

基于 logistic 模型商业银行借款企业违约概率的度量 ——以制造业上市公司为例

冯 蕾, 董继刚

(山东农业大学 经济管理学院, 山东 泰安 271000)

摘要: 新巴塞尔协议的要求以及利率市场化的逐步完成, 商业银行的风险管理已经成为重中之重, 其主要内容就是测算借款人的违约概率。文章运用 logistic 回归模型, 对 2011 年 - 2014 年我国 80 家制造业 A 股上市公司的 21 个指标进行比较分析, 得到以固定资产比率、流动负债率、每股收益以及资本积累率为显著影响变量的违约概率预测模型, 以期为商业银行建立起对借款企业违约风险的定量测试和评估系统以及完善贷款定价机制提供依据, 并提出相应的对策建议。

关键词: 商业银行; 违约概率; logistic 模型; 制造业企业

中图分类号: F830.33 **文献标识码:** A **文章编号:** 2095 - 0098(2016)06 - 0027 - 08

引 言

随着利率市场化改革的推进,《新巴塞尔协议》^①的着重要求、资本项目的逐步开放、以及银行业与国际接轨的不断深入, 商业银行加强风险管理变得愈加重要, 而我国在实际信贷管理中存在信贷评估与管理的错位和脱节, 对贷款风险尚没有合理的量化, 因此, 作为信用风险管理的主要内容, 能够建立合理考虑违约风险的贷款定价机制以及风险防范机制, 是亟待解决的现实问题。而关于商业银行借款企业违约概率的度量, 相对于国外先进银行仍处于一个相对落后的地位, 因此, 应深入研究贷款违约风险度量问题, 建立违约风险的定量测试和评估系统, 实行“一户一定”精细化贷款风险管理, 加强风险前控, 充实风险量化技术, 这有利于我国商业银行进行贷款风险定价, 对于提高我国商业银行的风险管理水平以及贷款定价能力, 有效防范信用风险, 均具有重要的理论及现实意义。

二、国内外研究现状

最初, 专家评定法、信用评级法和财务比率综合分析法等都是西方发达国家进行信用风险测量的常用古典信用风险度量方法。而信用风险定价模型的研究自 20 世纪 70 年代以来得到了长足的发展, 根据对违约的可预测程度以及企业的价值是否作为参数输入定价方程, 将信用风险定价的方法和模型分为结构化定价模型和简约化定价模型。Merton(1974)^[1]以 Black - Scholes 的期权定价技术为基础建立定价违约风险债务模型; Black&Cox(1976)^[2]在 Merton 模型的基础上, 提出了针对企业违约临界值的第一阶段模型; Longstaff&Schwarz(1995)^[3]等人提出第二代结构模型, 证明利率风险和违约风险是商业银行贷款定价研究的必要因素。

随着结构模型的发展, 基于强度的简约化模型迅速发展起来。Jarrow&Turnbull(1995)^[4]等人建立信用风险度量模型对金融衍生品的定价信用风险进行分析, 首次提出了简约化模型。简约化模型摒弃了结构化模型以企业资本结构来模型化违约风险的桎梏, 直接通过对债券价格与利率市场的数据进行分析, 来对违约

^① 2010 年 9 月 12 日, 由 27 个国家银行业监管部门和中央银行高级代表组成的巴塞尔银行监管委员会就《巴塞尔协议 III》的内容达成一致, 全球银行业正式步入巴塞尔协议 III 时代。

收稿日期: 2016 - 08 - 20

作者简介: 冯 蕾(1990 -), 女, 山东青州人, 硕士研究生, 研究方向为货币理论与政策;

董继刚(1961 -), 男, 山东郯城人, 教授, 博士生导师。

风险进行测量。以 Madan&Unal(2000)^[5]以及 zhou(2001)等学者为代表,更是将基于强度的简约化模型和基于期权方法的结构化模型进行整合,建立更加完善的信用风险定价模型。

