



经典实用的信用风险管理方法，国际领先的标准
评分卡模型开发技术



基于R语言的 证券公司 信用风险计量和管理

崔玉征◎编著

清华大学出版社



基于R语言的
证券公司
信用风险计量和管理



崔玉征◎编著

清华大学出版社
北京

内 容 简 介

本书共分为两部分,第一部分主要讲述信用风险评级模型的开发方法,主要包括自动建设信用风险数据库、信用风险标准评分卡模型、KMV 模型、Z-Score 模型等核心技术,并公布了全部模型的 R 语言源代码;第二部分主要讲述实用的信用风险管理方法,主要包括风险规避、风险缓释、风险收益匹配、经济资本管理等实用方法。

本书适合在银行、证券、保险、基金等金融机构从事信贷和债券投资等相关业务的从业者阅读。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

基于 R 语言的证券公司信用风险计量和管理/崔玉征编著. —北京:清华大学出版社,2017

ISBN 978-7-302-46349-8

I. ①基… II. ①崔… III. ①证券公司—业信用—风险管理—研究 IV. ①F830.39

中国版本图书馆 CIP 数据核字(2017)第 014234 号

责任编辑:张 伟

封面设计:李召霞

责任校对:王荣静

责任印制:杨 艳

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印装者:三河市吉祥印务有限公司

经 销:全国新华书店

开 本:170mm×240mm 印 张:20.5 字 数:389 千字

版 次:2017 年 3 月第 1 版 印 次:2017 年 3 月第 1 次印刷

定 价:69.00 元

产品编号:072849-01



我的同事崔玉征先生，集其多年来在国内外大型金融机构信用风险领域的理论探索与实践经验撰写了本书。他在谈及写作和出版目的时告诉我：国内相关的著作并不少，但缺乏一种可以在理论上把问题说清楚，同时可以据此实操的指南型书籍。同时他也希望让国内更多的同行分享他在信用风险研究方面的经验。

企业信用风险计量与管理能力是我国金融行业目前相对较弱的模块。其原因我认为有以下几个方面。

(1) 过去的 30 年里，中国的宏观经济形势持续向好，经济总量高速增长，资产规模快速膨胀，其增量可以有效地覆盖许多风险。

(2) 近 10 年来，地方政府与国企负债规模快速增长，不少金融机构以争夺该类资产作为风险管理的主要策略，而忽略打造面向市场的风险管理能力。

(3) 我国基础经济数据的质量一直没有得到有效的提高，迫使金融机构使用定性研究多于使用量化工具。

凡此种种，我们面临的一个现实是：中小金融机构在风险管理方面普遍存在短板。

从 2013 年开始，我国的宏观经济进入中低速发展的“新常态”，经济总量高速增长，资产价格不断攀升的局面不再持续，取而代之的是宏观经济的中低速增长。资产价格结束单边牛市，进入一个“波动市”。不少行业出现了较为严重的产能过剩，企业倒闭、破产、重组等现象，规模比过去更大、更频繁。简言之，企业的风险正在加速暴露，它带给金融行业的风险也在不断显现。

在此局面下，金融机构必须认识到这些风险的产生是金融行业的“新常态”。我们必须要去适应这种“新常态”，打造我们基于市场环境的风险管理能力。

时代在变化，一方面给我们带来前所未有的挑战；另一方面也给我们带来许多重要的战略性机遇。

互联网对人类生活的渗透日趋加剧，对企业的影响也是更加深入。在这种环境下形成的大数据，为我们的信用研究提供了更多的维度、更客观真实的

数据和更丰富的研究方法。

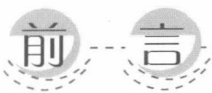
人工智能/机器学习在近年产生了革命性的进步。基于互联网数据的个人征信模型，已经进入实用阶段。它的高效率与可靠性已经大大超越人工信用管理。在企业的信用评价与风险管理方面，更是具有广泛的战略性应用前景。

