基于选择抽样的银行危机先行指标研究*

金洪飞，万兰兰
（上海财经大学 金融学院，上海 200433）

摘　要：文章在基于选择的抽样方法下，以前一期的各经济金融变量作为解释变量，采用 *Logit* 模型对银行危机的先行指标进行了经验分析。文章分别采用了加权与非加权的最大似然估计法，得到以下结果：实际 *GDP* 增长率、人均实际 *GDP* 以及实际利率的上升会降低银行危机的发生概率，而国内信贷与 *GDP* 之比、通货膨胀率以及广义货币与外汇储备比值的上升则会提高银行危机的发生概率，但政府财政盈余与 *GDP* 之比以及存款保险制度则对银行危机的发生概率没有显著影响。此外，非加权最大似然估计法的适用性比加权最大似然估计法更强，模型的拟合优度检验和预测准确率都表明，模型中统计上显著的以上6个变量是很好的银行危机先行指标。

关键词：银行危机；先行指标；基于选择的抽样；预测准确率

中图分类号：F831　**文献标识码**：A　**文章编号**：1001-9952(2014)05-0054-10

一、引　言

作为金融体系乃至整个经济体系中不可或缺的部分，银行业的平稳运行关系到整个经济的健康发展。近年来，金融危机不断爆发，作为金融危机的一种主要表现形式，银行危机因其频发性和破坏性而受到人们的高度重视。Laeven 和 Valencia(2008 和 2012)对1970年以来发生的所有系统性银行危机进行了整理和归纳，数据显示，1970－2011年，全球共有116个国家发生了147次银行危机，其中1990－1994年发生了43次，1995－1999年发生了31次，2008年有22个国家发生了银行危机。

银行危机的频发及其高昂的代价促使研究者对其成因或先行指标①进行了大量的经验研究，以降低银行危机发生的概率。传统的对银行危机形成因素的分析主要采用信号分析法和离散选择模型，其中离散选择模型的应用更为普遍。在美国次贷危机以前，Demirgüç-Kunt 和 Detragiache(1998)以及 Bussiere 和 Fratzscher(2006)采用基于面板数据的 *Logit* 模型研究了银行危机的先行指标以用于预警。在2008年美国金融危机后，银行危机的成因或预警又重新成为学术界的研究热点。Moshirian 和 Wu(2009)采用面板 *Logit* 模型，以1980－2001年18个发达国家和18个新兴市场国家为样本，研究了银行业股价收益的波动

收稿日期：2014-01-06
基金项目：教育部新世纪优秀人才支持计划(NCET-13-0889)；上海市曙光计划项目(09SG32)；上海市教委科研创新重点项目(10zs51)
作者简介：金洪飞(1970－)，男，浙江东阳人，上海财经大学金融学院教授，博士生导师；
万兰兰(1988－)，女，江西南昌人，上海财经大学金融学院博士研究生。

①银行危机的先行指标是指危机发生前的指标，本文采用的是前一期的各解释变量。

性与系统性银行危机之间的关系。他们发现，银行业股价收益的波动性可以很好地预测发达经济体的银行危机，但对新兴市场国家银行危机的预测能力较差。Klomp(2010)利用1970－2007年110个国家的面板数据，采用随机系数 *Probit* 模型研究了银行危机成因的异质性，发现高信贷增长率、低 *GDP* 增长率和高实际利率是银行危机发生的重要因素，但是没有一个因素能够在60%以上的银行危机中起到显著作用。Roy 和 Kemme (2012)的研究则表明，实际资产泡沫的破灭是引发银行危机的重要因素，这种资产泡沫的产生不一定伴随大量资本流入、公共债务积累以及低利率和结构冲击。

近年来，国内对银行危机生成机制的研究也不少，如谭福梅(2009)、董青马和卢满生(2010)采用离散选择模型研究了银行危机的成因或预警。此外，陈雨露和马勇(2013)分别从宏观经济、产业结构和制度等层面考察了一国金融危机的发生概率，研究结果表明，人均 *GDP* 增长会降低银行危机的发生概率，通胀率升高则会导致银行危机的发生概率增大，而在资本项目开放度和司法效率越高的国家，银行危机发生的概率越小。

