

基于 LS-SVM 的小微企业信用评估研究

肖斌卿^{1,2}, 柏巍¹, 姚瑶³, 李心丹¹

(1. 南京大学 工程管理学院, 江苏 南京 210093; 2. 纽约大学 Stern 商学院, 纽约 10012;
3. 江苏中诚信信用管理有限公司 评级部, 江苏 南京 210019)

[摘要] 构建并优化小微企业信用评估技术已经成为商业银行开展小微业务必然选择。基于小微企业内在特征, 设计以小微企业现金流信息为违约触发机制的小微企业信用评估指标体系, 构建最小二乘支持向量机模型 (LS-SVM), 运用某国有控股银行的小微企业贷款微观数据证实该模型能够相对提高预测精确度和稳定性。应用 LS-SVM 构建的小微企业信用评估指标体系以及评估模型有助于提升银行对小微企业的了解程度和小微企业风险管理能力, 减轻信息不对称, 在一定程度上能够化解供给型信贷配给导致的小微企业融资难问题。

[关键词] 小微企业; 信用评估; 企业信用评估; 现金流信息; LS-SVM; 小微金融; 商业银行; 信用配给; 信用评分
[中图分类号] F275 **[文献标识码]** A **[文章编号]** 1004-4833(2016)06-0102-10

一、引言

企业信用评估一直以来都是财务学和金融风险管理领域的热点问题。相对于研究以及业界实践已经比较成熟的大中型企业信用评估, 中小企业信用评估一致没有得到很好地发展, 而中小企业中的“小微企业”信用评估在研究上更是没有得到足够的重视, 小微企业信用评估理论研究成果还不多见。随着对小微金融盈利(小微金融可以产生可观的现金流和更低的风险)的重新认识^[1], 商业银行开始进军小微金融并开创了很多信用评估技术。在这些实践中, 很多商业银行已经将信用评分用于小微企业信用评价中, 包括美国富国银行、泰国泰华农民银行以及德国 IPC 公司等。国内业界在小微企业信用评价方面的实践起步较晚, 但也纷纷进行了相应的探索。例如民生银行以“信贷工厂”的建设为出发点, 开发出“小微企业评分卡”模式和“客户行为评分”模式; 包商银行开发了符合本土化要求的小微企业信贷业务打分卡系统。国内外商业银行的小微金融实践富有特色, 这些商业银行将信用评估应用于小微企业信贷业务, 高成本、高风险和信息不对称问题得到了明显的改观, 小微企业“融资难”和“融资贵”的压力在一定程度得到了缓解, 但从全局来看, 我国小微企业的信用评估技术目前正处于起步阶段, 构建有效的小微企业信用评估问题一直没有得到很好地解决。

构建小微企业信用评估模型对于商业银行开展小微企业信贷业务具有重要意义和价值。传统信贷市场上, 小微客户一直存在典型的信贷配给^[2]。除了价格型信贷配给, 还包括供给型和需求型信贷配给^[3]。供给型信贷配给是指由于利率限制和信息不对称, 银行改变非价格部分的贷款规模, 导致市场出清; 需求型信贷配给是缘于需求方自身原因的交易成本配给、风险配给以及社会资本配给等。构建小微企业信用评估模型有助于增加商业银行对企业信用状况的了解把握, 从而提升管理小微信贷风险能力, 应对供给型信贷配给问题, 在一定程度上化解小微企业融资难的困境。

[收稿日期] 2015-12-01

[基金项目] 国家自然科学基金(71271109; 71671083; 71203144); 教育部人文社会科学研究青年项目(13YJC790174)

[作者简介] 肖斌卿(1979—), 福建南靖人, 南京大学工程管理学院副教授, 纽约大学 Stern 商学院访问学者(2014—2015年), 博士, 从事金融工程与金融管理研究; 柏巍(1988—), 江苏盐城人, 南京大学工程管理学院博士研究生, 从事金融工程研究; 姚瑶(1990—), 江苏南京人, 江苏中诚信信用管理有限公司职工, 硕士, 从事金融工程研究; 李心丹(1966—), 湖南浏阳人, 南京大学工程管理学院教授, 博士, 从事金融工程与金融管理研究。

本文基于小微企业特征并结合理论与实践概括了影响小微企业信用情况的因素,架构了小微企业信用评估指标体系,利用某国有商业银行省分行小微企业信贷业务数据,构建了最小二乘支持向量机模型,提高了小微企业评估模型的预测精确度和稳定性。接下来,本文分别将包括现金流信息指标和不包括现金流信息指标的小微企业信用评估指标作为模型输入变量,进行对比分析,以验证现金流信息指标在小微企业信用评估中的重要作用。

二、文献回顾与评述

信用评分在消费者贷款和住房抵押贷款领域有着广泛和悠久的历史的应用,但是直到20世纪中期才被引入小企业贷款。学者们认为其中一个原因是小微企业贷款的多样化,另一个原因是小微企业提供的财务报表标准度不够且报表虚假的可能性较高。随后有研究发现,小微企业业主的个人信用行为与企业的信用表现有很强的相关性,这反映了小微企业与企业主体戚与共的关系。这些研究极大地激发了学术界和业界研究小微企业信用评估的热情^[4]。

国外学者们主要从小微企业信用评价指标完善以及构建评价模型两个视角展开研究。在评价指标体系完善上,学者们探索如何引入更加有效的评价指标。例如Minnis探讨了税收等指标在预测小微企业违约情况的有效性^[5]。学者研究认为,通过数据挖掘的方法来进行信用评估能得到较为稳健的结果。在数据挖掘的方法中,就企业信用评级而言,支持向量机方法与其他方法相比,具有明显的优势^[6]。国内学者对小微企业评价研究还主要以定性为主。学者普遍认为,目前小微企业融资难问题最重要的原因是银企之间的信息不对称,解决该问题的有效途径在于,以国外商业银行的实践经验为基础,结合本土实际,改进针对小微企业的信用评估方法。

