

基于 Logistic 模型的住房抵押贷款违约风险实证研究

王晓君, 骆 桦

(浙江理工大学理学院, 杭州 310018)

摘 要: 住房抵押贷款已成为我国商业银行拓展信贷营销业务、优化信贷资产结构的主要手段。但是随着中国住房抵押贷款的发展, 风险也日益突出。在实地调研的基础上, 利用 Logistic 回归模型对影响我国住房抵押贷款违约风险的因素进行了实证分析, 研究发现, 借款人每月偿还贷款金额与家庭收入之比越高, 违约率就越高; 借款人年龄越大, 违约率就越高; 借款人受教育程度越高, 违约率就越低; 借款人工作稳定性越高, 违约率就越低。这为商业银行在住房抵押贷款发放时管理违约风险提供了有益的帮助。

关键词: 住房抵押贷款; 违约风险; Logistic 回归模型

中图分类号: F830.572

文献标志码: A

0 引 言

近年来, 住房抵押贷款已成为我国商业银行具有强劲发展势头的“零售业务”, 被视为拓展信贷营销业务、优化信贷资产结构的主要手段^[1]。但随着中国个人住房抵押贷款的急剧膨胀, 不良贷款和不良贷款率等不良风险指标也日益突显, 借款人违约风险有继续增加的趋势。个人住房按揭贷款违约会给贷款银行等金融机构带来预期利润损失。更值得关注的是, 一旦出现个人住房按揭贷款大规模违约, 势必会给整个金融系统带来巨大的冲击, 而这种资产价格变化所带来的冲击在金融加速器的作用下, 会对宏观经济产生巨大影响^[2]。因此, 需要研究住房按揭贷款违约风险的主要影响因素及其影响程度。

1 住房抵押贷款违约风险的微观分析

随着住房成交量的上升, 银行住房抵押贷款的违约率也有提升的趋势。很多年轻人由于担心房价上涨, 在还没具有较大的经济实力时就选择买房, 结果一旦工作出了问题, 便直接影响他们按月还贷。

另外, 有很多房贷违约者把买房作为一种投资, 一旦投资失误, 就选择违约。因此, 银行在办理住房抵押贷款业务时要特别注意借款人的一些指标特征。

1.1 借款人特征维度

借款人的特征维度变量主要包括家庭收入、职业稳定性、年龄、婚姻状况、学历等。

a) 家庭收入。借款人的家庭收入影响其还贷情况, 是衡量借款人偿还抵押贷款能力的一个重要指标。我国已婚家庭绝大多数是双薪家庭, 借款人个人收入是家庭总收入的一部分, 总收入说明借款人的债务偿还能力^[3]。

b) 借款人职业。借款人收入的不稳定性可能导致住房抵押贷款违约风险。但是由于借款人未来收入的稳定性无法测定, 因此考虑用借款人的职业代替, 一般来说, 垄断行业、政府机构相对较稳定, 个体私营企业稳定性较弱。

c) 借款人年龄。人们往往认为不同年龄的借款人风险倾向和赚钱能力是不同的。一般情况下, 年轻人考虑不周全, 倾向于冒险, 而且刚开始工作收入较少。但是随着年龄的增加, 收入逐渐稳定, 考虑问题也更加周全。而当人到了退休年龄后, 思考问

题就比较全面,因而倾向于保守,但是其收入也会有一定的下降。因此,不同的风险倾向和收入水平使得各个年龄层的贷款风险有所不同。

d) 婚姻状况。婚姻状况的不同可能对应着不同的家庭收入来源。已婚家庭往往是双薪的,而单身的家庭,其收入可能只有单薪。因此已婚借款人家庭收入往往更为稳定。因而将婚姻状况作为一个重要变量予以考虑。

e) 最高学历。借款人的学历越高,往往预示着其工作越稳定,收入水平越高。而且学历高的人考虑问题更加全面,可能不会贸然违约^[4]。

1.2 贷款特征维度

贷款特征维度主要包括贷款金额、期限、月还款额占收入比等。

a) 贷款金额。贷款金额越多,表明其债务负担越大,当借款人财务状况恶化的时候选择违约的可能性就较大。而且对于银行来讲,单笔贷款金额越大,一旦借款人违约,银行遭受的损失也越大。

b) 贷款期限。贷款期限越短,借款人前期的还款使其在住房权益中累积的速度就越快,对借款人来说违约对其所造成的损失越大。所以,若借款人的贷款期限短,则其由于意外或者工作变动等原因导致的违约风险相对较小。

c) 月还款额占家庭收入比。月还款额占家庭收入比直接反映了借款人在该项债务上的负担情况,比例越大说明借款人债务负担越重,将来发生财务危机的可能性越大。Williams 等和王福林等在分析违约决策时发现,相比其他的借款人,那些月还款额占家庭收入比大于 30% 的借款人违约的可能性更大^[5-6]。

