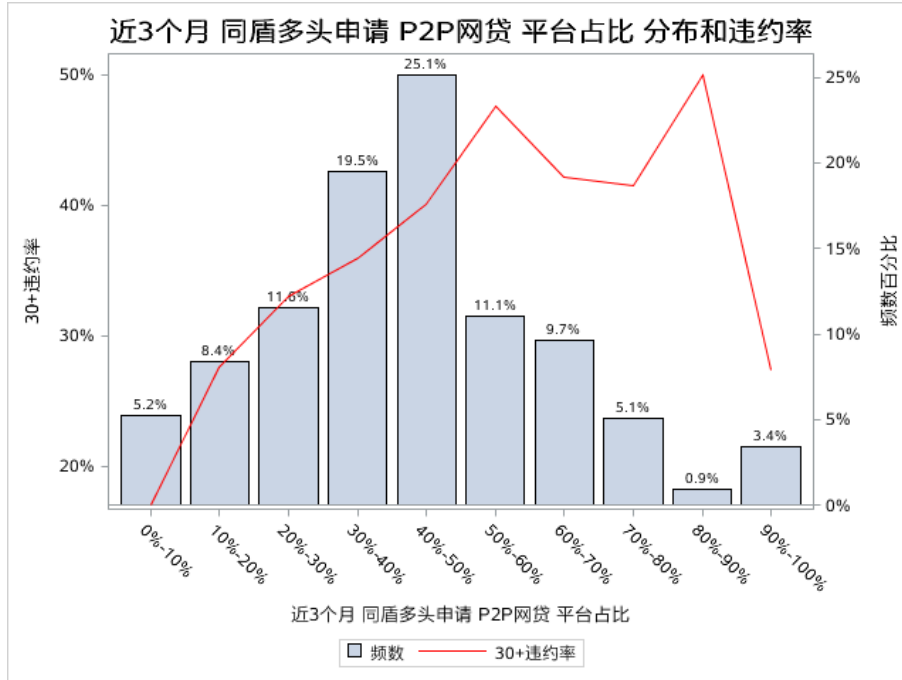
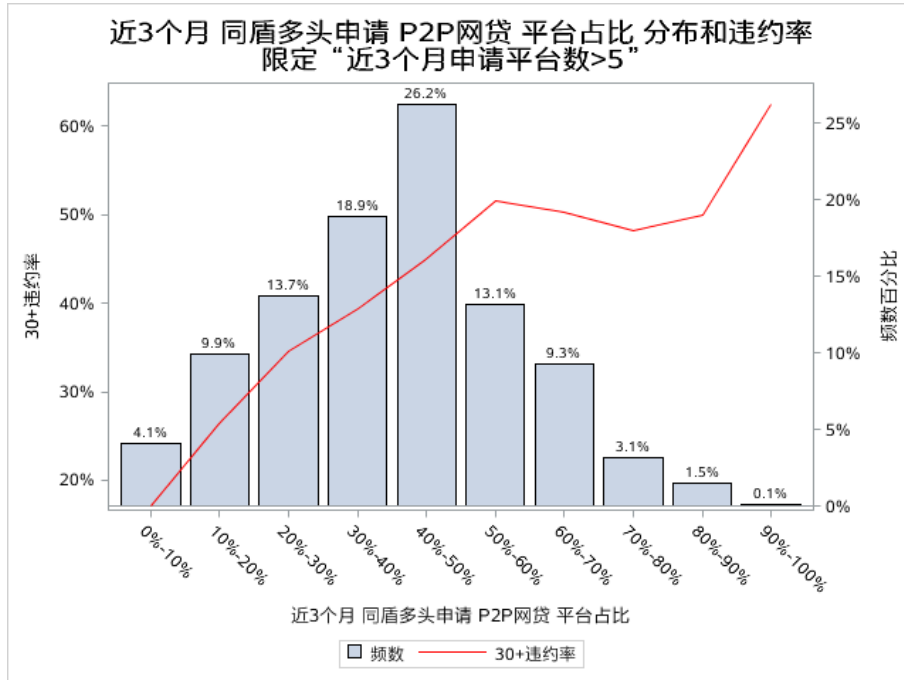
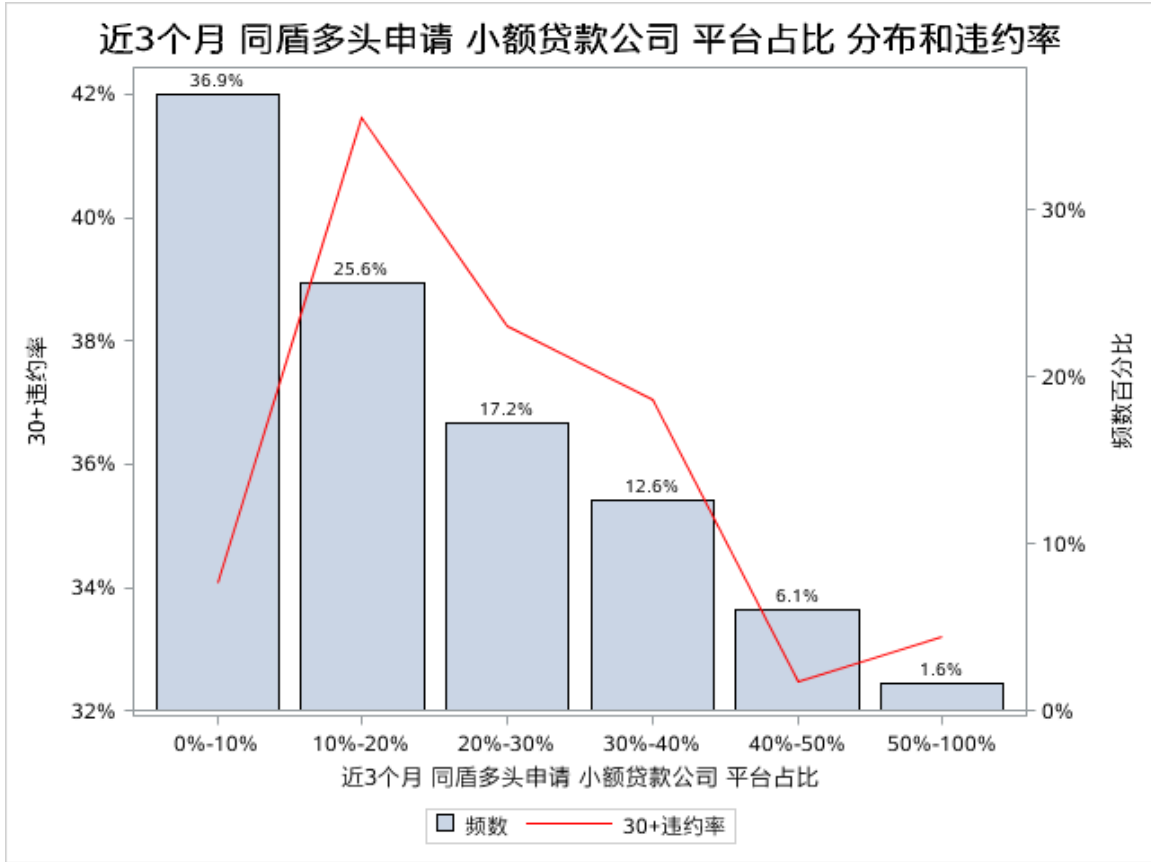


图 66: 近 3 个月 同盾多头申请 P2P 网贷 平台占比分布和违约率，限定“近 3 个月申请平台数>5”