以上关于银行危机成因的经验分析大多采用的是基于面板数据的二元离散选择模型。事实上，在某个年份发生银行危机的经济体占所有经济体的比例很低，而且对于某一个经济体，发生银行危机的年份占全部年份的比例也很低。这就使得无论是基于截面数据、时间序列数据还是面板数据，发生银行危机的观测值占全部观测值的比例非常低。如果银行危机发生时反映银行危机的虚拟变量取1，未发生银行危机时取0，那么得到的样本是严重左偏的，导致计量结果的可靠性较低。为了克服以上问题，本文采用了基于选择的抽样方法(*Choice-based Sampling*)，[①]这一方法能够解决观测值中被解释变量取1的概率远远低于取0的概率这一问题，本文样本中被解释变量取0和取1的概率相等。事实上，基于选择的抽样方法在各个领域已经得到广泛的应用。例如，患某种疾病的个体占总体的比例往往很低，医学研究者通常会根据医院的就诊记录把患病个体全部抽取出来，而在健康人群中随机抽取出数量接近的个体组成基于选择抽样的样本，这种抽样方法在医学研究中也被称为病例对照抽样(Siegel 和 Greenhouse，1973)；又如，Domowitz 和 Sartain(1999)、Platt 和 Platt (2002)等采用基于选择的抽样方法研究了企业破产的影响因素。但尚未有学者使用基于选择的抽样方法对银行危机的成因或先行指标进行经验分析。

二、计量模型与变量选择

(一)计量模型

本文的研究目的是使用实际数据通过计量模型来分析什么因素导致银行危机的发生或者影响银行危机发生的概率。为此，我们建立了如下二元 *Logit* 模型：

$$P(y_i=1|x_i)=F(x_i'\beta) \tag{1}$$

其中，y 为被解释变量，取值为1或0，x_i 为解释变量组成的向量，β 为对应的系数向量。式(1)的左边为给定解释变量向量 x_i 时 y_i 取1的概率，右边 $F(\cdot)$ 为 *Logit* 分布的累积分布函数，即：

$$F(x_i'\beta)=\frac{1}{1+e^{-x_i'\beta}}=\frac{e^{x_i'\beta}}{1+e^{x_i'\beta}} \tag{2}$$

①样本的选取并不是完全随机的，而是根据被解释变量的取值或属性进行抽样的方法即为基于选择的抽样方法(*Choice-based Sampling*)。

模型的对数似然函数为：

$$lnL=\sum_{i=1}^{n}y_i ln(F(x'_i\beta))+\sum_{i=1}^{n}(1-y_i)ln(1-F(x'_i\beta)) \quad (3)$$

对式(3)求解最大化问题即可得到最大似然法下参数向量 β 的估计值。

与随机抽样相比，在基于选择的抽样中，总体中低比例个体被抽中的概率被人为提高，而高比例个体被抽中的概率被人为降低。Manski 和 Lerman(1977)认为采用非加权的最大似然法得到的估计量是非一致的，因而建议使用加权最大似然法，其具体形式如下：

$$lnL=\sum_{i=1}^{n}\frac{Q_1}{H_1}y_i ln(F(x'_i\beta))+\sum_{i=1}^{n}\frac{Q_0}{H_0}(1-y_i)ln(1-F(x'_i\beta)) \quad (4)$$

其中，Q_1 为总体中银行危机个体所占比例，H_1 为所抽取的样本中银行危机个体所占比例，Q_0 为总体中未发生银行危机的个体所占比例，H_0 为所抽取的样本中未发生银行危机的个体所占比例。Q_1/H_1 和 Q_0/H_0 分别为银行危机个体和未发生银行危机的个体在模型估计中的权重，对式(4)求解最大化问题即可得到加权最大似然法下参数向量 β 的估计值。

对于非加权最大似然法和加权最大似然法孰优孰劣的问题，学术界并没有统一的定论。Nathan 和 Holt(1980)的经验结果表明，在某些情况下非加权最大似然估计量比加权最大似然估计量更有效，而加权最大似然法下的这种有效性缺失可能会严重影响分析结果，Scott 和 Wild(1986)的研究也论证了这一观点。Kish 和 Frankel(1974)则认为，在基于选择的抽样方法下，使用加权的估计量是一个比较合理的选择，Cameron 和 Trivedi(2005)也认为加权最大似然法能够保证估计结果的一致性。因此，本文将分别采用非加权最大似然法和加权最大似然法对模型进行估计，并比较两种估计方法下的经验分析结果。

（二）变量选择

1.被解释变量。本文的被解释变量为"是否发生银行危机"，对此我们设立了一个虚拟变量 BC，若发生银行危机，则 BC 取 1，否则取 0。

2.解释变量。借鉴 Demirgüç-Kunt 和 Detragiache(1998)对银行危机先行指标的研究，并考虑到各变量数据的可得性，本文选取了以下解释变量：

(1)实际 GDP 增长率($GDPG$)。作为经济运行中不可或缺的部分，商业银行的稳定运行关系到经济的平稳发展，而经济体的运行状况反过来也会影响银行体系的发展。一般来讲，如果 GDP 增长率较高，则银行危机发生的概率就较小。