2 Logistic 回归模型

住房抵押贷款违约风险问题,主要表现为借款人违约和不违约两种离散现象,笔者采用二分法表示贷款是否违约这一被解释变量,赋值为 0 和 1。而 Logistic 回归模型的最大优点在于它解决了因变量不连续的回归问题,特别是因变量是分类变量时非常适合使用该模型进行研究。所以,把二分类 Logistic 模型作为一种适当的估计模型,应用 Logistic 回归分析可识别自变量对住房抵押贷款违约风险的影响方向,达到识别影响住房抵押贷款违约风险主要因素的目的。

2.1 基本模型

笔者构建的回归模型是以借款人特征为解释变量,以是否违约为被解释变量,其中被解释变量是否违约,可采用二分变量法表示,赋值为 0 和 1。应用 Logistic 模型进行实证分析,得出不同因子对违约的解释程度和违约发生的概率。Logistic 模型的具体形式为:

$$\ln\left(\frac{P_i}{1-P_i}\right) = Z_i = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n + \mu_i$$

其中, P_i 是住房抵押贷款违约发生率, $1-P_i$ 是不发生违约的概率, β_i 是待估计系数, X_i 是自变量, μ_i 是随机误差项。Logistic 模型主要采用极大似然法进行模型的估计与拟合,回归系数检验采用 Wald 统计量检验,检验值越大表明自变量的作用越显著。当 $P_i \geq 0.5$ 时,发生违约的概率较大,容易发生违约^[7]。

Logistic 回归模型具体在住房抵押贷款违约研究中的应用,假设 Y 表示发放一笔住房贷款,用 $Y=1$ 表示到期后借款人违约, $Y=0$ 表示借款人不违约。我们的目的是利用收集的样本数据建立模型,对借款人违约(即 P 表示 $Y=1$) 的概率 P 进行预测。 X_i 表示描述与违约概率有关的自变量影响因素的一些指标。这些指标影响了解释变量 $\ln\left(\frac{P}{1-P}\right)$ 的方向和程度。利用样本数据对模型中的参数 β_i 进行估计,并对模型进行相关的统计检验。因此 Logistic 回归模型是处理这一类问题较好的方法^[2]。

2.2 样本选取与变量量化

根据有关理论和我国现状,笔者选择借款人的性别、年龄、婚姻状况、最高学历、职业、家庭月收入、贷款期限、贷款金额、月还款金额、月还款额占家庭收入比作为解释变量,选择以虚拟变量表示的借款人是否违约作为被解释变量,违约贷款赋值为 0,正常贷款赋值为 1^[8]。从实用角度,采用分类自变量,对所选的 10 个自变量进行量化,分别进行赋值(见表 1)。之所以选择分类变量是因为当自变量是连续型时,比如年龄与违约的关系,可以求出一个 OR 值,假设为 1.4,表明每增加一岁,违约的发生率就增加 1.4 倍,这在现实情况中,可能很难做到这么准确,而且可能更想了解的是中年人比青年人的违约情况。因此,把年龄等各个解释变量进行划分,这样解释起来就更容易,也更加符合实际。

表 1 变量及其量化方式说明

序号	变量	变量含义	变量赋值
1	Sex	性别	男=1,女=0
2	Age	年龄	25 岁以下=0,25~40 岁=1,40~50 岁=2,50 岁及以上=3
3	Marr	婚姻状况	单身=0,已婚=1,离异=2
4	Educ	最高学历	初中及以下=0,高中及中专=1,大专=2,本科=3,硕士及以上=4
5	Work	职业	个体、小型私营企业=0,一般中大型企业=1,事业单位=2,垄断行业=3,政府机构=4
6	Income	家庭月收入/元	2000 以下=0,2000~3000=1,3000~4000=2,4000~5000=3,5000 及以上=4
7	Term	贷款期限/年	10 以下=0,10~15=1,15~20=2,20~25=3,25 及以上=4
8	Loan	贷款总额/万元	10 以下=0,10~20=1,20~30=2,30~40=3,40 及以上=4
9	PayRa	月还款额/元	1000 以下=0,1000~2000=1,2000~3000=2,3000~4000=3,4000 及以上=4
10	MPGIF	月还款额占收入比	5%以下=0,5%~10%=1,10%~20%=2,20%~30%=3,30%及以上=4

2.3 变量的描述性统计分析

本文数据来源于 2012 年对浙江省杭州市、宁波市、绍兴市、嘉兴市、金华市多家商业银行的客户进行了问卷、电话、网络等调查,获得贷款数据 3 983 份,包括正常贷款 3 810 份,违约贷款 173 份。数据的描述性统计分析如表 2 所示。

