

# 一种新的风险价值(VaR)计算方法及其应用研究

周孝华 张 燕

(重庆大学经济与工商管理学院)

**摘要:** 应用格列汶科定理对历史模拟法进行改进,以提高历史模拟法的计算准确性,并利用该方法来确定POT模型中的阈值,以此新方法计算金融市场的风险价值(VaR);然后运用上证指数进行了实证分析,得到了较好的结果。

**关键词:** 风险值;历史模拟法;极值理论;POT模型

**中图分类号:** F830 **文献标识码:** A **文章编号:** 1672-884X(2008)06-0819-05

## New Calculation Method and Application of Value-at-risk (VaR)

ZHOU Xiaohua ZHANG Yan

(Chongqing University, Chongqing, China)

**Abstract:** A new method of threshold selection was investigated. According to Glivenko Theory, historical simulation was improved to made the calculation result more accurate and determine the threshold of POT model for calculating new VaR of finance market. Then, conducts the empirical analysis was made using Shanghai index and a good result was obtained.

**Key words:** value-at-risk (VaR); historical simulation; extreme value theory (EVT); peak-over-threshold models (POT)

自从J. P. Morgen公司于20世纪90年代提出VaR(value at risk)方法以来,VaR已在国际上获得了广泛的应用和重视,并逐渐成为金融市场风险度量的主流方法。计算VaR的方法分为参数法和历史模拟法,二者各有优缺点。参数法要求在资产或市场因素服从正态分布的假设下确定置信区间后再通过解析公式计算VaR值,但实际情况是金融资产的时间序列数据的分布与正态分布相比存在明显的尖峰厚尾现象,这会导致对VaR的低估。与参数法相比,历史模拟法有3个明显优势:①本质上非常简单,多数时候无需任何参数估计,市场因素序列都是观测出来的;②完全体现了市场的实际分布,分布的各阶矩以及市场因素之间的相关性集合均可模拟出来;③非常易于向高层管理者作出说明<sup>[1]</sup>。但是,如果市场因素的未来分布与历史分布差别过大,测算结果就会非常不准,并且历史模拟法需要大量的观测值,要承担很大的计算量。

由于VaR方法是对正常市场交易情况下风险的度量,因此在极端市场情况下(收益分布

的尾部),用基于正态分布假设等经典方法去度量VaR的误差较大。这就要求在运用VaR方法时应该运用压力测试进行补充。压力测试是基于历史或潜在的市场震荡数据,考察在极端条件下,市场价格大的变化对资产组合的价值变化的“最坏情景”,用于设定风险价值的标准或风险约束,确定资产组合风险水平是否在风险承受能力之内,包括识别那些会产生致命损失的情景,评估这些情景的影响并为风险管理提供额外的保证<sup>[2]</sup>。近年来,许多研究风险度量的学者纷纷利用极值理论(EVT)进行VaR估计。该理论不研究序列的整体分布情况,只关心序列的极值分布情况,利用广义帕累托分布或者广义极值分布来逼近损失的尾部分布情况。EVT所采用的模型主要有2种,在统计理论上均具有一定的优越性:①区块极大值模型(block maxima models,简称BMM模型);②越槛高峰模型(peak-over-threshold models,简称POT模型)<sup>[3]</sup>。极值理论的主要优点在于:可以准确地描述分布尾部的分位数;具有解析的函数形式,计算简便;在极端条件下,用极值理论

方法得到的 VaR 估计值与经验分布非常接近;提供了超越样本的预测能力,比常用方法具有更大的优越性;有完备的数学理论支持<sup>[2]</sup>。它在金融风险管理中的良好应用前景已经得到越来越广泛的关注。

## 1 VaR 方法与 POT 模型

VaR 是一种被广泛接受的风险度量工具。2001 年的巴塞尔委员会指定 VaR 模型作为银行标准的风险度量工具。它可以定义为在一定的置信水平  $p$  下,某一资产或投资组合在未来特定时间内的最大损失,或者说是资产组合收益损失分布函数的分位数点。假设  $X$  代表某一金融资产的收益,其密度函数为  $f(x)$ ,则 VaR 可以表示如下:

$$VaR_p = -\inf\{x | f(X \leq x) > (1-p)\}. \quad (1)$$

当密度函数  $f(x)$  为连续函数时也可写作

$$VaR_p = -F^{-1}(1-p),$$

式中,  $F^{-1}$  为分位数函数,它被定义为损失分布  $F(x)$  的反函数。但是 VaR 模型只关心超过 VaR 值的频率,而不关心超过 VaR 值的损失分布情况,且该模型在处理损失符合非正态分布(如后尾现象)及投资组合发生改变时表现不稳定,会出现式(2)的情况:

$$VaR_p(X+Y) \geq VaR_p(X) + VaR_p(Y), \quad (2)$$

不满足 ARTZNER 提出的一致性风险度量模型的次可加性<sup>[4]</sup>,因此,VaR 方法需要其他方法进行补充,常用的就是进行压力测试。

压力测试常采用极值理论方法进行度量。极值理论是测量极端市场条件下风险损失的一种方法,它具有超越样本数据的估计能力,可以准确地描述分布尾部的分位数。它主要包括两类模型:BMM 模型和 POT 模型,常用的模型为 POT 模型<sup>[5]</sup>。POT 模型的特点是:通过事先设定一个阈值,把所有观测到的超过这一阈值的数据构成一个数据组,以该数据组作为建模的对象,只考虑尾部的近似表达,而不是对整个分布进行建模;对数据要求的数量比较少;可以进行单步预测,给出在未来一段小的时间内的 VaR 估计值;前提条件是超限发生的时间服从泊松分布,超限彼此相互独立,服从 GPD(generalized pareto distribution)分布,且超限与超限发生的时间相互独立<sup>[6]</sup>。样本独立同分布可以保证 POT 模型的前提条件成立。

假设序列  $\{z_i\}$  的分布函数为  $F(x)$ ,定义  $F_u(y)$  为随机变量  $Z$  超过阈值  $u$  的条件分布函数,  $F_u(y) = P(Z-u \leq y | Z > u)$ ,  $y \geq 0$ 。

根据条件概率公式可以得到

$$F_u(y) = \frac{F(u+y)-F(u)}{1-F(u)} = \frac{F(z)-F(u)}{1-F(u)}, \quad z \geq u \quad (3)$$

$$\Rightarrow F(z) = F_u(y)[1-F(u)] + F(u).$$

**定理 1<sup>[7]</sup>** 对于一大类分布  $F$  (几乎包括所有的常用分布) 条件超量分布函数  $F_u(y)$ , 存在一个  $G'_{\xi,\sigma}(y)$  使得:

$$F_u(y) \approx G'_{\xi,\sigma}(y) = \begin{cases} 1 - \left(1 + \frac{\xi}{\sigma} y\right)^{-1/\xi}, & \xi \neq 0; \\ 1 - e^{-y/\sigma}, & \xi = 0. \end{cases} \quad (4)$$

当  $\xi \geq 0$  时,  $y \in [0, \infty]$ ; 当  $\xi < 0$  时,  $y \in [0, -\frac{\sigma}{\xi}]$ 。分布函数  $G'_{\xi,\sigma}(y)$  被称作广义的 Pareto 分布。根据式(4)可以得到广义的 Pareto 分布的概率密度函数  $g'_{\xi,\sigma}(y)$ , 因此对于给定的一个样本  $\{z_1, z_2, \dots, z_n\}$ , 对数似然函数

$$L(\xi, \sigma | y) = \begin{cases} -n \ln \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \ln \left(1 + \frac{\xi}{\sigma} y_i\right), & \xi \neq 0, \\ -n \ln \sigma - \frac{1}{\sigma} \sum_{i=1}^n y_i, & \xi = 0, \end{cases} \quad (5)$$

从而可以估计参数  $\xi$  和  $\sigma$ 。然后需要确定阈值  $u$ , DANIELSSON 等<sup>[8]</sup>给出了对阈值的估计方法, 一般有 2 种:

(1) 由 Hill 统计量

$$H_{k,n} = \frac{1}{k} \sum_{i=1}^k \ln \left( \frac{X(i)}{X(k)} \right)$$

构造 Hill 图。Hill 图定义为点  $(k, H_{k,n})$  构成的曲线, 选取 Hill 图形中尾部指数的稳定区域的起始点的横坐标  $K$  所对应的数据  $X_k$  作为阈值  $u$ 。

(2) 根据样本的超限期望函数, 其定义为

$$e(u) = \frac{\sum_{i=k}^n (X_i - u)}{n - k - 1}, \quad k = \min\{i | X_i > u\},$$