(2)通货膨胀率(INF)。在其他条件不变时，通货膨胀率反映价格体系的稳定性，稳定的价格体系对于银行体系的稳定是有利的。高通货膨胀率会损害储蓄者而有利于贷款者，使资金的供给下降、需求上升，导致名义利率上升，而这会增强金融体系的脆弱性，提高银行危机发生的概率。

(3)实际利率(RIR)。许多研究者认为高实际利率会增加银行危机发生的概率，但从理论上讲，实际利率对银行危机发生概率存在方向相反的两种影响。一方面，实际利率上升意味着融资成本增加，起到紧缩经济的作用，同时也会使企业财务状况恶化，并增加银行坏账，导致银行危机发生概率上升；但另一方面，实际利率可以看作是资本的实际报酬率，实际利率高意味着实体经济发展前景好，经济往往处于上升期，因此在其他条件不变时，银行危机的发生概率较低。

(4)$M2$/外汇储备($M2/RES$)。$M2$/外汇储备反映了银行体系的系统性风险，其值上升说明外汇储备相对于货币供应下降，这意味着银行本币负债的外汇保障下降，银行体系风险

增加，从而银行危机发生的概率上升。

(5)国内信贷与 GDP 之比(DC/GDP)。信贷的过度扩张会导致银行体系的脆弱性增加，银行危机爆发前往往存在信贷市场的过度繁荣，因此我们用国内信贷/GDP 来反映信贷的扩张情况。

(6)财政盈余占 GDP 比例(GS/GDP)。政府的财政状况往往决定了政府对银行危机的防治能力，同时也影响社会公众对银行体系的信心。如果公众认为即使发生银行危机，政府也能及时遏制，则他们不会对银行体系的稳定运行产生怀疑，从而银行危机发生的概率较小。而当一个经济体的财政赤字过大时，赤字往往会通过金融体系被货币化，如或明或暗地要求银行购买政府债券，为政府的赤字融资，这不可避免地将对银行业造成伤害。

(7)人均实际 GDP($GDPPC$)。人均实际 GDP 是衡量经济发展的一个指标，人均实际 GDP 越高，说明经济越发达，金融体系抵御风险的能力往往越强。

(8)名义汇率变化率(GEX)。汇率变化对银行危机的影响可以从两个方面考虑：一方面，当汇率贬值时，包括银行在内的企业以外币计价的债务的本币价值上升，导致资产负债表恶化，从而发生银行危机的概率上升；如果本币汇率贬值是资本外逃的结果，那么汇率贬值将是银行危机的一个先行指标。另一方面，如果一个经济体的汇率事先存在高估，在其他条件不变时，汇率贬值将使实际汇率趋向均衡水平，这有利于资源有效配置和经济增长，从而降低银行危机发生的概率。在本文中，由于汇率采用直接标价法，某个经济体的 GEX 为正表明该经济体的货币相对于外币贬值。

(9)是否存在明确的存款保险制度(DPI)。存款保险制度建立的初衷是为了防止发生银行挤兑，增强银行体系的稳定性。存款保险制度保障了储户的存款安全，但也带来了一定的银行体系道德风险问题。因此，存款保险制度对银行体系的稳定运行既有正面影响也有负面影响。

考虑到银行危机发生当年各变量已经受到危机的影响，我们采用危机发生前一年的各变量来研究其形成机制。

三、样本与数据

正如上文所指出的，在某个时间点发生银行危机的国家相对于没有发生银行危机的国家要少得多。因此在面板数据中，被解释变量 BC 的分布是严重左偏的，其偏度系数(*Skewness*)将远远小于 0，从而使回归结果的可靠性降低。结合数据的可得性，本文从 Laeven 和 Valencia(2008 和 2012)的数据库中选取了 97 个银行危机样本。[①] 对于未发生银行危机的样本，本文采取了基于选择的抽样方法。考虑到银行危机的持续期平均约为 3 年，[②] 令 T 期为银行危机发生当年，本文剔除 T 期至 $T+2$ 期的观测值后，得到了 917 个未发生银行危机的样本。我们进一步采用不放回抽样的方法从中随机抽取了 97 个作为研究样本。由此，本文的总样本包括 97 个银行危机样本和 97 个未发生银行危机的样本，即被解释变量取 1的概率和取 0 的概率相等，进而可以计算得到加权最大似然法中的权重 $Q_1/H_1=0.2764$ 和 $Q_0/H_0=1.7236$。

关于各变量的数据来源，被解释变量“是否发生银行危机”的取值来源于 Laeven 和 Va-

①我们将不同年份的同一经济体看作是不同的个体。

②根据 Reinhart 和 Rogoff(2008)的估计，银行危机的平均持续期约为 3 年。

lencia(2008 和 2012)的数据库;解释变量中的实际 *GDP* 增长率、实际利率、通胀率、人均实际 *GDP*、*M*2/外汇储备、国内信贷/*GDP* 等数据来自世界银行的世界发展指数(*WDI*)数据库;财政盈余/*GDP* 数据来自 *IMF* 的国际金融统计(*IFS*)数据库;名义汇率变化率数据来自 *WDI* 的官方汇率并经作者计算得到,其中欧元对美元的汇率来自中经网统计数据库;存款保险制度的数据来源于 Demirgüç-Kunt 等(2005)和 Lé(2013)。