表 2 违约组与非违约组数据统计描述

	违约组					非违约组				
	样本数	最小值	最大值	均值	标准差	样本数	最小值	最大值	均值	标准差
性别	173	0	1	0.76	0.437	3 810	0	1	0.53	0.500
年龄	173	1	3	2.00	0.612	3 810	0	4	1.88	0.795
婚姻	173	0	2	1.12	0.697	3 810	0	2	0.87	0.507
教育	173	1	3	1.88	0.600	3 810	0	4	2.18	0.906
职业	173	0	4	2.08	1.269	3 810	0	4	3.12	1.221
收入	173	0	4	2.65	1.169	3 810	0	4	2.32	1.390
贷款期限	173	1	4	2.47	1.007	3 810	0	4	2.80	1.028
贷款金额	173	0	4	2.24	1.480	3 810	0	4	1.95	1.131
月还款额	173	0	4	2.65	1.272	3 810	0	4	1.84	1.228
月还款额占收入比	173	2	4	3.35	0.702	3 810	0	4	2.82	1.225

从表 2 中可以发现,在性别、年龄、教育、职业、贷款金额、月还款额以及月还款额占收入比例等变量的均值方面违约样本和正常样本存在显著差别。在性别方面,违约样本均值为 0.76,正常样本均值为 0.53,可以看出男性借款人违约率较高。在年龄方面,违约样本均值为 2.00,正常样本为 1.88,正常样本低于违约样本,表明年龄相对较大的借款人违约率较高,这可能跟年龄大的人做房地产投资等情况有关。在教育方面,违约样本均值为 1.88,非违约样本均值为 2.18,违约样本总体学历低于非违约样本。在职业方面,违约样本均值为 2.08,非违约样本均值为 3.12,违约样本显著低于非违约样本。在贷款总额方面,违约样本均值为 2.24,正常样本均值为 1.95,违约样本均值高于正常样本均值。在月还款额方面,违约样本均值为 2.65,正常样本均值为 1.84,违约样本均值显著高于正常样本均值。在月还款额占收入比例方面,违约样本均值为 3.35,正常样本均值为 2.82,违约样本均值显著高

于正常样本均值。

2.4 Logistic 回归结果分析

用 SPSS16.0 软件对住房抵押贷款违约风险的微观影响因素进行二元 Logistic 回归^[9],用向前逐步回归法得到如下结果:SPSS 结果显示一共回归了 7 次,由于篇幅原因,仅列出第一次和最后一次的回归结果,见表 3。

表 3 Logistic 模型准确率预测

			已预测		百分比校正
			违约		
			不违约	违约	
Step 1	违约	不违约	380	0	100.0
		违约	17	0	0.0
	总计百分比				95.7
Step 7	违约	不违约	380	0	100.0
		违约	12	5	29.4
	总计百分比				97.0

首先,从表 3 看,总体预测准确率为 97%,可以说明该模型准确度不错,基本可以反映样本信息。

接下来对模型进行拟合优度检验。选择 Hosmer and Lemeshow 拟合优度指标对模型进行检验。由表 4 可以看出, Hosmer and Lemeshow 统计量的

概率 P 值小于给定的显著水平, $\alpha=0.05$, 模型的拟合度比较高。另外类 R^2 指标 Nagelkerke R Square 指标为 0.502, 也说明模型的系数显著性比较好。

表 4 Logistic 模型检验与系数检验

模型检验 Hosmer and Lemeshow Test			系数检验 Model Summary		
步骤	卡方 Chi-square	显著性 p	-2 对数似然值	Cox & Snell R^2	Nagelkerke R^2
1	16.857	0	123.529a	0.042	0.140
7	50.506	0	89.880b	0.270	0.502

由表 5, 可以看出年龄(Age)、教育(Educ)、职业(Work)、月还款金额占收入比例(MPGIF)这 4

个变量的 sig 值小于 0.02, 能较好通过模型的检验。由模型估计结构可得 Logistic 模型的函数为:

表 5 Logistic 模型估计结果

		回归系数 B	标准误差 S.E.	统计量 Wald	自由度 df	显著性 Sig.	Exp(B)
Step 1a	MPGIF	1.452	0.499	8.481	1	0.004	4.274
	Constant	-8.059	1.887	18.239	1	0.000	0.000
Step 7f	Age	1.742	0.506	13.251	1	0.000	0.158
	Educ	-1.610	0.422	14.564	1	0.000	0.200
	Work	-1.360	0.301	20.433	1	0.000	3.897
	MPGIF	1.794	0.557	10.373	1	0.001	6.014
	Constant	-8.103	2.412	11.285	1	0.001	0.000

$$\ln \frac{P}{1-P} = Z = -8.103 + 1.742Age - 1.610Educ - 1.360Work + 1.794MPGIF$$