国内学者对信用评分方法的研究主要通过两个途径展开,其一是针对某类特定的信用评分模型进行改进^[7-8]。国内学者改进特定信用评分模型进行实证研究的目的或是追求提升模型的预测力、稳定性等性能,或是追求满足新的监管要求或现实操作性要求。同时,学者们往往通过将其改进的模型与原始模型或其他模型进行对比检验的方式,证明改进模型的先进性。其二,改进信用评分模型的基础,如引入拒绝推论^[9]。从数据缺失机制角度来看,过往信用评分实证研究中所用的样本数据绝大部分都是商业银行通过的企业贷款申请数据,这导致建模过程中存在样本选择偏差问题,可能使得模型参数估计有偏,进而影响模型的精确度。目前国内学者尝试通过拒绝推论解决这个问题,但这方面研究还较少,且未引起广泛关注。

现有对小微企业信用评估方法的研究主要不足在于没有真正结合小微企业特点。目前的实证研究大都肯定了从小微企业特点出发构建小微企业信用评估模型的重要性,但是,其信用评估指标体系的构建往往是采取传统的、以大中型公司为范本的相关指标,其优化的信用评估模型只是以小微企业客户信息为数据样本,在构建过程中均未与小微企业特点相结合以突出特点。只有解答了“什么样的信用评估方法才是适合小微企业的”这一问题,才能使得这项研究具有现实意义。本文致力于从小微企业特点出发,构建针对小微企业的且具有实用价值的信用评估指标体系和模型,并借鉴机器学习最新发展技术来解决这一问题。

三、小微企业信用评估指标体系构建

小微企业是个体工商户、家庭式作坊企业、微型企业和小型企业的统称。本文研究的“小微企业”是符合《中小企业划型标准规定》中界定标准的小型 and 微型企业、家庭作坊式企业、个体工商户等类型的客户。小微企业客户主要有两个特点:(1)企业规模小,缺乏抵押品且财务信息不透明。小微企业因资产总值中固定资产占比较低而难以提供足够的抵质押品。此外,财务制度不健全,财务信息不透明情况严重^[4]。这阻碍了银行通过硬信息评价小微企业信用情况的途径,提高了银行错误判断申请人信用情况的可能性,在一定程度上造成了供给型信贷配给的产生。(2)组织形式简单,所有权

和经营权高度统一。多数小微企业以个人独资或合伙形式经营,股权结构单一,具有较强的家族式经营特征。所以在构建信用评价指标体系时需要重点考察企业主的特质。在小微企业信用评估中,小微企业独特个性对其信用状况的影响需要被着重考虑。结合已有研究,本文从企业基本面信息、企业财务信息、现金流信息和企业主特征四个方面进行小微企业信用评估因素的分析。

(1) 企业基本面信息。企业基本面信息是对企业信用状况最直接的体现。企业基本面状况主要从企业规模、企业生存年限、企业性质、行业门类等方面进行考察。企业规模是企业当前经营状况的直接体现。企业生存年限在一定程度上反映了企业的历史经营能力,而小微企业生存期限普遍较短。企业经济性质反映了企业所有权,一般认为国家控股和集体控股的企业信用情况较好。小微企业中,私人绝对或相对控股企业和个体工商户占了绝大部分。

在对企业产生影响的外部因素中,最重要和最直接的就是企业所处的行业状况。一般认为,处于成长期的行业和较好市场环境的行业的企业发展前景更为理想,信用状况也相对会更好。小微企业大多位于行业发展的末端,并且具有非常高的成长性和风险性。企业所处的行业门类状况一般可以从行业发展阶段及前景、行业发展环境、行业市场现状等角度进行考察。小微企业大部分分布在制造业、批发和零售业,在建筑业、住宿和餐饮业、租赁和商务服务业等行业也有少量分布。根据行业将小微企业进行划分,能够更好地提高模型的准确性和扩展性。

(2) 企业财务信息。因为小微企业的信用违约风险与大中型企业相比具有异质性特征,所以在分析小微企业的企业财务状况对企业信用状况的影响作用和反映作用时,不能完全比照大中型银行。Altman 和 Sabato 的研究表明,应用财务指标对小微企业样本构建的信用风险评估模型,其提前一年的预测精确度比全部企业构建的模型的预测精确度高出约 30%^[10]。Minnis 也支持了这一观点^[5]。由于我国小微企业具有财务信息严重不透明的显著特点,构建财务指标时不能完全套用大企业的构建方式。本文从传统的“四要素”(偿债能力、盈利能力、营运能力和发展能力四个方面)构建财务方面的指标体系。

(3) 现金流量信息。小微企业资产规模较小,资产流动性差,生产经营种类较为单一,现金储备有限,并且信息不对称情况严重,使得其外源性融资渠道较少。因此,小微企业一旦面临短期融资压力,或者陷入生产经营困境,通常就难以再次获得外源性资金的支持。在这种情况下,资产负债类指标无法全面反映企业信用情况的变化,也无法精确估计其还款能力。Kim 认为,与资产价值等滞后指标相比,预期现金流作为估计小微客户的违约概率的指标更为有效^[11]。一方面,在缺乏外源性融资的情况下,小微企业自身生产经营产生的内源性资金尤为重要,其信用状况的变化表现在企业经营现金流的变化上。另一方面,小微企业较低的固定资产占比意味着其投资周期较短,这代表企业资金能够快速流动和回笼的特点。并且,基于现金流的预测模型以难以额外获得的外源性融资为假设条件^[12-13],符合小微企业的特点。实践领域中,目前已有商业银行在处理小微企业贷款申请时,在关注资产价值类指标的同时,重视其现金流的变动情况。具体而言,我们采用现金流充足性水平、紧急变现能力、融资约束程度差异、现金流入与还款方式的匹配度等指标。

