

·方 略·

学生贷款违约风险影响因素的实证研究

廖茂忠

(中山大学 教育学院, 广东 广州 510275)

摘要:从国际比较视角观察,违约是学生贷款可持续发展的主要障碍。构筑有效的风险防范体系有赖于人们对违约因素的细致观察。本研究综合运用因子分析、判别分析、聚类分析和 Logistic 模型等方法,对 1690 个学生贷款样本进行了实证研究。研究发现,影响学生贷款违约风险的重要因素包括:学费债务、就业与年龄、家庭、偿还与时间以及还款期限。尽管在不同的违约群组中,影响学生贷款违约风险的重要因子存在一定的差异性,但总体而言,贷款的总债务水平是其中最为关键的因素。

关键词:学生贷款;违约;因子分析;判别分析;聚类分析;Logistic 回归

中图分类号:G467.8 **文献标识码:**A **文章编号:**1672-0059(2010)01-0062-06

An Empirical Study of Factors Influencing Student Loan Default

LIAO Mao-zhong

(School of Education, Sun Yat-sen University, Guangzhou 510275, Guangdong, China)

Abstract: From international perspective of observation, default composes the major obstacle to sustainable development of student loan. The establishment of an effective risk prevention system relies on meticulous observation and research on factors influencing loan default. Through a comprehensive use of factor analysis, discriminant analysis, cluster analysis and logistic model, the paper conducts an empirical study on 1690 student loan samples. It is found that the important factors influencing default risk of student loan include tuition debt, employment and age, family factor and repayment period. Although there are some discrepancies between different default groups, total debt level is the key factor that influences student loan default as a whole.

Key words: Student Loan; Default; Factor Analysis; Discriminant Analysis; Cluster Analysis; Logistic Model

一、引言

学生贷款违约是一个复杂的社会问题。它与社会环境、社会制度、文化传统、贷款设计,以及借款者个体的生活际遇和思想观念等密切相关。然而,贷款违约风险是一个彻彻底底的源于实践的问题。我国助学贷款的实施已有 10 年,有关国家助学贷款的研究成果也非常丰富。但是,一直以来,绝大多数研究所关注的重点是学生贷款制度的完善及如何推动学生贷款的实施,相对而言,对学生违约风险与贷款回收的研究重视不够。直到 2005 年 12 月,北京首次出现了大学生因贷款违约遭到起诉的案例,人们才开始真正直面贷款的违约与回收问题(金融机构除外)。另一方面,

中国的学生贷款研究行走在实践与政策之间,而且学生贷款研究在一定程度上属于“时政热点”研究的范畴。从目前所能收集到的文献来看,对学生贷款违约因素的规范实证研究少之又少,不同的研究者得出的结论或提出的对策往往存在惊人的一致性;在研究方法上,大多数停留在经验、感觉、思辨、推理的层面,很难深入。从总体上看,与国外同类研究相比,我们的研究还不成熟,对于现实世界的解释力贫乏。本文则尝试用规范的实证研究揭示影响中国学生贷款违约风险的主要因素。

二、研究模型的设计

人们在研究个人信贷违约风险中运用的研究方

法主要有:多元线性回归、Logistic 模型、Probit 模型、Tobit 模型、聚类分析、判别分析、期权分析等。在关于学生贷款的实证研究中,回归分析的研究方法被广泛采用。斯托克哈姆和赫斯埃尔登兹运用了判别分析的方法研究违约学生的个人特征,沃尔克等人在研究借款人的个人特征及所在院校特征对学生贷款违约的影响时采用了 Logistic 模型,格林等人采用了 Probit 模型和 Tobit 模型分析学生贷款违约的原因,而蒙蒂维德的相关研究则运用了生存分析的研究方法。