四、经验结果分析

(一)基于选择的 *MLE* 和 *WMLE* 结果比较

1.计量结果。表 1 分别给出了非加权最大似然法(*MLE*)和加权最大似然法(*WMLE*)下计量模型的估计结果。首先来看非加权最大似然法(*MLE*)下的模型估计结果,模型 1 包含了所有解释变量,其中国内信贷规模与 *GDP* 之比 *DC*/*GDP* 的系数在 5%的水平上显著为正,说明其他因素不变时,国内信贷规模扩张会增加银行危机发生的概率;反映存款保险制度的变量 *DPI* 在 10%的水平上不显著,说明是否存在明确的存款保险制度对银行危机发生的概率没有显著影响;实际 *GDP* 增长率 *GDPG* 的系数在 5%的水平上显著为负,说明给定其他条件,经济增长率提高会降低银行危机发生的概率;人均实际 *GDP* 的系数在 1%的水平上显著为负,说明其他条件不变时,人均收入水平越高的经济体爆发银行危机的可能性越低;名义汇率变化率 *GEX* 和财政盈余占 *GDP* 比例在 10%的水平上都不显著,说明汇率变化和财政状况不是引发银行危机的因素;通货膨胀率 *INF* 的系数在 1%的水平上显著为正,说明其他条件不变时,通货膨胀率上升会增加银行危机发生的概率;广义货币与外汇储备之比 *M*2/*RES* 的系数在 10%的水平上显著为正,说明其他条件相同时,货币超发会增加银行危机发生的概率;最后,实际利率 *RIR* 的系数在 1%的水平上显著为负,说明其他条件相同时,实际利率高的经济体发生银行危机的概率低于实际利率低的国家。以上对通货膨胀率、人均实际 *GDP* 以及存款保险制度的分析结果与陈雨露和马勇(2013)的结果一致,而对信贷指标的考察结果则有所不同,陈雨露和马勇(2013)讨论的是对私人部门的信贷对银行危机的影响,认为增加对私人部门的信贷可以减少银行危机的发生。

我们依次剔除模型 1 中不显著的变量后,分别得到了模型 2、模型 3 和模型 4。估计结果显示,模型 1 中显著的变量在模型 2、模型 3 和模型 4 中仍显著,而且系数的符号没有变化,除常数项外,系数的大小也几乎不变。此外,四个模型的对数似然值和 *McFadden* R^2 也基本保持不变,这进一步说明了模型的稳健性。四个模型的 *LR* 统计值都在 1%的水平上显著,说明这四个模型整体上都是显著的。

表 1 Logit 模型估计结果

	模型 1		模型 2		模型 3		模型 4		模型 5	
	MLE	*WMLE*	*MLE*	*WMLE*	*MLE*	*WMLE*	*MLE*	*WMLE*	*MLE*	*WMLE*
Constant	−0.024 (−0.05)	−1.733*** (−3.50)	−0.024 (−0.05)	−1.727*** (−3.49)	−0.008 (−0.02)	−1.717*** (−3.42)	−0.123 (−0.27)	−1.918*** (−4.09)	−0.108 (−0.72)	−1.956*** (−12.79)
DC/*GDP*	0.012** (2.31)	0.014*** (2.65)	0.012** (2.31)	0.014*** (2.63)	0.012** (2.35)	0.014*** (2.75)	0.012** (2.36)	0.015*** (2.88)		
DPI	−0.228 (−0.64)	−0.346 (−0.97)	−0.230 (−0.64)	−0.354 (−1.00)	−0.219 (−0.61)	−0.327 (−0.94)				
GDPG	−0.128** (−2.31)	−0.160*** (−2.75)	−0.128** (−2.32)	−0.161*** (−2.75)	−0.130** (−2.33)	−0.164*** (−2.74)	−0.128** (−2.34)	−0.160*** (−2.64)		
GDPPC	−0.540*** (−2.84)	−0.591*** (−3.17)	−0.542*** (−2.87)	−0.599*** (−3.18)	−0.555*** (−3.01)	−0.617*** (−3.26)	−0.566*** (−3.09)	−0.633*** (−3.34)		
GEX	0.001 (0.05)	0.004 (0.30)							0.026** (2.51)	0.032*** (3.56)

