金融发展与创新译丛　朱民 主编

梁晶工作室
LIANGJING PUBLISHING HOUSE

Credit Risk Modeling using Excel and VBA

# 信用风险模型

## 基于Excel和VBA平台

将信用风险模型与普遍使用的软件平台相结合
全面涵盖了信用风险管理领域的相关信息

信用风险模型
易于掌握的
应用指南
本书讲述如何衡量与预测信用风险

[美] 冈特·勒夫勒 Gunter Löffler　彼得·N.波施 Peter N. Posch　/著

中国财政经济出版社

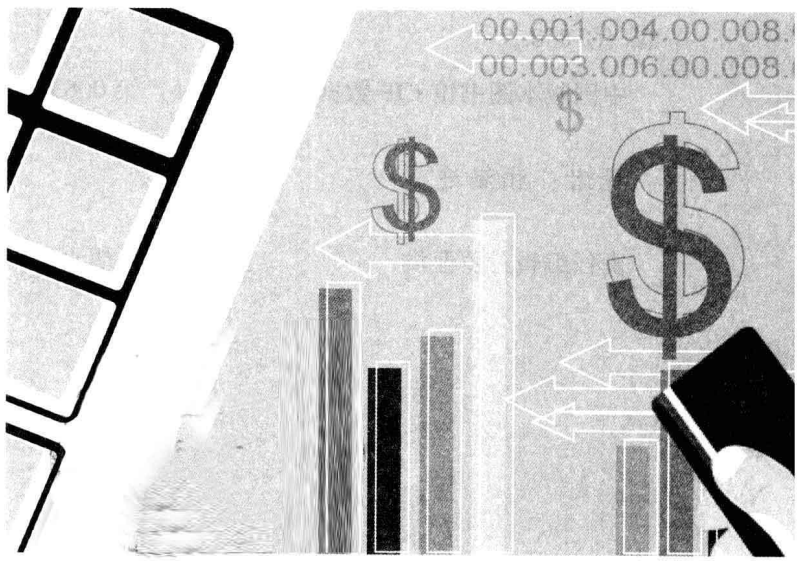
金融发展与创新译丛　朱民 主编

聚晶工作室
LIANGJING PUBLISHING HOUSE

Credit Risk Modeling using Excel and VBA

# 信用风险模型

## 基于Excel和VBA平台




[美] 冈特·勒夫勒 Gunter Löffler　彼得·N.波施 Peter N. Posch　/著

柏满迎 /译

中国财政经济出版社

**图书在版编目（CIP）数据**

信用风险模型：基于 Excel 和 VBA 平台/（德）勒夫勒（Loffler，G.），（德）波施（Poshch，P. N.）著；柏满迎译．—北京：中国财政经济出版社，2011. 12

（金融发展与创新译丛）

书名原文：Credit risk modelling using Excel and VBA

ISBN 978－7－5095－2682－8

Ⅰ. ①信… Ⅱ. ①勒… ②波… ③柏… Ⅲ. ①信用－风险管理－研究 Ⅳ. ①F830. 5

中国版本图书馆 CIP 数据核字（2011）第 006393 号

译者：柏满迎

责任编辑：罗亚洪　　责任校对：胡永立

中国财政经济出版社出版

URL：http：//www. cfeph. cn

E－mail：cfeph @ cfeph. cn



社址：北京市海淀区阜成路甲 28 号　邮政编码：100142

发行处电话：88190406　财经书店电话：64033436

北京财经印刷厂印刷　各地新华书店经销

787×1092 毫米　16 开　19. 5 印张　296 000 字

2011 年 12 月第 1 版　2011 年 12 月北京第 1 次印刷

定价：43. 00 元

ISBN 978－7－5095－2682－8/F·2282

图字：01－2009－0976

# 前　言

本书是一个现代信用风险方法的导论，也是一个指导如何把信用风险模型用于实践的“食谱”。我们希望这两个目的都能够实现。根据我们自己的经验，理解分析方法的最好办法是实践。