2) 小额贷款公司占比 (var\_xedk\_r) >10%时, 占比越高违约率会越低。

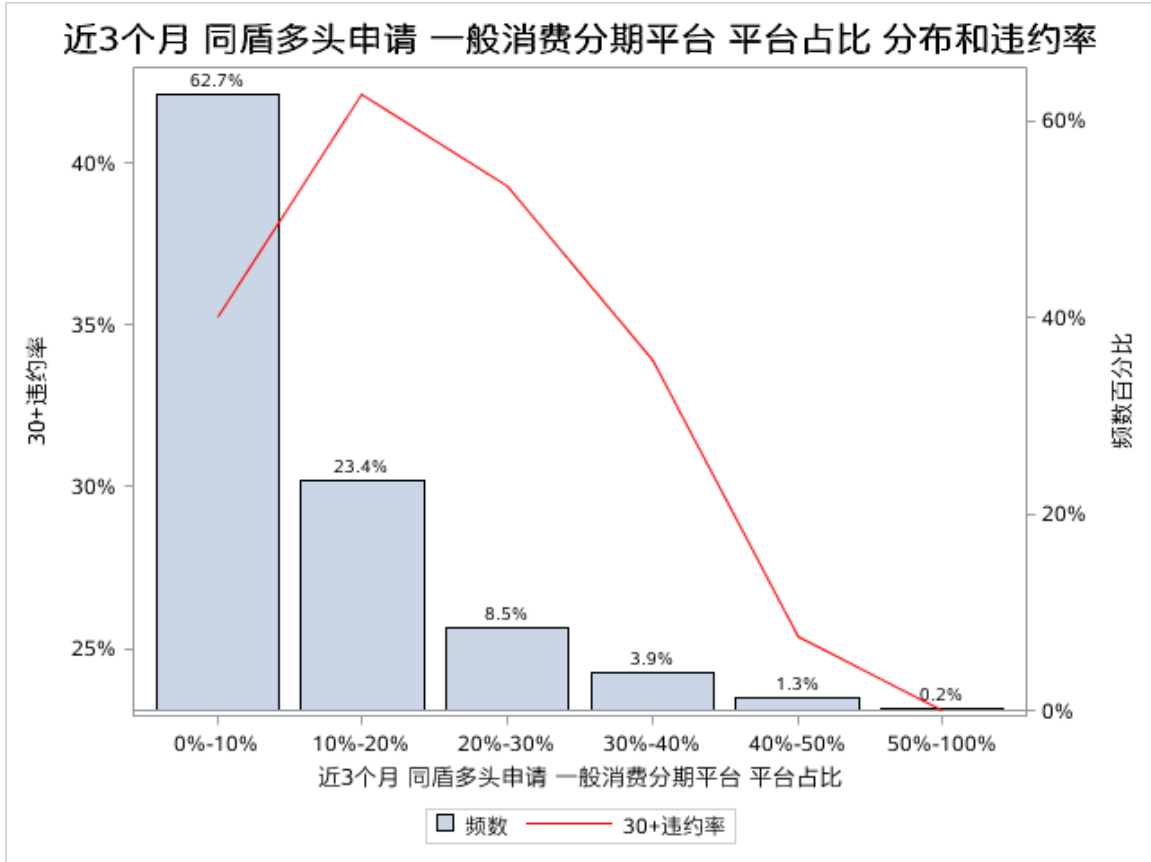
图 67: 近 3 个月 同盾多头申请 小额贷款公司 平台占比分布和违约率



3) 一般消费分期平台占比 (var\_ybxffq\_r) >10%时, 占比越高违约率会越低。

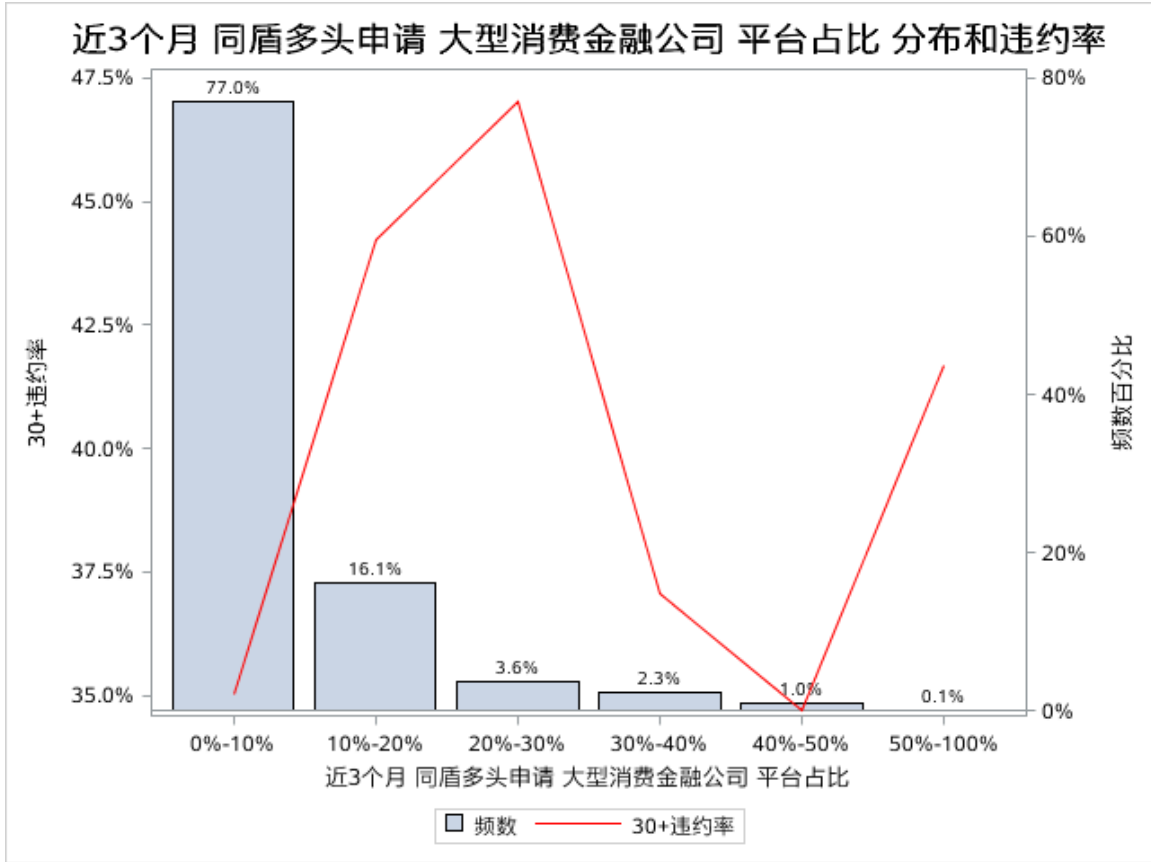
图 68: 近 3 个月 同盾多头申请 一般消费分期平台 平台占比分布和违约率





4) 大型消费金融公司占比 (var\_dxxfjr\_r) 与违约率没有显著的单调关系。

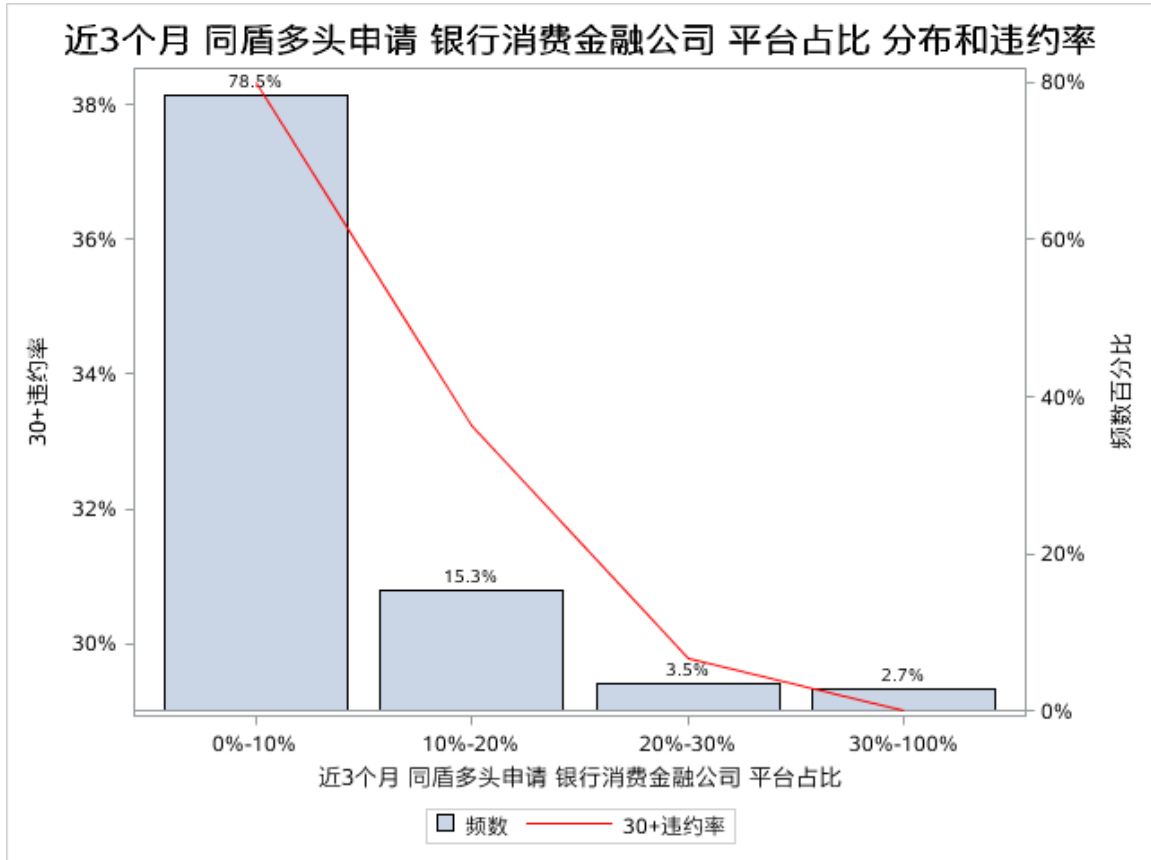
图 69: 近 3 个月 同盾多头申请 大型消费金融公司 平台占比分布和违约率



5) 银行消费金融公司占比 (var\_yhxfjr\_r) 越高, 违约率越低。有一次印证了银行在  
众多融资渠道中的“优质”标签, 即使是银行旗下的消费金融公司也比一般消费金

融公司获取更优质的客群。

图 70：近3个月 同盾多头申请 银行消费金融公司 平台占比分布和违约率



### 5.5.3 “融资渠道分布与信用风险关系”结论的稳定性分析

为验证“在正规金融机构（接入人行征信的）的融资占比越高，违约率越低”这一结论的稳定性，尽可能排除地区因素对结论的影响，我们对占比超10%的四个主要省份——广东省、浙江省、广西壮族自治区、四川省，分别统计同省份内“人行 近2年 正规金融机构 贷款 审批 查询次数占比分布”指标与违约率之间关系（如下图），同样发现两者负相关。尽管不同地区的融资渠道偏好有差异，例如四川省近2年正规金融机构贷款审批查询次数占比为80%以上客户比例为10%以上，而广东省对应的客户比例为6%左右，即四川省的小微商户

更偏好在正规金融机构申请贷款，但在各自省内仍然是正规金融机构的申请占比越高，风险相对越低。

图 71: 四个省份内，人行近 2 年正规金融机构 贷款审批 查询次数占比分布和违约率

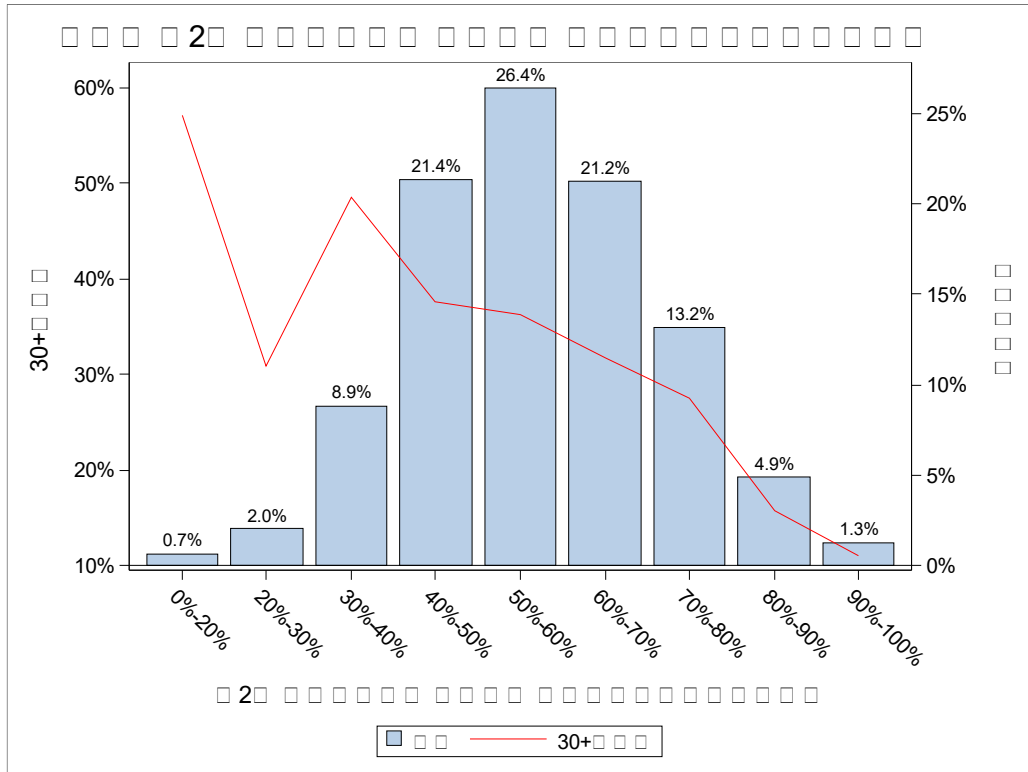


图 71(续): 四个省份内，人行近 2 年正规金融机构 贷款审批 查询次数占比分布和违约率

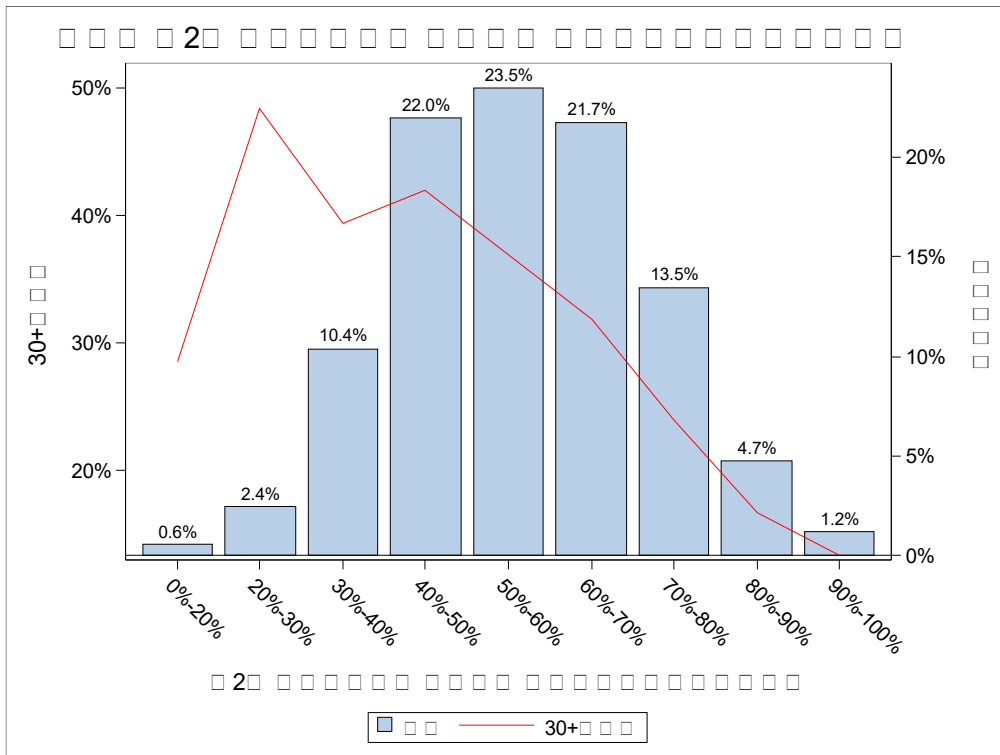
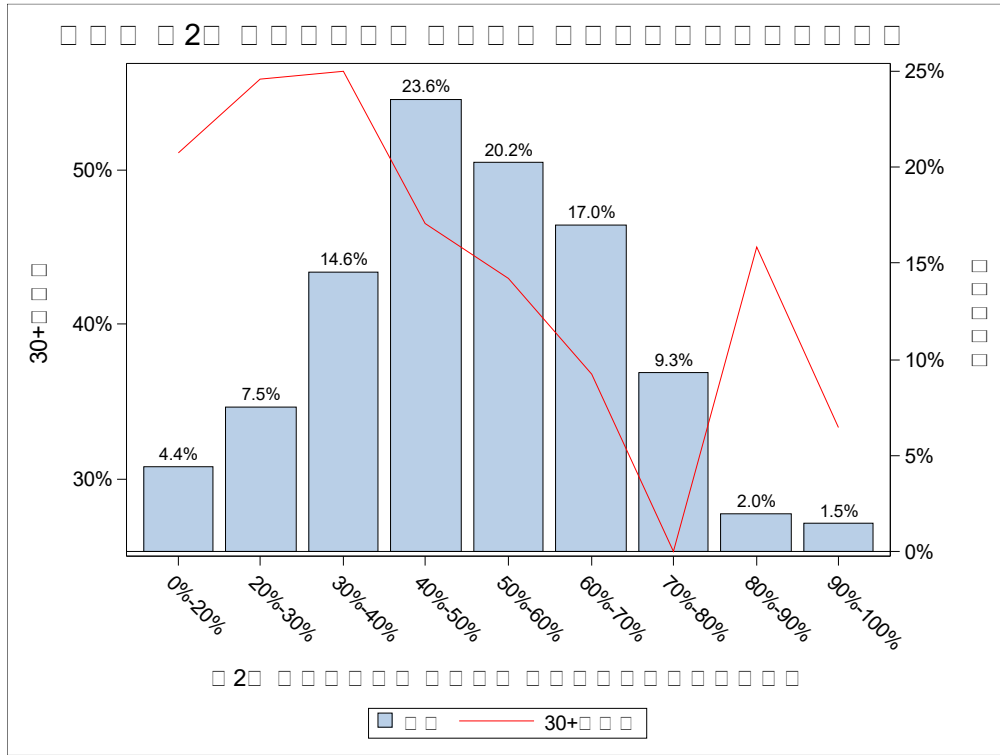
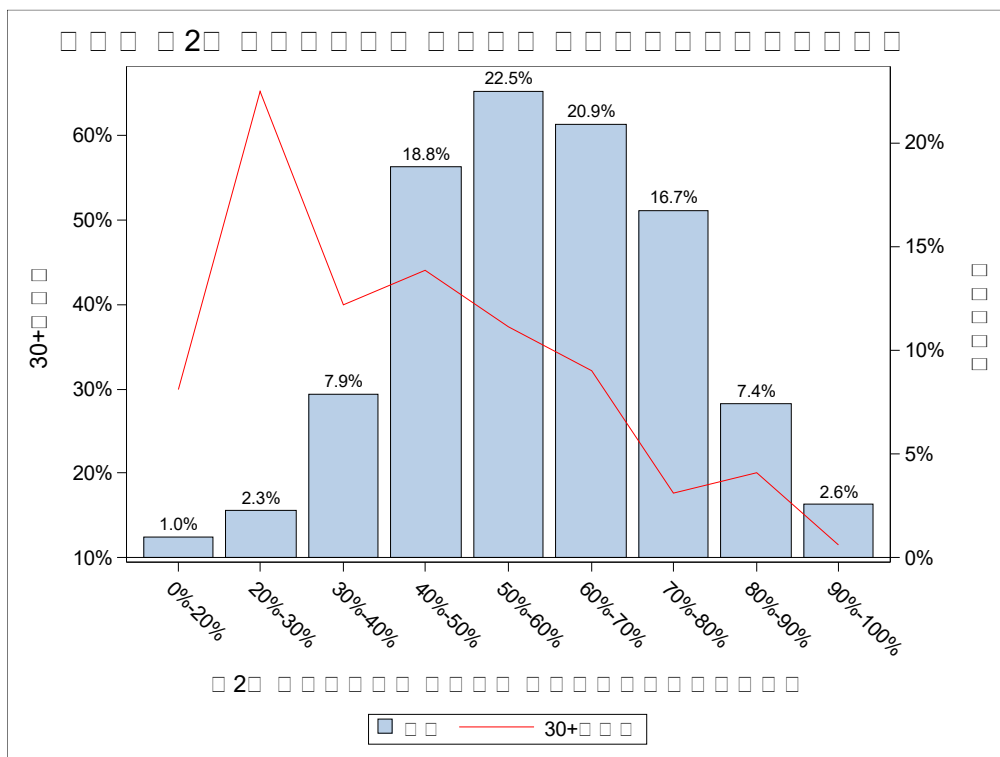