相较于国外,由于我国违约风险度量的研究起步晚以及国内金融环境的独特性,现在仍处于学习、消化吸收、借鉴西方前沿理论、经验与先进度量方法的阶段。于立勇(2003)^[6]建立了基于神经网络的信用风险违约率模型;石晓军、陈殿左(2004)^[7]应用 KMV 模型对上市公司进行违约概率分析;林杰新、罗伟其和庞素琳(2005)^[8]利用 Bayes 判别分析法建立违约率模型;石晓军、肖远文和任若恩(2006)^[9]建立了边界 Logistic 回归违约率模型;刘迎春(2010)^[10]验证了 Logistic 回归模型对违约风险判断预测的准确性;庞素琳(2012)^[11]结合项目成功概率及激励相容性,建立违约风险下的信贷决策模型与机制;蒙震(2014)^[12]基于财务实力、管理实力、社会实力三个因子,建立中小企业信用风险评估内容的度量模型。而经过国内大量学者的深入研究,发现基于我国当今实际金融环境,最为实用的则是 Logistic 模型的违约概率测度方法。

Logistic 回归模型通过对现有客户按照违约和不违约样本进行 0-1 分类,选取财务指标作为解释变量,构建模型,计算出借款企业的违约概率。Logistic 回归模型作为传统违约概率度量方法,适用于二分类变量,具有简单、参数较少,不需要对变量做限制性统计假设,效果好的特点,而且该模型采用最大似然法比其他模型采用的最小二乘法所求的估计量准确性更高。目前,我国商业银行债务人资信资料欠缺、客户数据不完善、信用评级不权威以及数据获取难等问题限制了其他方法的使用,则 Logistic 模型在前提假设、变量选择、模型构建以及数据采集等方面都非常适合我国当前金融市场的实际情况。因此,本文采用 Logistic 回归模型对商业银行借款企业的违约概率进行测度。

三、实证分析

(一) 数据来源及样本选择

1. 数据的来源

本文数据来源于国泰安研究服务中心数据库中国上市公司财务指标分析数据库、上海证券交易所统计年鉴,目标企业的资讯由国内两大交易所上市买卖的 A 股各板块制造类上市公司提供^①。A 股市场的 ST 制造类上市公司几乎涉及了制造业的各二级下属行业,如:食品制作类企业、饮料制作类企业、家具工艺企业、纸制品企业、石油炼化和核燃料操作企业、药物研发企业、非金属矿物加工企业、有色金属精炼以及压延制作企业、金属加工品业、通用器材加工企业、专用器材加工企业、交通运输器材加工企业、电气机械器材加工企业、通信工具计算机以及其它电子器材制作企业、仪器仪表和书写办公用器材制作企业等,这极大的保证了样本的全面性及代表性,提高了研究的准确度。

2. 样本的选择

鉴于借款企业信息的保密,商业银行不对外公布相关数据,这就导致违约数据难以获得。ST(*ST)上市公司通常是连续两年(三年)出现亏损或者最近一个会计年度每股净资产低于股票面值,而且上市公司的财务数据经过审计后,其数据信息相对于非上市公司更为可靠;而且据 2014 年资料显示,制造业上市公司占我国上市公司总数的比重约 80%,且占财务危机企业的比重较大,制造业企业在我国上市公司中具有一定的代表性。因此,为保证研究的可操作性和可重复性,本文在依据制造业企业在规模和分方面类似的原则,在沪深两市的制造业上市企业中确定了 2011 年至 2014 年之间,由于财务情况异常而特别处理的 40 家 ST 企业,以及与 40 家 ST 企业相呼应的 40 家非 ST 企业,具体选择的依据如下:

目标企业配比的确定。在实际中被 ST 的上市公司在数量上要显著的少于财务状况良好的上市公司,许多研究者对相配比例是 1:1、1:2 还是 1:3,哪种相配比例更能提高模型准确性进行了研究,但是目前仍缺少明确的理论能够证明 1:2 或 1:3 准确度更高。因此,出于数据的对称性以及使用的普遍性,本文选取 1:1 的相配比例。