超限期望图点  $(u, e(u))$  构成的曲线, 选取充分大的  $u$  作为阈值, 使得当  $x \geq u$  时  $e(x)$  为近似线性函数。当  $u$  确定以后, 利用  $\{z_i\}$  的观测值, 根据式(5)进行最大似然估计得到  $\xi$  和  $\sigma$  的估计值  $\hat{\xi}$  和  $\hat{\sigma}$ 。同时, 可得到  $\{z_i\}$  的观测值中比阈值  $u$  大的个数, 记为  $N_u$ , 根据式(1)用频率代替  $F(u)$  的值, 可以得到  $F(z)$  在  $x > u$  时的表达式:

$$F(z) = F_u(y)[1-F(u)] + F(u) = \begin{cases} \frac{N_u}{N} \left\{ 1 - \left[ 1 + \frac{\hat{\xi}}{\hat{\sigma}} (z - u) \right]^{-1/\hat{\xi}} \right\} + \left( 1 - \frac{N_u}{N} \right) = \\ \frac{N_u}{N} [1 - e^{-(z-u)/\hat{\sigma}}] + \left( 1 - \frac{N_u}{N} \right) = \\ \begin{cases} 1 - \frac{N_u}{N} \left[ 1 + \frac{\hat{\xi}}{\hat{\sigma}} (z - u) \right]^{-1/\hat{\xi}}, & \hat{\xi} \neq 0; \\ 1 - \frac{N_u}{N} e^{-(z-u)/\hat{\sigma}}, & \hat{\xi} = 0. \end{cases} \end{cases} \quad (6)$$

对于给定某个置信水平  $p$ , 可以由式(1)得到:

$$VaR_p = \begin{cases} u + \frac{\sigma}{\xi} \left[ \left( \frac{N}{N_u} p \right)^{-\xi} - 1 \right], & \xi \neq 0; \\ u - \sigma \ln \left( \frac{N}{N_u} p \right), & \xi = 0. \end{cases} \quad (7)$$

## 2 历史模拟法的改进

历史模拟法就是用收益的历史数据的分布来模拟未来风险值的方法。首先将收益的历史数据从小到大排列,然后按照给定的置信水平找到对应的分位点,把这一点的收益(损失)作为未来收益(损失)的风险值。MAHONEY 发现历史模拟法能够在所有置信水平下(可高达 99%)得到 VaR 的无偏估计<sup>[9]</sup>,但是历史模拟法也存在很多问题:需要的数据量较多,估计的结果完全依赖于历史数据集合的选取,另外历史数据区间长度的选择也很重要。

对于历史模拟法的这些缺点,可以在数据的选取方法和处理上加以改进。对于  $N$  个历史收益率数据,选择其中的  $M$  个,共有  $C_N^M$  种选法,记每次选出来的  $M$  个数据为  $\{a_{ij}, j=1, 2, \dots, M\}, i=1, 2, \dots, C_N^M$ ,记它的  $\alpha$  分位数为  $\{a_i^*, i=1, 2, \dots, C_N^M\}$ ,则可以得到  $C_N^M$  个  $\alpha$  分位数。根据对风险的不同敏感度,选择相应的分位数作为 VaR。它从某种程度上反映了收益率的真实分布中分位数的信息。

对于经验分布函数  $F_n(x)$ ,随着  $n$  的增大,  $F_n(x)$  越来越接近  $X$  的分布函数  $F(x)$ ,关于这一点,GLIVENKO 在 1953 年给出了理论上的论证<sup>[10]</sup>。

**定理 2** 若总体  $X$  的分布函数为  $F(x)$ ,经验分布函数为  $F_n(x)$ ,则对  $\forall x \in R$ ,有

$$P\left\{\lim_{n \rightarrow \infty} \sup_{-\infty < x < \infty} |F_n(x) - F(x)| = 0\right\} = 1,$$

也即  $F_n(x) \xrightarrow{a.e.} F(x)$ 。

该定理表明,  $F_n(x)$  以概率 1 一致收敛于  $F(x)$ ,即可以用  $F_n(x)$  来近似  $F(x)$ ,这也是利用样本来估计和判断总体的基本理论依据。因此可以用经验分布函数的分位数来逼近总体分布的  $\alpha$  分位数。由文献<sup>[11]</sup>有如下定理:

**定理 3** 设  $X_1, X_2, \dots, X_n$  是来自具有密度函数  $p(x)$  的总体的一个样本,对给定的  $p \in (0, 1)$ ,  $p(x)$  在总体  $p$  的分位数  $\xi_p$  处连续,且  $p(\xi_p) > 0$ ,又定义  $k$ ,使得  $k = np + o(\sqrt{n})$ ,则对样本的第  $k$  个次序统计量  $X_{(k)}$  有

$$\frac{\sqrt{n}(X_{(k)} - \xi_p)}{\sqrt{p(1-p)/p(\xi_p)}} \xrightarrow{L} N(0, 1).$$