总之，风险管理能力是金融机构最有价值的核心竞争力。期待本书的读者们在风险管理研究方面与我们共同成长。

禹荣刚

深圳正前方金融服务有限公司 董事长

2016年12月25日



我们先看看自从 2014 年 3 月 5 日国内第一只债券“11 超日债”发生违约以来资本市场信用风险事件发生的整体概况。

据 Wind 统计数据，2014 年全年共计 5 只债券发生了违约，合计违约金额 11.6 亿元；2015 年全年共计 19 只债券发生了违约，合计违约金额 117.1 亿元；截至 2016 年 6 月 30 日共计 36 只债券发生了违约，合计违约金额 235.91 亿元。通过这些数据可见，国内资本市场发生违约的事件和违约金额均呈爆发式增长的态势。

可是，从全球视野和信用风险评级模型开发的角度来看，目前发生违约的这些样本量仍然是太少，我们仍然无法采用科学、专业的方法来开发能够有效预测违约发生概率的信用风险评级模型。本书介绍的开发信用风险评级模型的实用方法，全都经过至少 3 年在国内资本市场的实践应用，并已经被事实证明了是行之有效的。

本书详细讲述的信用风险评级模型开发技术主要包括：

- (1) 适用于评估资本市场个人融资主体的 AHP 法模型开发技术；
- (2) 以互联网公开数据集 GermanCredit 为例，详细讲述信用风险标准评分卡模型开发技术；
- (3) 基于实质性违约样本，提取机构发生实质性违约前的“征兆”，建立“技术性违约”的标准，并以此来开发适用于机构的信用风险评级模型的技术。

针对机构的上述做法已经被实践证明是成功的，收到了很好的效果。在理论上上述做法同样是可解释的，因为我们可以把机构比喻成一个“生命体”，该“生命体”的血液即为企业的现金流，那么任何一个“生命体”一般都不会突然“病入膏肓”，即任何一个正常经营的企业一般都不会突然违约！在违约之前，一定会有一些“征兆”，这些“征兆”就是机构主体评级模型中各指标变化的趋势。如何选择这些“征兆”呢？目前，国内资本市场已经有 61 只债券发生了实质性违约，共涉及 37 个发债主体，通过研究这些实质性违约主体最近 3~5 年的财务数据及其他定性数据，我们不难从中发现最终导致实质性违约发生的“征兆”。

在信用风险评级模型开发时,违约通常被定义为因变量(D),而违约的发生是一系列因素(即自变量,包括定量和定性等指标)共同作用的结果。因此,因变量与自变量之间可以用如下函数关系式表示: $D=f(\text{定量指标、定性指标及其他综合因素})$ 。通过分析已经发生实质性违约的 37 家发债主体,我们不难提取一些导致其发生实质性违约的共同指标,通过分析这些指标变化的趋势,违约前的“征兆”也就不难发现了。我们将这些“征兆”定义为“技术性违约”,并以此来开发信用风险评级模型。

作者采用这种方法开发的机构主体评级模型,经过 3 年多在证券公司的实际应用,不论是模型的区分能力、预测能力,还是模型的稳定性方面,都表现出了很好的效果。

本书定位为在资本市场从事信用风险计量和管理的工具书。如果你在当前市场条件下开发信用风险评级模型时遇到困难,阅读本书你一定会找到实用性的解决方法;如果你在如何管理信用风险时遇到困难,阅读本书你也会找到实用的解决方法。

本书共分为两部分。第一部分重点讲述信用风险的计量方法和适用于证券公司的信用风险计量体系,主要包括第一章至第十章的内容,并附以大量的 R 实例代码。第二部分重点讲述适用于证券公司的信用风险管理方法,主要包括第十一章和第十二章的内容。

科学的信用风险计量和管理体系,需要先进的信用风险管理系统。衡泰软件的衡泰信用风险管理系统(xCRMS),能实现企业级信用风险计量、分析、管理和报表的全流程系统化管理。xCRMS 功能齐全,可以非常方便地配置本书中讲述的所有模型,实现评级模型的批量化运算、测试和管理。衡泰信用风险管理系统的强大功能已经在安信证券、招商证券、国泰君安证券、中投证券、光大证券、新时代证券、国都证券等众多证券公司中得到了良好的应用。

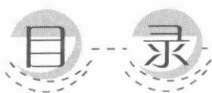
本书假定读者非常熟悉模型开发工具 R 编程语言,对于 R 语言新手来说,建议首先阅读附录 A 中的内容,以基本熟悉 R 语言的相关语法。

