

金融管理 研究

第8辑 | VOL.8

FINANCE
AND MANAGEMENT
RESEARCH

王周伟 主编



上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

金融管理研究

第8辑 | VOL.8

FINANCE
AND MANAGEMENT
RESEARCH

王周伟 主编



上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

内容提要

本书为研究金融管理的文集,包括 13 篇文章,内容涵盖网贷信用风险管理、平台交易、保险公司治理、打车手机软件、企业商业信用、企业社会责任、商业银行风险等论题。
本书适合金融、经济及管理类专业研究人员参考阅读。

图书在版编目(CIP)数据

金融管理研究.第 8 辑 / 王周伟主编. —上海:上海交通大学出版社,2016
ISBN 978-7-313-16116-1

I. ①金… II. ①王… III. ①金融管理-研究 IV. ①F830.2
中国版本图书馆 CIP 数据核字(2016) 第 266488 号

金融管理研究 (第 8 辑)

主 编: 王周伟	
出版发行: 上海交通大学出版社	地 址: 上海市番禺路 951 号
邮政编码: 200030	电 话: 021-64071208
出 版 人: 郑益慧	
印 刷: 上海盛通时代印刷有限公司	经 销: 全国新华书店
开 本: 710mm×1000mm 1/16	印 张: 12.5
字 数: 85 千字	
版 次: 2016 年 12 月第 1 版	印 次: 2016 年 12 月第 1 次印刷
书 号: ISBN 978-7-313-16116-1/F	
定 价: 45.00 元	

版权所有 侵权必究
告 读 者: 如发现本书有印装质量问题请与印刷厂质量科联系
联系电话: 021-61453770

编 委 会

学术委员会(以姓氏笔画为序)

王明进	张顺明	张寄洲
李百炼(美国)	李宝仁	周忠国(美国)
图马哲(法国)	岳荣先	庞 涛(美国)
林俊国	郭军义	

主 编

王周伟

执行主编

卓德保 崔光灿 赵金实

专业编辑(以姓氏笔画为序)

刘江会	张 震	李 刚	郑晓涛
姚亚伟	崔百胜	黄 静	敬志勇

前言

《金融管理研究》是经济、金融和管理理论研究与实践创新领域的中英文系列专业学术丛书。由上海师范大学商学院主办,每年出版2~4辑,每期20万字左右。本丛书秉持学术性、创新性与前瞻性,努力为经济与管理研究搭建一个具有国内外领先水平的学术交流平台。

《金融管理研究》丛书主要收录经济、金融与管理领域内,原创性强、高质量的规范、实证的学术论文,也可以是案例分析、文献综述、调研报告等。

第8辑包含经济管理类学术研究论文共13篇。具体包括:我国大学生P2P网贷信用风险度量研究、我国保险公司治理有效性研究、人口流动、户籍分割与区域经济结构、考虑流动性风险的开放式基金业绩评价、基于DEA模型的商业银行互联网转型绩效研究、打车类手机软件的现状和发展趋势分析、基于第三方投资者教育的“证券投资分析”课程教学改革初探、新股收益的影响因素研究、市场地位对我国企业商业信用的影响研究、供给侧改革下我国普惠金融深化改革的路径研究、信贷扩张与商业银行风险承担、企业社会责任对财务绩效的影响、基于重组信息的股价异动研究。

总而言之,本丛书集合了大量探索性、提炼性和创造性的科研成果,希望对中国经济与管理学界学术知识与实践经验的交流与发展有所裨益。

目 录

我国大学生 P2P 网贷信用风险度量研究	
——基于“速溶 360”平台交易数据	/陈 燕 夏人青 傅 毅/ 1
我国保险公司治理有效性研究	
——基于盈利能力视角的实证分析	/郝 臣 杨冬雪 崔光耀 秦晓天/ 23
人口流动、户籍分割与区域经济结构	/汤海燕 崔光灿 姜 巧/ 52
考虑流动性风险的开放式基金业绩评价	
——基于数据包络模型	/付淑换 何暑子/ 67
基于 DEA 模型的商业银行互联网转型绩效研究	/谭黎阳 母亚萍/ 83
打车类手机软件的现状和发展趋势分析	/龚秀芳 杜慧莲/ 99
基于第三方投资者教育的“证券投资分析”	
课程教学改革初探	/张 毅 伍 燕/ 110
新股收益的影响因素研究	
——基于新时期新股“一字涨停版”数量的研究	/杨竹清 张超林/ 119
市场地位对我国企业商业信用的影响研究	/李 晶/ 134
供给侧改革下我国普惠金融深化改革的路径研究	/陆岷峰 杨 亮/ 144
信贷扩张与商业银行风险承担	
——基于后危机时期中国银行业的实证研究(2010—2014)	/刘 伟 朱新颜/ 156
企业社会责任对财务绩效的影响:政治关联的作用	/彭 屹 刘春济/ 171
基于重组信息的股价异动研究	/孙志宾 殷玲华 唐奕蓉/ 182
《金融管理研究》(中英文)约稿启事	/ 189