信用风险文献大体划分为两个阵营：风险度量和定价。本书属于风险度量阵营。书中关于违约概率估算和信用组合风险的章节多于定价和信用衍生品的章节。我们对风险度量问题的涵盖也是有选择的。我们认为有选择性比涵盖很多议题但缺少细节更好，我们希望所提供的资料能为解决本书没有涉及的问题提供一个很好的基础。

我们选择 Excel 作为主要工具，因为它是一个非常普及和灵活的工具，为很多问题提供了不错的解决方法。就一些问题而言，即便是Excel的热爱者也会承认 Excel 不是他们首选。但是即使如此，在实施策略通常可以转换到其他编程环境的情况下，演示如何把模型投入使用也是赏心悦目的。我们尝试提供有效力的、具有普遍意义的解决方法，但这不是我们唯一的主导目标。鉴于本书的双重目的，我们有时会倾向于看起来更易掌握的解决方法。

以前对 Excel 有所了解的读者肯定会获益，例如，他们知道如何使用一个简单的函数，如 AVERAGE（）；了解 SUM（A1：A10）、SUM（$A1：$A10）之间的区别此类。对于经验较少的读者，我们在附录中提供了对 VBA 的介绍。

我们假设读者对基础统计知识（如概率分布）和金融经济学（如贴现、期权）都多少有所了解。不过，当我们认为至少一些读者会从中获益时，我们会解释基本概念。例如，我们针对最大似然估计或回归添加了附录。

我们十分感谢对手稿提出反馈的同事、朋友和学生，他们是：

Oliver Blümke, Jürgen Bohrmann, André Güttler, Florian Kramer, Michael Kunisch, Clemens Prestele, Peter Raupach, Daniel Smith 和 Thomas Verchow。一个匿名的评论者也提供了很多有帮助的评论。我们感谢 Eva Nacca 所做的格式校正工作。最后我们感谢编辑 Caitlin Cornish, Emily Pears 和 Vivienne Wickham。

我们同样对任何错误和无意的与最佳实践的偏差负责。我们欢迎您的评论和建议，请发邮件到 comment@ loefflerposch. com 或访问我们的主页 www. loeffler-posch. com

我们对家人亏欠很多。在努力寻找合适的词语来表达感激之前，我们更应停下来，给予他们最需要的东西——我们的时间。

# 一些疑难问题的提示

我们希望您在应用本书中的工作表、宏和函数时不会遇到问题。如果您遇到了，则可以考虑下面可能的原因：

- 我们反复使用 Excel 求解程序。如果求解程序的嵌入程序在 Excel 和 VBA 中没有被激活，则可能产生问题。如何激活嵌入程序，见附录 A2。从表象看，Excel 的版本不同可能会造成调入求解程序的宏无法运行的情况，即使在对求解程序的参照已经设定的情况下。
- 在第 10 章，我们使用嵌入的分析工具中的函数。同样，这需要被激活。详见第 9 章。
- 一些 Excel 2003 的函数（如 BINOMDIST 或 CRITBINOM）相对较早版本而言已经改变。我们已经在 Excel 2003 检验了我们的程序。如果你正在使用较早版本的 Excel，这些函数可能在一些情况下给出错误值。
- 所有的函数只针对演示的目的而被测试。我们还没有努力使它们能适用到人们会想到的大多数目的。例如：
  - 一些函数假设数据按某种方式排序，或者按行而不是按列排列；
  - 一些函数假设自变量是一个范围而不是一个数组。解决这一问题详见附录 A1。

全部函数（Excel 的以及用户自定义的）的目录以及完整语句和简短的描述见附录 A5 的结尾。

# 目　　录

**第 3 章**



**第 4 章**



**第 5 章**

**第 8 章**



**第 9 章**



**第 10 章**

**第 11 章**

第 1 章

# 用 Logit 模型测算信用评分

通常情况下，借款者的违约概率会受到许多因素的影响。对于零售客户，一般需要考虑的因素包括贷款申请人的收入、职业、年龄以及其他特性；而针对公司客户，人们往往会考虑公司的杠杆水平、盈利能力或现金流等。此外还有很多其他因素，这里不再一一列举。评分模型（scoring model）明确了如何结合不同的信息以获得对违约概率的准确评估，从而帮助金融机构实现对违约风险的自动化和标准化评估。