续表 1　Logit 模型估计结果

	模型 1		模型 2		模型 3		模型 4		模型 5	
	MLE	*WMLE*	*MLE*	*WMLE*	*MLE*	*WMLE*	*MLE*	*WMLE*	*MLE*	*WMLE*
GS/GDP	−0.007 (−0.32)	−0.010 (−1.04)	−0.007 (−0.32)	−0.010 (−1.04)						
INF	0.081*** (3.18)	0.085*** (3.13)	0.081*** (3.53)	0.088*** (3.59)	0.082*** (3.58)	0.090*** (3.65)	0.082*** (3.59)	0.089*** (3.53)		
M2/RES	0.026* (1.88)	0.026** (2.20)	0.026** (1.89)	0.026** (2.20)	0.030** (1.91)	0.026** (2.20)	0.025* (1.87)	0.024** (2.23)		
RIR	−0.053*** (−3.25)	−0.058*** (−3.35)	−0.053*** (−3.26)	−0.059*** (−3.45)	−0.05*** (−3.32)	−0.060*** (−3.53)	−0.053*** (−3.30)	−0.058*** (−3.28)		
对数似然值	−104.145	−58.888	−104.146	−58.902	−104.230	−59.020	−104.418	−59.218	−129.733	−74.588
McFadden R^2	0.226	0.244	0.226	0.244	0.225	0.243	0.223	0.240	0.035	0.043
LR 统计值	60.65***	38.08***	60.65***	38.05***	60.48***	37.81***	60.11***	37.42***	9.47***	6.68***
H−*L* 统计值	6.14		5.09		6.72		8.55		28.16***	
Andrews 统计值	11.02		9.82		11.25		12.85		23.99***	

注:括号内为 *z* 值,***、** 和 * 分别表示在 1%、5%和 10%的水平上显著。在加权最大似然估计中,两类个体的权重是不同的,从而无法计算 *H*−*L* 统计量和 *Andrews* 统计量。

值得注意的是,对于名义汇率变化率(即名义贬值率),根据描述性统计,无论是均值、中位数、最大值还是最小值似乎都说明汇率贬值应当是银行危机的一个诱因。此外,我们还发现在总样本的 194 个观测值中,汇率贬值率超过 5%的有 69 个,其中危机样本中有 43 个,而未发生危机样本中仅有 26 个,也存在明显的差异。于是,我们使用汇率变化率 *GEX* 作为唯一的解释变量进行了回归,结果见表 1 中模型 5。可以看到,变量 *GEX* 在 5%的水平上是统计显著的,其系数估计值为正,说明汇率贬值幅度越大的经济体越有可能发生银行危机,但模型 5 的 *McFadden* R^2 远低于其他模型,而 *GEX* 在模型 1 中不显著的原因可能是存在多重共线性问题。事实上,我们以 *GEX* 作为被解释变量、模型 1 中的其他解释变量作为解释变量进行 *OLS* 回归,发现 *DPI*、*GDPPC* 和 *RIR* 在 5%的水平上都是显著的,*GDPG* 在 10%的水平上显著,而且这四个变量的系数均为负,表明有明确的存款保险制度及较高的人均收入水平、实际利率和经济增长率都可以降低经济体的货币贬值率,这使得不包含变量 *GEX* 的模型 2、模型 3 和模型 4 在各个变量的显著性、*McFadden* R^2 等方面与模型 1 几乎都不存在明显差异。

再来看加权最大似然法(*WMLE*)下的模型估计结果,从表 1 中可以看到,*LR* 统计值同样在 1%的水平上显著,模型整体的显著性并没有发生变化。从模型的准拟合优度来看,各模型的拟合优度均略有提高。在解释变量的显著性方面,除常数项外,非加权最大似然法下显著的解释变量在加权最大似然法下同样显著,且国内信贷/*GDP*、实际 *GDP* 增长率、*M2*/外汇储备以及名义汇率变化率的显著性还略有提高,而存款保险制度和财政盈余/*GDP* 在加权最大似然法下依然不显著。

二元选择模型是一个非线性模型,*McFadden* R^2 与线性回归中的 R^2 不同,从而不能简单地理解为模型的拟合优度。在非加权最大似然估计中,我们可以采用拟合优度的 *Pearson* χ^2 检验,即 *H*−*L* 统计量和 *Andrews* 统计量。在模型的拟合值与实际值一致的零假设下,*H*−*L* 统计量和 *Andrews* 统计量都服从卡方分布。当统计值超过临界值时,可以认为模型的拟合值与实际值并不一致,从而模型的拟合优度较低;而当统计值低于临界值时,则无法拒绝零假设,可以认为模型的拟合优度较高。本文在进行 *H*−*L* 检验和 *Andrews* 检验时将样本分成了 10 组,因而在拟合值与实际值一致的零假设下,*H*−*L* 统计量和 *Andrews* 统计量分别服从自由度为 8 和 10 的卡方分布。表 1 中模型 1、模型 2、模型 3 和模型 4 的 *H*−*L* 值和 *Andrews* 值都不显著,说明这四个模型的拟合优度很高,可以认为模型的拟合值

