

An Empirical Credit Risk Study of SEMs in Small Loan Companies

—Based on Logistic Model and Probit Model

Jiamin Zhang, Jianjun Zhou

Department of Statistics, College of Mathematics and Statistics, Yunnan University, Kunming
Email: jjzhou@ynu.edu.cn

Received: Sep. 20th, 2014; revised: Oct. 25th, 2014; accepted: Nov. 7th, 2014

Copyright © 2014 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

SMEs (small and medium-sized enterprises) as small loan companies' main customer groups are objects of study in this paper. Based on the companies' day-to-day credit loan business, Logistic model and Probit model are applied to this empirical analysis of credit risk assessment. Adopt brainstorming to get eight required indicators with the help of experts on microcredit industry. These eight indicators as independent variables in the model can be divided into two parts according to content. One part is used to describe borrower's personal circumstances and another part is used to measure companies' business conditions. Data from Kunming KC small loan company are fitted with these two models. In the face of a specific loan, calculate the corresponding credit risk score and then make the decision. Upon examination, results of this study are feasible and effective.

Keywords

Small Loan Companies, SMEs, Credit Risk, Logistic Model, Probit Model

小额贷款公司中小企业信用贷款风险实证研究

—基于Logistic和Probit模型

张佳敏, 周建军

云南大学, 数学与统计学院统计系, 昆明

Email: jjzhou@ynu.edu.cn

收稿日期: 2014年9月22日; 修回日期: 2014年10月25日; 录用日期: 2014年11月7日

摘要

本文以小额贷款公司主要的客户群体中小企业作为研究对象,立足于公司日常的信用贷款业务,基于 Logistic 与 Probit 模型进行风险评估实证分析。在小额贷款行业专家的帮助下,采取头脑风暴法获得建立模型所需的八项指标。这八项指标作为模型的自变量,从内容上可分为两部分,一部分描述借款人个人情况,另一部分衡量企业经营状况。在昆明高新科创小额贷款公司数据支持下获得模型结果。在面对具体的贷款业务时,可以通过相应模型计算客户信用风险得分,最终做出是否予以贷款的决策。经检验,研究成果可行、有效。

关键词

小额贷款公司, 中小企业, 信用风险, Logistic 模型, Probit 模型

1. 引言

在我国经济发展方式转变、结构调整的大局中,中小企业的发展具有举足轻重的作用。据统计,我国中小企业占企业总数的 99%, 占工业产值的 60%, 占就业安排的 75%。但相对于其作用与数量而言,中小企业在融资上的困难已经成为制约其生存和发展的不争事实。在我国的银行融资格局中,大银行高度垄断制约了中小金融机构获得金融资源的能力,同时也限制了它们为中小企业服务的能力。而大银行追求贷款规模效益和风险平衡又不愿为中小企业提供贷款。这使得众多中小企业无法从银行获得贷款,因此在国家“鼓励引导和规范民间资本进入金融服务领域”的政策支持下,小额贷款公司像雨后春笋般地涌现[1]。

自 2005 年国家启动“商业性小额贷款公司试点”以来,短短几年时间,我国小额贷款公司从无到有,发展可以用突飞猛进来形容。据中国人民银行发布的数据,截止到 2014 年 3 月底,全国已经有小额贷款公司 8127 家,从业人数达到 98,888 人,实收资本金 7494.07 亿元,贷款余额 8444 亿元[2]。具有扶贫性质的小额贷款公司根据市场需求主要以中小企业为贷款对象,具有放贷门槛低、机制灵活、手续简便、放贷速度快、贷款期限灵活等优势,恰恰迎合了中小型企业的需求,为当地中小企业拓宽融资渠道提供了更多的选择,成为金融体制改革的一项重大创新与突破。

然而,由于向小额贷款公司寻求贷款的中小企业大多处在创立初期或者扩大生产规模的阶段,各方面条件均不完善,有的甚至连详细的财务记录都没有,使得获取中小企业信息的难度和成本都有所增加,从而给小额贷款公司正确评估信用贷款风险带来很大的挑战。同时,国内外关于小额贷款公司中小企业贷款信用风险评估模型的量化模型几乎没有。因此,本文提出了基于 logistic 模型和 Probit 模型的中小企业贷款风险的信用评估模型。