图 71(续): 四个省份内, 人行近 2 年正规金融机构 贷款审批 查询次数占比分布和违约率



同样的，为尽可能排除时间因素对“融资渠道分布与信用风险关系”结论的影响，我们对每个放款年份（2017-2019 年）分别统计该年度“人行近 2 年正规金融机构贷款审批查询次数占比分布”指标与违约率之间关系（如下图），可以发现尽管头尾波动较大（或与首尾样本较少有关），但每个年度内“在正规金融机构（接入人行征信的）的融资占比越高，违约率越低”这一结论仍然成立。

图 72：2017-2019 各个年度内，人行近 2 年正规金融机构贷款审批查询次数占比分布和违约率

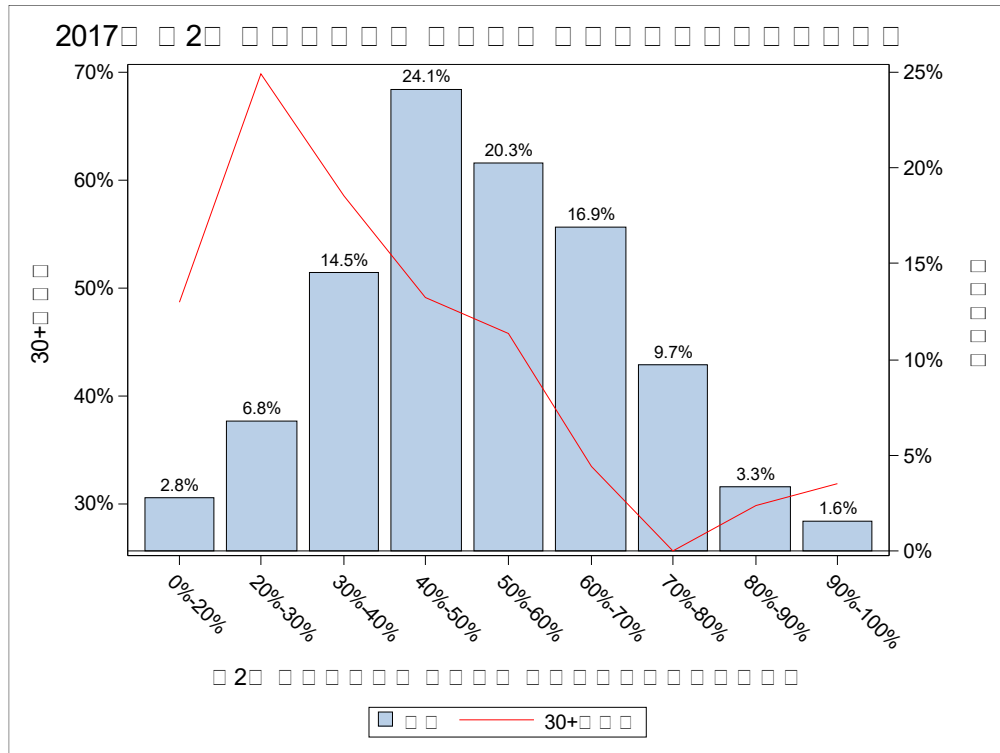
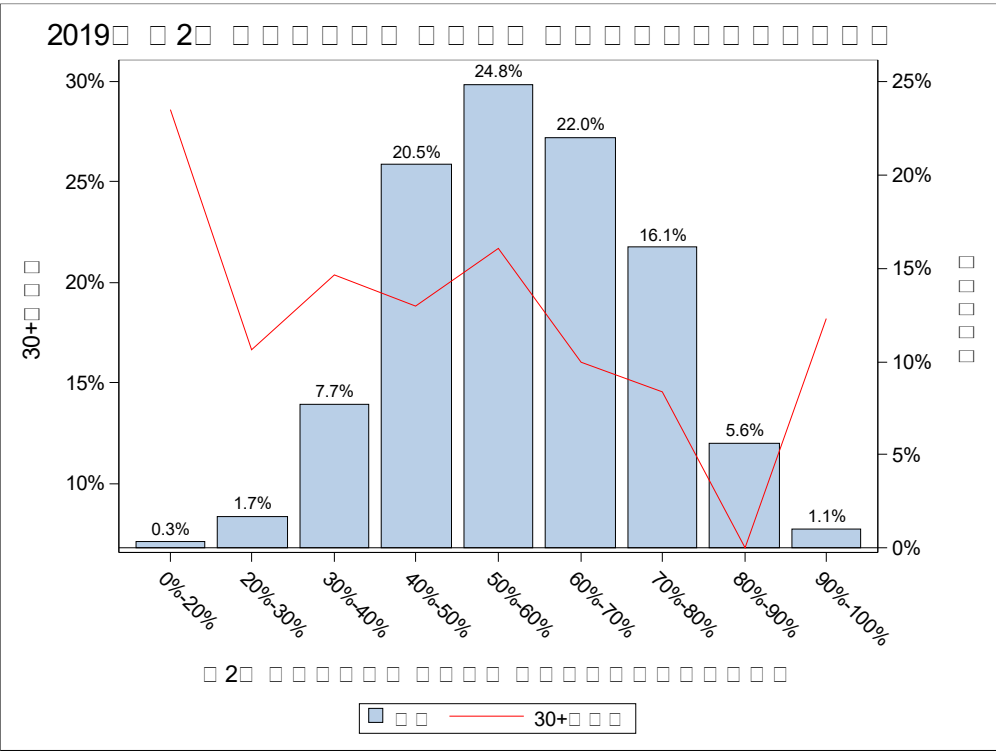
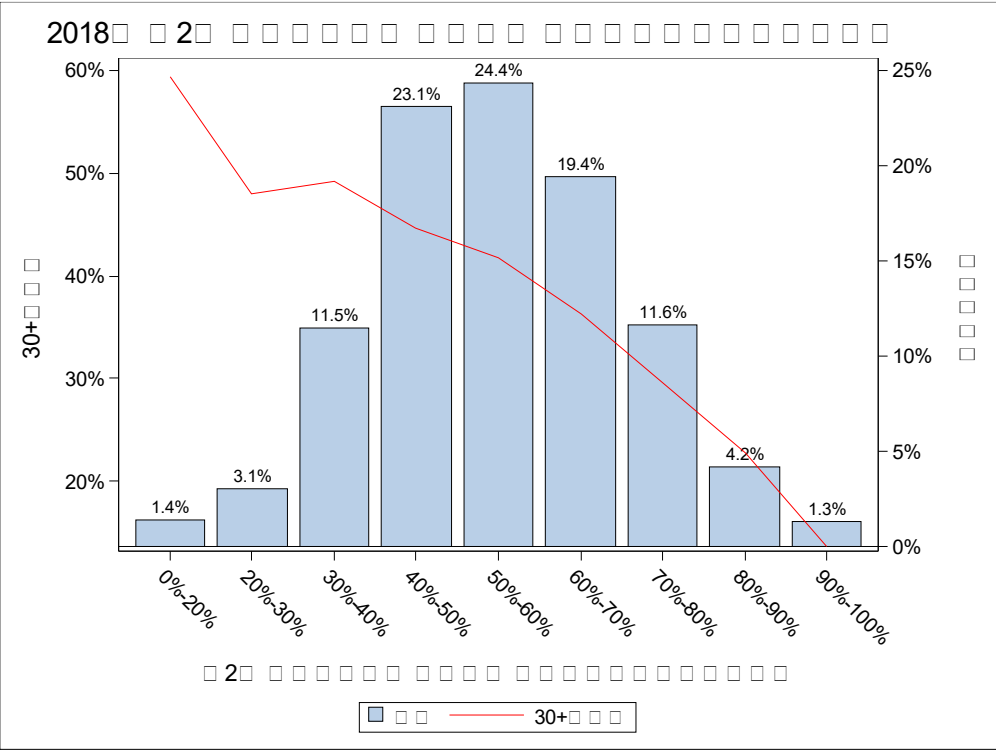


图 73(续): 2017-2019 各个年度内, 人行近 2 年正规金融机构贷款审批查询次数占比分布和违约率





总体而言，排除地区和时间因素，小微商户“在正规金融机构（接入人行征信的）的融资占比越高，违约率越低”这一结论仍然成立，即具有较好的稳定性。

## 5.6 融资行为对风险预测的单变量分析

我们已经基于人行征信和同盾多头报告衍生了众多变量，并初步分析了这些变量和违约率之间的单调关系，得到了以下主要结论：

- 1) 融资行为频率越高，违约率越高；
- 2) 融资渠道选择优先度：银行>消费金融公司≈小额贷款公司>P2P 平台
- 3) 在正规金融机构（接入人行征信的）的融资占比越高，违约率越低；
- 4) 向银行融资频率和占比越高，违约率越低；
- 5) 向 P2P 网贷平台融资频率和占比越高，违约率越高；

我们能否更进一步，定量地衡量这些变量对违约率预测的影响程度？答案是肯定的，以下我们将用“IV 值”和“决策树特征重要性”两个统计指标来对变量用于衡量违约率的效果进行排序，并且使用的样本为“同时获取有效人行征信和有效同盾多头报告”的 10081 个样本。

### 5.6.1 融资行为变量的 IV 值

我们在前文 3.3.4 中介绍了 IV 值（Information Value）的基本原理、计算方法、衡量尺度，知道了“单变量 IV 值越大，代表该变量对于 Y 的预测力越强影响越大”。这里的 Y 是取值为 0 或 1 的布尔值。回顾前文，Y=1 定义为发生过 30 天及以上逾期的客户，

Y=0 定义为已结清或还款满 6 期,且未发生逾期的客户。我们分析的 10081 个样本中,

Y=0 和 1 的占比分布如下:

表 18: 样本总体好坏客户分布

Y	频数	百分比	累积频数	累积百分比
0	6427	63.75%	6427	63.75%
1	3654	36.25%	10081	100.00%

将单变量分成最多 5 组, 整个变量的 IV 值是各组 IV 值的总和:  $IV = \sum_i^n IV_i$ , 而单个组的 IV 值计算如下:

$$IV_i = \left( \frac{g_i}{g_T} - \frac{b_i}{b_T} \right) \times \ln \left( \frac{g_i/g_T}{b_i/b_T} \right)$$

将  $IV \geq 0.02$  作为单变量具有预测力的基本门槛, 筛选得到 27 个变量 (如下表), 我们认为对预测 Y=0 或 1 的概率是有作用的。IV 值越高, 预测效果越强,  $IV \geq 0.1$  表示该变量有中等预测力, 有 10 个变量达到中等预测力。最后一列是该变量与违约率的单调关系, 正相关表示该变量值越大违约率越高, 负相关表示该变量值越大违约率越低。

表 19: IV 值较高的单变量及含义

Obs	变量名	中文含义	IV	正/负相关
1	var_rzzl_r	同盾_近 3 个月_融资租赁_申请平台占比	0.184	负
2	var_p2p	同盾_近 3 个月_P2P 网贷_申请平台数	0.181	正
3	var_rzzl	同盾_近 3 个月_融资租赁_申请平台数	0.158	负
4	var_p2p_r	同盾_近 3 个月_P2P 网贷_申请平台占比	0.145	正
5	QUERY_OWN_CNT_6M	人行_近 180 天_个人查询_查询次数	0.137	正
6	QUERY_L_R_24M	人行_近 2 年_贷款审批_查询次数占比	0.130	负
7	QUERY_OWN_CNT_12M	人行_近 360 天_个人查询_查询次数	0.125	正
8	QUERY_L_R_12M	人行_近 360 天_贷款审批_查询次数占比	0.112	负
9	QUERY_OWN_CNT_3M	人行_近 90 天_个人查询_查询次数	0.108	正
10	QUERY_L_R_6M	人行_近 180 天_贷款审批_查询次数占比	0.105	负
11	QUERY_OWN_CNT_24M	人行_近 2 年_个人查询_查询次数	0.090	正
12	QUERY_L_R_3M	人行_近 90 天_贷款审批_查询次数占比	0.063	负
13	TD_3M_CNT	同盾_近 3 个月_申请平台数	0.059	正
14	var_dxxfjr	同盾_近 3 个月_大型消费金融公司_申请平台数	0.045	正
15	QUERY_CNT_6M	人行_近 180 天_贷款审批和个人查询_查询次数	0.042	正

16	var_ybxffq	同盾_近 3 个月_一般消费分期平台_申请平台数	0.040	正
17	var_dxxfjr_r	同盾_近 3 个月_大型消费金融公司_申请平台占比	0.039	正
18	QUERY_CNT_12M	人行_近 360 天_贷款审批和个人查询_查询次数	0.037	正
19	QUERY_CNT_3M	人行_近 90 天_贷款审批和个人查询_查询次数	0.036	正
20	var_yhxwdk_r	同盾_近 3 个月_银行小微贷款_申请平台占比	0.032	负
21	var_ybxffq_r	同盾_近 3 个月_一般消费分期平台_申请平台占比	0.031	正
22	QUERY_BANK_R_24M	人行_近 2 年_银行_贷款审批_查询次数占比	0.027	负
23	var_xedk	同盾_近 3 个月_小额贷款公司_申请平台数	0.024	正
24	var_yhxfjr_r	同盾_近 3 个月_银行消费金融公司_申请平台占比	0.023	负
25	QUERY_XF_R_24M	人行_近 2 年_消费金融公司_贷款审批_查询次数占比	0.022	正
26	QUERY_XF_R_12M	人行_近 360 天_消费金融公司_贷款审批_查询次数占比	0.022	正
27	QUERY_CNT_24M	人行_近 2 年_贷款审批和个人查询_查询次数	0.022	正

从以上 27 个有预测力变量的含义以及与违约率的正/负相关性来看，我们可以得到以下关于 H 公司小微商户样本的结论：

- 1) 小微商户通过融资租赁方式申请贷款的行为，对违约率有中预测力，且负相关；
- 2) 小微商户在 P2P 网贷平台申请贷款的行为，对预测违约率有显著效果，且正相关；
- 3) 小微商户在未接入征信的非正规金融机构（人行征信查询原因为“个人查询”）申请贷款的频率和占比，对违约率有中预测力，且正相关；
- 4) 小微商户在接入征信的正规金融机构（人行征信查询原因为“贷款审批”）申请贷款的频率和占比，对违约率有中预测力，且负相关；
- 5) 小微商户近期申请贷款的频率，对违约率有弱预测力，且正相关；
- 6) 小微商户在非银行系金融机构（大型消费金融公司/一般消费分期平台/小额贷款公司）申请贷款的频率和占比，对违约率有弱预测效力，且正相关；
- 7) 小微商户在银行系金融机构（银行/银行消费金融公司/银行小微贷款）申请贷款的频率和占比，对违约率有弱预测力，且负相关；

