极值估计
在金融保险中的应用

EXTREME VALUE ESTIMATION AND ITS APPLICATIONS IN FINANCE AND INSURANCE

欧阳资生◆著

中国经济出版社
CHINA ECONOMIC PUBLISHING HOUSE

湖南商学院学术专著基金和湖南省自然科学基金（编号：05JJ40106）资助出版

极值估计在金融保险中的应用

欧阳资生 著

中国经济出版社
CHINA ECONOMIC PUBLISHING HOUSE
北京

图书在版编目（CIP）数据

极值估计在金融保险中的应用/欧阳资生著．—北京：中国经济出版社，2006.4

ISBN 7-5017-7433-1

Ⅰ．极...　Ⅱ．欧...　Ⅲ．①极值（数学）—数量统计—应用—金融②极值（数学）—数理统计—应用—保险　Ⅳ．①F83②F84

中国版本图书馆 CIP 数据核字（2006）第 022710 号

出版发行：中国经济出版社（100037·北京市西城区百万庄北街 3 号）
网　　址：www.economyph.com
责任编辑：许秀江
电　　话：68319291　13683308557
E-mail：xxj09@163.com
责任印制：石星岳
封面设计：中子画艺术
经　　销：各地新华书店
承　　印：北京君升印刷公司
开　　本：880mm×1230mm　1/32　**印张：**7.5　**字数：**200 千字
版　　次：2006 年 4 月第 1 版　**印次：**2006 年 4 月第 1 次印刷
印　　数：1—5000 册
书　　号：ISBN 7-5017-7433-1 / F·5989　**定价：**18.00 元

服务热线：68344225　68369586　68346406　68309176

序 言

怎样对极值风险进行合理、有效度量，是金融保险业和金融保险研究领域一个备受关注的研究课题。在学术研究和实际领域中，风险度量和保险保费的计算方法很多，但对如何度量像1987年10月股市崩盘这些极值风险和“9·11事件”这类巨灾保险的保费的方法却很少。从这个角度考察，选择这一课题进行研究的确充满挑战性和难度。如何在前人研究的基础上，梳理出清晰的研究思路，如何在理论上有所创新，如何对于极值风险度量和巨灾保险保费的计算提出真知灼见，都是一个研究者必须面对的难题。但是，我很高兴地看到，本书作者欧阳资生以自己的努力对上述问题做出了有价值的探索。

对于统计、风险管理和保险精算工作者而言，极值理论由于它的研究对象的不寻常性而具有特别的吸引力。作为一种对随机现象的研究，极值理论最早可以追溯到20世纪早期，但是直到20世纪50年代，才开始真正引起科研工作者的注意，并开始对它建模. 极值模型的应用始于工程设计，现已广泛应用于金融、保险、水利、气象等各个方面。本书的作者着重研究了极值理论，其中主要是极值估计在金融、保险中的应用。

金融与保险中的极值事件就是那些发生概率很小，但又

对金融和保险业造成重大影响的（有时是毁灭性的）事件。这些极值事件如果发生，一般都超过了单个保险公司和金融机构的承受力，或对其造成严峻的冲击。对这些公司而言，其结果都是灾难性的。

在利用极值理论进行风险管理时，首先必须对极值事件的统计规律进行分析，得出极值分布中各个参数和高分位数的估计，这时本书的极值估计方法就派上了用场。在作者理论与实践并重的研究中，我们可以看出作者对极值理论的全面、系统的理解。本书作者首先系统地研究了极值估计的方法，从数学证明和计算机随机模拟两个方面验证模型，然后，将估计理论应用于金融、保险中，对金融、保险中的极值事件建立模型，并以我国实际的股票收益率数据和医疗及巨灾保险索赔数据进行实证分析，达到了对金融、保险中的极值风险进行有效度量的目的。作者的上述研究以极值估计为基础，不仅有理论上的创新，也有历史经验数据的支持。从这个意义上，我认为本书作者所完成的研究工作的价值是怎样高估也不过分的。