(4) 企业主特征。小微企业所有权和经营权高度统一的特点决定了企业领导人即企业主对企业信用情况的影响程度较深。美国富国银行根据小企业主评分而对小企业信用进行自动评分。学术界也就企业主特征对小微企业信用情况影响的作用表示肯定。例如,Levenson 和 Willard 肯定了企业主信用特征的影响,认为企业主的还款意愿在很大程度上决定了小微企业的还款意愿^[14]。具体而言,我们采用企业主个人信用情况以及企业主个人信用情况与企业信用情况的紧密程度两个指标。

四、小微企业信用评估模型构建

小微业务信用评分的建模方法分为主观和客观两类。主观的方法是判断型,又称为专家型或经验型,即根据各方面专家对信贷政策、市场、客户、风险点的共同知识和经验,筛选出一组风险要素并

进行赋值,由此形成基于经验判断的信用评分模型。它一般适用于客户信息量化程度不高和样本量较小的情况。客观的方法是数据挖掘型,即运用数据挖掘技术在目标领域的大规模历史数据的基础上推演出来的信用评分模型。它主要包括多元线性回归、逻辑回归、分类树、神经网络、支持向量机等方法,不同的方法有各自的限制条件和适用情形。这些方法的好处在于从数据出发能够客观公正地进行判断,且机器学习模型通过对历史数据的不断学习,能够动态改变因子所配的权重,不断纳入新的信息。

支持向量机(Support Vector Machine, SVM)是由 Vapnik 等提出的一种基于结构风险最小化原则的学习理论^[15]。相比于传统的统计方法, SVM 不依赖于对所解问题的先验知识^[15]。支持向量机有很好的泛化能力,在处理具有小样本、非线性等特征的数据上有较好的表现。其主要思想是建立一个超平面作为决策面,使得正例和反例之间的间隔边缘实现最大化。它通过核函数,将输入变量映射到线性可分的高维空间,从而构造分类超平面。最小二乘支持向量机算法的出发点也寻求形如 $(wx) + b = 0$ 的划分超平面,但是其原始最优化问题为凸二次规划:

$$\begin{aligned} \min_{w, b, \xi_i} & \frac{1}{2} w^T w + C \sum_{i=1}^l \xi_i^2 \\ \text{ST. } & y_i = w^T \phi(x_i) + b + \xi_i \\ & i = 1, 2, \dots, l. \end{aligned} \quad (1)$$

该问题转化为对偶问题后,为求解一组线性等式,不要求解二次规划。最终我们得到 LS-SVM 的决策函数:

$$f(x) = \text{sgn}\left(\sum_{i=1}^l y_i \alpha_i^* K(x_i, x) + b^*\right) \quad (2)$$

该方法简化了计算复杂性,计算成本小,泛化功能好,且不易陷入局部最小,得到了广泛的应用。

小微业务的信用评估模型不仅需要满足预测稳定性和精确性要求,而且必须符合小微业务的特点和数据的实际情况^[16]。目前,我国商业银行自身累计的小微企业样本数据往往存在大量的缺失、噪声点和奇异值,在应用信用评估技术时,可作为模型的有效输入信息的数据样本较为有限。所以,适用于我国商业银行小微业务的信用评估模型需要在小样本条件下,满足预测稳定性和精确性要求。

本文借鉴已有的在信用评估方面的支持向量机模型研究成果,针对小微企业信用样本数据,通过对比研究,建立最小二乘支持向量机(LS-SVM)模型,以有效地提高预测稳定性和精确性。接着,本文将引入现金流信息指标的 LS-SVM 模型与未引入现金流信息指标的 LS-SVM 模型进行对比检验,以验证前文所论述的“现金流信息指标对小微企业信用评估具有重要意义”的观点。

五、数据预处理与实证研究

(一) 数据预处理与描述性统计

本文的数据来源为某国有商业银行省级分行小微业务首笔交易信息。本文筛选整理出了包含 629 个小微客户的数据。本文按照贷款风险分类标准进行好客户和坏客户的划分,将“关注”和“正常”两类客户作为未违约样本,标注为“0”;将“次级”“可疑”和“损失”三类客户作为违约样本,标注为“1”。在我们的样本中,样本违约率为 1.91%。本文数据中“现金流入与还款方式的匹配度”指标和企业主信息中的全部指标都无法获取。此外当相关系数的统计系数大于 0.6 时,两变量为强相关,我们仅选择了其中一个变量。经过筛选后的样本数据变量体系如表 1 和表 2 所示。

对于描述变量和模型输入变量,我们分别统计了每个变量违约和未违约客户的数量和所占百分比。具体情况见表 1 所示。

从小微贷款的整体情况来看,小微企业申请的银行贷款数额大多数小于或等于 20 万元,且违约客户也大多数属于这个区间。近九成小微企业申请贷款用于弥补流动资金,同时超过九成的小微贷

