

# 中国证券市场股指收益分布的实证分析

黄德龙, 杨晓光

(中国科学院数学与系统科学研究院, 中国科学院管理决策与信息系统重点实验室, 北京 100080)

**摘要:** 利用 1996 年至 2004 年上证综合指数和深证综合指数数据, 对股指收益的分布特性进行了多角度的实证考察. 在正态性假设被拒绝以后, 利用国际上考察股票收益分布所使用的几个分布函数——scaled- $t$  分布、逻辑斯谛分布、指数幂分布、混合正态分布、ARCH-M 模型、GARCH-M 模型——对股指收益数据分别进行拟合, 对拟合出来的分布函数运用拟合优度检验, 并比较各种拟合分布下 VaR 值与历史模拟的差别. 实证结果表明 scaled- $t$  分布能够较好地模拟股指收益, 有助于投资者正确估计市场风险. 此外, 对正态分布、scaled- $t$  分布与历史数据落在不同区间的概率进行了比较, 以期能够判断用正态分布模拟股指收益对高收益和高损失的低估可能性的大小.

**关键词:** 股指收益; 分布函数; 拟合优度; VaR 偏差

**中图分类号:** F830.91 **文献标识码:** A **文章编号:** 1007-9807(2008)01-0068-10

## 0 引言

研究中国股市收益分布, 不仅有助于认识中国证券市场的内在运行规律, 采取正确的证券市场监管措施; 而且可以帮助市场参与者进行资产定价与资产组合, 正确进行风险度量, 更好地使用现代风险管理技术; 因此有着重要的实践意义.

西方的计量经济学家们对于证券资产收益率分布的研究由来已久. 早在上个世纪 50 年代, Kendall 和 Osborne 就通过对英美股市收益率的数据分析研究并得出结论: 股票资产的收益率近似服从正态分布. 这种观点符合统计学中的大样本思想, 再加上正态分布的性质容易处理, 从而广为研究人员和业界所接受. 比如 1973 年提出的 Black-Scholes 公式就是以对数收益率满足正态分布为基础建立起来的; 资本资产定价模型 (CAPM) 假定收益是关于时间独立同分布的, 其联合分布为多变量正态分布; 1994 年 JP Morgan 公司推出的 VaR 系统 RiskMetrics, 实质是假设有价证券的收益率服从正态分布.

股票收益的正态分布假设被如此广泛地应用着, 但反对它的声音却从未间断过. Alexander 对 Osborne 的数据重新进行了分析, 认为尖峰、厚尾是证券资产收益率的基本特征, 用正态分布来描述金融资产的短期收益率是不太合适的. Peters<sup>[1]</sup> 发现 1928 到 1989 年的 S&P500 股票收益呈现负偏、尖峰、厚尾的特征. 从理论的角度去看, 当信息没有及时为整个市场所知或者投资者没有对信息做出及时反应的时候, 正态分布假设确实是值得怀疑的. 如果信息按照一簇一簇传到市场, 而不是按照线性方式及时地到达, 就会导致信息的分布呈尖峰态, 股票收益的分布将因此受到影响. 另外, 假设信息能够流畅地传到市场, 但投资者并没有及时做出反应, 而是等到一些信息堆积起来, 才去消化这些信息, 也会导致股票收益分布的尖峰态. 对此, Peters 给出了理论推导.

近年来很多学者, 对这一问题作了进一步的研究, 尝试了用一些各不相同的分布来描述股票资产的对数收益率, 从而考虑到它的尖峰、厚尾、负偏特征.

Smith<sup>[2]</sup> 首先提出用逻辑斯谛分布来模拟股

收稿日期: 2004-08-23; 修订日期: 2007-11-28

基金项目: 国家自然科学基金资助项目 (70425004; 700221001; 70331001).

作者简介: 黄德龙 (1981—), 男, 安徽人, 博士生, Email: dellonhwang@yahoo.com.cn

票收益,这种分布近似于正态分布,不过比正态分布厚尾.其后,Gray和 French<sup>[3]</sup>、Peiró<sup>[4]</sup>对逻辑斯谛分布的拟合优劣性作了进一步的分析.