目标公司类别的确定。由于不同细分行业的行业状况以及竞争格局存在差异,财务数据的代表性以及出现违约概率的程度就存在差异,因此,本文根据国内证监会《上市公司行业分类指引》的规定要求,确定和所选择制造业 ST 公司位于相同行业的运营良好的公司组合配对,若行业里缺少恰当的配对公司,就从相似

^① 由于国内 A 股上市企业与 B 股和 H 股所使用的会计规则及计量方式存在差异,这就造成了财务方面某些数据难以比较,因此本文只选取 A 股上市企业的目标公司。

的行业里确定。

目标公司资产规模相匹配。公司的资产规模的大小与公司的偿债能力具有相关性,因此,本文按照同一细分行业资产规模相近的原则,进行 ST 公司与非 ST 公司的配对,这样就能有效的避免公司因资产规模的差异而影响数据指标的准确性。

(二) 模型构建及指标选择

商业银行借款企业违约概率的度量,即对借款企业发生违约的概率进行测度,一般分为“违约”和“未违约”两种状况,商业银行应该理性的综合上市公司各种财务指标的基础上对借款企业的违约概率进行判断,这是一个典型的二元决策问题。因此,本文采用 logistic 回归模型对商业银行借款企业的违约概率进行测度,选择上市公司财务指标作为模型中的解释变量,以公司是否违约作为被解释变量(发生违约即 ST 公司则为 1,未发生违约即非 ST 公司则为零),为确定对借款企业具有显著性影响的财务变量,构建如下 Logistic 回归模型。用 Logistic 回归模型分析违约概率,首先需要对 ε_i 进行定义, ε_i 是一系列影响违约概率因素的线性组合,也就是:

$$\varepsilon_i = \alpha + \sum_{k=1}^m \beta_k x_{ki} ,$$

$$P(y_i = 1 | x_i) = \frac{1}{1 + \exp \left[- \left(\alpha + \sum_{k=1}^m \beta_k x_{ki} \right) \right]} ,$$

经过 Logit 变化可将其转变为线性函数:

$$\ln \left[\frac{p_i}{1 - p_i} \right] = \alpha + \sum_{k=1}^m \beta_k x_{ki}$$

上式中, β_k ($k=1, 2, \dots, n$) 为第 k 个解释变量的回归系数。 β_k 为正,表示第 k 个变量对借款企业发生违约的概率具有正向的影响; β_k 为负,表示第 k 个变量对借款企业发生违约的概率具有负向的影响。 α 为常数项。等式左边取零或 1,等式右边 x_k 是预测违约概率的相关因素,如表 1 所示,本文分为七个方面选取 21 个财务指标作为预测公司违约概率的相关因素指标。

表 1 Logistic 回归预测违约概率相关因素指标

财务指标综合能力	具体财务指标(Logistic 回归模型中的代码)
长期偿债能力指标	资产负债率(ZCFZL)
	流动资产比率(LDZCBL)
	固定资产比率(GDZCBL)
	权益对负债比率(QYDFZBL)
	利息保障倍数(LXBZBS)
	流动负债比率(LDFZBL)
短期偿债能力指标	流动比率(LDBL)
	速动比率(SDBL)
	营运资金比率(YYZJBL)
	营运资金对净资产比率(YYZJDJZCBL)
股东获利能力指标	每股收益(MGSY)
	每股净资产(MGJZC)
现金流量能力指标	现金流量比率(XJLLBL)
	营业收入现金净含量(YYSRXJBL)
	营业毛利率(YYMLL)
盈利能力指标	营业净利率(YYJLL)
	资产报酬率(ZCBCL)
	总资产净利润率(ZZCJLL)
	净资产收益率(JZCSYL)
发展能力指标	资本积累率(ZBJLL)
	总资产增长率(ZZCZZL)