款为短期贷款,这两种情况一方面决定了小微贷款需求“短、小、急、频”的特点,另一方面使得当企业流动资金链条出现断裂时,小微贷款容易出现违约风险。

表1 描述变量统计结果

一级指标	二级/三级指标	数值或内容类别	违约客户		未违约客户		合计		
			数量	百分比(%)	数量	百分比(%)	数量	百分比(%)	
贷款标识信息	发放金额	(0,20 万元]	34	5.41	40	6.36	74	11.76	
		(20 万元,100 万元]	45	7.15	43	6.84	88	13.99	
		(100 万元,+)	230	36.57	237	37.68	467	74.24	
		合计	309	49.13	320	50.87	629	100	
	业务种类	流动资金贷款	271	43.08	291	46.26	562	89.35	
		国内贸易融资	24	3.82	17	2.7	41	6.52	
		国际贸易融资	8	1.27	9	1.43	17	2.7	
		票据类	6	0.95	1	0.16	7	1.11	
		项目贷款	0	0	2	0.32	2	0.32	
		合计	309	49.13	320	50.87	629	100	
		贷款期限	短期	278	44.2	289	45.95	567	90.14
			中期	20	3.18	19	3.02	39	6.2
			长期	11	1.75	12	1.91	23	3.66
			合计	309	49.13	320	50.87	629	100

从申请贷款的小微企业整体情况来看,绝大多数小微企业为私人绝对控股,且绝大部分的违约样本发生在这个群体中。行业门类中,超过九成的小微企业集中于制造业与批发和零售业,且大多数违约样本也发生在这两个行业中,说明这两个行业中的小微企业的信用状况相对较弱,这与前文发现的“制造业与批发和零售业的生存能力最弱”这一现象相符合。

对于模型输入变量,本文分别统计了每个变量的有效样本量、缺失值数量、均值、标准差和极值,具体情况见表2。h27(投资活动现金流入/投资活动现金流出)和h29(筹资活动现金流入/筹资活动现金流出)的缺失值在半数左右,这是由于小微企业的企业活动以经营性活动为主,筹资活动和投资活动发生较少。当企业在向银行申请贷款前没有进行投资或不存在负债时,其投资活动现金流出量或筹资活动现金流出量为零,缺失值出现。h15(固定资产周转率)的数值缺失率约为14.5%,这是由小微企业资产规模较小的特点造成的,部分小微企业自身不拥有固定资产。h34(利息偿付现金比率)同样存在约为8.6%的数值缺失率,这是因为这部分小微企业在向银行提出贷款申请时尚未因负债产生利息费用。本文用平均值替代缺失值。此外,本文删除明显噪声数据。

从企业基本面信息来看,企业生存年限较短,平均约为五年。从企业财务信息来看,偿债能力方面,样本企业的现金比率较高,均值为2.04,且违约企业与未违约企业差距较大,说明小微企业注重现金及其等价物对流动资产的覆盖能力,短期偿债能力较好。盈利能力方面,销售利润率和总资产报酬率的均值分别约为9%和17%。营运能力方面,样本企业的存货周转率和应收账款周转率较高,说明企业的流动资产周转速度较快。此外,固定资产周转率的均值高达883.93,这是由于样本企业的固定资产规模较小造成的。发展能力方面,样本企业的营业收入增长率、营业利润增长率和总资产增长率分别约为16%、9%和17%,资本保值增值率为高达1.31,显示出很强的增长率。从现金流量信息来看,现金流充足性水平方面,样本企业的各指标表现较为平均,其中未违约企业的再投资现金比率显著高于违约企业。紧急变现能力方面,样本企业的投资活动现金流入/总现金流入均值为零,说明小微企业较少进行投资活动。融资约束程度差异方面,未违约企业的利息偿付现金比率为79.74,违约企业仅为1.34,说明小微企业的现金对偿付贷款利息的重要性。

(二) 支持向量机实证结果与模型评估

在实证分析之前,我们首先对数据集进行三项处理:(1)处理明显错误数据,例如企业生存年限(h03)