本研究借鉴前人研究的经验,综合运用因子分析、判别分析、聚类分析和 Logistic 模型对学生贷款违约风险的影响因素进行实证研究,如图 1。首先,由于样本中变量较多(包括了 27 个自变量),为了更加精确地揭示关键的影响因素和避免自变量的共线性,需要对自变量进行因子分析;其次,为进一步确定不同公因子在影响力方面的差别,我们对公因子进行判别分析;再次,通过聚类分析将加权后的公因子变量聚为不同的违约风险类别,从而为更加细致揭示影响因素奠定基础;最后通过建立 Logistic 模型揭示学生贷款违约的关键因素。

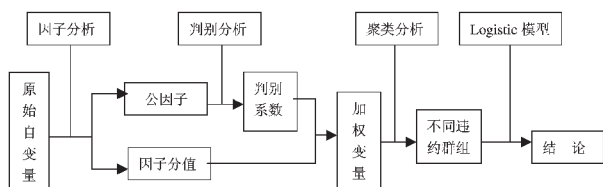


图 1

三、研究样本的选取及其描述性统计分析

1. 样本的选取。本研究所有数据均来源于我国东部某地方综合性本科院校。调查共获取样本 2280 例,经严格筛选,整理后最终确定的研究样本数为 1690 个,其中正常样本为 1287,违约样本为 403,违约率为 23.8%。样本中的借款人选取的范围是从 2001 年始到 2006 年 7 月进入还款期的借款学生,涵盖了不同学科专业、学制、学历,以及来自不同地域的学生。所选样本包含了借款人较为全面的个人基本信息。本研究选取了学生性别、学历、父母职业以及总债务负担等 27 个自变量。为方便统计分析,需要对这些变量进行合理化的量化。

2. 通过对调查样本进行描述性统计分析,得出学生贷款样本中的变量信息,初步研究学生贷款的违约情况。

从样本的描述统计中,我们发现:从借款学生性别来看,男生的违约比例远高于女生;从借款学生父母所从事的职业看,家长职业不同的学生其贷款状况

存在着一定的差异,但就均值而言并不明显;从借款学生是否属于师范专业看,师范专业的学生比非师范学生的还款状况要好一些;从借款学生离校的时间看,毕业后的时间越长,违约率偏向于更低,即随着借款学生毕业后时间的增加,贷款偿还状况趋向更好的方向发展;从是否在各学年度均贷款来看,各年度既贷了学费又贷了生活费的借款学生的违约率要明显高于非各年度均贷款学费和生活费的学生;从年度贷款额和债务总额来看,贷款额度和债务越大,违约的概率就越高;从借款学生季度需偿还额看,学生季度债务越重,违约情况就会越严重。为了更加深刻地揭示影响学生违约的关键因素,下面我们将对调查数据作进一步的模型处理。

四、模型估计的建立

1. 提取公因子与因子赋值

通过因子分析,从 27 个自变量中提出了 8 个公因子,分别为:学费债务因子(FAC_1),包括的变量有借款学生的总债务负担,学费贷款总额,学费贷款的总期数,第二、三学年度学费贷款,以及每季度的还款负担;生活费债务因子(FAC_2),包括的变量有生活费贷款总额,生活费贷款的总期数,第二、三学年度生活费贷款,是否各学年均贷款,以及是否首年度均贷款(包括学费贷款和生活费贷款);偿还与时间因子(FAC_3),包括的变量有借款学生已偿还的贷款比例,入学时间,已离校时长,以及还款截止时间;学历学制因子(FAC_4),包括借款学生的学历和学制两个变量;专业因子(FAC_5),包括的变量有借款学生是否师范专业和学科门类状况;还款期限因子(FAC_6),只涵盖了还款期限这一个变量;家庭因子(FAC_7),包括的变量有借款学生的父母职业状况和生源地状况;就业与年龄因子(FAC_8),涵盖了借款学生毕业时的年龄和就业状况。通过探索性因子分析,从这 27 个指标中找出了 8 个潜在因子。每个指标只受一个潜在因子的影响。根据因子得分系数矩阵和原始变量的标准化值来计算每个样本在相应因子上的得分。用数学形式表示为:

$$f_j = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ \mathbf{M} & & & \mathbf{M} \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_{1j} \\ \mathbf{M} \\ \beta_{nj} \end{pmatrix}$$

f_j 为赋值后的因子分值, β_{nj} 为因子得分系数矩阵。因子分值反映因子和各个变量间的密切程度,当各公因子间完全不相关时,因子负荷值就等于因子与变量的相关系数。