当 $P = \frac{1}{1+e^{-Z_i}} > 0.5$ 时, 住房抵押贷款很可能会发生违约; 反之则为正常贷款。

从上述模型的估计结果可得, 年龄(Age)、教育(Educ)、职业(Work)、月还款金额占收入比例(MPGIF)这 4 个因子是决定住房抵押贷款违约风险的关键影响因子。具体来说, 年龄因子的系数为 1.742, 说明借款人的年龄这个量化变量每增加 1%, 发生违约的机会比率就增加 1.742%; 教育因子的系数为 -1.610, 说明借款人的教育这个量化变量每增加 1%, 发生违约的机会比率就减少 1.610%; 工作因子的系数为 -1.360, 说明借款人的工作这个量化变量每增加 1%, 发生违约的机会比率就减少 1.360%; 月还款金额占收入比例因子的系数为 1.794, 说明借款人的月还款金额占收入比例每增加 1%, 发生违约的机会比率就增加 1.794%。这些结论对违约与否的影响在现实中也存在着相应原因, 借款人年龄越大, 其所购的可能不是第一套房, 可能为子女购房, 或者投机购房, 因而对该房的依赖性不会很强, 违约率可能较高; 借款人教育程度越高, 工作越稳定, 往往表明收入水平越高, 而且学历高的人考虑问题更加全面, 因而可能不会贸然违约, 因此其违约率可能就越低; 借款人月还

款金额占收入比越高, 表明借款人还款压力比较大, 一旦工作出现问题, 可能直接造成其还贷困难, 导致坏账。

3 结 论

笔者利用从浙江省杭州市、宁波市、绍兴市、嘉兴市、金华市调研得到的数据, 对住房抵押贷款违约风险的影响因素进行了实证研究, 研究结果表明: a) 教育程度与违约率成反比, 教育程度越高, 违约的可能性就越小; b) 工作稳定性与违约率成反比, 工作越稳定, 违约的可能性就越小; c) 年龄与违约率成正比, 年龄越大, 违约的可能性就越大; d) 月还款金额占收入比与违约率成正比, 月还款金额越大, 违约的可能性就越大。对违约率影响较大的因素依次是月还款额占家庭收入比、年龄、受教育程度、工作稳定性。从监管者角度来看, 应该在考虑上述各项违约影响因素的基础上, 制定较为严格的个人住房抵押贷款政策, 以此来规范商业银行房贷行为, 降低商业银行房贷不良比率, 保持稳健经营。

参考文献:

- [1] 郑 媛. 商业银行个人住房抵押贷款风险分析[J]. 集体经济: 财税金融, 2010(6): 101-102.
- [2] 钱争鸣, 李海波, 于艳萍. 个人住房按揭贷款违约风险研究[J]. 经济研究, 2010(增刊): 143-152.
- [3] Furstenberg V, George M, Green R J. Home mortgage

- delinquencies; a cohort analysis[J]. *Journal of Finance*, 1974, 29: 1545-1548.
- [4] Herzog J P, Earley J S. Home Mortgage Delinquency and Foreclosure[M]. New York: National Bureau of Economic Research, 1970.
- [5] Williams, Beranek A O W, Kenkel J. Default risk in urban mortgages; a Pittsburgh proto type analysis[J]. *American Real Estate and Urban Economics Association Journal*, 1974(2): 101-112.
- [6] 王福林, 贾生华, 邵海华. 个人住房抵押贷款违约风险影响因素实证研究: 以杭州市为例[J]. *经济学*, 2005, 4 (3): 739-751.
- [7] 何晓群. 应用多元统计分析[M]. 北京: 中国统计出版社, 2010.
- [8] 马 宇. 我国个人住房抵押贷款违约风险影响因素的实证研究[J]. *统计研究*, 2009, 26(5): 100-105.
- [9] 王济川, 郭志刚. Logistic 回归模型: 方法与应用[M]. 北京: 高等教育出版社, 2001.

An Empirical Study on Default Risk of Housing Mortgage Loans Based on Logistic Model

WANG Xiao-jun, LUO Hua

(School of Sciences, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: Housing mortgage loan has become a major means for Chinese commercial banks to expand credit marketing business and optimize credit assets structure. However, with the development of Chinese housing mortgage loan, its risks have become more severe increasingly. This paper conducts an empirical analysis on factors influencing the default risk of Chinese housing mortgage loan using Logistic regression model based on field research. The research finds that, the higher the ratio of monthly credit repayment amount to household income of borrower, the higher the default rate; the elder the borrower, the higher the default rate; the higher the education level of the borrower, the lower the default rate; the higher the work stability of the borrower, the lower the default rate. This provides favorable help for commercial banks to manage default risk while issuing housing mortgage loans.

Key words: housing mortgage loan; default risk; Logistic regression model

(责任编辑: 马春晓)