表 2 模型输入变量的描述性统计结果

一级指标	二级指标	三级指标	指标代码	未违约客户		违约客户		样本总量						
				均值	标准差	均值	标准差	均值	标准差	最小	最大	N	缺失值	
企业 基本面 信息	企业规模	注册资本	h01	783.66	1541.60	742.30	1202.55	763.34	1384.51	0.00	19000.00	629	0	
		职工人数	h02	105.32	127.31	84.66	107.71	95.17	118.44	4.00	1200.00	629	0	
	生存年限	企业生存年限	h03	6.14	5.03	5.27	5.15	5.71	5.11	-32.00	25.00	629	0	
		资产负债率	h04	0.33	0.17	0.33	0.18	0.33	0.18	0.00	0.94	629	0	
	偿债能力	现金比率	h08	2.27	20.66	1.78	17.73	2.03	19.27	-1.72	346.62	628	1	
		销售利润率	h09	0.09	0.09	0.09	0.12	0.09	0.11	-0.06	1.14	626	3	
	盈利能力	总资产报酬率	h10	0.16	0.11	0.16	0.13	0.16	0.13	-0.09	0.92	629	0	
		存货周转率	h12	18.83	125.25	19.24	79.20	19.04	105.25	0.00	2225.00	617	12	
	企业 财务信息	营运能力	应收账款周转率	h13	24.43	30.86	30.05	73.97	27.15	56.04	-18.32	1076.00	606	23
			固定资产周转率	h15	1077.59	6524.94	681.43	3941.80	883.93	5417.00	0.00	93220.00	538	91
总资产周转率			h16	2.89	2.81	2.74	2.02	2.82	2.46	0.00	36.19	629	0	
发展能力		营业收入增长率	h17	0.15	0.28	0.15	0.20	0.16	0.25	-2.55	1.00	626	3	
		营业利润增长率	h18	0.06	1.20	0.11	0.65	0.09	0.98	-19.85	3.70	627	2	
		总资产增长率	h19	0.18	0.23	0.15	0.29	0.17	0.27	-2.41	0.97	629	0	
		资本保值增值率	h20	1.30	0.35	1.32	0.52	1.31	0.44	0.78	7.43	629	0	
		净利润获现率	h23	-0.03	10.32	1.70	24.93	0.82	18.95	-180.67	435.70	627	2	
		现金流 充足性 水平	经营活动现金流量比	h24	1.03	0.10	1.03	0.23	1.04	0.18	-1.16	3.07	611	18
			总现金流量比	h25	1.02	0.10	1.01	0.14	1.02	0.12	-0.13	2.97	611	18
现金 流量信息	再投资现金比率	再投资现金比率	h26	-0.44	9.52	0.68	8.14	0.11	8.88	-170.12	128.38	629	0	
		投资活动现金流入/ 投资活动现金流出	h27	0.04	0.27	0.03	0.17	0.04	0.23	-0.02	2.82	295	334	
	紧急 变现能力	投资活动现金流入/ 总现金流入	h28	0.00	0.02	0.00	0.01	0.00	0.02	0.00	0.40	611	18	
		筹资活动现金流入/ 筹资活动现金流出	h29	5.82	84.03	2.22	39.30	4.07	66.10	-650.00	990.00	415	214	
	融资约束 程度差异	筹资活动现金流入/ 总现金流入	h30	0.06	0.14	0.06	0.22	0.07	0.19	-2.55	2.03	611	18	
		经营现金流量总负 债率	h31	0.23	1.13	0.91	7.00	0.57	4.98	-11.34	78.80	627	2	
	利息偿付现金比率	筹资净现金流所有者 权益比率	h33	0.05	0.30	0.04	0.29	0.05	0.30	-1.97	2.91	629	0	
		利息偿付现金比率	h34	82.61	566.14	74.10	438.23	78.52	508.14	-1019.00	5954.00	575	54	

应该都是大于零,因此数据中凡出现负数均以缺失处理;(2)以各指标的平均值补充缺失数据;(3)标准化处理。由于各指标的量级不一致,直接输入模型会导致训练学习过程受到偏差影响,因此对各指标进行标准化处理,将其转变为[0,1]之间的数据,统一量级。计算公式如下:

$$h_{01} = \frac{h - h_{min}}{h_{max} - h_{min}} \quad (5)$$

本文基于多项式核函数构建 LS-SVM 模型,将其与支持向量机中的 FQP(模糊二次规划)和 LSVM(拉格朗日 SVM)估计方法进行比较。由表 3 的结果可见,在样本内 LS-SVM 对小微企业是否违约的误判率最低(0.45276),这是 LSVM 综合错误成本的 60%,比 FQP 和 DQP(可得对偶二次规划)都低。这说明 LS-SVM 在样本内是较好的模型。而在样本外,LS-SVM 的均方误差与综合错误成本在四个模型中依旧是最低的,分别为 0.25361 和 0.45276。这说明 LS-SVM 模型具有较好的预测性和

表 3 LS-SVM、LSVM、DQP、FQP 估计方法的比较

模型说明	训练		验证		K-S 统计量 ²
	平均平方误差	平均平方误差	平均平方误差	综合错误成本 ¹	
LS-SVM	0.21638	0.25600	0.25361	0.45276	0.090
LSVM	0.23558	0.41867	0.27247	0.47244	0.080
DQP	0.21577	0.26133	0.25428	0.50394	0.062
FQP	0.20698	0.32000	0.26485	0.50787	0.075

注:1. 综合错误成本是通过模型预测的不同错误类型赋予差异性的成本来评估模型预测成本的方法。模型预测错误包括了两类:假阳性错误,即第一类错误(Type I errors)和假阴性错误,即第二类错误(Type II errors)。同下文 ROC 曲线所述,假阳性错误是指将好客户误分类为了坏客户,而假阴性错误是指将坏客户错误的分类为了好客户。对于银行而言,犯第一类错误将会丧失未来的盈利机会,而犯第二类错误将带来严重的违约风险,进而引致损失。因此,从损失规避的视角而言,第二类错误的成本更高。因此,需根据银行实际风险偏好设定每类错误相应的成本,并计算每个模型或算法的综合成本,从而对模型或算法的区分能力进行评价。2. K-S 统计量 P 值均大于 0.05,不拒绝原假设,可以认定两个组来自同一分布。

稳健性。

本文进一步根据核函数训练结果绘制 ROC 曲线(图 1)。ROC 是衡量不同模型好坏的一个重要指标。ROC 是综合了灵敏度和特异度所绘制出的曲线,具体而言就是以假阳性率(1 - 特异度)为横轴,真阳性率(灵敏度)为纵轴所描绘的曲线。相对于基线,曲线下面积反映了预测试验的价值大小,越接近 1,真实度越高,越接近 0.5,越没有预测价值。从图中可以看到,LS-SVM 在样本内 ROC 曲线在最左侧,表现最好。而在样本外,LS-SVM 也能获得相对较好的 ROC 曲线。综上,从均方误差和综合错误成本以及 ROC 分析可以发现,LS-SVM 模型较其他模型有较好的效果。