2. 求各公因子的判别系数值及其检验

通过 SPSS 判别分析,得到的结果,如表 1。

表 1 典型判别函数分析主要结果

因子	名称	Function	Wilks' Lambda	Tolerance	Sig.
FAC1	学费债务因子	.856**	.947	.994	.000
FAC2	生活费债务因子	.365**	.872	.997	.000
FAC3	偿还与时间因子	-.272**	.864	.999	.000
FAC4	学历学制因子	.059	1.000	1.000	.390
FAC5	专业因子	-.090	.999	1.000	.190
FAC6	还款期限因子	.422**	.878	.997	.000
FAC7	家庭因子	-.104	.999	1.000	.128
FAC8	就业与年龄因子	.179**	.859	.999	.009
常数		.000			

容许度(Tolerance)是方差膨胀因子(VIF)的倒数,当容许度值接近 1 时,表明自变量之间的共线性程度很低。表 1 中各公因子的容许度值都非常接近 1, 所以各公因子间的共线性程度特别低,判别分析有效。如表 1 所示,该判别函数中学费债务因子、生活费债务因子、偿还与时间因子、学历学制因子、专业因子、还款期限因子、家庭因子和就业与年龄因子的判别系数分别为:0.856、0.365、-0.272、0.059、-0.090、0.422、-0.104 和 0.179。

3. 通过聚类分析将样本聚为不同违约风险程度的类别

本研究借鉴学者在住房抵押贷款违约方面的研究方法,消除聚类时的不同量纲和多重共线性的影响,通过用判别函数系数的绝对值对因子变量进行加权生成新的聚类变量来做聚类分析。新聚类变量的生成的数学公式如下:

$$WFAC_i = |\beta| * FAC_i \quad (i=1,2,3,4,8)$$

其中, FAC_i 为判别分析中的因子得分值; β 是判别分析中计算出的典型判别函数的系数。

为了验证聚类分析的有效性,将总样本随机均分为两组。本研究采用离差平方和法 (Within Group) 和欧氏距离的平方 (Squared Euclidean distance) 对每组样本以不同类别的试聚类,反复对比各不同组别不同类别的聚类输出结果,最终,选取出最佳的聚类类数为 3 类。A、B 两组聚类学生贷款违约样本分布情况,如表 2 所示。

表 2 显示, A 组类别 1 的样本总量是 267 个,其中正常样本是 185 个,违约样本是 82 个,违约比例为 30.7%; B 组类别 2 的样本总量是 247,其中正常样本是 170 个,违约样本是 77 个,违约比例为 31.2%。因而,可以认为 A

表 2 A、B 两样本组聚类分析结果

组别	类别	总样本量	其中		违约比例
			# 正常	# 违约	
A	1	267	185	82	30.7%
B	2	247	170	77	31.2%
A	2	180	106	74	41.1%
B	1	197	116	81	41.1%
A	3	398	350	48	12.1%
B	3	401	360	41	10.2%
合计		1690	1287	403	23.8%

组类别 1 和 B 组类别 2 基本匹配。类似地, A 组类别 2 和 B 组类别 1 基本匹配; A 组类别 3 和 B 组类别 3 基本匹配。对比 A、B 两组的各类别,可以发现,从聚类结果的样本量和违约样本在各自分组类别中的比例来看, A、B 两组的聚类结果具有内在的一致性, 两组各类别间的样本信息基本能够匹配。为方便整体研究,将 A、B 两组的聚类结果进行合并。