欧阳资生同志在2001年9月考入中国人民大学统计学院，成为一名博士生，我有幸成为他的导师。2004年，他顺利毕业。在3年时间里，欧阳资生同志一直勤奋努力，在学业上取得了优异成绩。在校期间他就参加了多个项目的研究，在应用数理统计、金融和保险研究方面打下了扎实的基础，展现了出色的研究能力。在博士毕业论文写作时，他选择了“极值估计及其在金融保险中的应用研究”的课题进行研究，我认为这是一个难度较大的课题。但是他迎难而上，认真钻

研，不懈努力，终于以一篇十分优秀的博士论文完成了论文答辩。现在这篇论文又以专著的形式出版。作为一个教师，最愉快的就是看着学生成长、成熟，并在自己的专业领域学有所成。我衷心祝愿他以本书的出版为新的起点，在今后获得更大的成就。

吴喜之

2006 年 3 月 25 日于拉萨西藏大学

目 录

第三章 修正的 Pickands 估计门限值的自助估计方法

第四章 基于指数回归模型的极值估计的门限值的选取方法

第五章 一种概率分布的高分位数的最优估计

第六章 小样本情形下适度删失时的极值指数估计

第七章 极值估计在度量极值风险中的应用

第八章 大的索赔数据的广义 Pareto 分布拟合

第九章 贝叶斯极值估计及其在信用估计中的应用

第一章 引 言

§1.1 研究意义与国内外研究综述

1.1.1 研究意义

对于统计工作者而言，极值理论由于它的研究对象的不寻常性而具有特别的吸引力。作为一种对随机现象的研究，极值理论最早可以追溯到20世纪早期，但是直到20世纪50年代，才开始真正引起科研工作者的注意，并开始对它建模。极值模型的应用始于工程设计，现已广泛应用于金融、保险、水利、气象等各个方面。我们这里着重研究极值理论在金融、保险中的应用。

在保险中，如果我们考察1970年以来世界范围内所发生的金额最大的30起索赔和最严重的30起自然灾害（见表1.1和表1.2），我们不难从这些事件中找出一些共同的特征：

（1）它们对保险业和再保险业造成了相当大的影响；

（2）人们难以对它们做出远期预测；

（3）纵观整个保险业的历程，这些事件发生的概率很小，通常被称之为极值事件或稀有事件（Rare Event）。

粗略地讲，保险与金融中的极值事件就是那些发生概率

很小，但又对保险和金融业造成重大影响的（有时是毁灭性的）事件，这些极值事件如果发生，一般都超过了单个保险公司和金融机构的承受力，或对其造成严峻的冲击。由表 1.1 和表 1.2 可以清楚地看到，表中所列的每个小概率事件（极值事件）都超过了单个保险公司的赔付能力，或对其形成严重的冲击，对这些公司而言，其结果都是灾难性的。

事实上，如果我们选定一家保险公司，然后收集该公司的每次索赔额的历史数据进行分析，往往会发现在保险业中有一个有趣的现象：即“占总次数 20% 的那些索赔额的数额之和大约是公司历史索赔总额的 80%（有些公司还不止 80%）!”（参阅 Embrechts 等（1997））。因此如何准确地刻画这些极值事件便是统计工作者刻不容缓的任务。

在金融领域，极值理论也已受到越来越多的银行家、金融工作者的注意。对金融资产收益率，过去，人们常常利用正态分布建模，从而用方差来度量风险。然而，早在 1963 年，Mandelbort 在他的文章中就已指出：高额的金融资产收益率是非正态的，是“厚尾”的。所谓厚尾分布，意即它的极值实现值要比正态分布的大，并且更加频繁。此后的大量研究证明了 Mandelbort 的观点是正确的，并且发现传统的方差-协方差法、历史模拟法、蒙特卡罗模拟法在估计金融资产收益率的 VaR 值时的低效。事实上，VaR 只是损失分布的一个极端分位点。通过估计损失分布来估计分位点，在实际应用中存在许多影响估计效率的问题。一般来说，观测数据中，极端值出现得比较少，特别是越极端的值越难出现，因此使用历史模拟法、蒙特卡罗模拟法来获得模拟值，寻找满