此定理说明,经验分布的分位数就是以总体为均值呈正态分布的。

根据定理 2 和定理 3,可以推导出一种改进的历史模拟法。根据定理 2,对于选定的历史数据可以用其经验分布函数的  $\alpha$  分位数来逼近历史数据总体分布的  $\alpha$  分位数;再由定理 3 可知,选取不同历史数据得到的分位数组  $\{a_i^*, i=1, 2, \dots, C_N^M\}$  是以总体为均值呈正态分布的,因此可以用这些分位数的均值来近似总体的分位数值,从而得到 VaR 值。这样就对历史模拟法进行了改进,一定程度上避免了该方法的缺陷。

## 3 利用改进的历史模拟法结合 POT 模型计算 VaR

在 POT 模型中另一个重要的问题就是确定定理 3 的阈值。这个值非常重要,它是正确估计参数  $\xi$  和  $\sigma$  的前提。如果  $u$  取值过高,会导致超额数据量太少,使估计出来的参数方差很大;如果  $u$  取值过低,则不能保证超量分布的收敛性,会使估计产生大的偏差。根据极值理论来计算 VaR 处理的数据也是一些超过阈值的极端值。在极端条件下,用该方法得到的 VaR 的估计值与经验分布非常接近。根据定理 2 和定理 3,改进的历史模拟法是用经验分布函数  $\alpha$  的分位数逼近总体分布的  $\alpha$  分位数来得到 VaR,二者都是用经验分布进行拟合。而且历史模拟法处理的数据是收益率数据,根据该方法得到的分位数值实际上就是历史数据的正常值和极端值的一个临界点,根据阈值  $u$  的定义,可以选择改进的历史模拟法得到的 VaR 值作为阈值  $u$ 。

利用这个原理可以提出一种新的确定阈值  $u$  的方法来计算 VaR,即先利用改进的历史模拟法计算出一个在给定的置信水平  $p$  下的分位数值;然后把这个分位数值作为 POT 模型中的阈值;再根据 POT 模型计算出该置信水平下的 VaR 值。其计算过程如下:

**步骤 1** 设采用的数据个数为  $N$  ( $N > 100$ ),令  $M=100$ ,按时间先后顺序第 1 次取 1 到 100 个数据,第 2 次选取第 2 到 101 个数据, ..., 第  $N$  次选取第  $N$  个到第  $N+99$  个数据(选取方法有多种,本研究仅选用该方法,且以前的研究结果<sup>[3]</sup>表明选取 100 天的数据是比较适合 1% 到 5% 的置信水平的),以此类推,得到  $k$  组数据,然后给定置信水平  $p$ ,用历史模拟法分别计算各组数据的分位数,得到  $k$  个分位数值,根据定理 1 和定理 2,将选取的这些分位数的平均值记为  $b$ ,在这  $k$  个分位数值中找出与  $b$  最接近的分位数,把它记为  $u$ ,并以这个  $u$  作为 POT 模型中的阈值。

**步骤2** 以  $u$  为阈值对原始的  $N$  个收益率数据进行筛选得到观测值中比阈值  $u$  大的个数,记为  $N_u$ 。

**步骤3** 利用式(5)得到  $\xi$  和  $\sigma$  的估计值  $\hat{\xi}$  和  $\hat{\sigma}$ 。

**步骤4** 根据式(7)得到给定置信水平  $p$  下的 VaR 值。

这种新的阈值选取方法结合了历史模拟法和极值理论的优点,与一般的阈值选取方法有所不同。这种新的选取阈值的方法以格列汶科定理和历史模拟法的原理为理论依据,理论上是可行的,并且该方法选取阈值是以历史模拟法计算的分位数结果为基础,使得计算结果偏差较小,采用计算机编制程序,消减了使用 Hill 图和超限期望图来选取阈值的计算的繁杂性。

4 实证分析

采用上海证券交易所公布的日综合指数  $P$  为原始数据(数据来源:新浪网),样本空间选自1990年12月19日~2004年9月30日(仅选取到2004年的数据是为了使得数据的统计特征明显而便于分析)。样本容量为3391个。定义收益率为  $R_t = \ln P_t - \ln P_{t-1}$ ,取置信水平为1%。