综上，我们用于刻画 H 公司小微商户融资行为而衍生的指标，与信用风险之间有统计意义上的关联关系和预测力，可以考虑加入构造信用风险模型。

### 5.6.2 融资行为变量的决策树特征重要性

决策树特征重要性排序，是另一种衡量 X 单变量对 Y 预测力的统计方法。仍然使用 10081 个 H 公司既有有效人行征信报告又有有效同盾多头报告的小微商户样本，按 7:3 的比例划分训练集和验证集，并使用 SAS 统计软件的决策树模型功能计算得到单变量的特征重要性指标，结果如下：

图 74: 决策树特征重要性结果

变量重要性

变量名称	标签	拆分规则数	替代规则数	重要性	验证重要性	验证训练重要性的比率
var_rzzl_r		2	0	1.0000	1.0000	1.0000
var_rzzl		0	2	0.9890	0.9882	0.9993
var_p2p_r		1	2	0.9714	0.9288	0.9562
QUERY_OWN_CNT_12M		1	0	0.8791	0.4271	0.4858
QUERY_OWN_CNT_24M		0	1	0.8207	0.3987	0.4858
QUERY_CNT_12M		0	1	0.8012	0.3893	0.4858
var_p2p		3	1	0.5932	0.5752	0.9697
QUERY_L_CNT_24M		0	3	0.5027	0.3522	0.7006
TD_3M_CNT		0	3	0.4778	0.5162	1.0803
var_ybxffq		0	3	0.4455	0.4821	1.0821
QUERY_L_R_24M		2	0	0.4360	0.4350	0.9979
QUERY_L_R_12M		0	2	0.4031	0.4031	0.9999
QUERY_BANK_CNT_12M		1	1	0.3250	0.2263	0.6964
QUERY_BANK_CNT_24M		1	1	0.3248	0.2299	0.7078
QUERY_L_R_6M		0	1	0.2008	0.2792	1.3904
QUERY_BANK_CNT_6M		0	1	0.1869	0.1163	0.6222
var_dxxfjr_r		1	0	0.1864	0.1770	0.9498
var_dxxfjr		0	1	0.1844	0.1751	0.9498
QUERY_XF_R_24M		0	1	0.1576	0.1497	0.9498

可以发现，通过决策树变量选择得到的特征重要性结果，与 IV 值的分析结果非常接近，效果较好的变量同样是以下几类变量：

- 1) 融资租赁类平台的申请频率和占比；
- 2) P2P 网贷类平台的申请频率和占比；
- 3) 正规金融机构和非正规金融机构申请频率和占比；
- 4) 整体申请融资频率；
- 5) 银行/消费金融公司的申请频率和占比；

综上，通过决策树特征重要性指标也证明，融资行为相关的指标变量对于预测违约率和信用风险水平是有显著效果的。

### 5.7 基于融资行为的信用风险评分卡模型

经过 IV 值和决策树特征重要性的筛选，我们已初步得到对于预测违约变量 Y（是否发生 30 天及以上逾期）有效的融资行为相关 X 变量集。接下来，我们尝试用这些变量构

建一张基于融资行为变量的信用风险子评分卡模型，并评估验证这张子评分卡模型的违约率预测效果、风险排序效果和稳定性。

用于建立信用风险评分卡模型的样本，仍然是 10081 个 H 公司既有有效人行征信报告又有有效同盾多头报告的小微商户样本，按 7:3 的比例划分训练集和测试集，划分结果如下：

表 20：训练集和验证集样本量及好坏样本

样本集名称	样本量	总体百分比	Y=0 样本量	Y=1 样本量
训练集 (DEV)	7040	69.8%	4428	2612
验证集 (OOT)	3041	30.2%	1999	1042
总计	10081	100.0%	6427	3654

训练集样本用于训练模型，而验证集则用于评估训练出来的模型的泛化能力。由于两个数据集使用的样本无重合，因此当训练集得到模型在验证集上有几乎同样有效的预测力时，说明这个模型的泛化能力是较好的。

### 5.7.1 评分模型训练

评分卡模型是基于逻辑回归的基本原理训练开发的，因变量Y=1如上文中定义为发生过30天及以上逾期的坏客户，好客户Y=0则定义为已结清或还款满6期,且未发生逾期的客户。X变量则是我们在前文中根据人行征信报告和同盾多头报告衍生出来的用于刻画小微商户融资行为的特征变量，分为“融资行为频率”和“融资渠道分布”两类特征变量，并经过



特征分箱处理将每个单变量分为不超过5组的区间变量，分箱后的单变量，应满足每一箱中同时包含好坏样本，且Bad Rate（即每一箱中Y=1的违约率）呈单调走势，当有特殊值（如空值）时可单独为一箱。具体的分箱方法（等频分箱/等距分箱/最优分箱）我们在本文中不重点介绍，以单变量var\_P2P（“同盾\_近3个月\_P2P网贷\_申请平台数”）为例，通过最优分箱方法，我们将其分为5组：

表 21: WOE分箱示例（变量var\_P2P）

组号	下界	上界	组样本量	Y=0 样本量	Y=1 样本量	Bad Rate	WOE	IV_I
1	0	1	1888	1417	471	24.95%	-0.5736	0.0801
2	2	2	1393	945	448	32.16%	-0.2186	0.0091
3	3	3	1169	731	438	37.47%	0.0156	0.0000
4	4	7	2087	1128	959	45.95%	0.3655	0.0411
5	8	+	503	207	296	58.85%	0.8855	0.0590

说明：以组号=1为例，表示变量“同盾\_近3个月\_P2P网贷\_申请平台数”在[0,1]的样本量共1888个，占总训练样本数7040的32.00%，其中Y=0的好样本共1417个，Y=1的坏样本共471个，即Bad Rate=471÷1888=24.95%， $WOE_1 = \ln \frac{b_1/b_T}{g_1/g_T}$ ，其中 $b_1$ 是第1组内坏客户数471， $b_T$ 是训练样本总体坏客户数2612， $g_1$ 是第一组内好客户数1417， $g_T$ 是训练样本总体好客户数4428，计算得到 $WOE_1=-0.5736$ ， $IV_1 = \left(\frac{g_1}{g_T} - \frac{b_1}{b_T}\right) \times WOE_1$ ，计算得到 $IV_1=0.0801$ 。单变量“同盾\_近3个月\_P2P网贷\_申请平台数”的 $IV = \sum_1^5 IV_i=0.1894$ ，说明

该变量具有中等预测力。

单变量分箱后，我们从每个变量计算得到的IV值筛选出对Y有较显著预测力的指标，作为逻辑回归模型的训练入模变量，并采用“逐步回归法”（Stepwise Regression）得到最终入模变量如下：

表 22：逻辑回归最终入模变量

最大似然估计分析					
参数	自由度	估计	标准 误差	Wald 卡方	Pr > 卡方
Intercept	1	-0.5286	0.0261	410.6079	<.0001
QUERY_BANK_R_24M_WOE	1	0.599	0.13	21.2327	<.0001
QUERY_L_R_24M_WOE	1	0.6483	0.0849	58.3743	<.0001
QUERY_OWN_CNT_12M_WOE	1	0.555	0.0887	39.1513	<.0001
var_dxxfjr_WOE	1	0.461	0.1205	14.6361	0.0001
var_p2p_WOE	1	0.632	0.0666	89.9428	<.0001
var_rzzl_WOE	1	0.7818	0.0728	115.2376	<.0001
var_yhxwdk_WOE	1	0.7018	0.1848	14.4204	0.0001

7个入模变量的系数均为正，P值 $\leq 0.0001$ ，说明变量效果显著，变量含义如下：

表 23：逻辑回归最终入模变量含义

变量名	含义
QUERY_BANK_R_24M_WOE	人行_近2年_银行_贷款审批_查询次数占比_分箱

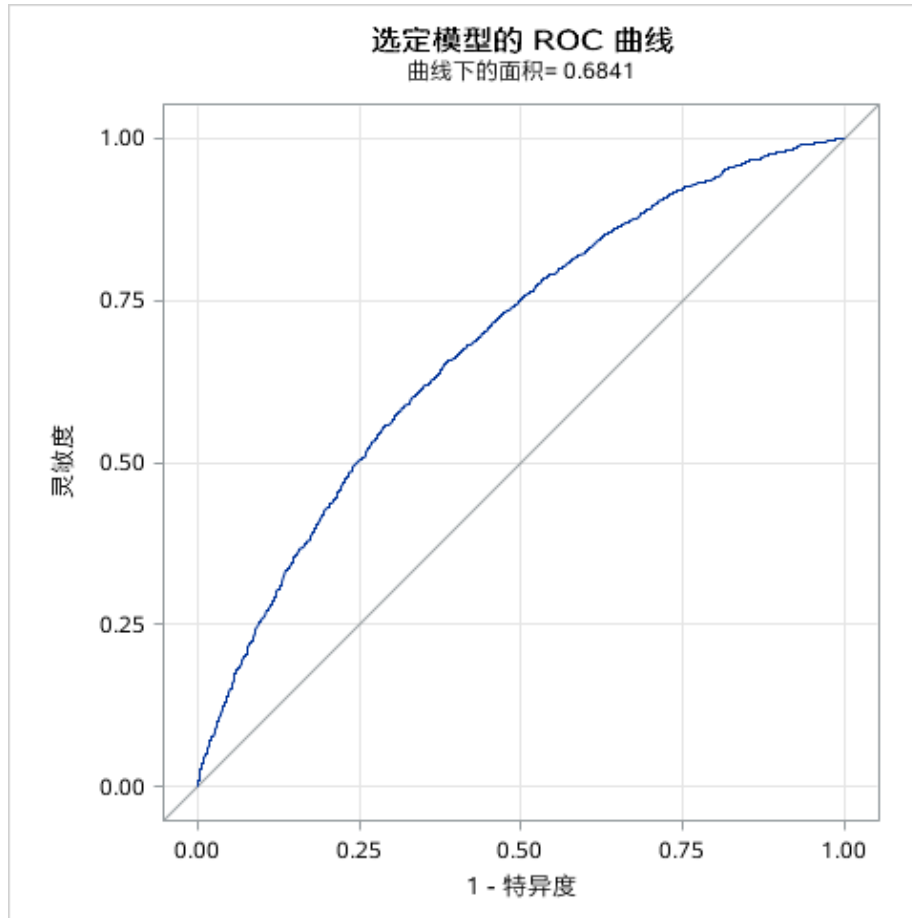
QUERY_L_R_24M_WOE	人行_近2年_贷款审批_查询次数占比_分箱
QUERY_OWN_CNT_12M_WOE	人行_近360天_个人查询_查询次数_分箱
var_dxxfjr_WOE	同盾_近3个月_大型消费金融公司_申请平台数_分箱
var_p2p_WOE	同盾_近3个月_P2P网贷_申请平台数_分箱
var_rzzl_WOE	同盾_近3个月_融资租赁_申请平台数_分箱
var_yhxwdk_WOE	同盾_近3个月_银行小微贷款_申请平台数_分箱

---

模型ROC曲线如下，AUC=0.6841，计算KS=26.5%，说明模型有一定预测力，即对于区分好坏客户有一定效果，但并不算非常强，可结合其他维度变量（比如客群属性、历史

履约、收入水平等）增加模型预测力。本文仅探讨融资行为变量构造的子评分卡模型，因此不再展开。

图 75：选定模型的 ROV 曲线和 AUC 值



将逻辑回归模型按照一定方法校准获得评分卡形式，结果如下：

表 24：“融资行为”子评分卡

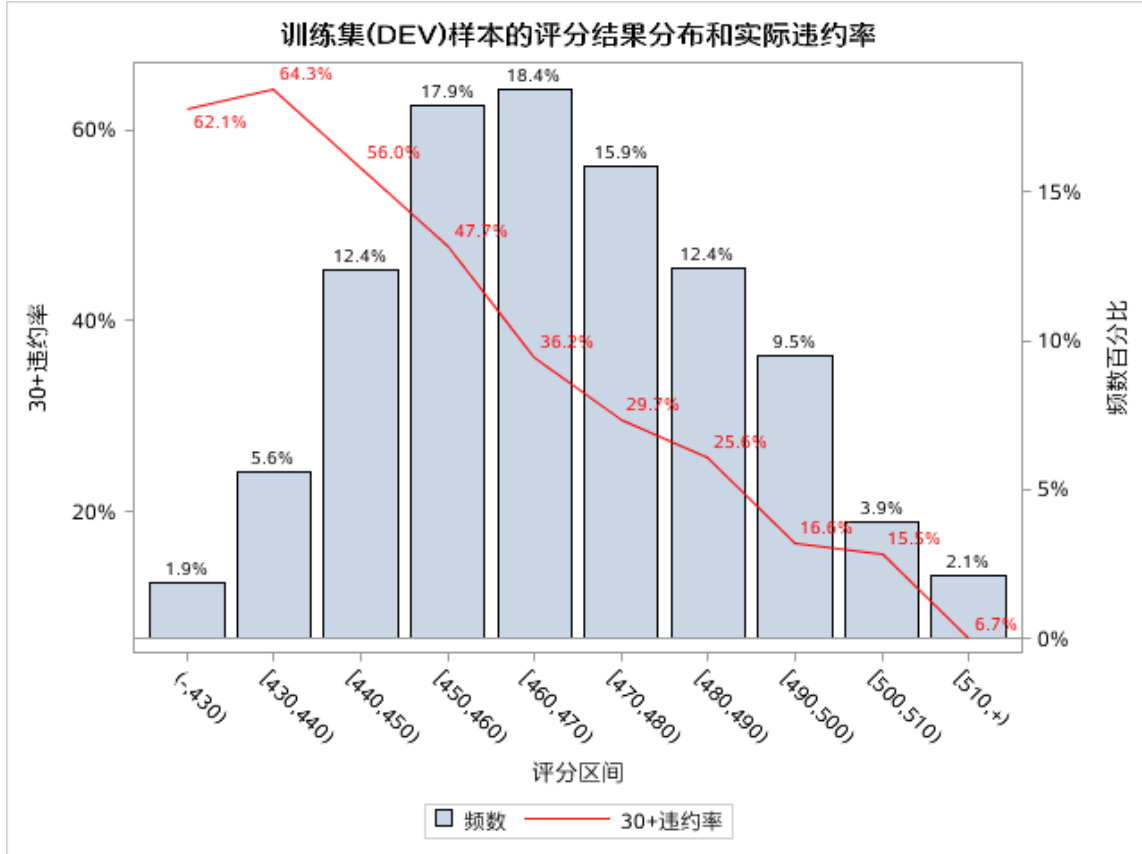
基准分		466
人行征信	QUERY_BANK_R_24M	(-, 0.36) -7
	人行_近2年_银行_贷款审批_查询次数占比	[ 0.36, 0.45) -3