(三) 模型检验结果分析

本文运用 SPSS 统计软件对样本公司进行实证分析,在 Logistic 回归模型中采用正向逐步选择法(forward stepwise)对解释变量进行检验,筛选出对违约概率具有显著性影响的指标因素。其中,对模型整体的检验是通过卡方统计量、Cox & Snell R Square 和 Nagelkerke R Square 进行判定,对模型的拟合度则根据似然比检验(likelihood Ratio Test),对变量系数的检验则根据基于卡方分布的 Wald 统计量。

1. 模型整体检验

表 2 样本处理摘要

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	320	98.8
	Missing Cases	4	1.2
	Total	324	100.0
Unselected Cases		0	0
Total		324	100.0

a. If weight is in effect, see classification table for the total number of cases.

如表 2 所示,本文选取了两组共 80 家 ST 或* ST 以及财务数据正常的制造业上市公司自 2011 年至 2014 年 324 组数据,根据数据分析的样本处理摘要显现,缺失了 4 组数据,这一般是因为数据中存在缺失值。

表 3 模型系数的混合检验

	Chi-square	df	Sig.
Step	5.823	1	.016
Block	133.812	4	.000
Model	133.812	4	.000

由表 3 可知,该模型的两个常用卡方统计量分别为:卡方值为 133.812,相应的自由度为 4, P 值为 0.00。利用 Excel 提供的 CHINV 函数计算显著水平为 0.05,自由度为 4,卡方分布的临界值为 9.488,而由上表可知本文计算的卡方值为 133.812 远远大于临界值,并且相应的 P 值小于 0.05,因此在显著性水平为 0.05 的情况下,每增加一个变量,增加变量后的方程仍然有意义,该模型系数的检验通过。

表 4 模型整体检验结果

统计量	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
检验值	209.802 ^a	.642	.756

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

表 4 给出最大似然平方的对数值 -2 Log likelihood = 209.802,该值在理论上服从卡方分布,由表 2 已知卡方临界值为 9.488,因此,最大似然对数值检验通过。判定系数 Cox & Snell R Square 和 Nagelkerke R Square 分别为 0.642、0.756,数值较大,说明该模型的拟合度较理想。

由表 2、表 3、表 4 的检验结果可以得到该模型的整体检验情况, χ^2 的检验结果为 133.812,在 1% 的水平下显著,说明该模型整体显著性很强,公司是否违约与反映公司长短期偿债能力、股东获利能力、现金流量能力、盈利能力、营运能力以及发展能力等七方面的财务指标变量之间存在显著关系。而判定系数 Cox & Snell R Square 和 Nagelkerke R Square 说明,公司发生贷款违约的 64.2% 的情形能够用以上反映公司七个方面的财务指标作自变量的估计的 Logistic 回归方程解释,作为 Cox & Snell R Square 的修正值 Nagelkerke R Square 为 0.756,则说明新增加变量将对模型具有较强的解释力度。总而言之,通过对 Logistic 回归模型的整体检验,该模型的整体拟合优度较好。

2. 判别效果

在实际应用中,通常采用分类表(Classification Table)判别 Logistic 回归预测模型的拟合效果,结果如下:

表 5 Logistic 回归的判别效果表

Observed		Predicted		Percentage Correct
		SFWY		
		未发生违约	发生违约	
SFWY	未发生违约	128	32	80.0
	发生违约	39	121	75.6
Overall Percentage				77.8

从 Logistic 回归的判别结果表 5 来看,该模型具有较强的预测能力,以预测概率 0.5 为判别分界点,总体预测准确度达到 77.8%。其中对于未发生违约的上市公司的 160 组数据中,判断正确的有 128 组,判断错误的有 32 组,准确度为 80%;对于发生违约的上市公司的 160 组数据中,判断正确的有 121 组,判断错误的有 39 组,准确度达到 75.6%。该模型判断的整体准确度是较高的。从两类公司数据的判别来看,模型对于出现违约的概率判断要稍低于对正常公司的判断,同时对公司违约有低估的可能。