Hsu、Gray和 French曾经讨论过指数幂分布,这种分布具有尖峰和厚尾的特征,尾部以指数级的速率缩小,因而可以给股票收益分布给出一个不错的拟合.

Press<sup>[5]</sup>认为证券收益由一个连续的扩散(布朗运动)和一个间断的跳跃(泊松过程)组成:前者造成了证券价格的连续变化,后者反映了消息面带来的较大的震动. Kon为这种混合正态分布找到了实证的证据.

Praetz<sup>[6]</sup>、B lattberg和 Gondes、Gray和 French、Felipe和 Javier<sup>[7]</sup>认为 scaled- $t$ 分布比其它分布更好地拟合了股票收益.特别地, Praetz在假设证券收益波动性是一个时变的随机变量的条件下,从理论上推导出证券收益的分布满足 scaled- $t$ 分布.

Engle<sup>[8]</sup>提出了自回归条件异方差(ARCH)模型,将方差和条件方差区分开来,并让条件方差作为过去误差的函数而变化,从而为解决异方差问题提供了新的途径. Bollerslev<sup>[9]</sup>在此基础上提出了广义自回归条件异方差(GARCH)模型.为了刻画时间序列受自身方差影响的特征, Engle、Lilier和 Robins<sup>[10]</sup>提出了 GARCH-M模型.而当需要刻画证券市场中的非对称效应时, Nelson<sup>[11]</sup>提出的 EGARCH模型能更准确地描述金融产品价格波动的情况.

近年来人们对中国股票市场收益的分布特征也进行了一些相应的分析与研究. 闫冀楠、张维<sup>[12]</sup>分别用指数幂分布、ARCH模型、混合正态分布拟合了 1990年至 1996年上证综指收益的分布,结果显示这三种分布都比正态分布更具刻画力,且其中混合正态分布为最佳. 闫冀楠、张维<sup>[13]</sup>针对 ARCH-M传统估计方法的不足提出了遗传算法的改进,并实证估计了上海股市的各种 ARCH-M模型,确证了在以周为时间刻度下上海股市中投资收益与投资风险之间确实存在正相关关系. 陶亚民、蔡明超、杨朝军<sup>[14]</sup>分别运用柯氏检验法和异方差的  $t$ 检验

法对 1990年至 1998年上证综指收益率的分布特征进行了实证分析. 研究认为在排除异常事件干扰的情况下,收益率服从正态分布. 陈启欢<sup>[15]</sup>认为通过对 A股市场 1992年至 2000年不同股指研究发现:收益率从整体上完全不符合正态分布,而大体上符合自由度 5~9 的学生  $t$ 分布. 林美艳、薛宏刚、赵凤群<sup>[16]</sup>用 JB 检验法,得出了 1996年至 2001年上证综指收益率分布与正态分布有明显的偏差,用学生  $t$ 分布对日收益率进行拟合的效果比较好. 有关中国股市收益分布的其它研究,还可参见封建强<sup>[17]</sup>、张维、黄兴<sup>[18]</sup>、李亚静、朱宏泉<sup>[19]</sup>、马玉林、施红俊、陈伟忠<sup>[20]</sup>、封建强、王福新<sup>[21]</sup>等.

实证研究是一个随历史数据发展而发展的研究领域. 本文利用能够获得的最新数据——1996年 1月 2日至 2004年 4月 30日的日交易数据,对上证综指、深证综指的分布特性进行考察. 首先对数据的基本统计量作些分析,然后对正态分布假设进行了 D检验. 接下来利用国际上考察股票收益分布所使用的几个分布函数——scaled- $t$ 分布、逻辑斯谛分布、指数幂分布、混合正态分布、ARCH-M模型、GARCH-M模型——对股指收益数据分别进行拟合. 对拟合的分布函数进行拟合优度检验,并比较各种拟合分布下 VaR值与历史模拟的差别,发现 scaled- $t$ 分布能够较好地模拟股指收益,特别地在尾部 scaled- $t$ 分布拟合效果很好,能够更好地帮助投资者正确估计市场风险. 最后,对正态分布、scaled- $t$ 分布与历史数据落在不同区间的概率进行比较,希望能够对用正态分布模拟股指收益可能带来的偏差给一个量的判断.