我国大学生 P2P 网贷信用风险度量研究^①

——基于“速溶 360”平台交易数据

陈 燕^② 夏人青^③ 傅 毅^④

摘 要:随着 P2P 网贷市场的不断细分,大学校园已成为众多网贷平台争夺的重要市场。本文以大学生 P2P 网贷信用风险为研究对象,采集了“速溶 360”平台的全量业务数据,通过数据重构,还原了每笔网贷发生时的业务场景数据。为了能够设计出度量大学生 P2P 网贷信用风险的评分卡,本文首先对数据的各个变量进行分箱,其次运用 Logistic 回归建立了大学生 P2P 网贷信用风险度量模型,通过对模型结果中的各个参数进行变换,最终得到了变量各分箱的信用分值,即评分卡。为检验评分卡的有效性,本文以评分卡规则对平台业务数据进行评分,将分值进行分段后我们发现,网贷业务的逾期率随着分值的提升而显著下降。因此,本文的研究成果能有效地度量大学生 P2P 网贷的信用风险,为该细分市场资金出借人进行决策提供参考依据。同时,运用本文评分卡开展网贷评分,也有助于网贷平台对具体业务进行分级管理,提高平台的整体信用管理水平。

关键词:大学生信用;P2P 网贷;信用风险

① 国家自然科学基金项目(71471117),上海市教委科研创新重点项目(13ZZ107),上海师范大学自科项目(SK201506),上海师范大学人文社科项目(A-3131-12-002050)。

② 陈燕(1981—),女,上海师范大学讲师,主要从事高等教育、大学生互联网行为研究。

③ 夏人青(1954—),女,上海师范大学高等教育研究所研究员、博士生导师,主要从事高等教育、比较高等教育研究。

④ 傅毅(1981—),男,上海师范大学商学院讲师,主要从事信用管理、互联网金融的相关研究。

一、问题提出

P2P 网贷是指借助专业网络平台帮助借贷双方确立借贷关系并完成相关交易手续的网络借贷,其中 P2P 是英文 peer to peer 的缩写,意即“个人对个人”。

P2P 网贷起源于欧美,2007 年首次进入中国。截至 2015 年 7 月 31 日,我国 P2P 网贷平台数量已达 2 806 家。仅 2015 年上半年,全国 P2P 网贷成交额近 3 000 亿元(数据来源:网贷之家 <http://www.wangdaizhijia.com>)。在经历了过去数年的高速增长之后,如今 P2P 网贷市场已开始细分,在众多的细分市场中,大学生 P2P 网贷异军突起,发展迅猛。

大学生 P2P 网贷的概念最早来源于一家名为 Social Finance(以下简称“SoFi”)的美国学生贷款公司。该公司成立于 2011 年,其主要业务为学生贷款再融资,即在 SoFi 平台上申请较低利率的新贷款,以偿还原有较高利率的学生贷款。SoFi 成立四年,其估值已高达 35 亿美元。因为国情不同,目前中国 P2P 平台的利率相较于政府主导的助学贷款利率毫无竞争优势,因此为学生提供的借款,用途并不局限于助学贷款,也包括创业、购物、旅游、教育培训等。根据“网贷之家”统计的 2015 年在校大学生 3 000 万左右推算,国内大学生网贷市场整体规模高达千亿级。目前我国专营大学生 P2P 网贷业务的 P2P 平台有速溶 360(<http://www.surong360.com>)、学贷网和利融网,其中速溶 360 在 2015 年初获得了起点创投基金的数千万元 A 轮风投资金。此外,业务涵盖大学生 P2P 网贷的综合性 P2P 平台有积木盒子、诺诺镑客、拍拍贷等。其中,诺诺镑客凭借 2013 年推出的名校贷,在 2014 年累计成交额高达 11.06 亿元,同比增长 163%,注册用户达 21 万人,人均交易达 7 500 元。