在聚类分析中已将 1690 个样本分为低、中、高三三个违约风险组,然而这三组的划分是否科学呢? 我们通过新生成的公因子分值来检验,如表 3。

以学费债务因子 (FAC_1) 为例,低违约组的 FAC_1 因子得分的均值为 -0.912, 中违约组的均值则达到 0.954,高违约组的均值为 0.633,总样本均值为 0.000 (这是因为在计算因子分值时对原始变量进行了标准化,各因子得分的均值为 0.000,标准差为 1.000,服从标准正态分布), 这表明不同违约组的学费债务因子得分差异显著。类似地,其他七个公因子的因子得分差异也很显著。这也进一步说明经聚类分析生成的三个不同违约组的可信度较高。

4. 高、中、低违约风险组的 Logistic 模型的建立

在前面因子分析、判别分析和聚类分析的基础上,将原有的 27 个原始自变量转化为 8 个有充分解释力的公因子,并将 1690 个样本按学生贷款的状况分为低违约组、中违约组和高违约组三类。低违约组

表 3 聚类分析后新生成的公因子分值比较表

	低违约	低违约	中违约	中违约	高违约	高违约	总样本	总样本
	均值	标准差	均值	标准差	均值	标准差	均值	标准差
FAC_1	-0.912	0.353	0.954	0.626	0.633	0.529	0.000	1.000
FAC_2	-0.135	0.507	-0.904	0.328	1.520	0.571	0.000	1.000
FAC_3	-0.099	1.278	0.091	0.622	0.086	0.675	0.000	1.000
FAC_4	0.023	1.028	-0.008	1.009	-0.036	0.926	0.000	1.000
FAC_5	0.012	0.976	-0.034	0.997	0.020	1.055	0.000	1.000
FAC_6	-0.200	1.249	0.214	0.720	0.133	0.553	0.000	1.000
FAC_7	-0.029	0.934	0.054	1.035	-0.011	1.082	0.000	1.000
FAC_8	-0.011	1.099	0.052	0.906	-0.046	0.895	0.000	1.000

包括了799个样本,占总样本的47.3%;中违约组包括了514个样本,占总样本的30.4%;高违约组则包括了377个样本,占总样本的22.3%。这为进一步细致研究贷款违约的影响因素奠定了基础。

本研究所考察学生贷款类型分为正常和违约,即因变量为二分类变量,而不属于连续变量,因此,对于二分类因变量的分析需要使用非线性函数。在二分类因变量分析中,最常用的函数是Logistic分布。Logistic回归分析是一种非线性分类的统计方法,适用于因变量中存在定性指标的问题,而且Logistic判别函数的建立方法——极大似然估计法有很好的统计特性,因此是分析违约风险的常用数学工具之一。本研究运用Logistic模型分析来识别影响学生贷款违约风险的主要因素,分析结果如表4。

高违约组Logistic模型估计中,学费债务因子(FAC₁)的系数估计值为1.244,其符号是正号,表明借款学生的学费债务水平越高,贷款发生违约的可能性就越大。学费债务因子的优势比估计值 $\text{Exp}(B)=3.470$,即在其他自变量值不变的情况下,学费债务因子分值每增加1,相应的违约的优势比的自然对数值为3.470,也就是说,学费债务因子分值每增加1,相应的贷款违约的概率就增加3.470倍。偿还与时间因子(FAC₃)的系数估计值为-0.625,其符号为负号,表明偿还与时间因子的分值越高,贷款发生违约的可能性就越小。偿还与时间因子优势比估计值 $\text{Exp}(B)=0.535$,即在其他自变量值不变的情况下,偿还与时间因子分值每增加1,相应的学生贷款违约的概率就减小0.535倍。由于偿还与时间因子是由季度已偿还比、入学时间、毕业时长以及还款截止时间等四个变量支配的,这四个变量在偿还与时间因子中的因子荷重分别为:0.835、0.807、-0.793和0.620,这表明季度已偿还比越高,入学时间越早、还款截止日期越早,学生贷款违约的可能性就越大,而毕业离校时间越长,贷款违约的