## 1 数据简介和正态性检验

本文中采用的原始数据为上证综指(999999)和深证综指(399106)每日收盘指数,时间区间为 1996年 1月 2日到 2004年 4月 30日. 股指收益定义为  $R_t = 100 * \lg(I_t / I_{t-1})$ , 其中  $I_t$  为  $t$ 时刻股票指数.

本文重点考察的 scaled- $t$ 分布有别于通常的  $t$ 分布(student's  $t$  distribution, 又作学生  $t$ 分布)和非中心  $t$ 分布(noncentral  $t$  distribution), scaled- $t$ 分布的密度函数见本文第三部分. 国内的部分文献中混用了  $t$ 分布这个名词. 本文中不考虑周一效应等日历效应问题.

表 1 股指收益的基本统计结果

Table 1 Statistical results of stock index returns

股指	最小值	最大值	均值	标准差	方差	偏度	峰度
上证综指	- 10.437 6	9.400 8	0.054 1	1.737 6	3.019 4	- 0.161 8	9.227 9
深证综指	- 10.588 7	10.478 9	0.065 1	1.874 8	3.514 8	- 0.329 0	8.610 0

从表 1 第 6 列可以看出上证综指、深证综指均为负偏,显示收益分布曲线向负轴方向倾斜,深证综指的偏斜程度更大一些。

从第 7 列可以看到 A 股股指收益均有很明显的尖峰特征,从而比正态分布厚尾。这说明了存在一些大幅度偏离均值的异常值。

表 1 的数据为股指收益偏离正态分布提供了证据,此外可以通过正态性检验来得到验证。我们采用国家标准 GB4882 - 85《正态性检验》中提供的 D 检验方法,D 检验的适用范围是样本量在 50 以上。D 检验的数据结果如表 2。

表 2 股指收益的正态性检验 (D 检验)

Table 2 Normality test of stock index returns(D test)

股指	样本量 n	统计量 Y	分位数 $Y_{0.005}$	分位数 $Y_{0.995}$	结论
上证综指	200 9	- 44.84	- 2.91	2.25	拒绝正态性假设
深证综指	200 9	- 41.12			拒绝正态性假设

表 2 的结果表明,与国际及国内的实证经验一样,本文的检验结果也说明股指收益的正态分布假设不能成立。这促使着人们去寻找更合理的分布假设。

是离散化参数。如果股指收益  $R_t$  满足逻辑斯谛分布,则  $E(R_t) = \mu, Var(R_t) = \frac{\pi^2}{3}$ 。

3) 指数幂分布 (exponential power distribution) 指数幂分布的密度函数可以写成

$$f(x) = \frac{\exp\left[-\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|^{1+\frac{2}{\alpha}}\right]}{2^{\frac{3+\alpha}{2}}\left(\frac{3+\alpha}{2}\right)}$$

其中,  $\mu$  ( $-\infty < \mu < \infty$ ) 是位置参数,  $\alpha$  ( $\alpha > 0$ ) 是离散化参数,  $\alpha$  ( $-1 < \alpha < 1$ ) 是描述形状的参数。反映了分布的峰度,  $-1 < \alpha < 0$  意味着低峰态;  $\alpha = 0$  时即为正态分布;  $0 < \alpha < 1$  表示尖峰态。如果股指收益  $R_t$  满足指数幂分布,则

$$E(R_t) = \mu, Var(R_t) = \frac{[3(1+\alpha)/2]}{[1+\alpha/2]}$$

4) 混合正态分布 (mixtures of two normal distributions) 混合正态分布的密度函数为

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}, & \text{以概率 } p \\ \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}, & \text{以概率 } 1-p \end{cases}$$