用 SPSS 统计分析软件对收益序列  $R_t$  进行统计分析。图1和表1给出了收益序列  $R_t$  的描述性统计量。可以看到序列具有明显的尖峰厚尾现象,从J-B检验可以显著地拒绝正态性假设。对收益序列进行单位根 ADF 检验(见表1),因为检验的统计量是-19.645 16,比显著性水平为1%的临界值-3.435 4小得多,所以拒绝原假设,序列不存在单位根,是平稳序列。

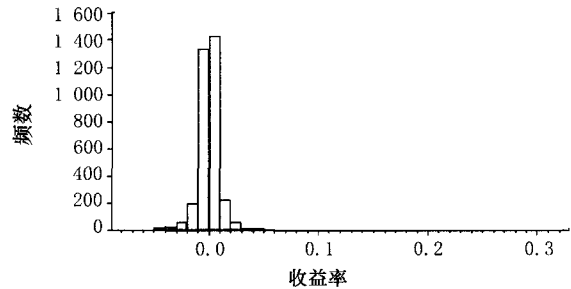


图1 收益序列  $R_t$  的描述性统计量

表1 序列的单位根检验

ADF Test Statistic	-19.645 16	1%	临界值*	-3.435 4
		5%	临界值	-2.862 9
		10%	临界值	-2.567 5

\* 单位根拒绝假设的 Mackinnon 临界值

对数据进一步做探索分析,考察数据的奇异性 and 分布特征。从盒形图(见图2)中可以看

到,收益率序列存在极端值情况,这与直方图(见图1)中显示的收益序列具有尖峰厚尾现象是一致的,因此,采用一般的 VaR 计算方法不能预测可能的最大损失,而必须运用极值理论来计算 VaR。

基于极值理论中的 POT 模型,需要利用充分大的阈值  $u$ ,对超限分布进行 GPD 拟合。首先采用改进的历史模拟法来求取阈值。

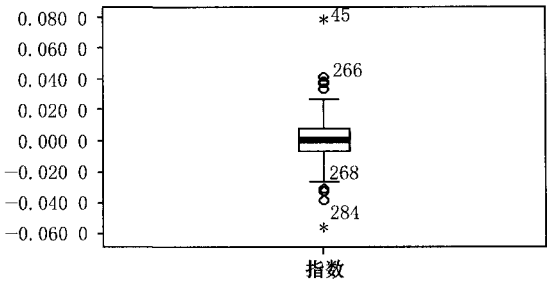


图2 收益序列  $R_t$  的盒形图

采用 Windows.net 编制程序,按照步骤1得到阈值  $u=1.897$ ,  $N_u=97$ 。由式(6)得到  $\hat{\xi}=8.326\ 14$ ,  $\hat{\sigma}=0.323\ 16$ 。对于  $\hat{\xi}$  的估计,EM-BRECHETS<sup>[11]</sup>认为金融序列的  $\alpha=1/\hat{\xi}$  的取值范围应当在3~4之间,而本研究的计算结果几乎不落在[3,4]的区域内,这主要是因为未考虑金融序列本身的分布,计算的是极端值,使得  $\hat{\xi}$  估计出来的值偏大,但这与 EMBRECHETS 的结论并不矛盾。再根据式(8)得到  $VaR_{0.01}=1.854$ 。阈值  $u=1.897$  是根据历史模拟法计算出的 VaR 值,  $VaR_{0.01}=1.854$  是历史模拟法和 POT 模型结合的方法计算所得,二者相差不大,这说明这种新方法是可行的。

再采用常用的阈值选取方法来计算 VaR,以便和新方法得到的 VaR 进行比对。这里只和采用超限期望图方法得到的结果进行对比。由图3可以发现样本的平均超限函数在  $u \geq 0.8$  时近似于一条直线,具有显著的 Pareto 分布特征。

当  $u=0.8$  时,收益率数据超过阈值  $u$  的个数  $N_u=369$ 。利用最大似然估计得到各参数的估计值为  $\hat{\xi}=0.229$ ,  $\hat{\sigma}=367$ ,  $VaR_{0.01}=1.967$ 。由此可见,采用改进的历史模拟法选取阈值的方法计算出的 VaR 值1.854和用超限期望图得到的 VaR 值1.967 差别不大,进一步说明了该方法可行,并且这种方法得到的 VaR 值更接近于用历史模拟法得到的 VaR 值1.897,故可以认为该阈值选取方法在一定程度上优于普通的阈值选取方法。