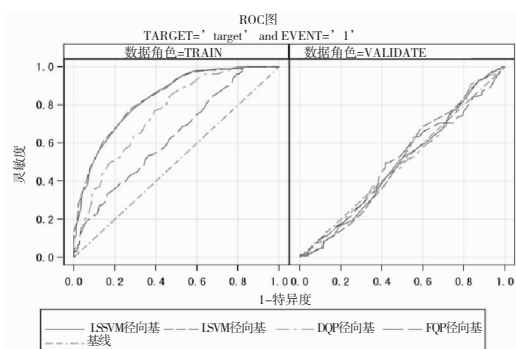


图 1 LS-SVM、LSVM、DQP、FQP 方法的 ROC 曲线

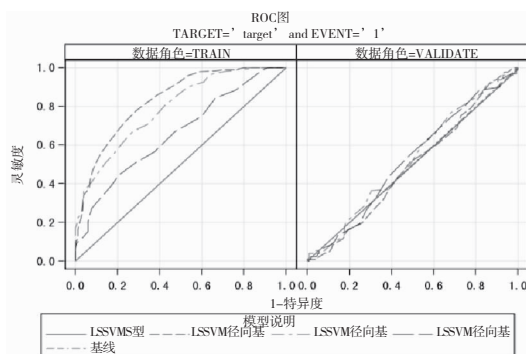


图 2 基于径向基、线性和多项式核函数的 ROC 曲线图

核函数直接决定了支持向量机的最终性能,为了获得更好的 LS-SVM 训练效果,我们对径向基、线性、多项式三种核函数分别进行了比较。训练结果显示径向基核函数的训练效果最佳,其综合错误成本最低(限于篇幅不再汇报具体结果)。图 2 是根据三种核函数训练结果绘制的 ROC 曲线。从 LS-SVM 拟合统计量看,LS-SVM 模型在训练集和验证集中的拟合结果基本稳定。

对预测模型进行评估,除了 ROC 曲线,还有响应率评估、提升度评估、增益评估,预测结果评估显示较为理想。对预测模型进行评估,除了 ROC 曲线,还有响应率评估、提升度评估、增益评估,预测结果评估显示较为理想,限于篇幅不再汇报具体结果。

综上所述,基于 LS-SVM 模型的预测结果表明,引入非财务指标的小微企业信用评级指标体系对小微企业违约具有更强的预测能力。改进的支持向量机方法在预测小微企业违约方面具有准确性和稳定性优势。

实证研究结果也进一步表明,将信用评分技术应用于小微企业信贷业务是有效的,因为运用 LS-SVM 模型对小微企业违约预测能力更强,能够弥补传统逻辑回归的不足。对银行来说,信用评分方法具有降低信贷成本、提高风险管理水平、拓展业务空间、优化贷款定价和有利于信贷资产证券化业务的开展等优点,从而在一定程度上从供给角度缓解小微企业融资困难。这一结果也意味着,大中型商业银行通过信用评分方法同样可以开展小微业务,银企规模不需要完全匹配^[17]。从银行规模角度出发,由于信息、计算机技术的发展和自身的技术、人员优势,与小银行相比,大型商业银行可以在小微企业信贷业务上更加有效地使用信用评分技术。从贷款技术角度出发,信用评分方法具有将小微企业的部分“软信息”转变为“硬信息”、降低信贷成本、拓展业务空间等优势,与其他贷款技术相比,更适合大型银行应用于小微企业信贷业务领域。

(三) 现金流信息指标重要性检验的实证结果

为了验证现金流量数据的缺失是否对分类判断产生影响,本文根据原始小微企业信用评估数据分别构建了包含现金流量信息的数据集和不包含现金流量信息的数据集。我们将两个数据集分别输入到 LS-SVM 模型中,其余所有参数设置相同,进行训练和学习。结果显示(见表 4 和图 3),包含现

金流量信息的数据集训练出的结果误分率明显低于不包含现金流量信息的数据集训练出来的结果,且误分率降低了 1.70%。这从数据方面证明了现金流指标缺失会影响 LS-SVM 模型的判断性能。

小微企业因为规模小而导致信息不对称情况严重,外源性融资渠道少,一旦面临短期融资压力或者陷入生产经营困境,通常就难以再次获得外源性资金的支持,因此资产负债类指标无法全面反映企业信用情况的变化,也无法精确估计其还款能力,故预期现金流在预测其违约概率上将更为有效。本文运用 LS-SVM 模型的结果支持了 Kim^[11] 等学者的研究,即与资产价值等滞后指标相比,预期现金流作为估计小微客户的违约概率的指标更为有效。

(四) 基于 LS-SVM 模型预测结果的讨论

综合实证结果可以发现,LS-SVM 模型(径向基核函数)能够较理想地对小微企业客户的违约与未违约情况进行分类,结果稳定。现金流量数据能够反映小微企业现金流入流出的重要信息,充足的现金流量数据能够保证分类和预测准确性的提高。

通过 LS-SVM 模型的预测结果看,统计学习方法在应对小微企业违约预测上确实具有一定优势。传统的 Logit 和 Probit 模型的主要特点是简单实用,但他们并不能处理非线性样本的问题。而非线性样本的处理正是机器学习的特点。事实上,已经有许多学者将机器学习模型应用于信用评估中,常见的机器学习模型如神经网络^[18] 和支持向量机^[19]。由于银行小样本、非线性、大量缺失值、奇异值等特点,所以使用神经网络进行建模往往会陷入过度拟合,泛化能力不足。以统计学习理论为基础,支持向量机通过有效地学习,具有较好的推广性和分类精度^[16];它在解决非线性和高维问题中具有独特优势,并被认为是小样本建模和预测的最佳理论^[7-8]。Baesens 等^[20] 学者的研究也表明,SVM 的预测结果要优于其他几种不同的信用风险评估模型。本文的研究结果与国内其他学者^[21] 的研究发现是一致的,即 SVM 与 Logit 回归以及神经网络相比要有更好的预测精度,而且基于径向基核函数的 SVM 预测精度要更优于其他核函数的 SVM 预测。