本书在撰写过程中得到了家人、朋友和业内专业人士的大力支持,在此对他们一一表示感谢。他们分别是深圳正前方金融服务有限公司董事长禹荣刚先生、杭州衡泰软件有限公司董事长徐行先生、美国穆迪董事总经理邵可弟先生、远光软件股份有限公司董事会秘书戴文斌先生。再次对他们表示感谢。

读者若有问题与作者交流,请发邮件至 yuzheng.cui@qq.com。

作 者

2016 年 10 月



第一部分 证券公司信用风险计量体系

第一章 信用风险计量体系的构成	3
第一节 信用风险概述	3
第二节 主体计量体系	3
第三节 债项计量体系	5
第二章 环境配置及数据库建设	7
第一节 数据库配置	7
第二节 建模工具 R 的安装和配置方法	21
第三节 自动获取建模所需数据	28
第三章 信用风险评级模型的开发过程	35
第一节 信用风险评级模型的类型	35
第二节 信用风险评级模型开发流程概述	35
第三节 明确要解决的问题	37
第四节 数据准备及数据预处理	38
第五节 变量选择	43
第六节 模型开发	52
第七节 主标尺设计及模型验证	57
第八节 模型实施	59
第九节 模型监测与报告	60
第四章 逻辑回归	64
第一节 基本原理	64
第二节 似然方程	66
第三节 Hessian 矩阵及参数估计	67
第四节 模型拟合统计量	68

第五章 个人主体违约概率计量	69
第一节 层次分析法及 R 源代码	69
第二节 基于逻辑回归的标准评分卡方法及 R 源代码	81
第三节 实用的数据预处理方法及 R 源代码	142
第六章 机构主体违约概率计量	151
第一节 机构主体信用风险标准评分卡模型及 R 源代码	151
第二节 KMV 模型及 R 源代码	215
第三节 Z-Score 模型及 R 源代码	222
第七章 违约损失率计量	238
第一节 违约损失率的定义及概述	238
第二节 实用的 LGD 模型开发方法	239
第八章 违约风险敞口计量	247
第一节 投融资类业务违约风险敞口计量	247
第二节 交易对手违约风险敞口计量	247
第九章 主标尺及模型验证	250
第一节 违约概率校准及主标尺设计	250
第二节 模型验证体系及 R 源代码	257
第十章 信用风险经济资本计量	264
第一节 经济资本概述	264
第二节 单笔融资类业务经济资本的计算	268
第三节 投资组合经济资本的计算	271

第二部分 证券公司信用风险管理体系

第十一章 信用风险管理体系	277
第一节 风险规避	277
第二节 风险缓释	278
第三节 风险收益匹配	278
第十二章 信用风险经济资本管理	280
第一节 经济资本分配	280
第二节 限额管理	282

第三节 风险定价.....	282
第四节 绩效考核.....	283
附录 A R 语言基础	285
附录 B 大数据技术在信用评级中的应用及 R 源代码	297

第一部分

证券公司信用风险计量体系

本部分主要介绍证券公司的信用风险计量体系及信用风险评级模型开发技术,主要包括信用风险计量体系的构成、数据库建设、评级模型开发的基本流程、逻辑回归的基本原理、个人和机构主体评级模型开发技术、违约损失率的计量、违约风险敞口的计量、主标尺的设计及模型验证体系和信用风险经济资本的计量十个章节。这十个章节的有机结合共同构成了适用于证券公司信用风险的计量体系。

本章分为三节，逐步讲述信用风险管理的核心要素及信用风险计量体系的组成部分，主要包括信用风险概述、主体计量体系和债项计量体系三部分。

第一节 信用风险概述

信用风险是指融资人或交易对手(以下简称主体)不能按照约定按时偿还债务的可能性。当主体不能按时履约时,我们通常定义为主体发生了违约。因此,信用风险管理有两个核心要素:第一个是对主体是否会发生违约进行预测管理;第二个是假设主体违约后对证券公司产生的后果进行预测管理。简而言之,这两个核心要素中第一个是违约(对应风控指标 PD, probability of default);第二个是对违约后果的管理(对应风控指标 LGD 和 EAD, loss given default and exposure at default)。这两个核心要素中,第一个要素是信用风险管理的重点;其次才是对违约后果的管理。因为,信用风险的源头是违约,即如果主体不发生违约,信用风险就不会发生,也就没有必要管理违约的后果。在数学上也可以给出严谨的证明,证券公司所有融资类业务的最大盈利点,一定是在所有主体均不违约时达到。

那么,我们怎么管理违约及违约的后果呢?