足条件的分位点时，常常会低估 VaR，同时这些估计对极端观察值的反应也很灵敏。而使用方差—协方差法也会因正态假设不能反映金融数据的厚尾性而常常使得 VaR 被低估。这正如 McNeil（1997）所指出的“在某种意义上，对极端值的分析中，从来就不可能有足够的数据。……由于仅仅很少的一些点进入尾部区域，我们对分布尾部的推断就更不确定。……此外推断对大的损失观察值也很灵敏，新的极值损失引入数据集也许对分析有很强的影响”。而且在实际的风险管理中，人们往往对金融资产收益率大起大落时的情况更为关心。这正如 J. B. Philippe（2000）所指出的“对于极值事件，从来没有证明价格波动的高斯定理成立，这是因为中心极限定理仅能应用于分布的中心区。现在很清楚，所有金融领域最关心的是这些极端风险，首先要控制的也是它们。最近几年，国际监管当局一直试图制定一些规定以限制银行暴露在这些极端风险面前。……简单地去掉这些极值事件的影响的做法是相当愚蠢的”。为了解决这一问题，以便更精确地估计 VaR，统计学家很自然地将统计极值理论引入到这个问题的研究中。

从以上分析可以看出，极值统计作为一种研究极端现象的工具，具有不可替代的作用。而在极值统计中，关键的问题是极值分布中各个参数和高分位数的估计。因此，我们的研究无论从理论上，还是在实践上都具有重要的意义。

1.1.2 国内外研究综述

在极值分析中主要有两类模型，一类是极值定理模型

(EVT)。这类模型主要对组内最大值建模，即所谓的区组最大方法（Block Maxima method，BMM）。例如我们知道某一资产若干年的损失值，BMM 则可用来分析月度、季度、年度的最大值的统计规律。在 BMM 中，极限型定理保证了组内最大的极限分布不外乎 Fréchet、Weibull 和 Gumber 分布之一，或者它们的一般形式——广义极值分布（GEV）。另一类是广义 Pareto 分布模型（GPD）,这一模型也称为 POT 模型（Peaks - Over - Thresholds），它对观察值中所有超过某一较大门限值(threshold）的数据建模。由于广义 Pareto 模型有效地使用了有限的极端观察值。因此，通常认为在实践中很有实际意义。无论是 EVT 模型还是 GPD 模型，关键的问题是对模型中的参数，事实上，主要是极值指数的估计。根据估计的方法，我们还可进一步划分为两类不同的研究方法，即围绕 Pickands 估计，Hill 估计和矩估计展开的半参数模型及基于 GEV 和 GPD 的参数模型。对这两种模型的理论的系统介绍可归功于 P. Embrechts，（1997），R. D. Reiss，和 Thomas（2000），S. Coles，（2001）和 S. Kotz 和 S. Nadarajah，（2000）等。

在极值估计中，极值指数和高分位数的估计构成了极值估计的主要内容。在估计 GEV 和 GPD 模型的极值指数和高分位数时，不论我们采用参数方法还是半参数方法，首先是确定门限值，找出门限值以上的观察数据；或者，换句话说，就是对所观察到的样本值的顺序统计量进行有效分割，得到用于估计的观察数据，然后才能用参数和半参数方法估计极值指数和高分位数。但是，值得指出的是，如何确定门限值，对样本进行最优分割，一直是困扰极值工作者的一个难题。