大学生相较于其他网贷借款人,有着身份易确认(通过学信网核实学籍 www.chsi.com.cn)、违约成本高、家庭是其潜在担保人等先天信用优势。速溶 360 在其官网上发布该平台从 2013 年 10 月至 2015 年 2 月的坏账率,其坏账率仅为 1.5%^①。即便如此,投资大学生网贷也并非高枕无忧,其主要风险在于:

^① 数据来源:http://newspaper.jfdaily.com/jfrb/html/2015-02/03/content_64045.htm。

网贷平台履约风险。由于涉足大学生网贷的平台日益增多,平台间恶性竞争日益激烈,获客成本不断提升,平台盈利渠道单一,从而影响了平台的正常运营。一旦平台倒闭,其信息中介的职能将无法履行,出借人也将面临借贷关系无法确认的窘境。此外,部分平台违规增信,私建资金池,非法集资,卷款逃跑,给出借人带来了巨大的损失。

网贷信息非对称风险。一方面,由于出借人只能通过浏览网贷平台获得贷款相关信息,而平台的信息发布又缺乏相关部门的监管,使得部分贷款信息无法获得及核实。另一方面,受各种业务需求的影响,大多数平台并不公开自身的坏账率和逾期率,即便公开也鲜见其统计方法和统计口径。因此,在 P2P 网贷业务中,出借人往往是信息获得的弱势群体,这也将直接影响出借人对贷款风险的判断。

大学生信用意识风险。由于受社会环境、家庭环境和学校教育等影响,我国大学生的信用意识水平参差不齐,且大学生整体信用意识不容乐观。2015 年 8 月 28 日,北京宜信致诚信用评估有限公司与中国人民大学信用管理研究中心联合发布的《全国大学生信用认知调研报告》显示,超过八成大学生没听说过或不了解个人信用报告,知道信用报告中逾期记录保存几年的大学生只有 12%,甚至有超三成大学生不知道个人信用报告里的逾期记录会影响未来的金融生活。针对使用过贷款的大学生的还款情况的调查则显示,有四成大学生曾逾期还款,而使用信用卡的大学生里,有 2% 的大学生经常不按时还款。

因此,如何针对大学生群体特征和 P2P 网贷行业特点,基于平台公布的数据,研究和构建大学生 P2P 网贷信用风险度量模型,为大学生 P2P 网贷平台完善风控体系,为出借人的借款项目选择提供决策依据,并进而有效管理大学生 P2P 网贷细分行业风险,已经成为业界和学术界共同关心的重要问题。

与本文相关的前期研究工作大致可分为两类:第一类是 P2P 网贷信用风险与出借人行为的实证研究。Klaft(2008)对 Prosper 平台上的数据进行分析,该研究的结论认为信用评级对贷款利率的影响远大于贷款人债务收入比,是最大的影响因素,而其他信息(如银行账户信息、是否自有房产等)对借款利率则几乎不造成影响。Puro(2010)站在借款人的角度,给出了借款人决策辅助策略,从而帮助借款人决策初始借贷利率和借贷金额。Lin 等人(2012)研究了