信用风险主体评级模型是对主体未来一段时间内是否会发生违约的前瞻性预测;信用风险债项评级模型是假设主体发生违约后,对证券公司可能面临的损失程度的前瞻性预测。因此,我们用主体评级模型去管理违约,用债项评级模型去管理违约的后果。

证券公司需承担信用风险的业务主要包括融资融券、股票质押、约定购回、债券投资、信托投资、资产证券化及其他非标融资类业务、场外衍生品业务、代销金融产品等。

第二节 主体计量体系

由第一节的分析可知,我们采用主体评级模型来预测主体发生违约的概率。那么,主体的类型包括自然人、法人、控股股东、实际控制人、机构等,并且机构又

分为不同的行业,我们需要对各种类型的主体均开发不同的主体评级模型吗?我们知道个人和机构的属性有着本质的不同,个人的主观随意性较强,即使还款能力很强的个人,也时常出现违约的情况;而机构是否违约基本是由其客观的还款能力所决定,即机构基本是“按章办事”,债务到期时如果还款能力足够,就不会发生违约,受主观随意性影响很小。

机构有不同的行业之分,每个行业的特性均存在差异。因此,证券公司的主体计量体系可设计为图 1.1 所示的架构,这也是国际通用的做法。

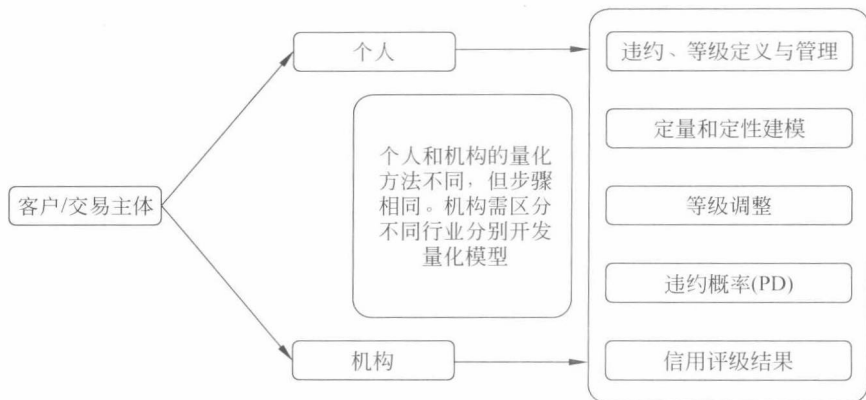


图 1.1 对主体量化的架构设计

对个人和机构及行业分类的认定,需做如下说明。

(1) 个人主体评级模型适用于自然人客户,即不需要给其他人“发工资”的个人融资客户,均适用于个人主体评级模型。

(2) 控股股东、实际控制人客户,按其控制的上市公司所属行业进行主体评级,即给人“发工资”的个人融资客户,由于其还款能力受其控制企业的盈利状况影响很大,因此按机构的主体评级模型评估其信用等级,这也是国际通用的做法。

(3) 对机构主体评级模型的开发需要区分不同的行业,分别开发不同的评级模型。需要重点指出的是,行业的划分没有确定的标准,通常根据数据收集的情况及企业实际的主营业务进行划分。但是,就目前国内资本市场的实际情况来看,我们通常需要将同一产业链中相同属性的上中下游企业进行合并处理,以获取足够的用于模型开发的样本总体。笔者曾按照这种做法开发了国内 19 个行业的主体评级模型,经过 3 年多的实际应用,取得了很好的实践效果。再者,这种合并产业相同属性的上中下游企业以获得足够建模样本的方法,不论在理论层面还是在实践层面都是非常合理的。比如,目前产能严重过剩的钢铁行业,其上游是铁矿石、煤炭等矿产行业,钢铁行业的过剩直接导致了采矿行业的不景气,且这些行业的经营模式也是基本相同的。因此,在做模型开发时,完全可以