与实际值是一致的；而模型 5 的 $H-L$ 值和 $Andrews$ 值超过了临界值，因此其拟合优度不高，可以认为模型的拟合值与实际值存在较大差异。

2.预测准确率。对于二元选择模型，模型的预测准确率也是检验模型统计充分性的一个指标。由于 $Logit$ 模型的左边是银行危机发生的概率，所谓预测值就是将解释变量的具体数值代入到拟合方程中计算得到的概率值。当 y 的观测值为 1 时，若预测的概率值大于 0.5，则认为预测是准确的，若预测的概率值小于 0.5，则认为预测不准确；当 y 的观测值为 0 时，若预测的概率值小于 0.5，则认为预测准确，反之则预测不准确。非加权最大似然法和加权最大似然法下各模型的预测准确率分别见表 2 和表 3。在非加权最大似然法下，模型 1、模型 2 和模型 4 的预测准确率完全相同，在未发生危机样本的 97 个观测值中预测准确的有 75 个，预测准确率为 77.32%，在危机样本的 97 个观测值中预测准确的有 66 个，预测准确率为 68.04%，总体预测准确率为 72.68%。模型 3 对危机样本的预测准确率与模型 1、模型 2 和模型 4 相同，但在未发生危机样本的 97 个观测值中预测准确的有 76 个，预测准确率为 78.35%，总体预测准确率为 73.20%。总的来说，模型 1、模型 2、模型 3 和模型 4 的预测准确率没有多大差异，预测准确率都较高。模型 5 对未发生危机样本预测准确的有 69 个，预测准确率为 71.13%，对危机样本预测准确的仅有 45 个，预测准确率为 46.39%，总体预测准确率为 58.76%。此外，对于每一个模型，未发生危机样本中的预测准确率都高于危机样本中的预测准确率。

表 2　非加权最大似然法下模型的预测结果和预测准确率

	模型 1		模型 2		模型 3		模型 4		模型 5	
	$y_i=0$	$y_i=1$	$y_i=0$	$y_i=1$	$y_i=0$	$y_i=1$	$y_i=0$	$y_i=1$	$y_i=0$	$y_i=1$
$P(y_i=1)<0.5$	75	31	75	31	76	31	75	31	69	52
$P(y_i=1)>0.5$	22	66	22	66	21	66	22	66	28	45
准确率(%)	77.32	68.04	77.32	68.04	78.35	68.04	77.32	68.04	71.13	46.39
总准确率(%)	72.68		72.68		73.20		72.68		58.76	

表 3　加权最大似然法下模型的预测结果和预测准确率

	模型 1		模型 2		模型 3		模型 4		模型 5	
	$y_i=0$	$y_i=1$	$y_i=0$	$y_i=1$	$y_i=0$	$y_i=1$	$y_i=0$	$y_i=1$	$y_i=0$	$y_i=1$
$P(y_i=1)<0.5$	97	69	97	69	97	69	95	72	97	90
$P(y_i=1)>0.5$	0	28	0	28	0	28	2	25	0	7
准确率(%)	100	28.87	100	28.87	100	28.87	97.94	25.77	100	7.22
总准确率(%)	64.44		64.44		64.44		61.86		53.61	

从总体预测准确率来看，加权最大似然法不如非加权最大似然法，加权最大似然法下模型 1、模型 2 和模型 3 的总体预测准确率只有 64.44%，比非加权最大似然法下的 72.68%和 73.20%要低很多，模型 5 的总体预测准确率也比非加权最大似然法下的准确率低 5 个百分点。在加权最大似然法下，未发生银行危机的预测准确率在模型 1、模型 2、模型 3 和模型 5 中均为 100%，在模型 4 中也高达 97.94%，可见加权最大似然法对未发生银行危机样本的预测准确率非常高。但该方法对银行危机样本的预测准确率却不太乐观，在模型 1、模型 2 和模型 3 中的预测准确率仅为 28.87%，在模型 4 和模型 5 中则更低。本文对银行危机先行指标的分析主要是为银行危机预警体系的建立提供参考，因而对银行危机样本的预测显得尤为重要，我们有理由认为非加权最大似然法比加权最大似然法的适用性更强。