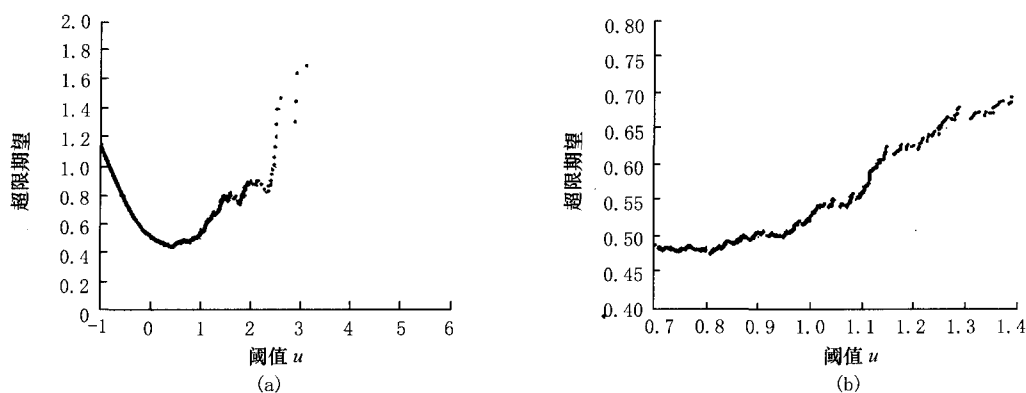


图3 超期望图

## 5 结语

历史模拟法是一种常见的非参数方法,但在极端值条件下计算 VaR 缺乏稳定性,而极值理论(EVT)方法是一种极端值条件下计算 VaR 的较好的参数方法,二者各有优缺点。本研究将历史模拟法加以改进并与 POT 法相结合来计算金融市场风险 VaR。这种推导出的新方法经过实证研究得到了较好的结果。这说明在利用 EVT 理论计算金融市场风险 VaR 时,阈值的选取非常重要,好的阈值选取方法能更加科学可靠地预测金融风险。当然,这种方法也存在一些瑕疵,如 POT 模型中要求数据是独立同分布的,而在实际中数据之间可能具有相关性;对 POT 模型中参数的估计方法有多种,不同的参数估计方法会影响结果的准确性等,这些正是本研究下一步所要做的工作。

## 参 考 文 献

- [1] 乔瑞 F. 风险价值 VaR[M]. 陈跃,译. 北京:中信出版社,2005.
- [2] 王春峰,万海晖,张维. 金融市场风险测量模型——VaR [J]. 系统工程学报,2000,15(1):746~751.
- [3] 郑文通. 金融风险管理的 VaR 方法及其应用[J]. 国际金融研究,1997(9):58~62.
- [4] 詹原瑞,田宏伟. 极值理论在汇率受险价值计算中的应用[J]. 系统工程学报,2000,1(15):44~53.
- [5] 田新时,郭海燕. 极值理论在风险度量中的应用——基于上证180 指数[J]. 运筹与管理,2004(1):64~69.
- [6] 欧阳资生,龚曙明. 广义帕累托分布模型:风险管理工具[J]. 财经理论与实践,2005(3):51~62.
- [7] 杨虎,刘琼荪,钟波. 数理统计[M]. 北京:高等教育出版社,2004.
- [8] DANIELSSON J, DE VRIES C G. Value at Risk and Extreme Returns[R]. London School of Economics, Financial Markets Group Discussion, 1997.

- [9] MANFRED G, KELLEZI E. An Application of Extreme Value Theory for Measuring Risk[R]. Department of Econometrics, University of Geneva and FAME, 2003.
- [10] LONGIN F M. From Value at Risk to Stress Testing: The Extreme Value Approach [J]. Journal of Banking and Finance, 2000(24):1 097~1 130.
- [11] EMBRECHETS P. Extreme Value Theory in Finance and Insurance [J]. Journal of Mathematics, 1999(13):33~60.

(编辑 张光辉)

通讯作者:周孝华(1965~),男,湖南武冈人。重庆大学(重庆市 400044)经济与工商管理学院教授、博士研究生导师。研究方向为金融工程、金融市场风险管理、财务与金融。

## “全球化背景下的中国产业竞争力 与可持续发展”国际研讨会

会议日期:

2009 年 4 月 16~17 日

会议地点:

上海市黄浦区

主办单位:

复旦大学管理学院产业经济学系

英国兰卡斯特大学管理学院中国管理研究中心

英国布鲁耐尔大学制度与经济发展研究中心

会务联系:孙霁

电子邮件:sunpei@fudan.edu.cn

信息来源: <http://www.fdns.fudan.edu.cn>