网络社交圈对 P2P 网络信贷的影响,分别用 Probit 模型和 Heckman 两步法模型分析了借款成功概率和借款利率与社交网络圈的关系,其结论表明网络社交圈可以增大借贷成功概率,降低借款利率。Chen 和 Han(2012)对中美两国的 P2P 借贷现状进行了比较分析,并发现两国的 P2P 资金借出方所关注的信用信息是不同的。Lee(2012)基于韩国最大 P2P 借贷平台的数据,研究发现该平台上的出借人存在着显著的羊群行为,并且随着投标过程的进行,这种羊群效应呈现出边际递减效应。曾江洪与杨帅(2014)以拍拍贷借贷平台数据为样本,研究发现平台资金出借人的投资决策易受前期出借人的影响,存在显著的羊群行为。王会娟与廖理(2014)通过采集“人人贷”数据,验证了平台的信用认证机制能够有效地揭示信用风险,同时也指出了平台评级指标的单一性使得其风险揭示作用具有局限性。另一类是对于大学生贷款信用问题的相关研究。张强等(2009)基于四所高校的 172 个贷款样本,应用改进模糊算法进行了信用评级,探讨了对学生信用评级有重大影响的因子。吴开俊和陈宇红(2010)以广东省四所高校为研究样本,分析了国家助学贷款违约风险的影响因素,并研究了违约风险的控制对策。张凤奕(2014)基于大学生助学贷款的信用问题,研究了大学生助学贷款信用问题的影响因素和后续危害,并提出了对策建议。Hillman(2014)对于学生贷款违约问题进行了深入研究。通过对于大一新生的调查发现,累计贷款金额与违约率是非线性关系,学校类型、家庭收入、就业情况以及学业完成度对于贷款违约率有较强的影响。Karen L. Webber 和 Sharon L. Rogers(2014)针对 1 399 个四年制非营利美国教学机构样本,分析了不同机构特征对于 2010 年联邦学生贷款违约率的影响。

本文在先前工作的基础之上,通过自行开发代码与商用网络爬虫软件相结合,在速溶 360 网络平台上共采集了 63 071 条网贷业务数据和 315 545 条月还款数据。通过数据分箱和 WOE(Weight of Evidence)替换,本文以网贷按时还款与否作为被解释变量,运用 Logistic 回归建立了大学生 P2P 网贷信用风险度量模型,并设计了大学生 P2P 网贷信用风险评分卡,通过评分测试,验证了评分卡的有效性,最后依据本文模型结果,给出了关于大学生 P2P 网贷市场发展的相关建议。

二、理论分析

1. 信用风险度量的理论框架

由于受网络技术、平台数据库管理、借贷双方投机主义等因素的影响,在 P2P 网络平台的借贷过程中,借款方与出借方获取的业务信息并不完全一致,即信息不对称。在这种信息不对称的市场环境下,逆向选择问题应运而生。通常情况下,借款人为了获得贷款,会按照 P2P 平台的要求,将相关信息自行上传到平台。随后,P2P 平台会对上传数据和相关历史数据进行审核并处理,并将处理过的数据在平台上发布。由于数据都已脱敏处理(屏蔽了姓名、电话、身份证等识别信息),因此出借人只能通过浏览网站来获取相关信息,而对信息的真实性无从判断,相反,借款人则对借贷的真实信息十分了解。因此出借人只能根据借款的平均信用水平来为自己做投资决策。很显然,这样会使得信用良好的借款人因不愿支付高利率而退出市场,相反只有信用较差的借款人才会进入市场,并最终使得 P2P 网贷市场的整体信用水平下降。借贷信息不对称产生的另一个问题就是道德风险。在 P2P 网贷中最常见的道德风险就是贷款用途变更。虽然在 P2P 网贷申请过程中,借款人需要描述借款用途,但是一旦网贷成功后,其贷款用途几乎不受监管,因此借款人在各种利益驱使下,很可能在出借人不知情的情况下,变更借款用途,并最终导致还款困难。逆向选择和道德风险都会使得 P2P 网贷的参与方利益受损,降低整个 P2P 网贷市场的效率。

信用风险度量可以在很大程度上缓解上述问题。出借人可以参考 P2P 网贷的信用风险度量值(评级、评分等)决定自身的投资策略,有效地避免逆向选择。对于借款人而言,信用风险度量则能综合当前借款、历史借还款等各项信息,有效地反映当前借款的信用风险水平,并直接影响借款成功与否,因此能在一定程度上控制借款人的道德风险。

2. 信用风险度量指标

目前世界范围内最为成熟的信用风险度量业务当属美国的个人信用评级,该业务主要由 Equifax、Experian 和 TransUnion 三家公司提供,它们都使用 FICO 信用评分来量化个人信用质量和风险。FICO 信用评分体系遵循 5C 原