其中,  $\mu_i$  ( $-\infty < \mu_i < \infty$ ) 是位置参数,  $\sigma_i$  ( $\sigma_i > 0$ ) 是离散化参数。上述分布意味着证券收益以的概率  $p$  服从均值  $\mu_1$ 、标准差  $\sigma_1$  的正态分布,以  $1-p$  的概率服从均值  $\mu_2$ 、标准差  $\sigma_2$  的正态分布。如果  $R_t$  服从上述分布,则

## 2 股指收益拟合的备择模型

为了分析究竟何种分布更确切地描述了股指收益,下面将考虑几种国际上常用的估计股指收益分布的模型,并对其拟合的效果进行分析。

1) scaled- $t$ 分布

scaled- $t$ 分布的密度函数可以写成

$$f(x) = \frac{\left(\frac{\nu+1}{2}\right)}{\left(\frac{\nu}{2}\right) \sqrt{(\nu-2)^2} \left[1 + \frac{(x-\mu)^2}{(\nu-2)^2}\right]^{\frac{\nu+1}{2}}}$$

其中,  $\Gamma(\cdot)$  表示伽玛函数,  $\mu$  ( $-\infty < \mu < \infty$ ) 是位置参数,  $\frac{\nu}{2}$  ( $\nu > 0$ ) 是离散化参数,  $\nu$  是自由度。如果股指收益  $R_t$  满足 scaled- $t$ 分布且  $\nu > 2$ ,则  $E(R_t) = \mu, Var(R_t) = \frac{3}{\nu-2}$ 。

2) 逻辑斯谛分布 (logistic distribution)

逻辑斯谛分布密度函数可以写成

$$f(x) = \frac{\exp\left(\frac{x-\mu}{\sigma}\right)}{\left[1 + \exp\left(\frac{x-\mu}{\sigma}\right)\right]^2}$$

其中,  $\mu$  ( $-\infty < \mu < \infty$ ) 是位置参数,  $\sigma$  ( $\sigma > 0$ )

$$E(R_t) = \mu = \mu_1 + (1 - \alpha)\mu_2,$$

$$\text{Var}(R_t) = \alpha [(\mu_1 - \mu)^2 + \sigma_1^2] + (1 - \alpha) \times [(\mu_2 - \mu)^2 + \sigma_2^2]$$

$$k = \frac{[\alpha(\mu_1 - \mu)^2 + 3\alpha(\mu_1 - \mu)\sigma_1^2] + (1 - \alpha)[(\mu_2 - \mu)^2 + 3(\mu_2 - \mu)\sigma_2^2]}{\{\alpha [(\mu_1 - \mu)^2 + \sigma_1^2] + (1 - \alpha) [(\mu_2 - \mu)^2 + \sigma_2^2]\}^{3/2}}$$

5) ARCH-M 模型、GARCH-M 模型

GARCH-M (GARCH-in-mean) 模型就是在 GARCH 的主模型上加上一项  $\alpha_1$ , 表明收益跟波动率存在一定的相关性. 采用的 GARCH-M 模型表达式为

$$y_t = \mu + \alpha_1 \sigma_t + e_t$$

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \sigma_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

其中,  $\sigma_t^2 = \text{var}(e_t | \mathcal{F}_{t-1})$ ,  $\mathcal{F}_{t-1}$  是时刻  $t-1$  及  $t-1$  之前的全部信息,  $e_t$  独立同分布, 且参数满足:  $E(e_t) = 0, D(e_t) = 1, E(e_t e_s) = 0 (t \neq s); \omega > 0, \alpha_i \geq 0, \beta_j \geq 0, \sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1. q = 0$  时的 GARCH-M 模型就是 ARCH-M 模型.

主模型旨在解释一项金融资产的回报率, 那么增加  $\alpha_1$  的原因是每个投资者都期望资产回报率是与风险度密切联系的, 而条件方差  $\sigma_t^2$  代表了期望风险的大小. 所以 GARCH-M 模型适合于描述那些期望回报与期望风险密切相关的金融资产.