本章将介绍如何运用被称为*逻辑回归*（或简称“logit”）的统计技术来确定评分模型。实质上，这相当于对信息赋予一个特定的值（例如，用负债/资产来衡量杠杆水平），然后求解出最佳的、能够解释历史违约行为的因素组合。

在明确了信用评分与违约概率之间的关系后，我们还将介绍如何估算并解释 logit 模型。然后，我们会对实际应用中的重要问题进行讨论，即探讨如何对异常值进行处理以及如何对变量与违约之间的函数关系进行选择。

在建立和运行一个成功的评分模型的过程中，验证（validation）是十分重要的一步。鉴于验证技术不仅被应用于评分模型，而且在机构评级（agency ratings）和对违约风险的其他度量方面也有着广泛应用，我们将在第 7 章对其进行单独讨论。

## 连接信用评分、违约概率以及观察到的违约行为

信用评分模型概括了影响违约概率的因子中所包含的信息。标准的

信用评分模型一般采用最直接的方式，即将这些因子线性地结合起来。令 $x$ 代表因子（数目为 $K$），$b$ 为这些因子的权重（或系数）；我们在评分实例 $i$ 中表示评分如下：

$$\text{Score}_i = b_1 x_{i1} + b_2 x_{i2} + \cdots + b_K x_{iK} \tag{1.1}$$

简捷的表达方式是将 $b$ 和 $x$ 表示为向量形式 $\mathbf{b}$ 和 $\mathbf{x}$，我们可以将（1.1）重写为：

$$\text{Score}_i = b_1 x_{i1} + b_2 x_{i2} + \cdots + b_K x_{iK} = \mathbf{b}'\mathbf{x}_i,\ \mathbf{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{iK} \end{bmatrix},\ \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_K \end{bmatrix} \tag{1.2}$$

如果该模型包含常数项 $b_1$，则令每个 $i$ 的 $x_{i1} = 1$。

简言之，假设我们已经确定了因子向量 $\mathbf{x}$，剩下需要确定的是权重向量 $\mathbf{b}$。通常以观察到的违约行为为基础进行估算。[①]设想我们已经收集到各个公司的因子值和违约行为的年度数据，表 1.1 就显示了这样一个数据集。[②]

**表 1.1**

**因子值和违约行为**

| 评分实例 $i$ | 公司 | 年份 | 次年的违约指标 $y_i$ | 年末因子值 $x_{i1}$ | $x_{i2}$ | | $x_{iK}$ |
|---|---|---|---|---|---|---|---|
| 1 | XAX | 2001 | 0 | 0.12 | 0.35 | … | 0.14 |
| 2 | YOX | 2001 | 0 | 0.15 | 0.51 | … | 0.04 |
| 3 | TUR | 2001 | 0 | -0.10 | 0.63 | … | 0.06 |
| 4 | BOK | 2001 | 1 | 0.16 | 0.21 | … | 0.12 |
| … | … | … | … | … | … | … | … |
| 912 | XAX | 2002 | 0 | -0.01 | 0.02 | … | 0.09 |
| 913 | YOX | 2002 | 0 | 0.15 | 0.54 | … | 0.08 |
| 914 | TUR | 2002 | 1 | 0.08 | 0.64 | … | 0.04 |
| … | … | … | … | … | … | … | … |
| N | VRA | 2005 | 0 | 0.04 | 0.76 | … | 0.03 |

① 在定性信用评分模型中则是由专家确定权重。

② 用于评分的数据通常都是以年度为基础的，当然我们也可以选择其他的数据收集频率和违约的观测时限。

表中如果包含同一家公司的若干年信息，则该公司会重复出现。一旦违约，公司通常会持续数年处于违约状态中。这种情况下，我们不采用违约当年之后的观察值。如果一家公司摆脱了违约状态，我们会再次将其纳入数据集中。