		[ 0.45, 0.64 )	-1
		[ 0.64, 0.86 )	+2
		[ 0.86, + )	+6
	QUERY_L_R_24M	( -, 0.37 )	-11
	人行_近2年_贷款审批_查询次数占比	[ 0.37, 0.44 )	-6
		[ 0.44, 0.62 )	-1
		[ 0.62, 0.81 )	+7
		[ 0.81, + )	+19
	QUERY_OWN_CNT_12M	( -, 3 )	+13
	人行_近360天_个人查询_查询次数	[ 3, 6 )	+5
		[ 6, 10 )	-1
		[ 10, 13 )	-3
		[ 13, + )	-7
同盾多头	var_dxxfjr	[ 0, 1 )	+2
	同盾_近3个月_大型消费金融公司_申请平台数	[ 1, 2 )	-3
		[ 2, + )	-8
	var_p2p	[ 0, 2 )	+10
	同盾_近3个月_P2P网贷_申请平台数	[ 2, 3 )	+4
		[ 3, 4 )	+0
		[ 4, 8 )	-7
		[ 8, + )	-16
	var_rzzl	[ 0, 1 )	-5
	同盾_近3个月_融资租赁_申请平台数	[ 1, + )	+15
	var_yhxwdk	[ 0, 1 )	-1
	同盾_近3个月_银行小微贷款_申请平台数	[ 1, 2 )	+5
		[ 2, + )	+17

按评分卡逻辑对7040个训练样本打分，并等距分为10个区间，分别计算每个评分区间内的样本量占比以及实际Bad Rate（逾期30+天的违约率），得到下图的频率分布直方图和实际违约率走势图。可以发现，样本按评分基本呈中间高两边低的正态分布，而实际违

约率则呈现随分值增高而降低的走势（除[430,440)分值段略有上翘），最低分值段的Bad Rate高达60%以上，而最高分值段的Bad Rate仅为6.7%，说明评分卡模型风险排序效果好且有显著好坏区分度。综上，融资行为子评分卡模型基本达到预期风险分层效果。

图 76：训练集(DEV)样本的评分结果分布和实际违约率

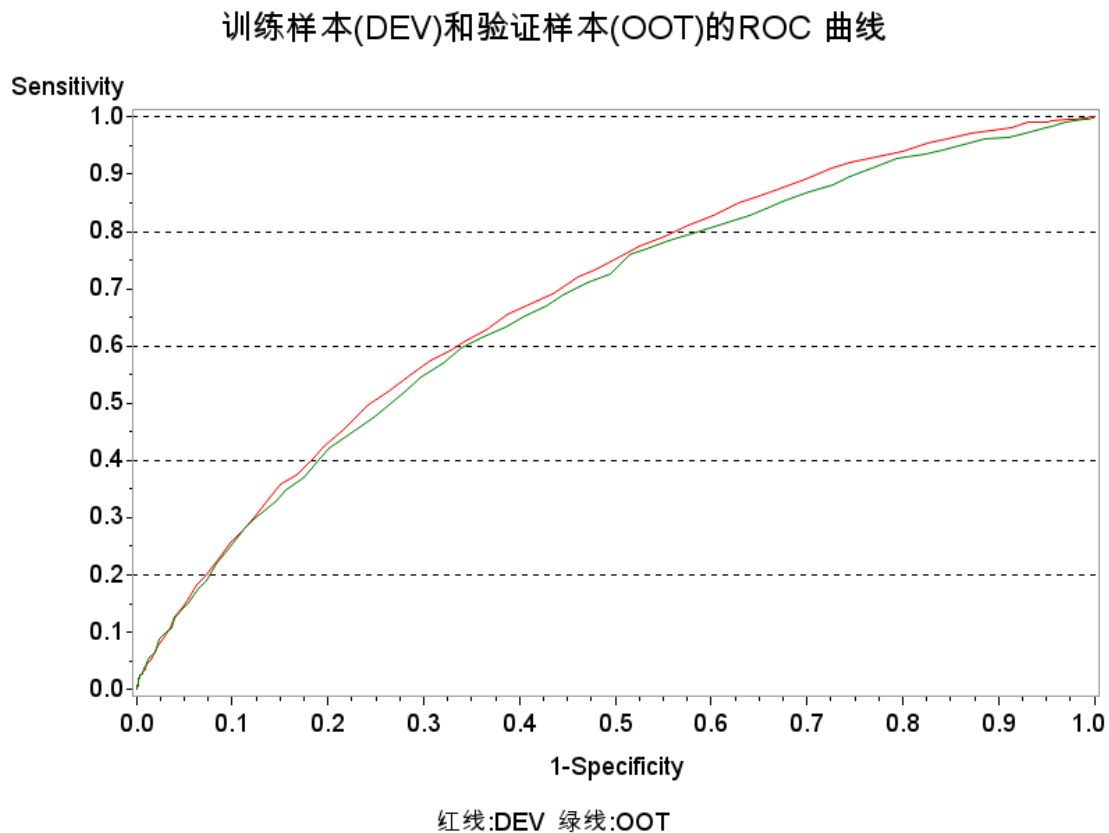


## 5.7.2 评分模型验证

10081 个样本总体中 7040 个为训练样本（占比 70%），另外 3041 个为验证样本（占比 30%）。我们基于训练样本得到了“融资行为”子评分卡，接下来我们要在验证样本中评估子评分卡的泛化能力。

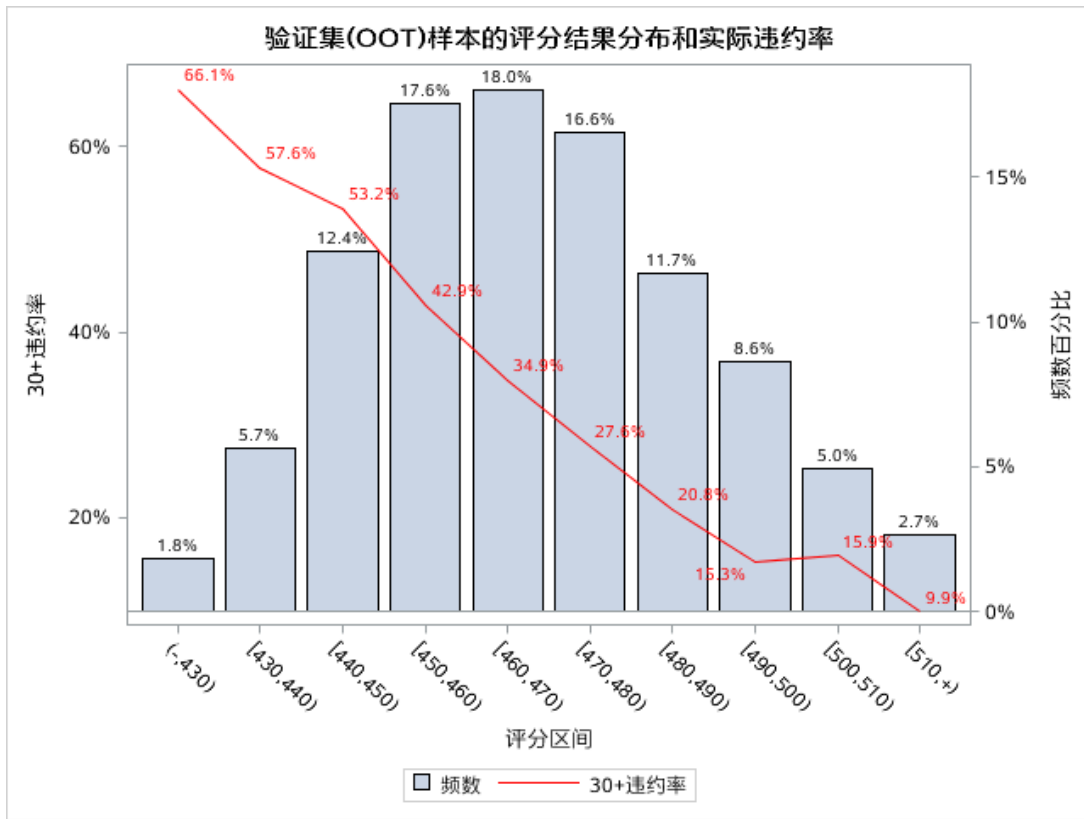
首先，从下图训练样本和验证样本的 ROC 曲线对比可以看到，两者的曲线接近重合。计算得到验证样本的  $KS=25.4\%$ ，对比训练样本的  $KS=26.5\%$  比较接近，说明训练得到模型的预测力在验证样本上同样得到了接近的效果，泛化能力符合预期。

图 77: 训练样本(DEV)和验证样本(OOT)的 ROC 曲线



按训练得到的“融资行为”子评分卡给验证样本打分，同样划分为 10 个等级的评分段，并绘制 10 个评分段的频率分布直方图和实际违约率走势图，可以看到验证样本的模型评分分布与训练模型相近，且风险排序效果同样较好，最低分值段的 Bad Rate=66.1%，而最高分值段的 Bad Rate=9.9%，好坏区分效果显著。

图 78：验证集(OOT)样本的评分结果分布和实际违约率



综上，我们仅基于融资行为相关的特征变量，就构建了具有一定风险分层和预测能力的子评分卡，这说明小微商户的融资行为频率和渠道构成分布，对于贷前信用风险的评估是有显著价值的。毫无疑问，如果结合其他维度数据（比如客群属性、历史履约、收入水平等）一定能帮助我们获得更精准的风险预测模型，完善小微商户风控体系，给予更合理的额度和利率定价。



## 5.8 融资行为变量对信用风险模型的提升效果

前文中我们仅用融资行为变量，就构造了一个具有显著预测力的信用评分模型，从而证实了这些变量对于信用风险预测的效果。接下来为了更客观衡量贷款实务中融资行为变量对信用风险模型的提升效果，我们将分以下两类 X 变量集合来构造评分卡模型：

**模型 1：**贷款实务中常用的模型特征变量，主要包括最基础的客户基本特征（性别、年龄、婚姻状况等）、征信衍生特征（从征信汇总信息、贷记卡记录、贷款记录等衍生得到）；

**模型 2：**融资行为变量+模型 1 中的所有特征变量；

最终通过比较模型 1 和模型 2 的风险分层能力和预测坏客户精准度，来评估融资行为变量对实务中信用风险模型的特省效果。

### 5.8.1 模型 1：不含融资行为变量的评分卡模型效果

和 5.7 章节中一样，使用 10081 个既有有效人行征信报告又有有效同盾多头报告的小微商户样本，按 7:3 的比例划分训练集和测试集，分别进行评分卡模型的训练和测试，但不同的是 X 变量集使用实务中常用的模型特征变量，共 588 个变量。其中，征信汇总类（信用历史汇总，贷款汇总、贷记卡汇总、逾期汇总等）变量 50 个，贷款明细衍生类变量 419 个，贷记卡明细衍生类变量 119 个。

经训练得到模型 1（不含融资行为变量）的评分卡逻辑如下：

表 25: 模型 1 (不含融资行为变量) 评分卡

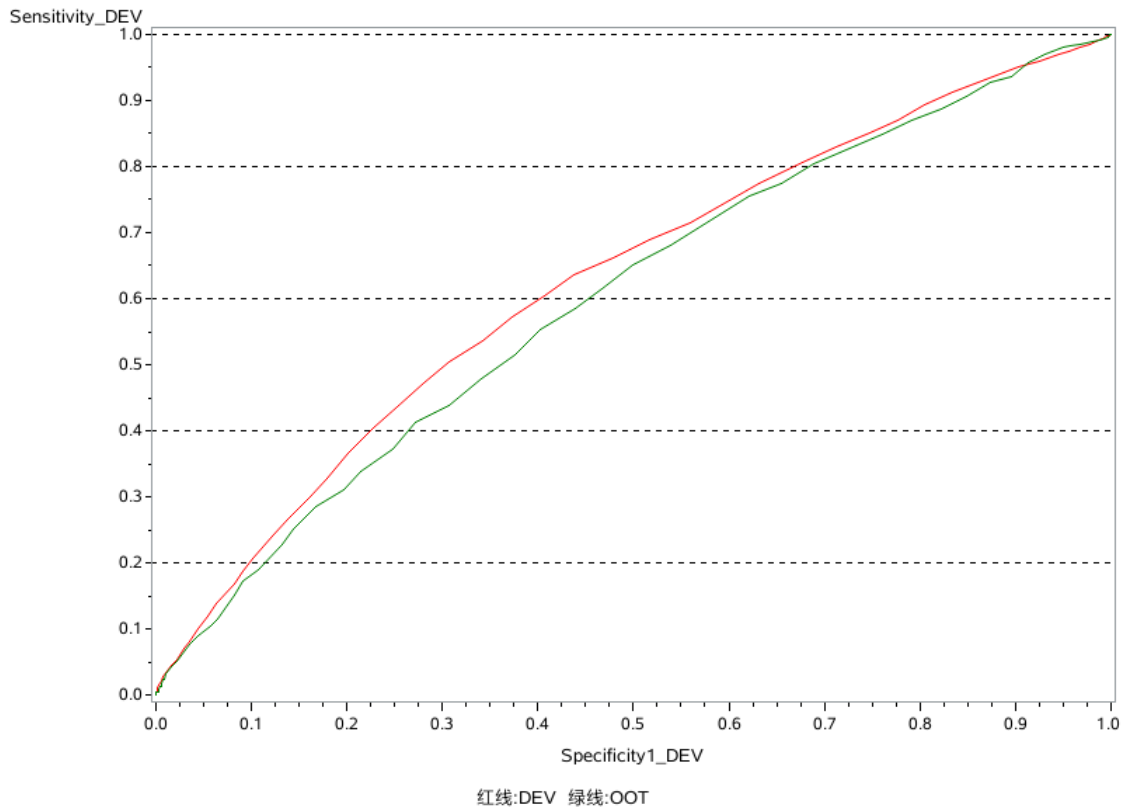
维度	变量	区间	评分	
基准分			472	
人行征信 (不含融资行为变量)	CARD_CNT_12M	(-, 1)	+5	
	人行_贷记卡_近360天_新增_张数	[ 1, 2 )	+2	
		[ 2, + )	-4	
		CARD_OPEN_MONTH	(-, 20)	-8
		人行_贷记卡_首张开卡日_距申请月数	[ 20, 50 )	-4
			[ 50, 131 )	-1
	[ 131, 171 )		+3	
		[ 171, + )	+7	
	LOAN_ACT_MAX_AMT	(-, 249880)	-4	
	人行_贷款_在还的_发放金额_最高金额	[ 249880, 489360 )	+5	
		[ 489360, 728840 )	-1	
		[ 728840, 4560520 )	-3	
		[ 4560520, + )	-7	
	LOAN_CNT_BANK_12M_R	.	-7	
	人行_贷款_近360天_新增银行发放_占历史银行发放_笔数比例	[ 0, 0.4 )	+2	
		[ 0.4, 0.65 )	+0	
		[ 0.65, + )	-4	
LOAN_CNT_XD_12M_R	.	-4		
人行_贷款_近360天_新增小贷公司发放_占历史小贷公司发放_笔数比例	[ 0, 0.25 )	+3		
	[ 0.25, 0.4 )	+2		