### 3 参数估计和拟合优度检验

使用正态分布和第 2 节中列出的几种模型拟合股指收益, 得到数据结果如表 3 所示.

在考虑的正态分布和上述四种分布中, 混合正态分布是唯一考虑到股指收益的有偏性的. 上述混合正态分布的偏度系数为

表 3 各种分布假设下的参数估计

Table 3 Parameter estimation under different return distributions

分布	参数	上证综指	深证综指
正态	$\mu$	0.051034	0.077393
	$\sigma^2$	1.152806	1.267086
scaled-t	$\mu$	0.052446	0.079337
	$\nu$	2.615304	2.599203
逻辑斯谛	$\mu$	0.052686	0.077544
	$\sigma^2$	0.705374	0.776809
指数幂	$\mu$	0.038680	0.091485
	$\alpha$	0.750387	0.850864
	$\beta$	0.642133	0.611827
混合正态分布	$\mu_1$	0.056551	0.067396
	$\mu_2$	0.752921	0.870832
	$\sigma_1^2$	0.046864	0.137959
	$\sigma_2^2$	1.895692	2.217070
	$\alpha$	0.477964	0.529573
	$k$	-0.006969	0.049278
ARCH-M	$\mu$	-0.004969	-0.003185
	$\alpha_1$	0.339355	0.208738
	$\alpha_0$	0.000187	0.000183
	$\beta_1$	0.432670	0.558377
GARCH-M	$\mu$	-0.002676	-0.002703
	$\alpha_1$	0.209137	0.204547
	$\alpha_0$	1.30e-05	6.95e-06
	$\beta_1$	0.227159	0.150535
	$\beta_0$	0.756677	0.839246

使用了 ARCH 模型、GARCH 模型对上证和深证股指收益数据进行模拟计算, 但是所得到的 ARCH(1) 模型和 GARCH(1, 1) 模型不如最后得到的 ARCH(1)-M 模型、GARCH(1, 1)-M 模型. 为节省篇幅, 文中只介绍 ARCH-M 模型和 GARCH-M 以及它们的相关结果.

正态分布、逻辑斯谛分布、scaled-t 分布、指数幂分布、混合正态分布下的参数估计采用的是最小二乘法, 对于 ARCH-M、GARCH-M 的参数估计采用的是 Marquardt 优化算法. 我们也用最大似然方法对参数值进行了估计, 得出与下文类似的结论. 为节省篇幅, 在此不把估计结果具体列出.

ARCH-M 和 GARCH-M 模型的各参数均通过  $\chi^2$  检验, 用 Marquardt 优化方法估计出来的模型的检验指标 AIC、L、SC、F 统计量等指标均非常良好. 以上证综指为例, ARCH-M 和 GARCH-M 下的 AIC 值分别为 -5.4147、-5.5624, L 值分别为 5.443、5.592, 均优于闫冀楠、张维文中用遗传算法对周收益率估计出的模型. 同时也验证了闫冀楠、张维 (1999) 一文末尾提出的问题: 该文的结论对于深证股市、对于日收益率依旧成立.

从表 3, 至少可以看出以下三点: 第一, scaled- $t$ 分布拟合股指收益的自由度估计值较小 (2 335和 2 436). 自由度较大时 scaled- $t$ 分布趋同于正态分布, 而实证得到的较小自由度表明股指收益分布确实存在明显偏离正态, 尤其在尾部. 第二, 用指数幂分布拟合后参数  $\alpha$  的估计值均大于 0 (0 642和 0 612). 参数  $\alpha$  反映出峰度高低.

$\alpha = 0$ 时就是正态分布, 随着尾部变厚,  $\alpha$  上升 (上界为 1).  $\alpha$  的估计值较大, 也表明股指收益的分布有明显的厚尾特征. 第三, 在 ARCH-M 和 GARCH-M 模型下,  $\beta$  的估计值均为正且通过  $t$  检验, 显示资产回报率与风险度密切相关, 因此在 GARCH 模型基础上加上一项  $\beta$ , 是有必要的.