则,即道德品质(Character)、还款能力(Capacity)、资本实力(Capital)、担保(Collateral)和经营环境条件(Condition)五个方面的指标。FICO 信用评分选取的这些指标全面地反映了影响借款人信用情况的各种因素。与欧美国家的征信业务相比,我国 P2P 网贷征信尚处于起步阶段,但发展迅速。2013 年 7 月,中国人民银行授权上海资信搭建的网络金融征信系统(NFCS)正式上线,要求将全国 P2P 企业纳入该系统,此举也是央行征信延伸至非银行领域的前瞻性举措,目前完成数据接入的 P2P 平台已经超过 400 家。然而当前 NFCS 主要是解决跨平台的信息共享问题,至于如何确定信用风险度量指标,有效地开展 P2P 网贷各细分领域的征信业务,仍然需要结合具体业务特征来建立其风险度量体系。本文参考 5C 原则,结合大学生 P2P 网贷业务特征,分别从个人基础信息、当前网贷信息、历史网贷信息三个维度采集不同指标来度量网贷业务信用风险。那么这三个维度的指标是否对信用风险产生显著的影响,本文将在后面的实证研究中予以检验。

三、研究设计

1. 数据采集与变量定义

1) 业务数据的采集

本文研究的目的是通过量化分析大学生 P2P 网贷信用风险的影响因素,建立该类网贷业务的信用风险度量模型。本文首先通过 R 语言编写数据采集器与八爪鱼数据采集器的交叉使用,采集了速溶 360 平台上自 2013 年 10 月至 2015 年 6 月的全量网贷业务数据(业务编号 No.1~No.77000),剔除审核未通过以及无法打开对应编号网页的网贷业务,总计获得 63 071 条网贷业务数据(数据均不涉及个人隐私),删除部分变量缺失的数据 1 953 条,剩余 61 118 条。

为了更为客观和精确地度量大学生网贷的信用风险,避免由于因还款未到期而低估了逾期网贷的比例,我们进一步筛选出所有应还款日期早于数据采集日期的业务,共计 46 598 条,并随机抽取其中 80% 的数据作为训练数据来建立 Logistic 模型,20% 的数据作为测试数据,用于后期模型判别准确性的检验数据。

2) 业务历史场景的重构

由于数据的采集时间为 2015 年 7 月,因此我们无法采集到每笔业务发生时的“累积还款金额”等动态变量。为了使得模型的解释变量结构更为完整,我们进一步采集了该平台网贷业务对应的全量还款信息数据,共计 315 545 条。通过全量还款信息数据,我们统计了每笔业务发生前借款人的“累积还款金额”“累积逾期金额”等动态变量,还原了每笔网贷业务发生时的历史动态场景数据。

3) 变量定义

受实际坏账率和网站数据公开程度的影响,本文将还款逾期作为信用风险识别指标,其中“还款逾期”是指借款人超过约定的最后还款期限,且未及时足额归还约定的借款本息。需要说明的是,逾期未必会最终形成坏账,但是逾期依然属于违约行为,会导致出借人无法按期收回投资和收益,造成相应的损失。

通过统计还款情况,对每笔网贷业务增加一个“还款情况”变量(变量编号:Y),若该笔网贷业务的所有还款中未出现“逾期”,则该变量赋值为 1,否则赋值为 0。各自变量及其定义见附表 1。

2. 研究模型构建

本文采用二分类 Logistic 回归对前期采集的大学生网贷业务数据进行分析,来研究多个自变量($X_1, X_2, \dots, X_{17}$)对因变量(Y:还款情况)的影响。

二分类 Logistic 回归模型可以表示为

$$P = \frac{e^{(C + \beta_1 X_1 + \dots + \beta_m X_m)}}{1 + e^{(C + \beta_1 X_1 + \dots + \beta_m X_m)}} \quad (1)$$

或者
$$\ln\left(\frac{P}{1-P}\right) = C + \beta_1 X_1 + \dots + \beta_m X_m \quad (2)$$

式中, C 是与各自变量 X_i 无关的常数项; $\beta_1, \beta_2, \dots, \beta_m$ 是回归系数,表示各自变量对因变量的贡献量。

由于本文 Logistic 回归模型自变量的选择是从个人基础信息(PI)、当前网贷信息(BI)和历史网贷信息(HI)三个维度进行选取和采集,因此回归模型也可表示如下:

$$Z = \ln\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = C + V_1 \cdot \text{PI} + V_2 \cdot \text{BI} + V_3 \cdot \text{HI} + \epsilon \quad (3)$$