将钢铁、煤炭、采矿等合并为一个新的行业——矿产能源行业。

(4) 本书所指的主体评级模型均是指信用风险标准评分卡模型,该类模型是目前全球在信用风险管理领域使用最广泛的一类模型,而主体评级模型又分为申请者评级、行为评级、催收评级和欺诈评级四类。

第三节 债项计量体系

由第一节的分析可知,债项计量体系共包括两个量化指标,分别是违约损失率(LGD)和违约风险敞口(EAD)。在模型分类上,LGD模型通常根据企业的融资用途开发不同的计量模型,我们通常根据主体的融资用途,将债项评级模型分为企业融资模型、现金流融资模型、项目融资模型等。EAD模型不需要分别开发,只需要根据违约风险敞口等于融出本金加上应收利息及必要的货币时间价值调整项计算即可。适用于证券行业的EAD模型与银行业的EAD模型有较大的不同,因为证券行业的融资类业务一般是一次性还本付息,而银行业一般是每月分期偿还本息。

需要特别说明的是,在进行债项评级模型开发之前,需首先开发并评估主体的信用等级。这主要是因为信用风险管理体系中,主体是否违约是因,对违约后果的管理是果,且主体评级与债项评级正相关(主体评级越高、债项评级也越高,主体的违约概率越低、其债项的违约损失率也越低)。

综上分析,债项的量化架构设计如图1.2所示。

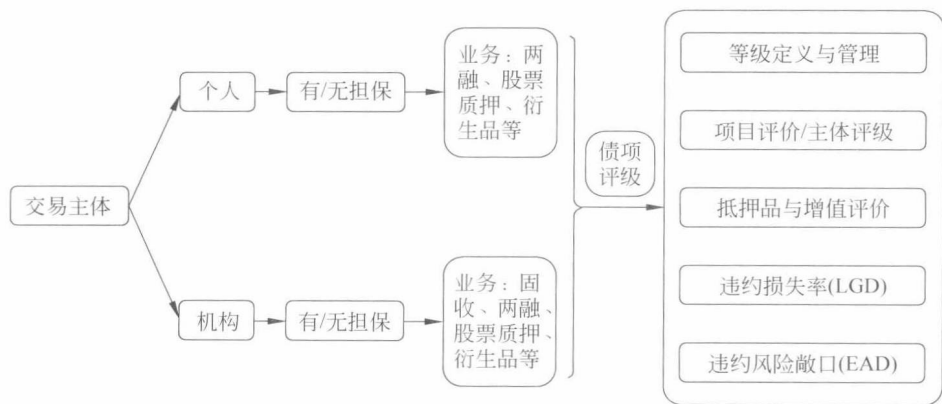


图 1.2 债项的量化架构设计

综合主体和债项的量化架构设计,证券公司的整个信用风险计量和管理体系如图1.3所示。每个主体对应一个唯一的违约概率(PD),且PD的计量与债项无关;同一主体开展不同的业务产生的债项,对应不同的违约损失率(LGD)和违约风险敞口(EAD);每个主体对应一个预期损失(EL),等于其违约概率

(PD)乘以其所有债项的汇总 $[EL = PD \times \sum (LGD_i \times EAD_i)]$ 。有了上述信用风险基础计量指标的计算结果,我们可以对信用风险管理的高级方法——经济资本管理法的核心指标进行计量,如非预期损失(UL)、经济资本(EC)、风险调整后资本收益(RAROC)等。