为了比较上述各种分布的优劣性, 下文将进行拟合优度检验. 将收益数据分为 20 组, 分属于区间,  $(- \infty, - 9), [- 9, - 8], \dots, [9, \infty)$ . 用  $v_i$  表示对数收益率数据落入上述第  $i$  个区间的个数, 用  $p_i$  表示满足拟合分布的变量落入第  $i$  个区间的概率. 考察的统计量为  $V = \sum_{i=1}^{20} (v_i - np_i)^2 / np_i$ , 根据概率和频率的关系, 如果拟合的比较好  $v_i/n$  和  $p_i$  应该差不多,  $V$  就是用来刻画  $v_i/n$  和  $p_i$  接近程度的量. 拟合优度检验的数据结果如

(1. 0e + 012) \* {2. 790 7    0 001 0    0 000 0    0 000 0    0 000 0    0 000 0  
 0 000 0            0 000 0    0 000 0    0 000 0    0 000 0    0 000 0  
 0 000 0            0 000 0    0 000 0    0 000 0    0 000 0    0 002 1    0 333 9}

可以看出  $V$  之所以很大, 主要大在两端. 中间的拟合效果好得多. 正态分布解决不了厚尾的问题, 拟合的优度便不会好到哪里去.

用 scaled- $t$ 来拟合上证综指和深证综指时, 拟

0 868 8    0 007 7    0 684 1    0 159 9    0 026 7    0 030 0    1 691 5  
 0 191 8    0 024 0    0 195 0    0 014 8    0 010 4    0 327 1    0 018 3  
 0 044 3    0 746 2    0 905 5    0 391 5    0 953 2    2 579 4

不管  $i$  从 1 到 20,  $(v_i - np_i)^2 / np_i$  都处于一个较低的位置, 不存在数值很大的异常点, 可见 scaled- $t$ 分布对于股指收益的各个区间的拟合都比较好. 注意到: 左尾部  $(- \infty, - 9)$  上的值为 0. 868 8, 右尾部  $[9, \infty)$  的值为 2 579 4, 表明拟合结果和实际分布接近, 较好地体现了股指收益分布尖峰厚尾的特征.

表 4:

表 4 拟合优度检验数据结果

Table 4 Results of goodness of fit test

分布	V(上证综指)	V(深证综指)
正态	3. 127 4e + 012	2. 434 6e + 010
scaled- $t$	9. 870 1	10. 294 9
逻辑斯谛	3. 785 7e + 003	1. 840 4e + 003
指数幂	1. 123 8e + 003	677. 073 0
混合正态分布	1. 549 6e + 004	1. 297 8e + 003
ARCH-M	2. 567 3e + 006	3. 146 0e + 006
GARCH-M	2. 272 8e + 006	3. 123 9e + 006

注: 统计量  $V$  均服从自由度为 19 的卡方分布, 置信水平为 0. 05 的临界值为 30. 1, 置信水平为 0. 01 的临界值为 36. 2

从表 4 可以看出, 正态分布提供的拟合效果最差, ARCH-M 模型次差, GARCH-M 模型略微好于 ARCH-M 模型. 此外, 混合正态分布、逻辑斯谛分布的拟合优度也很低, 指数幂分布相对好一点. 但不管是在 0. 01 的水平下还是在 0. 05 的水平下, 只有 scaled- $t$  分布没有被拒绝.

对于拟合优度的差异, 可以从更细致的数据中看出.  $V$  的每一项  $(v_i - np_i)^2 / np_i$  是用来刻画在该区间上  $v_i/n$  和  $p_i$  接近程度的量. 对于上证综指, 对应着正态分布拟合的  $(v_i - np_i)^2 / np_i$  分别为

合的效果很好, 在 0. 01 的水平上不能拒绝

scaled- $t$  分布. 对于上证综指, 对应着 scaled- $t$  分布

的  $(v_i - np_i)^2 / np_i$  分别为

#### 4 不同分布下拟合的 VaR 值

VaR (Value at Risk) 是对风险暴露的一种统计度量方法. VaR 定义为给定一定的时间区间和

这里 1. 0e + 012 \* {·} 表示括号内每一项都与 (1. 0e + 012) 相乘.