3. 参数估计及检验

在 Logistic 回归中检验一个解释变量是否应该包含于模型当中,检验每个解释变量在回归方程中的重要性,所采用的检验统计量为 Wald 统计量。直接比较 Wald 统计量(或 Sig 值)的大小,Wald 统计量大者(或 Sig 值小者)显著性高,也就更重要。另外,Wald 统计量近似服从于自由度等于参数个数的卡方分布。其表示如下:

$$Wald = \frac{b^2}{\text{var}(b)}$$

检验结果如表 6-1、表 6-2 所示。

表 6-1 Logistic 回归参数估计及检验

Variables in the Equation	B	S. E.	Wald	df	Sig.	Exp(B)
GDZCBL	6.726	1.327	25.698	1	.000	834.036
LDFZBL	2.950	.997	8.755	1	.003	19.114
MGSY	-4.331	.766	31.926	1	.000	.013
ZBJLL	-.355	.181	3.861	1	.049	.701
Constant	-3.630	1.001	13.141	1	.000	.027

表 6-2 Logistic 回归参数估计及检验

Variables not in the Equation ^a	Score	df	Sig.
LDBL	975	1	323
SDBL	952	1	329
YYZJBL	926	1	336
YYZJDJZCBL	451	1	502
ZCFZL	511	1	475
LDZCBL	1.003	1	317
QYDFZBL	1.084	1	298
LXBZBS	062	1	803
MGJZC	1.238	1	266
XJLLBL	1.047	1	306
YYSRXJBL	681	1	409
YYMLL	1.093	1	296
YYJLL	287	1	592
ZCBCL	514	1	474
ZZCJLL	485	1	486
JZCSYL	070	1	791
ZZCZZL	174	1	676

a. Residual Chi - Squares are not computed because of redundancies.

如表 6-1 所示, GDZCBL(固定资产比率)、LDFZBL(流动负债比率)、MGSY(每股收益)、ZBJLL(资本积

资产积累率(ZBJLL)的系数为 -0.355 ,对制造业企业发生违约贷款具有负向的影响。资本积累率即股东权益增长率,该指标越高,表明企业的资本积累越多,企业资本保全性越强,应付风险、持续发展的能力越大,该企业发生违约贷款的可能性就越小,是企业发展强盛的标志。

4. 预测模型正确性检验

图1所示是观测值和预测概率分布图。该图以0和1为符号,横坐标是个案属于1的隶属度,这里称为预测概率,纵坐标是个案分布频数,反映个案的分布。如果逻辑回归预测完全准确,那么该坐标途中预测概率是 $0-0.5$ 之间的个案都应该是0, $0.5-1$ 之间的个案都应该是1。本文实证生成的图中预测概率在 $0-0.5$ 之间的个案基本都是0, $0.5-1$ 之间基本都是1,这说明本文的逻辑回归预测模型基本正确。

四、结论及对策建议

(一) 结论

本文通过实证发现,商业银行借款企业的违约概率与固定资产比率、流动负债比率、每股收益、资本积累率四个变量的关系显著,这四个变量分别反映了偿债能力、股东获利能力、盈利能力以及发展能力等方面的情况。所得到的预测模型既有利于商业银行对借款企业的违约概率及预期损失进行更精确的度量,也有利于商业银行进一步对其经济资本的计算提供依据。值得注意的是,实证中采用上市公司的财务数据是有效的,但是由于样本选取的局限性以及上市公司披露虚假财务数据情况的可能,这可能对违约风险模型的准确性造成影响,从而影响对商业银行借款企业违约概率度量的研究。

(二) 对策建议

构建贷前风险评估体系。随着利率市场化的逐渐完成,获得自主定价权的商业银行要想在金融市场日趋激烈环境中生存和发展,其在制定贷款价格时就必须考虑借款企业的信用风险,而本文通过对借款企业违约风险进行度量,有利于商业银行对借款企业进行信用风险评估,构建贷前风险评估体系,进而为商业银行对借款企业实行差别化的贷款定价策略提供依据,增强定价灵活性,也有利于对客户实行精细化营销,在竞争中占据优势。