概率则越低。还款期限因子(FAC₆)的系数估计值为0.927,且符号为正号,表明还款期限越长,贷款发生违约的可能性越大。优势比估计值 $\text{Exp}(B)=2.527$,即在其他自变量值不变的情况下,还款期限因子分值每增加1,相应的学生贷款违约的概率就增加2.527倍。FAC₁、FAC₃、FAC₆的系数值对应的P值都<0.01,表明它们都在99%的置信水平上显著。与此同时,生活费债务因子、学历学制因子、专业因子、家庭因子,以及就业与年龄因子的系数估计未能通过显著性水平检验(95%的置信水平检验),所以它们对学生贷款质量的影响并不明显。因此,比较各因子系数估计值的绝对值及其显著性水平,可以得出:在高违约组中,学费债务因子、偿还与时间因子和还款期限因子是影响学生贷款违约风险的重要因子;按影响度排序为:学费债务因子、还款期限因子和偿还与时间因子。

类似地,中违约组Logistic模型估计中,学费债务因子、偿还与时间因子、还款期限因子和就业与年龄因子是影响学生贷款违约风险的重要因子;按影响度排序为:学费债务因子、偿还与时间因子、还款期限因子和就业与年龄因子。低违约风险组中,学费债务因子、生活费债务因子、还款期限因子和家庭因子是影响学生贷款违约风险的重要因子;按影响度排序为:学费债务因子、生活费债务因子、还款期限因子和家庭因子。

从模型的准确性看,Logistic模型估计的预测准确率较好,如表5所示。高违约风险组Logistic模型整体预测准确率为67.9%,其中正常学生贷款预测的准确率为79.7%,违约样本预测的准确率为51.0%,所以,从总体上看,模型预测准确程度较好;中违约风险组整体预测准确率为86%,其中正常学生贷款预测的准确率为93.2%,违约样本预测的准确率为69.8%,这说明模型预测准确程度较好;低违约风险组模型的整体预测准确率为89.0%,其中正常学生贷款预测的准确率

表4 Logistic模型分析结果

	高违约			中违约			低违约		
	B	Sig.	Exp(B)	B	Sig.	Exp(B)	B	Sig.	Exp(B)
FAC ₁	1.244	.000	3.470	1.026	.000	2.789	1.484	.001	4.412
FAC ₂	.437	.101	1.548	.230	.493	1.259	.669	.004	1.953
FAC ₃	-.625	.004	.535	-.843	.000	.430	-.214	.086	.808
FAC ₄	.047	.705	1.048	.057	.574	1.059	.064	.578	1.066
FAC ₅	-.193	.090	.825	-.055	.601	.947	.016	.902	1.016
FAC ₆	.927	.002	2.527	.484	.004	1.622	.628	.000	1.874
FAC ₇	.033	.746	1.034	-.098	.345	.906	-.417	.008	.659
FAC ₈	.206	.108	1.229	.288	.013	1.333	.113	.256	1.119
常数	-1.937	.000	.144	-1.760	.000	.172	-.848	.037	.428

表5 Logistic 模型预测效果表

	高违约			中违约			低违约		
	正常	违约	准确率	正常	违约	准确率	正常	违约	准确率
正常	177	45	79.7	331	24	93.2	710	0	100.0
违约	76	79	51.0	48	111	69.8	88	1	1.1
总样本	253	124	67.9	379	135	86.0	788	1	89.0

为100%,违约样本预测的准确率为1.1%,违约样本预测准确度很低的一个重要原因可能是该组样本中违约样本所占的比例太低(10.2%)。不过,从总体上评价,上述 Logistic 模型预测准确率较为理想。