（二）全样本 MLE 结果

为了验证基于选择抽样方法的有效性，本文还将全部未发生银行危机的917个样本和97个银行危机样本放在一起进行回归，得到全样本 *MLE* 估计结果。[①] 结果显示，实际 *GDP* 增长率、实际利率以及 *M2*/外汇储备等依然对银行危机发生概率具有显著影响，而之前显著的通胀水平、人均实际 *GDP* 以及国内信贷/*GDP* 等则在全样本估计结果中并不显著。从模型的预测准确率来看，全样本对未发生银行危机样本的预测准确率为99.89％，对银行危机样本的预测准确率则仅为7.22％。因此，相对于全样本的分析，基于选择的抽样是一个比较好的方法。上文中基于选择抽样的 *WMLE* 估计对危机样本的预测准确率很低，其原因就在于加权最大似然法实际上通过增加权重取到了与总体比例相当的未发生危机个体数，因此预测功能与全样本模型类似。

（三）稳健性检验

为了进一步验证非加权最大似然法下模型整体的稳健性，对于未发生银行危机的样本，本文在不放回抽样方法下又进行了10次随机抽样，将其分别与银行危机样本放在一起再次进行估计。在10次抽样的估计结果中，实际 *GDP* 增长率对银行危机发生概率的影响有9次在5％的水平上显著为负，有1次在10％的水平上显著；实际利率水平也表现出显著的负向影响，且有5次抽样的结果在1％的水平上显著，另5次在5％的水平上显著；通胀水平对银行危机发生概率的影响在10次抽样中均在1％的水平上显著为正；人均实际 *GDP* 的影响均在1％的水平上显著为负；*M2*/外汇储备的影响在3次抽样中在5％的水平上显著为正，在其他抽样中在10％的水平上显著为正；国内信贷/*GDP* 对银行危机发生概率的影响则均在5％的水平上显著为正；对于名义汇率变化率，当模型中包括所有解释变量时，它依然不显著。

与上文一样，我们在这10次抽样研究中也单独考察了名义汇率变化对银行危机发生概率的影响，结果显示其影响均在5％的水平上显著为正；另外，政府财政/*GDP* 以及是否存在明确的存款保险制度这两个变量对银行危机发生概率仍不存在显著影响。模型的 *LR* 统计值均在1％的水平上显著，准拟合优度处在0.186到0.239之间，总体预测准确率保持在70.62％－74.74％之间，对银行危机样本的预测准确率保持在62.89％－70.10％之间，模型的稳定性较强。综上所述，本文基于选择抽样方法的估计结果具有较高的稳健性。

五、结　论

本文在基于选择的抽样方法（*Choice-based Sampling*）下，结合相关数据的可得性，选取1970－2011年的97个银行危机个体和97个未发生银行危机个体作为研究样本，采用 *Logit* 模型对银行危机的先行指标进行了分析，得到以下结论：

1.实际 *GDP* 增长率和人均实际收入水平的提高可以降低银行危机发生的概率；国内信贷规模与 *GDP* 之比、通货膨胀率以及广义货币与外汇储备之比等指标的上升则会增加银行危机发生的概率；银行危机的发生与是否存在明确的存款保险制度没有必然的联系，经验数据表明，存款保险制度并不能有效阻止银行危机的发生；财政盈余与 *GDP* 之比对银行危机发生的概率也没有显著影响，财政赤字高未必引发银行危机。

2.从理论上讲，实际利率对银行危机发生概率具有方向相反的两种作用。一方面，实际利率的提高会增加企业的融资成本，加重企业的财务负担，从而可能导致银行坏账增加，银

①限于篇幅，文中未报告全样本 *MLE* 估计结果和下文稳健性检验的结果，如有需要可向作者索取。

行危机发生的概率上升;但另一方面,实际利率可以看作是资本的实际回报率,高实际利率往往出现在经济扩张期,因此实际利率高也可能降低银行危机发生的概率。本文未发生危机样本的实际利率水平明显高于危机样本,而且估计结果表明实际利率水平的上升会显著降低银行危机发生的概率,由此推断实际利率的第二种作用占主导。

3.本国货币的贬值程度与银行危机发生的可能性存在正相关关系,发生银行危机经济体的货币贬值率平均而言要高于未发生银行危机的经济体。但本币贬值率与人均实际收入水平、实际 *GDP* 增长率、实际利率水平以及是否存在明确的存款保险制度等变量之间存在多重共线性,因此当将这些变量和汇率变化率放在一起作为解释变量时,汇率变化率并不显著,但这并不意味着汇率变化率与银行危机发生的概率没有关系。

4.加权最大似然法下解释变量的显著性虽然较高,但该方法对银行危机样本的预测能力较低。因此,相对于加权最大似然法,非加权最大似然法的适用性更强,对银行危机样本的预测准确率更高。

5.全样本估计结果论证了基于选择抽样方法的优越性,多次抽样的估计结果也验证了本文结果的稳健性。模型的预测准确率表明,国内信贷与 *GDP* 之比、实际 *GDP* 增长率、人均实际 *GDP*、通货膨胀率、广义货币与外汇储备之比以及实际利率水平等指标同时作为解释变量可以较好地预测是否发生银行危机,因此这些指标可以作为银行危机的预警指标加以监控,以避免银行危机发生或降低其发生概率。

---

* 本文还得到上海财经大学研究生创新计划项目(CXJJ-2013-326)的资助。

主要参考文献:

[1]陈雨露,马勇.大金融论纲[M].北京:中国人民大学出版社,2013.