置信水平,测量预期最大损失(或最小收益)的方法.用公式表示如下:

$\text{prob}(P_t, \text{VaR}) = C$ , 其中  $P_t$  表示在  $t$  时间内可能的收益,  $C$  是置信水平.

如何根据历史数据计算 VaR, 是风险分析与管理中一个重要的基本问题. 目前有多种计算 VaR 的模型<sup>[22, 23]</sup>, 问题的关键在于如何由历史数据来拟合数据的真实分布. 在引言中提到的 JP Morgan 的风险度量模型 RiskMetrics, 其基本假定是收益率序列服从正态分布. 但从上文的实证分析可以看出, A 股收益率的分布并不是正态的, 而

是负偏、尖峰的, 具有厚尾性. 因而在正态假定下所计算的 VaR 值, 常常是低估实际风险, 这从下文中的数据也可看出. RiskMetrics 系统同时也允许选择和指定各种不同的收益率分布, 本文的研究目的之一就是要发现究竟什么样的收益率分布更适合我国的证券市场, 以便投资者在使用 RiskMetrics 等工具时可以正确选择分布.

对正态分布以及第 2 部分所给备择分布的拟合结果计算各自的 VaR 值, 然后与历史数据相对照, 可得数据结果如下:

表 5 不同分布对上证综指收益拟合结果的 VaR 值对照

Table 5 VaR of Shanghai composite index under different return distributions

置信水平	1%	5%	10%	95%
正态	- 2.630 7	- 1.845 2	- 1.426 3	1.947 2
	(0.530 0)	(0.247 4)	(0.154 6)	(- 0.275 9)
scaled- $t$	- 5.651	- 2.577 2	- 1.703 8	2.682 1
	(- 0.009 7)	(- 0.051 1)	(- 0.009 8)	(- 0.002 7)
逻辑斯谛	- 3.188	- 2.029 4	- 1.497 2	2.129 6
	(0.430 4)	(0.172 3)	(0.112 6)	(- 0.208 1)
指数幂	- 3.682	- 2.288 1	- 1.647	2.365 4
	(0.342 1)	(0.066 8)	(0.023 8)	(- 0.120 4)
混合正态	- 3.880 1	- 2.433 7	- 1.675 7	2.527 9
	(0.306 7)	(0.007 4)	(0.006 8)	(- 0.060 0)
ARCH-M	- 3.823 9	- 2.682 4	- 2.074 8	2.824 6
	(0.316 8)	(- 0.094 1)	(- 0.229 7)	(0.050 3)
GARCH-M	- 3.732 4	- 2.616 7	- 2.022 8	2.765 8
	(0.333 1)	(- 0.067 3)	(- 0.198 9)	(0.028 4)
历史数据	- 5.596 7	- 2.451 8	- 1.687 2	2.689 3

注:分布函数中的参数值均来自于表 3 对上证综指收益分布的拟合结果. 括号内数据表示该分布下的 VaR 值估计偏差与历史数据之比(以下简称偏差比), 偏差比为正意味着该分布以及该置信水平下对 VaR 值的估计大于历史数据, 该值为负意味着对 VaR 值的估计小于历史数据.

表 5 显示在同一置信水平下(1%, 5% 或 10%), 正态分布拟合的左尾部 VaR 的值均被大幅高估, 即投资者会大大低估最大损失; 换一个角度说, 在同一 VaR 值下, 置信水平将被大幅低估, 即投资者将会低估较大损失的可能性. 对于逻辑斯谛分布和指数幂分布拟合亦出现同样的现象.

对于混合正态分布拟合, 在 5% 和 10% 水平下 VaR 值与历史数据的偏差均小于 1%, 显示混

合正态分布拟合在这两个置信水平下较为精确地刻画了较大损失; 但是在 1% 水平下偏差达到 30.67%, 这表明混合正态分布对收益分布的厚尾现象依旧难以刻画.