门限值越大，可以分析的数据越少。这时，被分析的数据比较接近分布的极端，分析的偏差减少，但由于数据过少，估计的方差增加；反之，门限值过小，被分析的数据增加，分析的方差减少，但偏差却增加了。对这个问题的研究，统计工作者提出了许多方案。Embrechts（1997）建议使用模拟法，通过研究在不同门限值的情形下极值指数的形状来确定门限值的大小；D. J. Dupuis（1998）建议从参数的稳健性出发来确定门限值；但更多的作者，如 Beirlant（1996，2002），Danielsson（2001），A. Guillou（2001），A. Ferreira（2002b），G. Matthys 和 J. Beirlant（2003）等建议使用最小化某一均方误差或渐近二阶矩来获得门限值。在这一准则下，G. Matthys 和 J. Beirlant（2003），J. Beirlant，G. Dierckx and C. Starica（2002）对 GEV 模型建立了一种指数回归模型，得到了极值指数的较好的估计。A. Ferreira（2002b），J. Danielsson（2001）等利用自助法得到了 GEV 模型的门限值的渐近结果。S. Resnick（1998）利用光滑的矩估计对极值指数建模，得到了它的估计，并研究了它的极限性质。P. Groeneboom 等人（2003）利用核型估计对极值指数进行了建模和估计，比较了由 Hill 方法、矩估计方法、拟似然估计和核型估计得到的极值指数的性质。R. Huisman，K. Koedijk，和 C. Kool（2001），I. Aban 和 M. M. Meerschaert，（2004）利用最小平方法分别对 Hall 族和幂指数型的尾的极值指数进行了估计，得到了理想的结果。V. Brazauskas 和 G. Serfling（2003）对 Pareto 分布的极值指数构造了一种新的稳健估计，并提出了在实际中对 Pareto 分布的诊断，检验方法。

在估计极值指数后，往往还要进一步估计高分位数。前面已经说到，VaR 只是损失分布的一个极端分位点或者说高分位数点。它的估计值事实上就是通过描述损失分布的尾部而得到。L. De Haan 等人（1993）构造了高分位数估计量，并讨论了其大样本性质。Danielsson（1997a）引入 k 阶矩率估计量，借助于自助法，讨论了高分位数和超出概率的问题。Bermudez 等人（2001）利用贝叶斯方法对高分位数和尾概率同时进行了估计。Ferreira（2002a）同样利用自助法讨论了高分位数的逆问题，尾概率的估计问题，Ferreira（2002b）研究了利用矩估计方法求高分位数时的样本分割问题，他采用的方法是自助法。G. Matthys 和 J. Beirlant（2001）则通过建立指数回归模型得到高分位数估计量。McNeil（1998，1999），F. M. Longin（2000），R. L. Smith（2000），A. J. Mcneil（2000），T. G. Bali（2003）等分别对 GEV 和 GPD 模型进行了建模，通过极大似然估计对极值指数进行估计，然后估计高分位数，导出 VaR 的表达式，并利用西方几个著名股指进行实证分析。G. Ramazan 和 S. Faruk（2001）对怎样将极值理论中广义 Pareto 模型应用于风险管理和 VaR 计算进行了系统的介绍，利用 GPD 模型，GARCH（1，1），GARCH（1，1）-t。模型同时对 S& P500 股票指数建模并进行比较，说明了利用广义 Pareto 分布建模的可行性和优越性。当然，对于高分位数的估计还有一些其他方法，详细的介绍可参见 P. Embrechts（1997）。

对于二元极值模型，作为理论研究，已开始引起统计学家们的注意，McNeil（1999）结合 Copulas 理论，首先提出了

多元极值，Copulas，S. Kotz（2000）则对多元广义 Pareto 模型作了系统介绍。

对于极值模型的检验只有为数不多的文献，早期的成果 R. D. Reiss（2001）作了一个简单的回顾。D. Dietrch（2002）提出了对数据是否满足 GEV 分布的条件进行检验的方法。V. Choulakian（2001）则对 GPD 模型应用的条件提出了检验的理论依据。

目前国内对极值估计的理论研究主要是程士宏，潘家柱（1998，2000）等。程士宏，潘家柱利用弱收敛和正则变换等工具对已有的估计的渐进性质进行了研究。应用研究集中在广义 Pareto 模型及其在股票收益率中的应用，具体可参见田宏伟（2001），封建强（2002）等。但其研究主要集中在对广义 Pareto 模型及极值指数已有估计方法的介绍，然后利用该模型对股票收益率进行实证分析。但他们并没有对如何选取门限值，进一步提出新的极值指数估计进行研究，而这些也正是本书所希望回答的问题。

§1.2 本书研究的主要内容、方法和创新

1.2.1 本书研究的主要内容

本书一共分为九章，具体内容安排如下：

第一章阐述了本书研究的理论与现实意义；介绍了国内外主要研究成果，概述了本书研究内容、研究方法和创新之处。