2. 指标选择和模型

2.1. 指标选择

要对中小企业建立信用贷款评分模型就必须找到评价借款人各项相关条件的若干指标,这些指标在原则上应当与借款人的信用状况有较强相关性。同时,考虑到小额贷款公司在日常工作中一般均通过实

地调查获得指标信息，因此在指标选择时，还应充分考虑到操作的可行性，尽量选择相关性强且容易获得的指标。

在小微金融领域专家的帮助下，采取头脑风暴法，得到了建立中小企业信用贷款风险评估模型的八项指标，见表 1。这八项指标分为两部分，一部分是描述借款人个人情况的指标，包括：借款人年龄、学历、人行逾期记录、婚姻状况及居住状况；第二部分是关于企业经营状况的指标，包括：反映长期偿债能力的资产负债率，反映短期偿债能力的流动比率，反映企业获利能力的销售毛利率。当然，除此之外还有一些指标也与信用状况具有较强相关性，例如：借款者是否拥有不良嗜好。因为在实际贷款业务中，如果借款人有诸如赌博等不良嗜好，其还款状况就难以保证。可是这类指标在实际操作中很难获得，或者获得需要消耗大量的人力、物力、财力，因此未将这类指标选入模型内。

1) 年龄。国外学者的研究表明，年龄与信用风险之间存在曲线关系。虽然这一曲线关系可能并不具有普遍适用性，但至少表明年龄是信用风险的影响因素之一。

2) 学历。一般认为较高的学历伴随着较高的受教育程度和较多的工作机会，因而违约的可能性也较小。该指标与信用风险成负相关关系。

3) 人民银行逾期记录。贷款记录反映了一个人过去贷款的还款情况，以往逾期记录不良表明借款者的还款意愿或者还款能力存在问题，贷款风险增大。

4) 婚姻状况。一般认为拥有良好家庭状况的人追求稳定生活的意愿更为强烈，对于家庭的责任也使他们努力工作赚钱，因此信用情况更加良好。

5) 居住状况。一个人的居住状况与其资产情况有一定关系，而资产与违约概率负相关。

6) 资产负债率。资产负债率是一项反映企业长期偿债能力的指标，它是企业负债总额占企业资产总额的百分比。计算公式为： $\text{资产负债率} = (\text{负债总额} / \text{资产总额}) \times 100\%$ 。该指标值越大，信用风险也就越大。

7) 流动比率。流动比率是指企业流动资产与流动负债的比率，她反映企业的短期偿债能力。计算公式为： $\text{流动比率} = \text{流动资产} / \text{流动负债} \times 100\%$ 。该指标值越大，信用风险越小。

8) 销售毛利率。销售毛利率是毛利占销售收入的百分比，计算公式为： $\text{销售毛利率} = (\text{销售收入} - \text{销售成本}) / \text{销售收入} \times 100\%$ 。销售毛利率能反映企业的竞争力和获利潜力。该指标值越大，信用风险越小。

2.2. 模型选择

对于风险的研究主要有两种量化模型：Logistic 模型和 Probit 模型。基于这两类模型，我们对中小企业贷款风险的信用风险进行量化评估。

Table 1. Model indicators
表 1. 模型指标

指标类型	指标	指标原始值
借款人个人情况	年龄	实际值
	学历	小学及以下；初中；高中；技校；大专；本科及以上
	人民银行逾期记录	逾期次数实际值
	婚姻状况	未婚；已婚；离异；丧偶
	居住状况	自有；按揭/抵押；租住；借住亲戚朋友家
企业经营情况	资产负债率	实际值
	流动比率	实际值
	销售毛利率	实际值

Logistic 回归(Logistic Regression)被广泛用于对因变量为二分类变量进行回归建模。该模型对数据的需求量相对较少且不要求数据服从正态分布, Ohison(1980)将 Logistic 模型用于信用风险评估领域并获得 92% 以上的判别正确率[3]。

设 p 个自变量分别为 X_1, \dots, X_p , Y 表示借款人的违约倾向。若 $Y=1$, 表示借款人未按时还款, 发生违约; 若 $Y=0$, 表示借款人按时还款。如果令二分类因变量 $Y=1$ 的概率为 π , 则对违约概率 π 和自变量 X_1, \dots, X_p 可以建立如下模型:

$$\ln \frac{\pi}{1-\pi} = \text{logit}(\pi) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p (\pi), \quad (1)$$