违约信息储存在变量 $y_i$ 中，如果某公司在搜集因素值年份的次年违约，则 $y_i$ 取值为 1，否则取值为 0。观察值的总数用 $N$ 表示。

信用评分模型会对违约公司的违约概率给出一个较高的预测值，而对未违约公司给出的违约概率预测值较低。为了选择适当的权重 $\mathbf{b}$，我们首先需要把信用评分与违约概率联系起来，这里可以把违约概率表示为信用评分的函数 $F$：

$$\text{Prob}(\text{Default}_i) = F(\text{Score}_i) \tag{1.3}$$

像违约概率一样，函数 F 应该被限制在 0 ~ 1 的区间内，而每一个可能的信用评分值应该对应一个违约概率，累计概率分布函数可以满足以上要求。为实现这一目的，通常考虑的是 logistic 分布。logistic 分布函数 $\Lambda$（$z$）被定义为 $\Lambda(z) = \exp(z)/(1 + \exp(z))$。代入到（1.3），我们可以得到：

$$\text{Prob}(\text{Default}_i) = \Lambda(\text{Score}_i) = \frac{\exp(\mathbf{b}'\mathbf{x}_i)}{1 + \exp(\mathbf{b}'\mathbf{x}_i)} = \frac{1}{1 + \exp(-\mathbf{b}'\mathbf{x}_i)} \tag{1.4}$$

运用 logistic 分布函数将信息和概率联系起来的模型叫做 logit 模型。

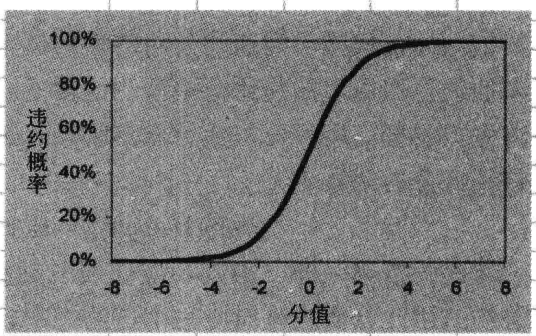
在表 1.2 中，我们列出了一些评分值以及与其对应的违约概率，并将它们之间的关系用图表示出来。可以看到，高分值对应着高违约概率。但是，在许多金融机构中，信用评分具有相反的特性：信用风险较低的借款人评分较高。此外，信用评分也被限制在某个特定的区间内，例如 0 ~ 100，这是很容易实现的。例如：如果我们用（1.4）来定义一个评分为 -9 ~ 1 的评分系统，但是希望用 0 ~ 100 的评分来替代（100 为最好），我们可以将最初的评分通过公式 $myscore = -10 \times score + 10$ 来转换。

在收集了因子 $\mathbf{x}$ 并选择了分布函数 $F$ 之后，为了估算出权重 $\mathbf{b}$，我们很自然地会考虑选择极大似然法（ML）。根据极大似然法的原理，我们选择能够使观察到的违约行为的似然值最大化的数值作为权重（见附录 A3 有关极大似然估计的更加详细的介绍）。

**表 1.2**

**logit 模型中的评分和违约概率**

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | 评分 | 违约概率 | | | | | | |
| 2 | -8 | 0.03% | =1/(1+EXP(-A2)) | | | | | |
| 3 | -7 | 0.09% | (可将公式复制到单元格B3:B18) | | | | | |
| 4 | -6 | 0.25% | | | | | | |
| 5 | -5 | 0.67% | | | | | | |
| 6 | -4 | 1.80% | | | | | | |
| 7 | -3 | 4.74% | | | | | | |
| 8 | -2 | 11.92% | | | | | | |
| 9 | -1 | 26.89% | | | | | | |
| 10 | 0 | 50.00% | | | | | | |
| 11 | 1 | 73.11% | | | | | | |
| 12 | 2 | 88.08% | | | | | | |
| 13 | 3 | 95.26% | | | | | | |
| 14 | 4 | 98.20% | | | | | | |
| 15 | 5 | 99.33% | | | | | | |
| 16 | 6 | 99.75% | | | | | | |
| 17 | 7 | 99.91% | | | | | | |
| 18 | 8 | 99.97% | | | | | | |