	[ 0.4, 0.8 )	-1
	[ 0.8, + )	-6
LOAN_JY_MIN_AMT	.	-4
人行_贷款_经营性贷款_最低金额	( 0, 50213 )	+2
	[ 50213, + )	+4
LOAN_MIN_XF_24M	.	+3
人行_贷款_近720天_新增消费贷款_最低金额	( 0, 80540 )	-1
	[ 80540, 147490 )	+2
	[ 147490, + )	+9

模型1评分卡的训练样本KS值为19.8%，验证样本KS仅14.6%，均不具备显著的预测力，ROC曲线如下图：

图 79：模型 1 的训练样本(DEV)和验证样本(OOT)的 ROC 曲线

训练样本(DEV)和验证样本(OOT)的ROC 曲线



模型 1 评分卡的风险排序效果在 500 分以下的区间段较好，训练样本和验证样本的实际违约率均随着评分的升高而降低。但在高分段（500 分以上）却出现排序异常：

图 80：模型 1 的训练集(DEV)样本的评分结果分布和实际违约率

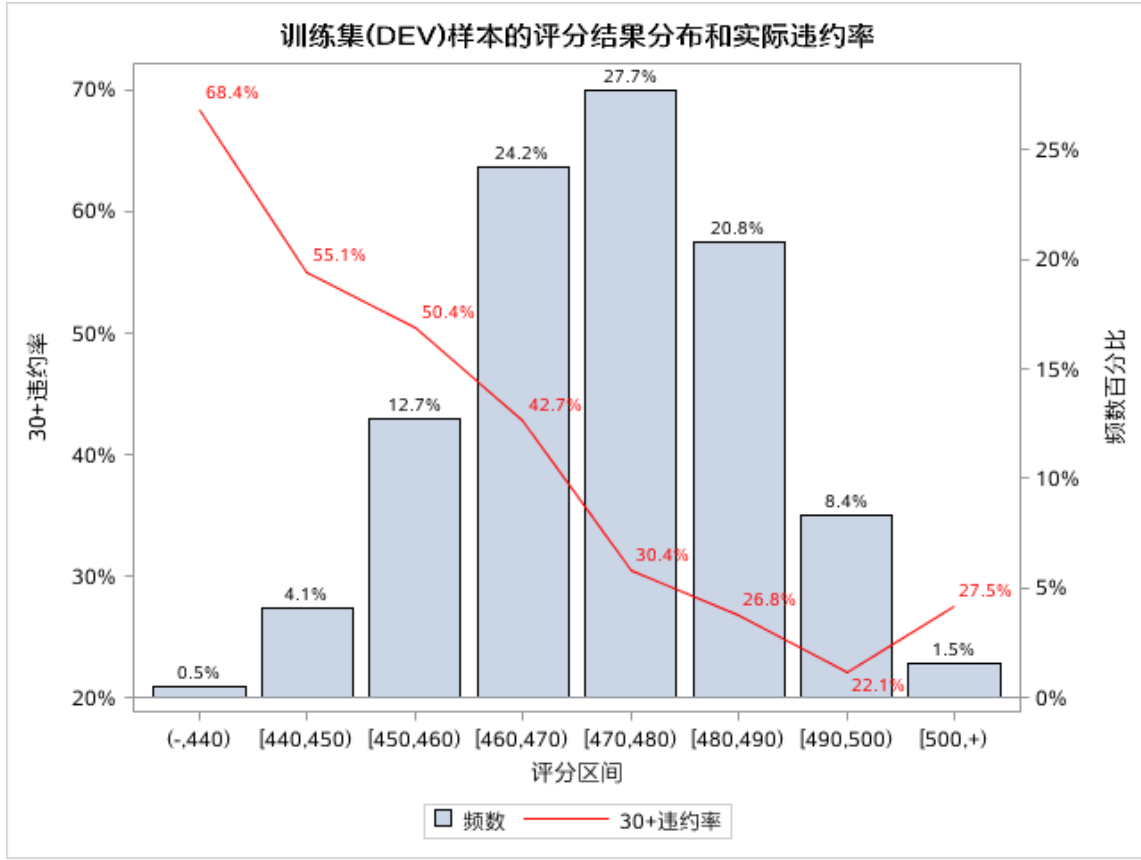
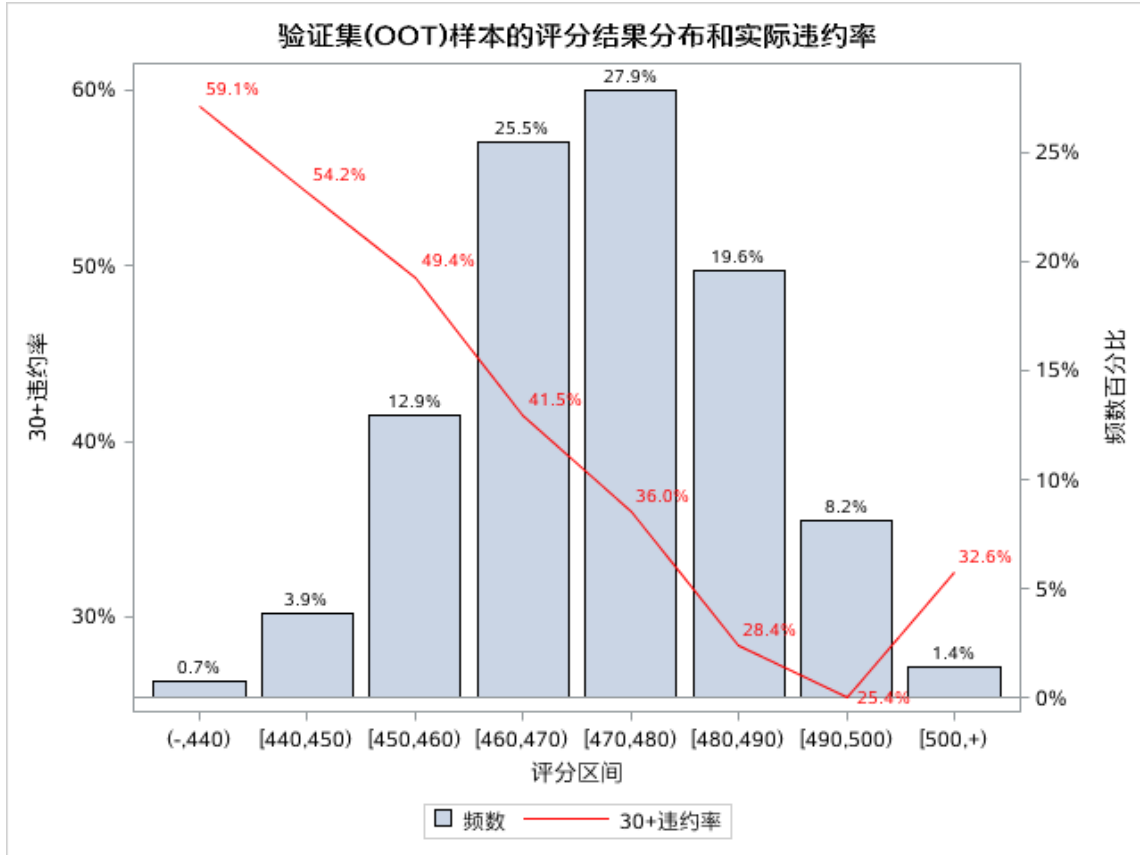


图 81: 模型 1 的验证集(OOT)样本的评分结果分布和实际违约率



### 5.8.2 模型 2：含融资行为变量的评分卡模型效果

为排除样本集和其他不同导致的模型效果比对偏差，模型2仍采用模型1中的同一批样本10081个，按7:3分为训练集和验证集。但模型2的X变量集，为5.7中“融资行为变量”和模型1中X变量集的并集，共792个特征变量。

经训练得到模型2（含融资行为变量）的评分卡逻辑如下：

表 26：模型2（含融资行为变量）评分卡

维度	变量	区间	评分
基准分			464
模型1	CARD_HIGHAMT_ORG_CNT	(-, 2)	-6
人行征信 (人行_贷记卡_单卡额度达10万及以上_发放机构数 不含融资行 为变量)		[ 2, 4)	-2
		[ 4, 9)	+3
		[ 9, +)	+6
	CARD_MIN_ACT_USE_R	(-, 0.10)	+1
	人行_贷记卡_单卡额度使用率_最小值	[ 0.10, 0.35)	-1
		[ 0.35, 0.94)	-3
		[ 0.94, +)	-16
	CARD_OPEN_MONTH	(-, 20)	-8
	人行_贷记卡_首张开卡日_距申请月数	[ 20, 50)	-4
		[ 50, 131)	-1
		[ 131, 171)	+3
		[ 171, +)	+7
	LOAN_CNT_BANK_6M_R	.	-7
	人行_贷款_近180天_新增银行发放_占历史银行发 放_笔数比例	[ 0, 2)	+2
		[ 0.2, 0.5)	+0
		[ 0.5, +)	-5
	LOAN_CNT_XD_12M_R	.	-4
	人行_贷款_近360天_新增小贷公司发放_占历史小 贷公司发放_笔数比例	[ 0, 0.25)	+3
		[ 0.25, 0.4)	+2
		[ 0.4, 0.8)	-1
		[ 0.8, +)	-6
	LOAN_NH_MAX_AMT	.	+1
	人行_贷款_农户贷款_最高金额	( 0, 184500)	-8
		[ 184500 , 453750)	+0
		[ 453750, +)	+5
	QUERY_BH_CNT_24M	(-, 1)	-1

	人行_近2年_保后管理_查询次数	[ 1, 3 )	+1
		[ 3, 4 )	+3
		[ 4,+ )	+7
人行查询	QUERY_CNT_1M_3M	( -, 0.15 )	-6
	人行_近30天查询次数_占近90天查询次数_比例	[ 0.15, 0.4 )	-1
		[ 0.4, 0.8 )	+2
		[ 0.8,+ )	+8
	QUERY_C_CNT_24M	( -, 4 )	+6
	人行_近2年_信用卡审批_查询次数	[ 4, 5 )	+4
		[ 5, 8 )	+0
		[ 8,13 )	-3
		[ 13, + )	-5
	QUERY_OWN_CNT_12M	( -, 3 )	+14
	人行_近360天_本人查询_查询次数	[ 3, 6 )	+6
		[ 6, 8 )	+0
		[ 8, 13 )	-3
		[ 13, + )	-9
	QUERY_PCT_OWN_24M	( -, 0.12 )	+11
	人行_近2年_本人查询_次数占比	[ 0.12, 0.19 )	+3
		[ 0.19, 0.29 )	-2
		[ 0.29, 0.43 )	-9
		[ 0.43, + )	-19
同盾多头	var_p2p	( -, 2 )	+10
	同盾_近3个月_P2P网贷_申请平台数	[ 2, 3 )	+5
		[ 3, 4 )	+1
		[ 4, 7 )	-6
		[ 7, + )	-15
	var_rzzl	[ 0, 1 )	-5
	同盾_近3个月_融资租赁_申请平台数	[ 1, + )	+12

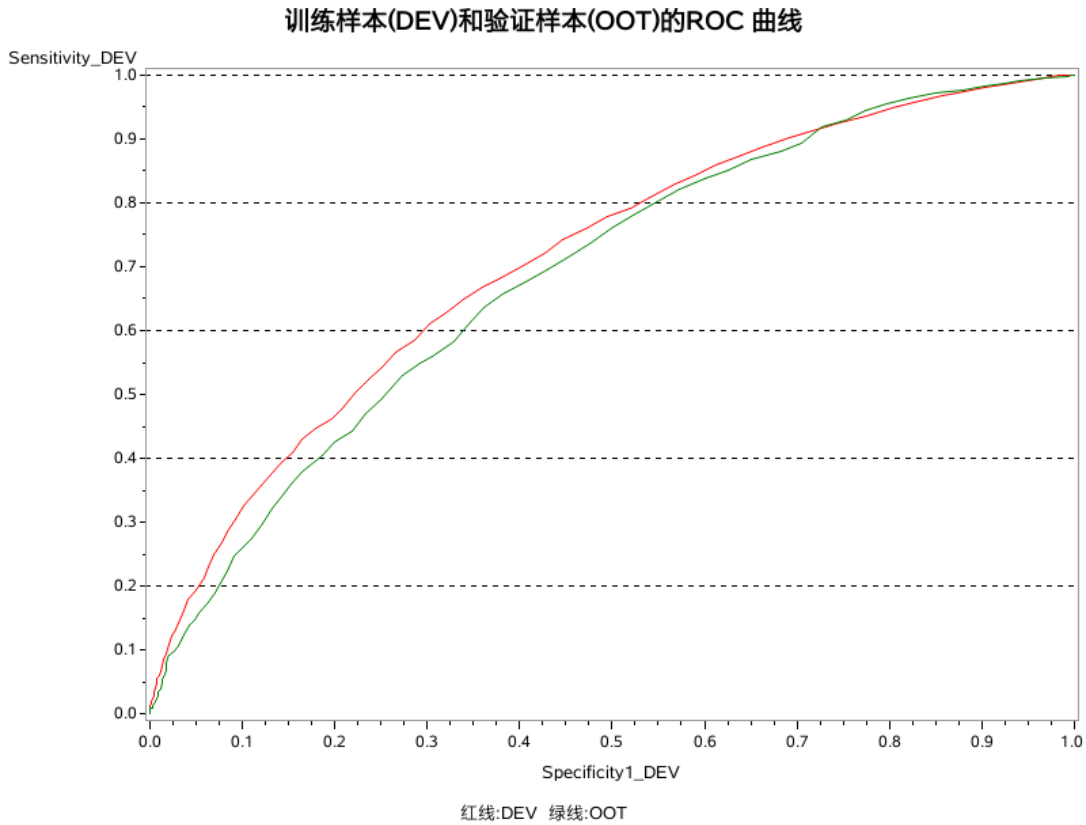
最终入模的13个变量中，7个来自模型1中的X变量，6个来自融资行为变量（4个来自人