式中, C 为常数项; V_1, V_2, V_3 表示回归系数向量; ϵ 表示随机误差。

四、实证结果与分析

1. 描述性统计

本文对所有变量按照连续性分为分类变量和连续变量,并分别给出了频数统计和描述性统计(见附表2和附表3)。通过观察附表2,我们不难发现,不同性别的借款人在参与网络借贷的人次上有着较大的差异,男生的网贷人次要远高于女生,占比达76%,这主要是因为在大学期间男生往往接触网络的时间更长,同时也更愿意尝试网络中的新鲜事物。从学历层面来看,本科生毫无疑问是借贷的主力军,占比达71.5%,而硕士和博士的借贷人次较少,两者之和未超过3%。此外,东部地区的借款人次最多,占比达40.3%,东北部地区借贷人次最少,占比达8.7%。从借款用途类型来看,购物是最主要的借贷需求原因,占比达27.5%。创业排在其后,占比达16.7%。

附表3展示了连续变量的描述性统计结果。其中,项目金额的平均数为961.29元,标准差为458.92元,这说明大学生P2P网贷几乎都在3000元以内,金额较小。同时我们也发现网贷的平均收益率(即利率)为23.28%,这是一个比较高的负担,对于没有稳定收入的大学生而言,容易造成还款压力。此外,累积已还次数为17.66次,由于平均周期为5.25个月,这意味着相当比例的借款人存在着2次以上的借款行为。累积逾期次数的平均数为1.9次,这也印证了《全国大学生信用认知调研报告》的调研结论——大学生的信用意识不容乐观。

2. 变量分箱的WOE转换与IV计算

本文参考Nebojsa Nikolic等(2013)的工作,应用信息熵方法,对变量进行WOE变换,并计算其IV。这里WOE表示证据权重(Weight of Evidence),是根据按期还款占比和逾期还款占比计算得到的,与逾期比例变动方向相反。IV为信息价值(Information Value),由WOE计算得到,用以描述该变量的判别能力,信息价值越高,说明该变量的判别能力越强。WOE与IV的计算公式如下:

$$WOE_{i,j}=\ln(\frac{G_{i,j}}{G})-\ln(\frac{B_{i,j}}{B}) \tag{4}$$

$$IV_i=\sum_j^M[\frac{G_{i,j}}{G}-\frac{B_{i,j}}{B}]\times WOE_{i,j}, i\in\{1,2,\cdots,N\}, j\in\{1,2,\cdots,M_i\}$$

这里 $WOE_{i,j}$ 表示第 i 个变量中第 j 个分箱的 WOE 值, IV_i 表示第 i 个变量的 IV 值。 $G_{i,j}$ 为第 i 个变量中第 j 个分箱中的按期还款业务数, $B_{i,j}$ 为对应的逾期还款业务数。 G 为总按期还款业务数, B 为总逾期还款业务数, N 为变量个数, M_i 为第 i 个变量的分箱数。

计算 WOE 值,就是找出每个变量合理的区间划分,即分箱。下文将对所有变量进行分箱操作,其中,性别、大学类型、借款类型和担保这 4 个变量,我们按照网贷平台原有分类进行分箱并 WOE 赋值。对于学历变量,本文按照专、本、研进行分类。对于户籍省份,为减少分箱数,参考国务院经济区划分规则,我们将其分为东部、中部、西部和东北部四个分箱,并对其进行 WOE 赋值。

其次,对于连续变量,本文则直接运用 IBM SPSS Statistics 22 中的最优离散化功能进行分箱,并运用公式(4)对各个分箱进行 WOE 赋值,具体见附表 4。

3. Logistic 回归结果与信用风险识别检验

本文以 IBM SPSS Statistics 22 为数据分析平台,调用其二元 Logistic 模型对数据进行建模,其中因变量选取还款情况,即网贷业务逾期与否。自变量为 X_1 至 X_{17} ,且自变量都已 WOE 赋值,方法选择向前逐步回归,分割值设定为 0.81,显著性水平为 5%,模型输出如表 1~表 3 所示。

表 1 模型系数的 Omnibus 检验

步骤 14	卡方	df	P-val
步骤	10.336	1	0.001
区块	26 783.081	12	0.000
模型	26 783.081	12	0.000