极大似然估计的第一步是建立似然函数。如果借款人违约（$Y_i = 1$），则观察到这一行为的概率是：

$$\text{Prob}(\text{Default}_i) = \Lambda(\mathbf{b}'\mathbf{x}_i) \tag{1.5}$$

如果借款人不违约（$Y_i = 0$），我们观察到这一行为的概率是：

$$\text{Prob}(\text{No default}_i) = 1 - \Lambda(\mathbf{b}'\mathbf{x}_i) \tag{1.6}$$

运用一点小技巧，我们可以将这两个公式合二为一，并自动给出正确的似然值，无论其是否违约。由于任何数字的 0 次方为 1，观察值 $i$ 的似然值可被写做：

$$L_i = (\Lambda(\mathbf{b}'\mathbf{x}_i))^{y_i}(1 - \Lambda(\mathbf{b}'\mathbf{x}_i))^{1-y_i} \tag{1.7}$$

假设违约是独立发生的，则一组观察值的似然值刚好是单个似然值的乘积[①]：

$$L = \prod_{i=1}^{N} L_i = \prod_{i=1}^{N} (\Lambda(\mathbf{b}'\mathbf{x}_i))^{y_i}(1 - \Lambda(\mathbf{b}'\mathbf{x}_i))^{1-y_i} \tag{1.8}$$

为了更加方便地求解极大值，取对数 $\ln L$：

① 如果有些年份违约率较高，而其他年份较低，人们会考虑独立性假设是否合适。而且还要求的输入因子能够刻画违约风险的波动。在很多应用中，独立性假设是合理的。

$$\ln L = \sum_{i=1}^{N} y_i \ln(\Lambda(\mathbf{b}'\mathbf{x}_i)) + (1 - y_i)\ln(1 - \Lambda(\mathbf{b}'\mathbf{x}_i)) \quad (1.9)$$

令关于 **b** 的一阶导数为 0，从而求解其极大值。这一导数（与 **b** 相同，是一个向量）如下：

$$\frac{\partial \ln L}{\partial \mathbf{b}} = \sum_{i=1}^{N} (y_i - \Lambda(\mathbf{b}'\mathbf{x}_i))\mathbf{x}_i \quad (1.10)$$

使用牛顿法（见附录 A3）求解（1.10）效果显著。使用这种方法，我们还需要求 **b** 的二阶导数，具体如下：

$$\frac{\partial^2 \ln L}{\partial \mathbf{b} \partial \mathbf{b}'} = -\sum_{i=1}^{N} \Lambda(\mathbf{b}'\mathbf{x}_i)(1 - \Lambda(\mathbf{b}'\mathbf{x}_i))\mathbf{x}_i\mathbf{x}'_i \quad (1.11)$$

## 用 Excel 估算 logit 系数

由于 Excel 不包含估算 logit 模型的程序，我们将简要介绍如何建立一个由用户自定义的函数来完成这个任务。我们将完整的函数称为 LOGIT。LOGIT 命令的句法与 LINEST 命令相同：LOGIT（$y$，$x$，[const]，[statistics]），其中［ ］表示可选参数。

第一个参数是因变量，在我们的例子中因变量就是代表违约的指示变量 $y$；第二个参数是解释变量；第三和第四个参数分别是关于包含一个常数项（若包含常数项则为 1 或省略，否则为 0）和回归统计量计算（若为 1 表示将进行统计计算，否则为 0 或省略）的逻辑值。函数给出的是一个数组，因此，它必须在一系列单元格的基础上执行，并通过［Ctrl］+［Shift］+［Enter］输入。