行征信查询及，2个来自同盾多头)。从入模变量首先可见，融资行为变量个数的占比几乎达到一半。

模型2评分卡的训练样本KS值为30.0%，验证样本KS仅27.3%，均达到非常显著的预测力，ROC曲线如下图：

图 82：模型2的训练样本(DEV)和验证样本(OOT)的ROC曲线



模型2评分卡的风险排序效果如下图所示非常理想，呈现严格单调递减。且高低分段的实际违约率拉开明显，训练样本中最好的520分以上客群实际违约率近1.8%，而最差的400分以下客群实际违约率高达85.7%。验证样本美中不足在于低分段风险排序性有瑕疵，但不影响低分段识别的客户仍为违约率超50%的高风险人群。

图 83：模型 2 的训练集(DEV)样本的评分结果分布和实际违约率

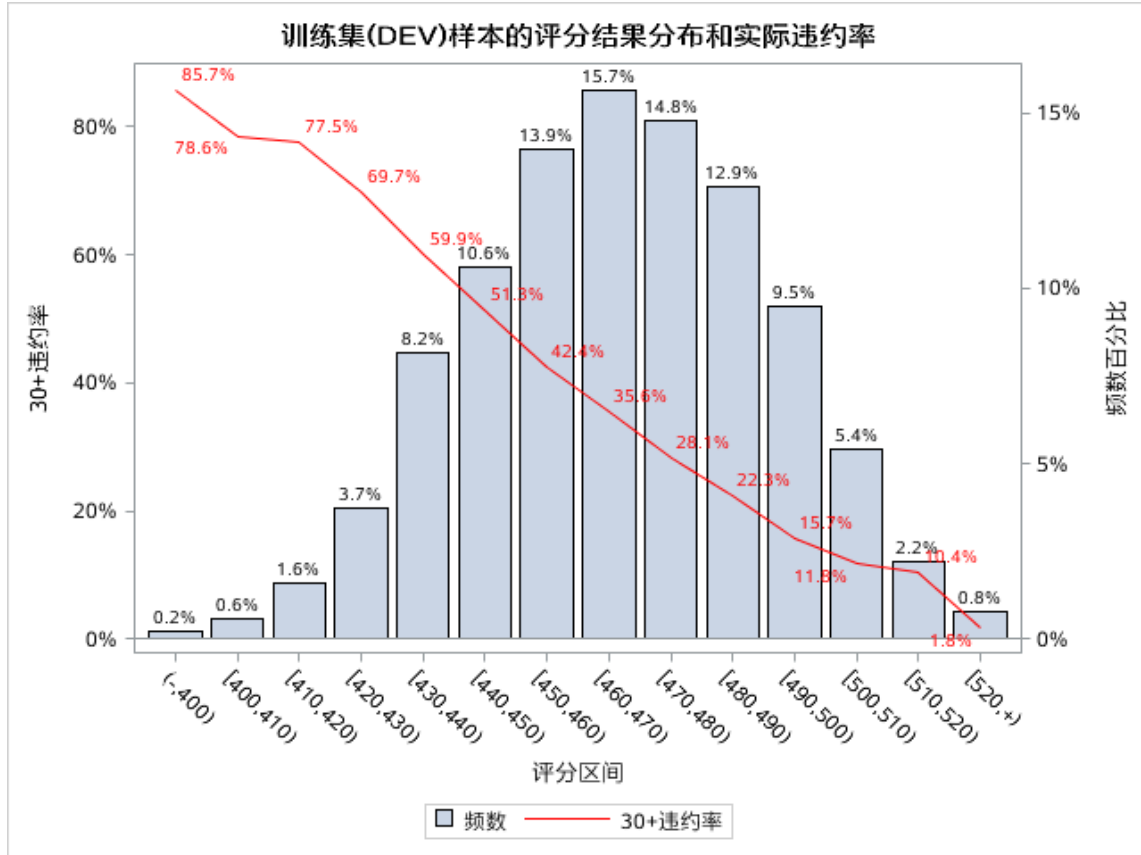
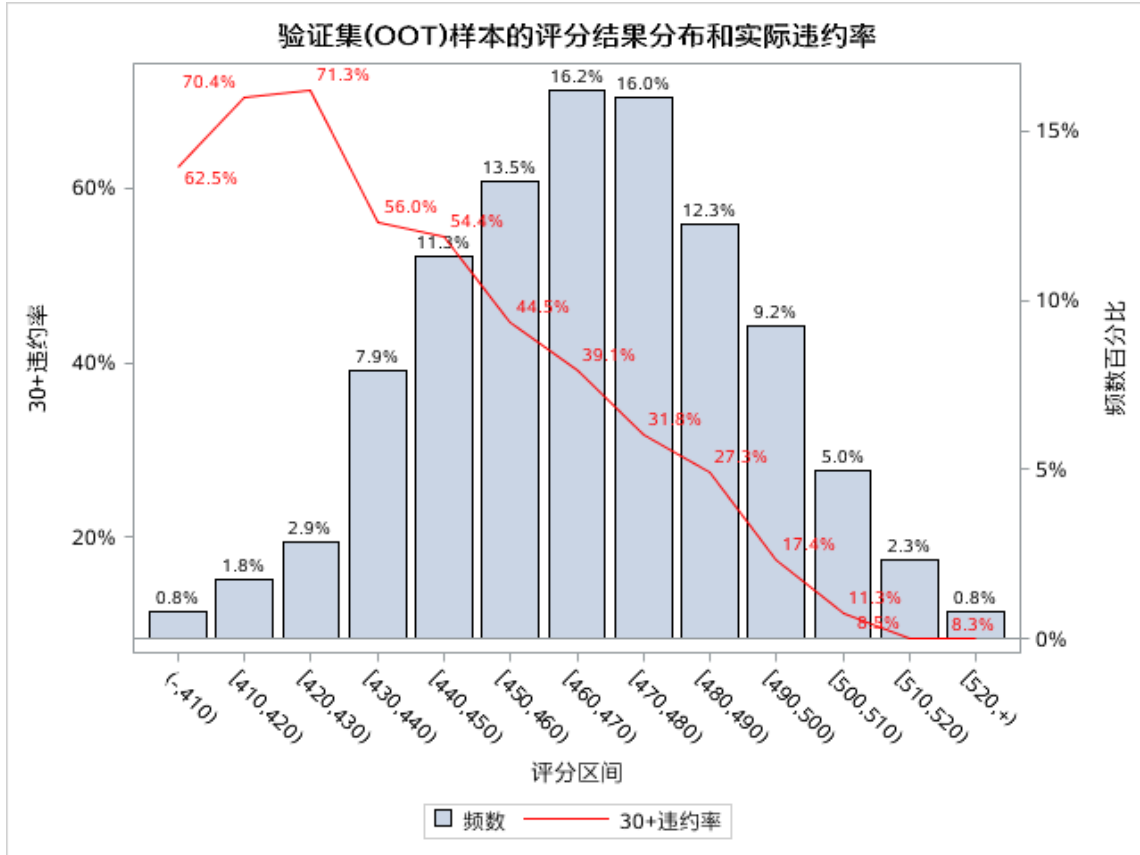


图 84：模型 2 的验证集(OOT)样本的评分结果分布和实际违约率



综上，模型2相对于模型1，KS从19.8%立即提升到30.0%，即模型预测力从不显著提升到效果显著，风险排序性和好坏客户区分度也有非常显著的改善。这正说明，融资行为类变量对于实务中信用风险模型的提升效果是极为可观的。不仅仅是常规的人行查询次数变量，更包括查询原因、查询机构类型、次数增长趋势等其他更细化的衍生变量。

### 5.9 融资行为与信用风险暴露期限的关系探索

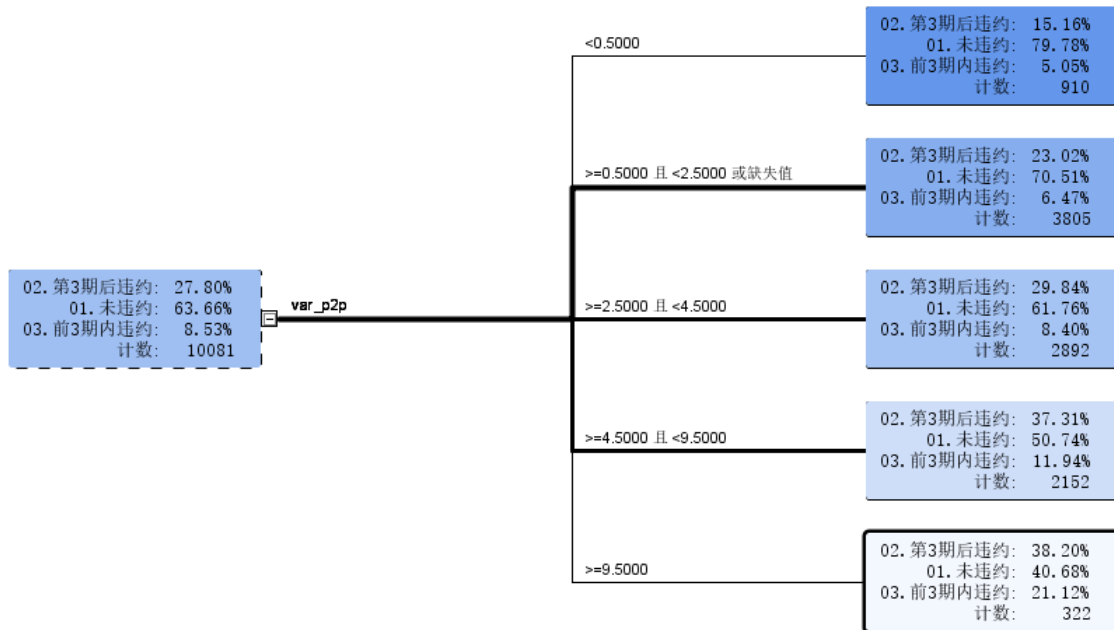
前文中无论是逻辑回归模型还是决策树模型，我们都用来解决一个  $Y=0$  或  $1$  的两分类问题，但事实上，除了“是否会发生违约”，我们还关心放款后多久发生违约，“短期内违约/中长期内违约/未违约”三类中属于哪一类。

### 5.9.1 决策树分析探索

我们根据 4.2.2 中对因变量的定义，将 Y 分为“01.未违约”、“02.第 3 期后违约”、“03.前 3 期内违约”三类，并仍然使用融资行为相关的特征变量作为 X，构造决策树模型，探索不同的融资行为特征下，小微商户的信用风险暴露期限长短是否有显著不同。

我们以单变量分析中预测力最为显著的变量之一 var\_P2P（同盾\_近 3 个月申请\_P2P 网贷\_平台数）为例，取值越大，“01.未违约”、“02.第 3 期后违约”、“03.前 3 期内违约”的分布逐渐下沉。当 var\_P2P $\geq 10$  时，前 3 期内违约的客户占比高达 21.1%，而未违约的客户占比仅为 40.7%。这说明小微商户的融资行为变量同样影响着信用风险暴露期限，在类似 P2P 这样的下沉融资渠道申请融资越频繁，越容易在早期就发生严重违约。

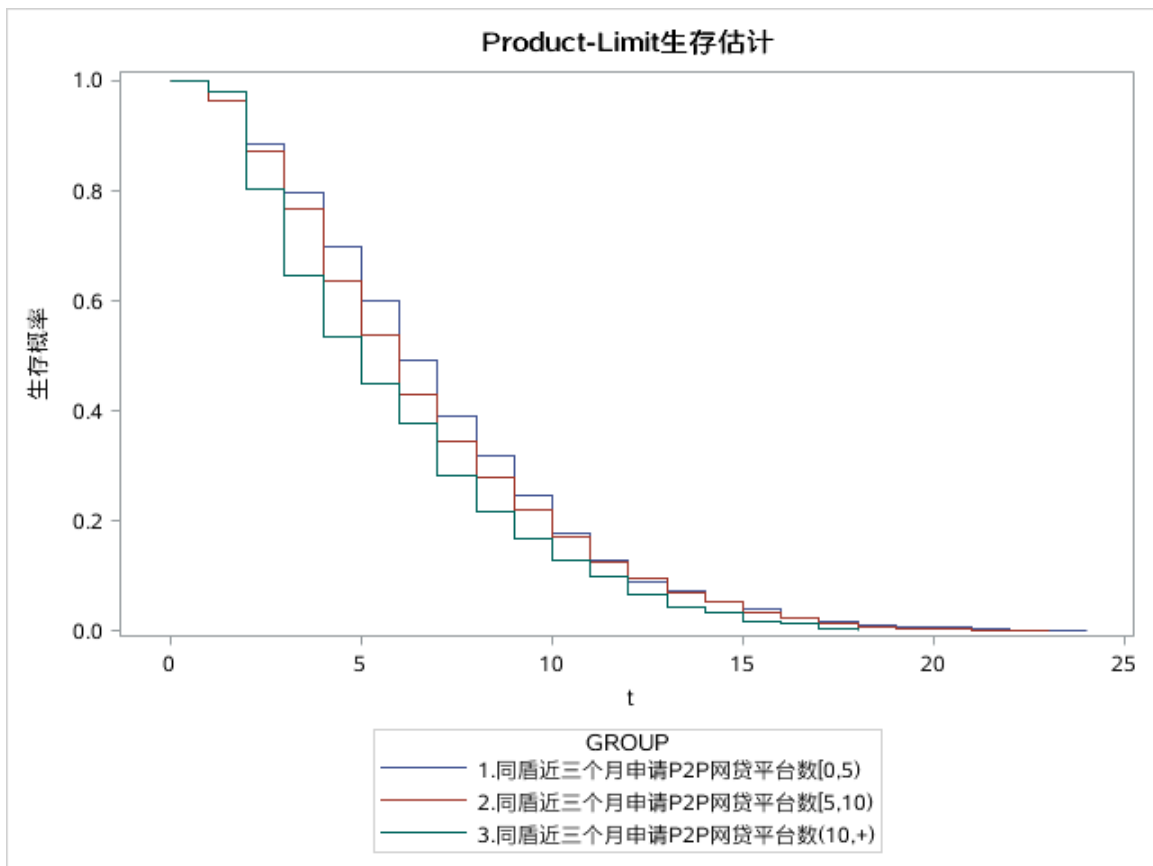
图 85: var\_P2P 与信用风险暴露期限关系决策树



### 5.9.2 生存曲线分析探索

同样以 var\_P2P（同盾\_近 3 个月申请\_P2P 网贷\_平台数）为例，将小微商户样本根据该变量取值从低到高分为 3 组：[0,5) / [5,10) / (10,+)，分别绘制出三组生存曲线，从而比较不同的 P2P 网贷平台申请数的小微商户，在还款周期中每个账单日的生存率。这里的终点事件，是“发生 30 天以上违约”。删失案例(censored)，是截止本文研究时点仍未发生 30 天以上违约的存活案例。由下图基于 var\_P2P 分组下的生存曲线可见，同盾\_近 3 个月申请\_P2P 网贷\_平台数越多，生存率随时间的下降速度越快。