其中 $\text{logit}(\pi)$ 代表将 π 变换为 $\ln \frac{\pi}{1-\pi}$, logit 变换使得在 $[0,1]$ 范围取值的 π 变换到 $(-\infty, \infty)$, 当 π 趋向于 0 时, $\text{logit}(\pi)$ 趋向于 $-\infty$; 当 π 趋向于 1 时, $\text{logit}(\pi)$ 趋向于 $+\infty$ 。这种违约概率 π 与自变量之间的回归关系就是 Logistic 回归模型:

$$\pi = \frac{\exp(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}{1 + \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)} \quad (2)$$

当有一组样本 $(X_{i1}, \dots, X_{ip}, Y_i)$ 时, 可以通过极大似然估计得到未知参数的估计, 从而可以得到第 j 个借款人是否违约概率的估计 $\hat{\pi}_j$, 如果 $\hat{\pi}_j > 0.5$, 则认为该借款人会违约; 否则会按时还款。

Probit 模型也被广泛用于对因变量为二分类变量进行分析。它主要假设违约事件发生的条件概率服从累计正态分布函数, 该模型在信用评估中有着广泛应用。该模型的优势在于模型用的是点估计, 即使随机变量的分布不能满足正态条件, 仍能得到无偏的估计值[4]。

设每一个样本 i 都存在一组变量 $X_i = (X_{i1}, \dots, X_{ip})^T$, 这些变量的线性组合可以使每一个样本得到一个分数 Y_i^* :

$$Y_i^* = \sum_{j=1}^p \beta_j X_{ij} = X_i^T \beta \quad (3)$$

Y_i^* 代表某个内在变量或是隐藏变量, 在中小企业信用贷款风险评估研究中, 代表的是借款人发生违约的倾向。当 $Y_i^* > 0$ 时, 因变量 Y_i 取 1(借款人未按时还款, 发生违约); 当 $Y_i^* \leq 0$, 则 $Y_i = 0$ (借款人按时还款), 用概率模型表示如下:

$$P_i = E(Y_i = 1 | X_i) = P(Y_i^* > 0) = F(X_i^T \beta), \quad (4)$$

其中 $F(\cdot)$ 表示标准正态分布的累积分布函数—— $F(X_i^T \beta) = \int_{-\infty}^{X_i^T \beta} f(z) dz$ 。取上述分布函数的逆函数得到 Probit 模型:

$$Y_i^* = F^{-1}(P_i) = X_i^T \beta, \quad (5)$$

利用最大似然法估计上式中的参数[5]。

3. 实证分析

3.1. 样本来源

本文的样本来自于昆明高新科创小额贷款公司, 该公司是经云南省人民政府金融办公室批准, 在工商行政管理部门依法登记的专业小额贷款服务机构, 同时也是昆明市具有相当规模和影响力的小额贷款公司之一。本文的数据主要来自该公司创建的客户管理信息系统, 共 150 个样本, 其中 14 个样本存在缺

失值，予以剔除，剩余的 136 个样本用于模型的建立。根据一般的研究惯例，推导一个参数至少需要 10 个样本，本文共需要推导 9 个参数，故满足条件。同时用于建模的 136 个样本中，有 36 个违约案例和 100 个未违约案例。

3.2. 指标预处理

由于指标中既有定性指标，又有定量指标，不能直接用收集到的数据建立模型，必须对指标进行预处理。对于定性指标，邀请专家结合行业背景及领域相关知识将其指标值按照与信用风险的关系进行量化，详见表 2。

3.3. Logistic 模型

使用 SPSS 软件对样本数据进行初次拟合，由于参数的显著性检验结果显示在所有八个自变量中， X_1, X_6 没有通过显著性检验，因此，剔除变量 X_1, X_6 ，用剩余 6 个变量 $X_2, X_3, X_4, X_5, X_7, X_8$ 重新建立 Logistic 模型，称为修正 Logistic 模型，其参数估计结果见表 3。

从表 3 可以发现，所有的未知参数都通过了显著性检验，因而可得到最终的信用风险评估 Logistic 模型：

Table 2. Definition of variables
表 2. 变量的定义

变量	指标	定义
X_1	年龄	小于 30 岁，取 2；30 ≤ 年龄 < 40，取 3；40 ≤ 年龄 < 50，取 4；50 及以上，取 5
X_2	学历	小学及以下 1；初中 2；高中 3；技校 4；大专 5；本科及以上 6
X_3	人民银行逾期记录	逾期次数实际值
X_4	婚姻状况	已婚 1；未婚 2；离异 3；丧偶 4
X_5	居住状况	自有 1；按揭/抵押 2；借住亲戚朋友家 3；租住 4
X_6	资产负债率	实际值
X_7	流动比率	实际值
X_8	销售毛利率	实际值
Y	违约倾向	违约，未按时还款 1；未违约，按时还款 0