[2]董青马,卢满生.金融开放度与发展程度差异对银行危机生成机制影响的实证分析[J].国际金融研究,2010,(6):79-85.

[3]谭福梅.系统性银行危机早期预警系统有效吗?——基于 Logit 模型的实证分析(1980-2007)[J].当代财经,2009,(12):49-54.

[4]Bussiere M, Fratzscher M.Towards a new early warning system of financial crises[J].Journal of International Money and Finance, 2006, 25(6): 953-973.

[5]Cameron A C, Trivedi P K. Microeconometrics: Methods and applications[M].Cambridge: Cambridge University Press, 2005.

[6]Demirgüç-Kunt A, Detragiache E.The determinants of banking crises in developing and developed countries[J].IMF Staff Papers, 1998, 45(1): 81-109.

[7]Demirgüç-Kunt A, Karacaovali B, Laeven L. Deposit insurance around the world: A comprehensive database[R].World Bank Policy Research Working Paper No.3628, 2005.

[8]Domowitz I, Sartain R L.Determinants of the consumer bankruptcy decision[J].Journal of Finance, 1999, 54(1): 403-420.

[9]Kish L, Frankel M R.Inference from complex samples[J].Journal of the Royal Statistical Society, Series B (Methodological), 1974, 36(1): 1-37.

[10]Klomp J.Causes of banking crises revisited[J].North American Journal of Economics and Finance, 2010, 21(1): 72-87.

[11]Laeven L, Valencia F.Systemic banking crises: A new database[R].IMF Working Paper No.224, 2008.

[12]Laeven L, Valencia F.Systemic banking crises database: An update[R].IMF Working Paper No.163, 2012.

[13]Lé M.Deposit insurance adoption and bank risk-taking：The role of leverage[R].International Academic Conference "Risk Taking in Financial Institutions，Regulation and the Real Economy"，Paris，2013.

[14]Manski C F，Lerman S R.The estimation of choice probabilities from choice-based samples[J].Econometrica，1977，45(8)：1977－1988.

[15]Moshirian F，Wu Q.Banking industry volatility and banking crises[J].Journal of International Financial Markets，Institutions and Money，2009，19(2)：351－370.

[16]Nathan G，Holt D.The effect of survey design on regression analysis[J].Journal of the Royal Statistical Society，Series B(Methodological)，1980，42(3)：377－386.

[17]Platt H D，Platt M B.Predicting corporate financial distress：Reflections on choice-based sample bias[J].Journal of Economics and Finance，2002，26(2)：184－199.

[18]Reinhart C M，Rogoff K S.Banking crises：An equal opportunity menace[R].NBER Working Paper No.14587，2008.

[19]Roy S，Kemme D M.Causes of banking crises：Deregulation，credit booms and asset bubbles，then and now[J].International Review of Economics and Finance，2012，24：270－294.

[20]Scott A J，Wild C J.Fitting logistic models under case-control or choice based sampling[J].Journal of the Royal Statistical Society，Series B(Methodological)，1986，48(2)：170－182.

[21]Siegel D G，Greenhouse S W.Multiple relative risk functions in case-control studies[J].American Journal of Epidemiology，1973，97(5)：324－331.

Leading Indicators of Banking Crises Based on Choice-based Sampling

JIN Hong-fei，WAN Lan-lan

(*School of Finance*，*Shanghai University of Finance and Economics*，*Shanghai* 200433，*China*)

Abstract：Based on choice-based sampling and taking economic and financial variables in the former period as explanatory variables，this paper employs Logit model to conduct an empirical analysis of leading indicators of banking crises. It uses weighted and non-weighted maximum likelihood estimation methods respectively to arrive at the following results：the increase in real GDP growth rate，real GDP per capita and real interest rates leads to the reduction in the probability of banking crises，while the increases in the ratio of domestic credit to GDP，inflation rate and the ratio of M2 to foreign exchange reserves results in the increase in the probability of banking crises；and the ratio of fiscal surplus to GDP and deposit insurance system have no significant effects on the probability of banking crises. Besides，the applicability of non-weighted maximum likelihood estimation method is stronger than the one of weighted maximum likelihood estimation method and both of goodness-of-fit test and prediction accuracy show that six variables above-mentioned which are statistically significant are good leading indicators of banking crises.

Key words：banking crisis；leading indicator；choice-based sampling；prediction accuracy

（责任编辑　康　健）