进一步,从小微企业信用评级指标体系来看,本文的研究进一步完善了小微企业信用评级指标,特别是纳入了现金流指标。传统的违约风险模型基本侧重于对客户的资产变化情况进行刻画,以客户的资产价值(包括有形或无形资产)低于违约边界作为违约的触发条件,对客户的违约概率进行预测^[22],即所谓的基于价值的模型(Value-Based Model)。然而,即使债务人具有盈利能力,或者资产总额仍然大于债务总额,但是短期现金流不足带来的流动性问题会导致其无法满足短期债务偿付的要求,从而导致债务人的违约^[11-12]。Norden 和 Weber^[13] 的研究也表明关注中小微企业以及个人的现金流变动情况对于估计和监测他们的违约概率有重要意义。He 和 Xiong^[23] 的研究也认为企业现金流越紧张,短期到期的偿付比例越高,违约的风险越大。本文引言中所提及的在实践中许多银行对小微客户违约风险采用的管控方法也证实了基于现金流的违约触发机制对这类客户的适用性^[24]。如德国 IPC 公司率先提出的基于客户现金流评测的微贷管理技术,目前已在全球范围内得到有效的应用。此外,产业链金融模式是另一种基于现金流的违约管控技术,针对中小微企业所处的产业链,通过评

表 4 现金流量信息存在与否的综合错误成本

模型说明	训练		验证	
	平均平方误差	平均平方误差	平均平方误差	平均平方误差
LS-SVM 含现金流信息	0.21638	0.25600	0.25361	0.45276
LS-SVM 无现金流量信息	0.22167	0.28267	0.25344	0.46063

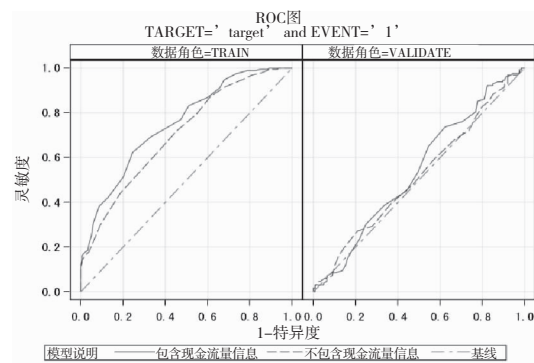


图 3 有现金流量信息和无现金流量信息的 ROC 曲线图

估上下游企业间稳定的现金流,并辅以信息流和物流的信息对客户进行违约风险评估,在对中小微客户违约风险管理方面取得了显著的成效。从现金流的视角出发,对中小微客户的违约风险进行刻画和建模,有助于真实地反应这类客户的违约触发机制,从而更加准确地预测违约事件发生的概率。

六、总结与展望

本文从小微企业“融资难、融资贵”问题和商业银行对小微业务普遍关注现象入手,从银行的角度,以我国小微企业特点为基本立足点,建立科学有效的小微企业信用评估指标体系,其中特别提出了现金流信息指标在小微企业信用评估中的重要作用,并利用对比分析方法构建小微企业信用评估模型。本文通过对小微企业客户数据样本特点的分析,选择支持向量机作为建模方法,并且通过实证检验,证明构建的 LS-SVM 模型(径向基核函数)在预测准确性和稳定性方面的优势。

本文的理论贡献可能在于,一是以小微企业特点为基本出发点,构建出针对小微企业的且具有实用价值的信用评估指标体系和评估模型;二是在以往文献基础上,构建更加全面的现金流信息指标体系,从实证角度论证了现金流信息在小微企业信用评估中的作用;三是根据我国商业银行小微业务的特点和数据的实际情况,构建最小二乘支持向量机模型,提升了预测准确性和稳健性。

面对复杂多变的市场环境和经济、金融背景,以及日益蓬勃发展的商业银行小微业务,本文的研究对于银行减轻银企信息不对称、从供给型信贷配给视角解决小微企业“融资难”问题具有一定的理论和应用价值。商业银行开展小微业务一直以来受到小微企业信息不对称所导致的违约风险高问题以及运营成本高问题。基于以上研究,我们认为对商业银行开展小微金融业务首先要打破“大银行对大企业,小银行对小企业”传统观念,以交易型信贷技术为核心的企业信用风险评估方法能够在一定程度上解决以关系型信贷技术的高运作成本问题。通过构建小微企业信用评价指标体系,将“软信息”进行“硬化”,借助机器学习的方法,不仅能够有效评估小微企业信用风险,还能通过进一步开发打分卡系统批量化处理小微企业业务,从而降低运作成本。其次,以德国 IPC 公司为代表的小微企业现场调查下的现金流分析和评估技术对于评价小微企业信用风险具有重要的实践指导价值。回归到信贷风险管理的本源,关注小微企业还款意愿和还款能力,有助于打破当前商业银行的“抵押物崇拜”和“担保崇拜”,从而真正解决小微企业“融资难”的问题。

未来研究,我们可以完善评估指标体系,获取现金流入与还款方式的匹配度信息和企业主信息,还可以进一步从行业发展阶段及前景、行业发展环境、行业市场现状等角度考察行业对小微企业信用情况的影响;针对信用评级模型中的数据缺失和样本选择偏差问题,引入拒绝推论方法;在构建小微企业信用评级模型时纳入被拒绝企业的样本信息,以提升模型的预测能力;构建具有动态特征的信用评估方法,发展行为信用评级技术,对小微企业信贷进行动态管理。