Table 3. Estimate of unknown parameters in Logistic model
表 3. Logistic 模型参数估计

	B	S.E.	Wald	df	Sig.	Exp(B)
X_2	-1.886	0.690	7.477	1	0.006	0.152
X_3	2.012	0.570	12.470	1	0.000	7.477
X_4	5.483	2.201	6.204	1	0.013	240.510
X_5	3.298	1.221	7.298	1	0.007	27.048
X_7	-1.023	0.352	8.452	1	0.004	0.360
X_8	-26.294	8.638	9.265	1	0.002	0.000
Constant	-3.773	2.823	1.787	1	0.181	0.023

a. Variable(s) entered on step 1: $X_2, X_3, X_4, X_5, X_7, X_8$.

$$\pi = \frac{\exp(-3.773 - 1.886X_2 + 2.012X_3 + 5.483X_4 + 3.298X_5 - 1.023X_7 - 26.294X_8)}{1 + \exp(-3.773 - 1.886X_2 + 2.012X_3 + 5.483X_4 + 3.298X_5 - 1.023X_7 - 26.294X_8)} \quad (6)$$

3.4. Probit 模型

使用 STATA 软件对样本数据进行初次拟合,结果同样显示在八个自变量中, X_1, X_6 没有通过显著性检验,因此剔除 X_1, X_6 ,用剩余 6 个变量 $X_2, X_3, X_4, X_5, X_7, X_8$ 重新建立 Probit 模型,称为修正 Probit 模型,拟合结果如下:

根据表 4 的结果,用 $X_2, X_3, X_4, X_5, X_7, X_8$ 六个自变量建立的 Probit 模型中各个自变量均通过了显著性检验。该信用风险评估 Probit 模型为:

$$Y^* = -1.7193 - 1.0698X_2 + 1.1297X_3 + 2.8920X_4 + 1.7793X_5 - 0.5724X_7 - 14.0730X_8 \quad (7)$$

3.5. Logistic 模型与 Probit 模型预测比较

为了比较 Logistic 模型和 Probit 模型在中小企业信用风险评估的预测效果,我们从公司的客户管理信息系统中再随机抽取 50 个客户资料,分别用两个模型对其进行信用风险的评估,并与实际结果相比较。检验结果如表 5 所示。

从表 5 的结果可以看出, Logistic 模型与 Probit 模型预测的正确率总体上相差不大,其中 Logistic 模

Table 4. Estimate of unknown parameters in Probit model

表 4. 修正 Probit 模型参数估计

	B	S.E.	Wald	df	Sig.	Exp(B)
X_2	-1.069802	0.3696592	-2.89	0.004	-1.79432	-0.3452829
X_3	1.129691	0.2951752	3.83	0.000	0.5511583	1.708224
X_4	2.891952	1.132011	2.55	0.011	0.673251	5.110652
X_5	1.779307	0.6144626	2.90	0.004	0.5749824	2.983632
X_7	-0.5724082	0.1812207	-3.16	0.002	-0.9275942	-2.2172222
X_8	-14.07299	4.349068	-3.24	0.001	-22.59701	-5.548978
_cons	-1.719339	1.509495	-1.14	0.255	-4.677896	1.239218

Table 5. Contrast of prediction of Logistic and Probit model

表 5. 模型预测比较

模型	分类预测	观察值		预测正确数	
		1	0		
Logistic	预测值	1	28	3	28
		0	7	12	12
	预测正确率	80%	80%	80%	80%
	预测错误率	20%	20%	20%	20%
Probit	预测值	1	26	2	26
		0	9	13	13
	预测正确率	74.3%	86.7%	78%	78%
	预测错误率	25.7%	13.3%	22%	22%

型在预测违约事件发生时准确率更高，而 Probit 模型在预测无违约风险发生时的准确率更高。因此在实际应用中，可以将二者结合起来，如果两个模型判断的结果都认为会违约，则不予贷款；如果两个模型判断的结果都认为不会违约，则做出贷款的决策意见；如果两者判断结果不一致时，则需要针对具体问题进一步分析。