进一步提高风险计量水平。用量化的数学模型对风险进行模拟与预测是商业银行进行信用风险管理的必然趋势,而我国商业银行现阶段对与信用风险的计量仍大多停留在定性分析,采用的量化工具也比较落后,远不能满足现代风险管理的要求,因此,商业银行应进一步提高风险计量水平,加快从经验判断法和财务指标分析法向模型分析法转变,侧重对违约概率和违约损失率测度模型的研究、开发和应用,形成与国际接轨的现代化信用风险管理制度。

检验信用风险度量模型的合理性。目前已有不少学者对商业银行借款企业的违约概率进行度量,但是所得到模型的合理性以及准确性仍需要进行更加深入的研究与鉴定,这样才能更快的投入到商业银行贷款定价及信用风险评估的实际应用中,增强商业银行的风险防范能力。

参考文献:

- [1] Merton Robert C. On the Pricing of Corporate Debt: The Risk Structure of Interest Rates [J]. Journal of Finance, American Finance Association, 1974(29): 449-470.
- [2] Black F, Cox J. C. Valuing corporate securities: some effects of bond indenture provision [J]. Journal of Finance, 1976, 31(2): 351-367.
- [3] Longstaff F, Schwartz E. A Simple Approach to Valuing Risky Fixed and Floating Rate Debt [J]. Journal of Finance, 1995, 50(3): 789.
- [4] Jarrow Robert A. and Stuart M. Turnbull. Pricing Derivatives on Financial Securities Subject to Credit Risk [J]. Journal of Finance, 1995, 50(3): 53-86.
- [5] Madan D, Haluk U. A. Factor Hazard Rate Model for Pricing Risky Debt and the Term Structure of Credit Spreads [J]. The Journal of Financial and Quantitative Analysis, 2000, 35(1): 43-65.

- [6]于立勇. 商业银行信用风险评估预测模型研究[J]. 管理科学学报 2003(5).
- [7]石晓军, 陈殿左. 债权结构、波动率与信用风险——对中国上市公司的实证研究[J]. 财经研究 2004(9).
- [8]林杰新, 罗伟其, 庞素琳. Bayes 判别信用评价模型及其应用研究[J]. 统计与决策 2005(2).
- [9]石晓军, 肖远文, 任若恩. 边界 Logistic 违约率模型 Bayes 分析及实证研究[J]. 中国管理科学 2006(8).
- [10]刘迎春. 基于 Logistic 回归的中国上市公司信用风险度量研究[J]. 黑龙江对外经贸 2010(11).
- [11]庞素琳. 违约风险下的信贷决策模型和机制[J]. 管理科学学报 2012(4).
- [12]蒙震. 基于信用风险评估的商业银行中小企业贷款定价影响机制研究[J]. 对外经济贸易大学学报, 2014(5).

The Measurement of the Probability of Default of the Borrowing Enterprises in Commercial Bank based on Logistic Model

——Take Manufacturing Listed Companies for Example

FENG Lei , DONG Jigang

(School of Economics and Management ,Shandong Agricultural University ,Tai'an ,Shandong 271000 ,China)

Abstract: With the requirements of the Basel III and the gradual completion of the interest rate market ,the venture management of commercial banks becomes a top priority; its main content is to measure the probability of default of the borrower. The thesis used the logistic regression model to compare and analyze the 21 indicators of 80 A – share listed companies in manufacturing industry in China from 2011 to 2014 ,get the default probability prediction model which had significant impact on variables such as fixed assets ratio ,current liabilities ratio ,earnings per share and the capital accumulation rate. This paper is aim to help commercial banks establish a quantitative test and evaluation system for the loan default risk and improve the loan pricing mechanism ,then put forward the corresponding countermeasures and suggestions.

Key words: commercial banks; probability of default; logistic model; manufacturing enterprise

(责任编辑: 沈 五)