参考文献:

- [1]加布里埃尔,欣克利,瓦格斯. 微型金融:方法与案例[M]. 游春,陈允宏,译. 北京:中国出版集团东方出版中心,2016.
- [2]Stiglitz J, Weiss A. Credit Rationing in Markets with Imperfect Information[J]. American Economic Review, 1981, 71(3): 393-410.
- [3]Boucher S R, Guirking C. Risk rationing and wealth effects in credit markets: theory and implications for agricultural development[J]. American Journal of Agricultural Economics, 2008, 90(2):409-423.
- [4]邓超,胡威,唐莹. 国内外小企业信用评级研究动态[J]. 国际金融研究,2010(10):85-91.
- [5]Minnis M, Sutherland A. Financial statements as monitoring mechanisms: evidence from small commercial loans[R]. Chicago Booth Research Paper, 2015.
- [6]Lisowsky P, Minnis M, Sutherland A G. Credit cycles and financial statement verification[R]. Chicago Booth Research Paper, 2014.
- [7]姜明辉,袁绪川. 个人信用评估 PSO—SVM 模型的构建及应用[J]. 管理学报,2008(4):511-515.
- [8]陆爱国,王珏,刘红卫. 基于改进的 SVM 学习算法及其在信用评级中的应用[J]. 系统工程理论与实践,2012(3):515-521.
- [9]邓超,胡威,唐莹. 基于拒绝推论的小企业信用评级模型研究[J]. 国际金融研究,2011(4):68-76.

- [10] Altman E I, Sabato G. Modelling credit risk for SMEs: evidence from the US market[J]. *Abacus*, 2007, 43(3): 332 – 357.
- [11] Kim I J, Sundaresan S. Does default risk in coupons affect the valuation of corporate bonds? [J]. *Financial Management*, 1993, 22(3): 117 – 131.
- [12] DeAngelo H, DeAngelo L, Wruck K H. Asset liquidity, debt covenants, and managerial discretion in financial distress: the collapse of LA Gear[J]. *Journal of financial economics*, 2002, 64(1): 3 – 34.
- [13] Norden L, Weber M. Credit line usage, checking account activity, and default risk of bank borrowers[J]. *Review of Financial Studies*, 2010, 23(10): 3665 – 3699.
- [14] Levenson A R, Willard K L. Do firms get the financing they want? Measuring credit rationing experienced by small businesses in the US [J]. *Small Business Economics*, 2000, 14(2): 83 – 94.
- [15] Vapnik V N. An overview of statistical learning theory. [J]. *IEEE Transactions on Neural Networks*, 1999, 10(10): 988 – 999.
- [16] Shin K S, Lee T S, Kim H. An application of support vector machines in bankruptcy prediction model[J]. *Expert Systems with Applications*, 2005, 28(1): 127 – 135.
- [17] 粟勤, 田秀娟. 近二十年海外银行中小企业贷款研究及新进展[J]. *浙江大学学报(人文社会科学版)*, 2011(1): 84 – 93.
- [18] Khashman A. Neural networks for credit risk evaluation: investigation of different neural models and learning schemes[J]. *Expert Systems with Applications*, 2010, 37(9): 6233 – 6239.
- [19] Yu L, Yao X, Wang S, et al. Credit risk evaluation using a weighted least squares SVM classifier with design of experiment for parameter selection[J]. *Expert Systems with Applications*, 2011, 38(12): 15392 – 15399.
- [20] Baesens B, Van Gestel T, Viaene S, et al. Benchmarking state-of-the-art classification algorithms for credit scoring[J]. *Journal of the operational research society*, 2003, 54(6): 627 – 635.
- [21] 张奇, 胡蓝艺, 王珏. 基于 Logit 与 SVM 的银行业信用风险预警模型研究[J]. *系统工程理论与实践*, 2015(7): 1784 – 1790.
- [22] Giesecke K. Default and information[J]. *Journal of Economic Dynamics and Control*, 2006, 30(11): 2281 – 2303.
- [23] He Z, Xiong W. Dynamic debt runs[J]. *Review of Financial Studies*, 2010, 25(6): 1799 – 1843.
- [24] 王明虎, 王小韦. 企业规模融资约束与资本结构波动[J]. *南京审计大学学报*, 2015(2): 12 – 18.

[责任编辑:杨志辉]

Small and Micro Businesses Credit Evaluation Research Based on LSSVM

XIAO Binqing^{1,2}, BO Wei¹, YAO Yao³, LI Xindan¹

(1. School of Engineering and Management, Nanjing University, Nanjing 210093, China;

2. Leonard N. Stern School of Business, New York University, New York 20012, USA;

3. Department of Ranking, Jiangsu Zhongcheng xin Credit Management Co. Ltd., Nanjing 210002, China)

Abstract: Based on the characteristics of small and micro businesses, we combine the theory with commercial bank practice to sort out and analyze the factors influencing the credit situation of small and micro enterprises. The index of small micro enterprise credit evaluation is constructed, which we add the factor of cash flow information. Based on a sample of actual borrowers from a large China commercial bank, the results reveal that LSSVM outperform the other techniques with the measure of ROC. In addition, this paper proves that cash flow index plays an important role in the credit evaluation of the small and micro enterprises through the empirical test. This study provides theoretical support and practical experience for China's commercial bank screening small and micro businesses credit evaluation index and building credit evaluation model.

Key Words: small and micro businesses; credit evaluation; enterprise credit evaluation; cash flow information; LS-SVM; small and micro-finance; commercial banks; credit rationing; credit scoring