图 86: var\_p2p 分组下的生存曲线



## 六、总结陈述

本文以国内广大小微商户为研究对象，客观陈述了小微融资难融资贵的当前困境，尤其是在银行等传统正规金融机构较难准入和获得足够授信的情况下，转而求助 P2P 网贷、小额贷款公司、消费金融公司等新型融资渠道。当客户有了更多的选择，他们就能从额度利率、快捷便利性、准入门槛、安全可靠等多种角度去制定适合他们的融资方案。反过来，我们也可以合理推测，客户的融资行为也反映了他们的信用风险水平。比如一个正常经营资质较好的小微商户更容易在银行获得授信，从融资成本和正规性考虑他们也会更倾向于向银行贷款，融资行为相对不频繁。而一个经营困难资金周转不佳的小微商户则可能向一切能获得资金的平台申请贷款，包括 P2P 网贷和民间借贷，融资频率也会更高。

“小微商户的融资行为与信用风险存在一定关联性”，基于这样的推理假设，我们分析了 H 公司 2017-2019 年间的 1 万多户小微商户样本，并希望得到这种关联关系的数据上的验证。

新型融资渠道的出现，不仅从一定程度缓解了小微商户的融资难题，更激发了另一新领域的活跃，即类似同盾、百融等风控大数据公司。这些大数据公司能补充人行征信体系在银行等持牌机构以外的信用数据缺失，以本文为例，同盾的多头报告记录了个人客户在近期的申请融资行为和融资渠道，为我们的研究提供了非常有用的数据来源。我们收集了 H 公司这 1 万多户小微商户的有效人行征信和同盾多头报告，并衍生除了融资行为相关的几十个 X 变量，主要分为“融资行为频率”和“融资渠道分布”两类。我们也根据 H 公司

小微商户的还款表现定义了目标坏客户——发生过 30 天及以上逾期的违约客户即为

$Y=1$ ，从未发生过逾期的客户为  $Y=0$ 。

通过对 X 和 Y 的数据分析，我们可以得到以下结论：

- 1) 融资行为频率和违约率正相关；
- 2) 融资渠道选择优先度：银行>消费金融公司≈小额贷款公司>P2P 平台；
- 3) 在正规金融机构（接入人行征信的）的融资占比与违约率负相关；
- 4) 向银行融资频率和占比与违约率负相关；
- 5) 向 P2P 网贷平台融资频率和占比与违约率正相关；

通过构建逻辑回归模型和决策树模型，我们得到以下结论：

- 1) 融资行为变量对于预测违约率有显著效果；
- 2) 用融资行为变量构建的子评分卡有较好的风险排序效果；
- 3) 含融行为变量的评分卡，比不含融资行为变量的评分卡，KS 提升度非常显著；
- 4) 融资行为变量可用于制定风控规则剔除局部高风险客户；
- 5) 融资行为变量除了可以预测是否违约还与信用风险暴露期限有关；

以上结论验证了我们一开始的核心假设“小微商户的融资行为与信用风险有显著关系”，充分利用融资行为变量在实务中可帮助我们评估小微商户处于什么样的资金需求阶段（正常或迫切）以及融资渠道（正规或下沉），从而反映未来一段时间内的信用风险水平，在实务中帮助我们构建更准确的风险评估体系和风险防范规则，为广大小微商户提供合理的额度授信和利率定价，实现真正的“普惠金融”。

呼应本文 1.3 中提出的五个子问题，我们通过详实的数据分析和统计测算，可以得到以下研究结论：

1) 小微商户融资行为如何刻画？如何获取数据并建立合理指标？

本文中，我们从“融资行为频率”和“融资渠道分布”两个维度刻画小微商户融资行为，以商户申请贷款时授权查询的人行征信报告和同盾多头报告作为数据来源，最终衍生出不同时间切片下的 70 余个特征指标，更细致客观地描述了 H 公司 1 万多个小微商户样本的近期融资情况。

2) 小微商户申请融资的渠道（银行/P2P/小贷公司等）数量和分布情况？

从人行征信报告可以分析得到：近 90 天内人均被 2.59 家接入征信的金融机构（银行、小额贷款公司、消费金融公司等）因“贷款审批”而查询过，且查询次数为 2.80 次，且其中银行查询占比达 60%以上，小额贷款公司占 20%，消费金融公司占 13%。同时，“个人查询”为原因的人均次数为 2.38 次，意味着这批客户近 90 天内向未接入征信的融资渠道（P2P 网贷平台、典当行、民间借贷等）申请贷款的人均次数也在 2 次以上。

从同盾多头报告可以分析得到：95%的小微商户近 3 个月申请平台数 $\geq 2$ ，且其中 P2P 网贷平台查询占比达 46%，小额贷款公司 20%，各类消费分期和消费金融公司占比 20%以上。由于同盾多头报告的接入机构主要为非银行机构，因此其中的银行查询占比远低于实际水平。

3) 高风险的小微商户和低风险的小微商户，在申请融资行为上有什么区别，是否能



找到具有区分力的特征指标？

融资行为频率与信用风险呈正相关，即融资频率越高，小微商户的风险越高；融资渠道分布同样与信用风险显著相关，在正规金融机构（接入人行征信的）的融资占比越高，小微商户的风险越低。

通过对融资行为特征的IV值以及决策树变量重要性分析，我们发现同盾\_近3个月\_P2P网贷\_申请平台数、同盾\_近3个月\_P2P网贷\_申请平台占比、人行\_近180天\_个人查询\_查询次数、人行\_近2年\_贷款审批\_查询次数占比等特征变量，对小微商户高低风险的区分度有显著效果。

4) 小微商户融资行为的相关特征指标，对预测客户短期和长期的违约率效果如何（用统计指标衡量）？

小微商户的融资行为变量影响着信用风险暴露期限，在类似P2P这样的下沉融资渠道申请融资越频繁，越容易在早期就发生严重违约。这个结论通过决策树分析和生存曲线分析均可得到验证。

5) 如何将融资行为维度的特征变量，用于建立有效的风控模型和规则，最终提升小微商户的业务水平？

将融资行为变量与常规信用风险评估模型变量一起作为新模型的筛选变量，可以发现含融资行为变量的新模型相对于不含融资行为变量的常规模型，KS 从 19.8%提升到 30.0%，即模型预测力显著提高，风险排序性和好坏客户区分度也有明显改善。这说明，融资行为类变量对于实际业务中信用风险模型的优化是极为有效的。

受限于此次研究样本中小微商户的融资行为数据，采集时点均为贷前申请阶段且仅有一次记录，本文仅对小微商户的融资行为与信用风险之间的关联性做了分析验证，而对因果关系等更多内生性问题并未涉及。在未来研究方向上，须更注重收集小微商户全信贷生命周期（贷前、贷中、贷后）的融资行为面板数据，寻找合适的工具变量并进一步探究融资行为与风险的因果关系，透彻分析产生小微商户融资难融资贵问题的本质原因和内在机理，为政府和金融机构解决这一社会性难题提供可靠思路。

## 参考文献

- [1]石宝峰,王静,迟国泰.普惠金融、银行信贷与商户小额贷款融资——基于风险等级匹配视角[J].中国管理科学,2017,25(9):28-36.
- [2]吾兰.商户贷款的信用风险评价实证研究[D].辽宁:大连理工大学,2014.
- [3]雷晓敏,宋家顺,徐峰岩.中小企业家信用评价模型指标构建分析[J].财经问题研究,2008,(1):119-124.
- [4]孟斌 迟国泰 龚玲玲.商户小额贷款信用评价模型[J].技术经济,2014,33(12).
- [5]刘子琪.互联网金融视角下小微企业的融资策略研究[J].商丘职业技术学院学报,2019,18(2):51-54.
- [6]田秀娟.我国农村中小企业融资渠道选择的实证研究[J].金融研究,2009,(7):146-160.
- [7]梁琦,于津平,吴崇.民营企业生命周期融资规律探析[J].南京社会科学,2005,(5):86-92.  
DOI:10.3969/j.issn.1001-8263.2005.05.014.
- [8]中国工商银行. 关于印发《中国工商银行小企业法人客户信用等级评定办法》的通知[R]. 中国工商银行, 工银发[2005]78号, 2005: 5-7, 13-21.
- [9]中国农业银行. 中国农业银行“三农”小企业客户信用等级评定管理办法[R]. 中国农业银行, 2010.
- [10]Niklis D, Doumpos M, Zopounidis C. Combining market and accounting-based models for credit scoring using a classification scheme based on support vector machines[J]. Applied Mathematics & Computation, 2014, 234(C):69-81.
- [11]Harris T. Credit scoring using the clustered support vector machine[J]. Expert Systems with Applications, 2015, 42(2):741-750.
- [12]Horata P, Chiewchanwattana S, Sunat K. Robust extreme learning machine[J]. Neurocomputing, 2013, 102(2):31-44.
- [13]张璐昱,王永茂.电商大数据金融下小微企业融资模式研究——基于蚂蚁金服与京东金融的比较[J].西南金融,2018, No.444(07):54-60.
- [14]冯雪莲,王培.互联网金融与小微企业融资的协同发展[J].时代金融(下旬),2018, 000(002):109,112.

- [15]张苗苗. 我国小微企业融资困境与对策研究[D]. 2019.
- [16]王家辉. 互联网金融和小微企业融资模式的创新思考[J]. 中国乡镇企业会计, 2019, 000(008):43-44.
- [17]余青. 缓解小微企业融资困境的支持政策研究[D]. 2019.
- [18]陈凌白. 辽宁地区小微企业融资问题调查分析[C]// 第十六届沈阳科学学术年会论文集(经管社科). 2019.
- [19]王文海. 小微企业融资困境与对策研究[J]. 中小企业管理与科技, 2018.
- [20]陈隆, 闫真宇, 邓舒仁. 对当前小微企业融资问题的若干思考[J]. 浙江金融, 2018, 000(001):17-23.
- [21]李素红, 方洁, 蔡韡. 互联网金融背景下科技型小微企业融资平台建设[J]. 科技管理研究, 2018, 038(011):195-201.
- [22]刘晓静. 小微企业融资模式探析[J]. 现代营销, 2018, 000(003):20-21.
- [23]韩曦莹. 我国小微企业融资难问题研究[J]. 商场现代化, 2018, 000(004):67-68.
- [24]常磊. 政府补助与小微企业融资困境——基于微观数据的分析[J]. 新金融, 2019, 000(006):58-64.
- [25]王娜. 基于互联网金融的小微企业融资研究[J]. 现代经济信息, 2018(02):329.
- [26]Xiong W , Yu J , Fan L H , et al. Enterprise Quality Credit Rating System Building[J]. Applied Mechanics & Materials, 2013, 281:677-681.
- [27]PAN Zhen-yuan. Discussion on credit rating system for small enterprise[J]. information technology, 2012.
- [28]Shankar S . The Role of Credit Rating Agencies in Addressing Gaps in Micro and Small Enterprise Financing: The Case of India[J]. Adbi Working Papers, 2019.
- [29]Johnny Djauhary, Prof. Dr. Jogiyanto HM., MBA. ANALISIS CREDIT RISK RATING (CRR) PADA SMALL MEDIUM ENTERPRISE (SME) YANG BERDAMPAK PADA NON PERFORMING LOAN (NPL) DI PT BANK ABC, TBK.[J]. universitas gadjah mada, 2013.
- [30]GRAY, RICHARD. It's important to pay attention to your business' credit rating, too[J]. enterprise/salt lake city, 2014.

- [31]谭新莹, 姜林枫. 互联网+时代小微企业信用风险评估体系的研究[J]. 时代金融, 2018, No.700(18):96+98.
- [32]周良增. 商业银行小微企业信用风险管理研究[J]. 时代金融, 2018, 712(30):91+96.
- [33]夏晗. 基于支持向量机回归集成的小微企业信用风险度评估模型研究[J]. 征信, 2019, 037(004):21-27.
- [34]李伟, 赵娜, 潘奎新. 客户信用评价指标体系研究——基于融资担保公司视角[J]. 山东农业工程学院学报, 2018, 35(11):81-86.
- [35]钟琳. 金融机构如何解除小微企业融资困局[J]. 今日财富:中国知识产权, 2018, 000(010):P.34-34.
- [36]朱文超. 小微企业贷款财务风险评估研究[J]. 新商务周刊, 2018, 000(015):99.
- [37]张建红. 大数据时代云会计下的小微企业融资策略探讨[J]. 经贸实践, 2018, 000(022):166.
- [38]周超. 小微企业信贷问题研究:基于行为金融和商业银行异质性视角[J]. 金融理论与实践, 2018, 000(003):27-32.