表2 模型摘要

	-2 Log Likelihood	Cox & Snell R ²	Nagelkerke R ²
步骤 14	14 815.728	0.512	0.762

表3 模型参数

变量编号	变量名	参数	S.E.	Wald	df	P-val	Exp(B)
X ₁	性别	0.665	0.190	12.231	1	0.000	1.944
X ₂	学历	0.631	0.046	189.206	1	0.000	1.880
X ₃	大学类型	0.409	0.091	20.238	1	0.000	1.506
X ₄	户籍省份	0.564	0.130	18.819	1	0.000	1.757
X ₅	借款用途	0.225	0.053	18.140	1	0.000	1.253
X ₆	项目金额	0.966	0.059	266.671	1	0.000	2.628
X ₈	收益率	0.704	0.028	650.354	1	0.000	2.023
X ₉	周期	0.802	0.106	57.326	1	0.000	2.229
X ₁₂	累积还款金额	1.233	0.030	1 697.045	1	0.000	3.432
X ₁₃	累积逾期金额	0.091	0.028	10.394	1	0.001	1.096
X ₁₅	累积逾期次数	0.741	0.021	1 249.055	1	0.000	2.098
X ₁₆	累积逾借金额比	0.356	0.023	229.978	1	0.000	1.427
	常数	1.045	0.027	1 500.040	1	0.000	2.845

表1显示了 Logistic 回归方程显著性检验的总体情况,由于本文显著性水平为5%,因此 P 值小于显著性水平,拒绝原假设,并认为解释变量的全体与 $\text{Logit}(P)$ 之间线性关系显著。表2显示两个 R^2 都大于0.5,其中 Nagelkerke R^2 达到0.764,说明本文的 Logistic 模型有着较好的拟合优度。表3给出了模型中的所有参数以及参数检验,所有变量的 P -val 都小于5%,通过检验。将这些参数代入式(2),则本文的信用风险度量模型可表示如下:

$$\ln\left(\frac{P}{1-P}\right)=0.665\times X_1+0.631\times X_2+0.409\times X_3+0.564\times X_4+0.225\times X_5+0.966\times X_6+0.704\times X_8+0.802\times X_9+1.233\times X_{12}+0.091\times X_{13}$$

$$+0.741 \times X_{15} + 0.356 \times X_{16} - 1.045 \quad (4)$$

其中 P 表示大学生 P2P 网贷业务正常还款概率,即网贷信用风险水平。 P 值越小意味着该笔网贷的信用风险越大。需要说明的是,模型中所有参数都是正数,这是由于所有的变量在数据处理中都已 WOE 赋值,该值与优势比保持着正方向。收益、总还款本息、累积已还次数、累积逾期次数比 4 个变量由于不显著被移出。

通过观察式(4)和附表 4,容易发现,在其他因素确定的情形下,男生的信用风险要大于女生,博士和硕士研究生的网贷信用水平优于其他学历的高校学生,重点高校学生的网贷信用水平优于其他高校。此外,大学生的户籍所在地区也是一个重要影响因素,东部地区大学生的网贷信用水平最高。

在网贷业务信息变量中,项目金额、收益率与信用水平是反方向的,其变量取值越大就意味着还款压力越大,自然就降低了该笔业务的信用水平。周期、月还本息对于信用风险的影响不是单调的,其对应的信用风险呈现出中间高、两边低的态势。值得注意的是,通过观察借款用途变量,我们发现创业网贷的信用风险是最大的。虽然我们不能完全排除借款人变更借款用途等潜在可能,但这一实证结果提醒我们,在鼓励大学生创业的同时,应进一步完善创业周边的金融生态,拓展大学生创业的融资渠道,毕竟通常大学生在创业阶段的盈利水平未必能达到 20% 左右的网贷利率。

在网贷历史信息变量中,所有变量的分箱水平与信用水平都是单调的,即高水平的累积还款金额和低水平的累积逾期金额对应着较低的信用风险,而高水平的累积逾期次数和累积逾借金额比则对应着较高的信用风险。这些指标反映了历史信息与当前网贷的信用水平是显著相关的,是平台和出借人对网贷信用风险度量的重要指标。

为了进一步检验本文建立的 Logistic 模型对于信用风险识别的有效性,我们分别将式(4)应用于训练数据和测试数据。通过计算和整理,本文模型对于训练和测试数据的预测成功率都接近 90%,该结果说明本文模型在实际应用中对于信用风险的识别有很高的精确度(见表 4)。