五、结论与讨论

通过以上实证研究的方法,我们从微观的层次研究了影响学生贷款违约的因素。

对上述研究进行归纳,我们发现:(1)学生违约受多重因素的影响。从总体上看,学生贷款规模(贷款的额度大小)是影响学生违约的关键因素。具体而言,判别分析表明,学费债务对违约的影响力要高于生活费债务的影响力。学生学费的贷款额占了贷款总额的大部分,即是说,从一般意义上讲,学费债务水平可以代表总的债务水平。相对于不同的违约风险的人群,违约的影响因素存在着一定的差异性。(2)在判别分析中(如表1),学历、专业、学位和家庭相关因子都未能通过显著性检验。这表明,借款学生的学历、专业、学业和家庭的差异对学生违约的影响并不显著。在 Logistic 模型分析中,对于低违约风险组而言,家庭因子通过了显著性检验,但在高违约风险组和中违约风险组该因子都未能通过显著性检验。所以,总体看来,借款学生的学历、专业、学业和家庭状况对学生违约的影响不大。(3)借款学生沉重的偿还负担是导致贷款违约的重要因素。判别分析和 Logistic 模型分析一致表明,借款学生期还款负担越重,学生违约率就越高。在描述统计中也显示,在高违约样本中月均需偿还的贷款额接近400元,这远远高于国际上普遍认同的还款负担率水平。(4)还款期限与违约成反比,这是本研究比较特殊的发现。从表面上看,这似乎不太正常。还款负担越重违约率越高,而影响还款负担率的主要因素有三个(债务总额、收入水平和还款期限),所以还款期限越长,还款负担率就越低。但这是在“债务总额和收入水平”不变的前提下成立的。通过进一步的研究分析发现,还款期限比较短的借款学生,往往贷款总额比较少,因此,相对的债务负担就较适度,所以这类学生的还款情况要好一些。(5)与国外的许多研究的结论相区别,本研究发现,借款学生的学历、

学制和专业对学生贷款违约有一定的影响,但并不显著。我们认为,这主要是因为不同学历、学制和专业的学生在就业和收入方面的差异并不是特别显著,这可能与本研究中研究生学历、五年制学生的样本数量偏少有关。(6)借款学生的家庭因素对学生贷款违约有一定的影响,但并不显著。这主要与目前我国银行贷款的对象绝大多数是来自经济困难的家庭,同质性较高有关。

本研究从较大规模的定量分析中得出了上述重要结论,其中最为核心的是:借款学生的债务水平是影响学生是否按时偿还贷款的最关键因素。基于此,我们认为偿还能力是学生还贷行为理论的物质基础。学生的收入水平是影响学生偿还能力的重要力量。只有较稳定、较充足的经济收入,才能保障还贷的能力。从广义上讲,个人的收入状况受制于当下的社会经济发展程度,工资薪酬制度,以及地区间和行业间的收入水平。然而,教育的质量,或者说是大学毕业生在学校所积累的价值观、专业素养和技能,以及社会竞争力等,对于大学毕业生当下就业质量(选择)及其未来收入水平的提高却有深刻而持久的影响。由于教育的特殊性,学生在学习期间的学业状况既不完全取决于学生本人,也不能完全归结于学校,而是双方共同努力的结果。但是学校却承载着学生、家长乃至社会对教育价值的希望和使命——这也是高校存在的一个基础。也就是说,学校在提高学生的知识、技能和竞争力等方面负有不可推卸的责任。高校不再是远离学生生计问题的净土,高校必须为学生的将来谋福利。对于借款学生来说,这一点显得尤为必要。经济困难只是他们面临的困难之一,除此之外,他们还时常面临知识基础、学习能力、社会认同以及社会交往等方面的困难。因此,需要对借款学生倾注更多地关注,以增强借款学生的学习信心和学习能力,从而帮助他们获得学业上的成功,只有这样才能最终提高借款学生的就业竞争力和还贷能力。此外,高校的专业与课程设置甚至教学方式等也应该始终围绕培养学生能力进行不断地完善和改革。归根结底,高校教育质量的提高是学生偿还能力提高的根本保障。

(下转第74页)