4. 结论与展望

4.1. 研究结论

本文利用 Logistic 模型与 Probit 模型对小额贷款公司中小企业信用贷款风险进行了定量分析，得到了如下结论：

第一， X_1, X_6 在两个模型中均没有通过参数的显著性检验。这说明，在中小企业的小额贷款中，贷款者的年龄和企业的资产负债率对信用贷款风险的影响不显著。虽然前文提到有国外学者研究表明年龄与信用风险之间存在一定关系，但是在小额贷款中借款客户的年龄对其贷款风险的影响并不大。这也印证了由于不同贷款机构信用风险特征和形成机理的差异，关于信用风险的结论很难达成高度的一致性的结论，因此在研究时应该结合机构自身特征、信用风险特殊性和实际运作特征进行具体分析。

资产负债率作为企业长期偿债能力的重要指标对贷款风险的影响也不显著，分析原因在于由于小额贷款“高利率、高风险”的特征，因此借贷双方出于自身利益的考虑都更倾向于一年内的短期贷款，而实际情况也确实如此，因此反映短期偿债能力的经济指标流动比率通过了显著性检验而资产负债率没有。

第二，在通过参数显著性检验的六个自变量 $X_2, X_3, X_4, X_5, X_7, X_8$ 中， X_2, X_7, X_8 之前的系数符号均为负，而 X_3, X_4, X_5 的系数均为正。这说明学历、流动比率、销售毛利率越高，信用风险越小；人民银行逾期记录、婚姻状况、居住状况越差，信用风险越高。这与之前的分析结果相一致，说明模型参数能很好体现自变量与因变量之间的关系。

4.2. 研究展望

虽然本文针对小额贷款公司不同发展阶段的实际情况建立不同的信用风险模型，致力于为公司决策者提供量化的数据参考，但是信用风险模型的建立并不是一劳永逸的，还有很多方面需要改进。

第一，考虑到样本量的限制，本文在指标预处理时对定性指标值的量化采取的是专家意见。事实上这种处理方法具有一定的主观性。伴随着小额贷款公司的发展，业务的扩大，信息管理的规范，可被利用的有效数据越来越多，在以后的研究中可以考虑在模型中引入虚拟变量，让数据说话，以便更客观、更精确的评估各定性指标与信用风险之间的关系。

第二，本文只选取了 Logistic 和 Probit 两个操作性较强的模型来评估信用贷款风险。事实上，信用风险的评估方法非常多，不仅有单一的模型，通过有机组合还有众多的组合模型，在以后的研究中可以尝试用其他更多的模型来做信用风险的评估，从而提高预测的准确性。

第三，本文对于客户的信用状况只分为两类：违约和未违约，在以后的研究中，可以考虑将客户的信用状况做更加详细的分级，提高判断的精度。

客户的信用状况受很多因素的影响，大到社会风气，小到个人心理，因此没有一个信用风险评估模型是永久适用的，必须要根据实际情况的发展变化不断调整。每隔一段时间，应该对模型的适用性进行检验，如果发现模型在预测上存在规模上的较大偏差，应当重新进行数据拟合，甚至改进现有模型。

致 谢

感谢编辑和审稿人对本论文提出的宝贵意见，感谢国家自然科学基金委和云南省教育厅对本研究的

支持。

基金项目

该研究得到了国家青年自然科学基金项目(11301464)和云南省教育厅科学研究基金(2013Y360)的资助。

参考文献 (References)

- [1] 陈赛 (2012) 完善小额贷款公司信用管理机制研究——以 A 小额贷款公司信用管理机制为例. 硕士学位论文, 华东师范大学, 上海.
- [2] (2014) 2014 年一季度小额贷款公司数据统计报告.
<http://www.pbc.gov.cn/publish/goutongjiaoliu/524/2014/20140422154234421754813/20140422154234421754813.html>
- [3] Ohlson, J. (1980) Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, **18**, 1.
- [4] 刘琪 (2011) 小额贷款公司个人贷款信用风险评估研究——Logistic 和 Probit 组合模型运用. 硕士学位论文, 扬州大学, 扬州.
- [5] 郑昱 (2009) 基于 Probit 模型的个人信用风险实证研究. *上海金融*, **10**, 85-86.