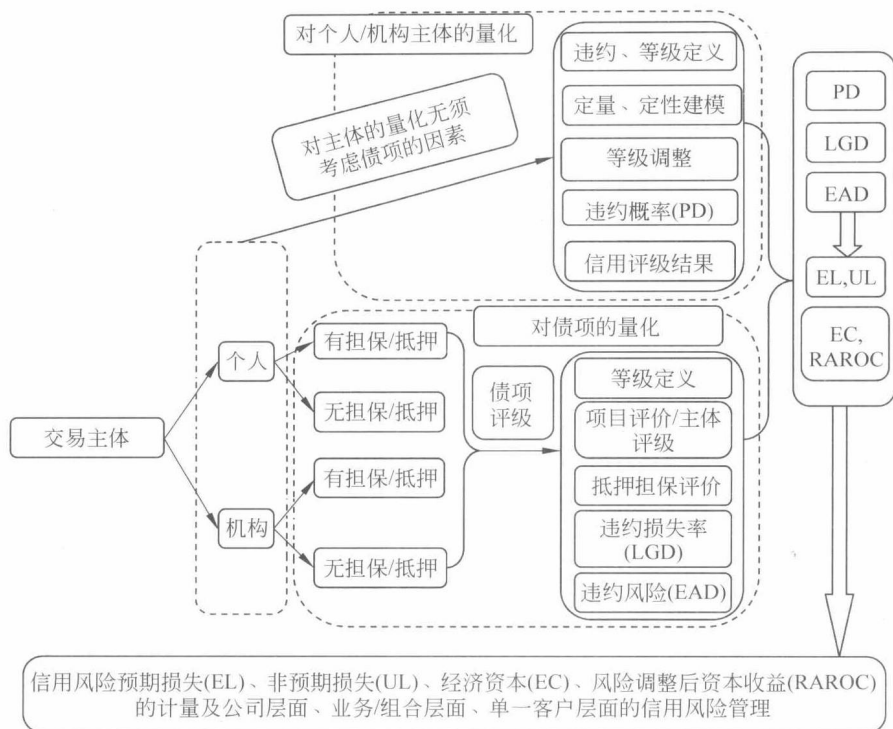


图 1.3 证券公司的整个信用风险计量和管理体系

第二章

环境配置及数据库建设

开发信用风险评级模型首先要有数据,本章重点讲述自动获取相关数据的数据库建设方法,共分为三节,第一节讲述数据库安装与配置方法,第二节讲述建模工具 R 软件的安装与配置,第三节讲述自动获取数据的实例。

第一节 数据库配置

本书以开源、免费的数据库软件 MySQL 为例,来逐步讲述其安装和配置过程,这也是建设信用风险数据库的必要前提。各步骤的详细过程如下所示。

(1) 访问 MySQL 首页(<http://www.mysql.com/>),如图 2.1 所示。

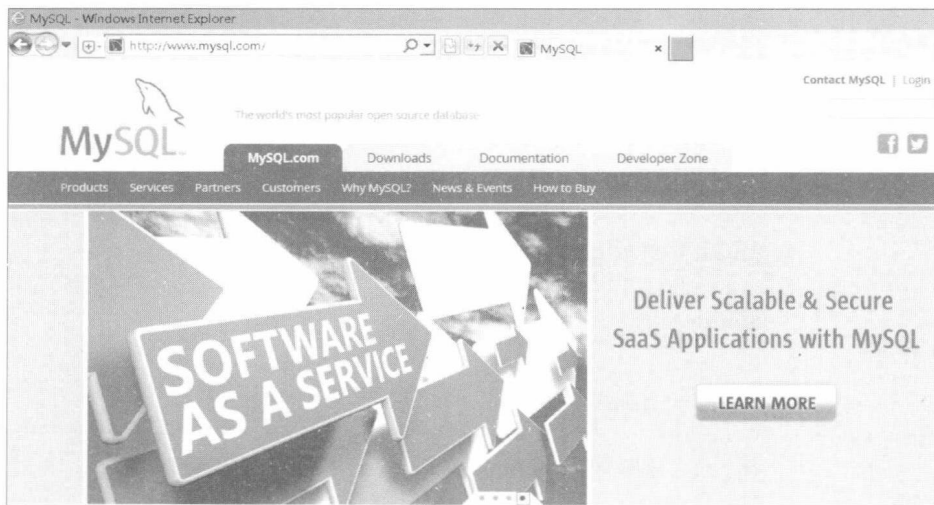


图 2.1 MySQL 网站首页

(2) 下载免费的 MySQL 社区版,选择免费的社区版 MySQL 服务器 (Community Server),单击“DOWNLOAD”,如图 2.2 所示。

(3) 选择 Windows 平台的版本,在“Select Platform”下拉列表中选择“Microsoft Windows”平台,并根据自己电脑的配置选择 32 位或 64 位平台(本书以 32 位机器为例),单击“Download”下载,如图 2.3 所示。