在开始编程之前，让我们先来看看该函数如何应用于例子中的数据集。[①] 我们搜集了违约信息以及用于预测违约的五个变量：营运资本（Working Capital，WC）、留存收益（Retained Earnings，RE）、息税前收益（EBIT）和销售额（S）这四个变量与总资产（Total Assets，TA）的比值，以及权益的市场价值（Market Value of Equity，ME）与总负债（Total Liabilities，TL）的比值。除了市场价值是根据发行的股票数目乘以股价获得的，其他数值都是从公司的资产负债表和损益表中获取。这五个比率来自于奥尔特曼（Altman，1968）提出的著名的 Z－score 法。

① 数据虽然是虚拟的，但是反映了美国公司的数据结构。

WC/TA 用来度量公司的短期流动性，RE/TA 和 EBIT/TA 分别度量历史和当前的盈利能力。S/TA 代表公司的竞争环境，ME/TL 是基于市场的、对杠杆水平的测度。

当然，人们也会考虑其他变量，比如现金流与债务付息额的差值、销售额或总资产（代表规模）、收益波动性、股价波动性等。另外，同一基本因子（underlying factor）可以用多种方法进行获取，例如当前利润可以用 EBIT、EBITDA（等于 EBIT 加上折旧和摊销）或净收入来衡量。

在表 1.3 中，数据被汇集在 A ~ H 列中。公司的名称和年份不需要估算。LOGIT 函数被应用于区域 J2 到 O2 内，LOGIT 函数中的违约变量数据位于区域 C2 到 C4001 内，而因子 $x$ 的数据在区域 D2 到 H4001 中。需要注意的是，不同于 Excel 中的 LINEST 函数，系数返回的顺序与变量输入的顺序相同，常数项（如果包含的话）被安排在最左边。解释系数 $b$ 的符号需要依据较高评分代表较高违约概率这一事先设定。例如，EBIT/TA 的负系数表明违约概率随着盈利能力的增加而减少。

**表 1.3**

**对一个包含违约信息和五个财务比率的数据集应用 LOGIT 命令**

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 公司名称 | 年份 | 违约 | WC/TA | RE/TA | EBIT/TA | ME/TL | S/TA | | 常数 | WC/TA | RE/TA | EBIT/TA | ME/TL | S/TA |
| 2 | 1 | 1999 | 0 | 0.50 | 0.31 | 0.04 | 0.96 | 0.33 | b | -2.543 | 0.414 | -1.454 | -7.999 | -1.594 | 0.620 |
| 3 | 1 | 2000 | 0 | 0.55 | 0.32 | 0.05 | 1.06 | 0.33 | {=LOGIT(C2:C4001,D2:H4001,1,0)} | | | | | | |
| 4 | 1 | 2001 | 0 | 0.45 | 0.23 | 0.03 | 0.80 | 0.25 | (可将公式应用于单元格J2:O2) | | | | | | |
| 5 | 1 | 2002 | 0 | 0.31 | 0.19 | 0.03 | 0.39 | 0.25 | | | | | | | |
| 6 | 1 | 2003 | 0 | 0.45 | 0.22 | 0.03 | 0.79 | 0.28 | | | | | | | |
| 7 | 1 | 2004 | 0 | 0.46 | 0.22 | 0.03 | 1.29 | 0.32 | | | | | | | |
| 8 | 2 | 1999 | 0 | 0.01 | -0.03 | 0.01 | 0.11 | 0.25 | | | | | | | |
| 9 | 2 | 2000 | 0 | -0.11 | -0.12 | 0.03 | 0.15 | 0.32 | | | | | | | |
| ... | | | | | | | | | | | | | | | |
| 108 | 21 | 1996 | 1 | 0.36 | 0.06 | 0.03 | 3.20 | 0.28 | | | | | | | |
| ... | | | | | | | | | | | | | | | |
| 4001 | 830 | 2002 | 1 | 0.07 | -0.11 | 0.04 | 0.04 | 0.12 | | | | | | | |

现在让我们进一步探讨 LOGIT 编程中的一些重要部分。在函数的第一行，我们分析其中的输入数据，以定义数据的维度，即观察值 N 的总数和解释变量（包括常数项）K 的数目。如果包含常数项（正常情况下应该包含），我们必须在解释变量矩阵中加入全为 1 的向量。这