对于 ARCH-M 模型拟合和 GARCH-M 模型拟合, 在 1% 水平下 VaR 的值被大幅高估, 这与正态分布、逻辑斯谛分布和指数幂分布拟合下的结果类似, 但是在 5% 和 10% 水平下 VaR 的值却

被低估.

scaled-*t*分布拟合的左尾部 VaR 值与历史数据的偏差比很小 (分别为 0.97%, 5.11%, 0.98%), scaled-*t*分布将帮助投资者正确估计各置信水平下的较大损失,由此正确评估市场风险.

同样计算置信水平为 95% 下的 VaR 值,可以发现在正态分布拟合的右尾部 VaR 的值被低估,即投资者会大大低估最大收益;换一个角度说,在同一 VaR 值下,置信水平将被大幅高估,即投资者将会低估较大收益的可能性.在逻辑斯谛分布、指数幂分布、混合正态分布下,这种对最大收益可能性的低估依次递减,但是偏差比的绝对值均高于 5%.相反的是,在 ARCH-M 模型和 GARCH-M 模型下,对最大收益可能性却有了一定幅度的低估,偏差比分别为 5.03%、2.84%. scaled-*t*分布下计算出来的 VaR 值与历史数据几乎相同 (偏差仅为 0.27%),可见 scaled-*t*分布将帮助投资者正确

评估较大收益的可能性.

总体来说,从数据结果可以看出,在正态假定下所计算的 VaR 值,会造成投资者低估实际风险;相对于正态分布、逻辑斯谛分布、指数幂分布、混合正态分布、ARCH-M 模型以及 GARCH-M 模型来说,按 scaled-*t*分布假定下所计算的 VaR 值和历史数据有着最好的契合,因而有助于帮助投资者正确评估市场风险.

### 5 正态分布和 scaled-*t*分布与历史数据的偏差对比

在这一部分,将尝试量化用正态分布来模拟股指收益而带来的偏差.在表 3 给出的拟合分布结果下,对照所给正态分布 (或 scaled-*t*分布) 变量落在对应区间上的概率,数据结果如表 6

表 6 特定收益区间上概率的对比:正态分布与 scaled-*t*分布

Table 6 Probability in specific intervals: Normal VS scaled-*t*

区间	上证综指			深证综指		
	正态	scaled- <i>t</i>	历史数据	正态	scaled- <i>t</i>	历史数据
$[A, A + S]$	0.433 4	0.397 9	0.398 2	0.433 1	0.399 7	0.397 2
$[A + S, A + 2S]$	0.064 2	0.072 6	0.074 7	0.069 2	0.076 0	0.084 1
$[A + 2S, A + 3S]$	0.001 3	0.016 7	0.014 4	0.001 6	0.017 1	0.015 9
$[A + 3S, A + 4S]$	3.024 5e - 6	0.005 8	0.006 0	4.728 6e - 06	0.005 8	0.006 0
$[A + 4S, A + 5S]$	8.100 9e - 10	0.002 6	0.002 5	1.723 4e - 09	0.002 5	0.003 0
$[A + 5S, A + 6S]$	2.363 8e - 14	0.001 3	0.001 5	7.432 2e - 14	0.001 3	0.000 5

注:每一个数值表示所给正态分布 (或 scaled-*t*分布) 变量落在对应区间上的概率.正态分布和 scaled-*t*分布都是对称分布,在相反方向上可以作类似分析. A、S 分别为股指收益数据的均值和标准差 (见表 2).

从表 6 可以看到:按正态分布模拟的股指收益落在  $[A, A + S]$  中的概率较大,在接下来的五个区间里,按正态分布模拟的股指收益落在其中的概率均小于 scaled-*t*分布,且差距越来越大;按正态分布模拟的股指收益落在各区间中的概率偏离历史数据的程度显著大于 scaled-*t*分布的偏离程度.

比如对于上证综合指数,在偏离均值一个和两个标准差的区间,按 scaled-*t*分布模拟的收益落在该区间的概率比历史数据落在该区间的概率低 2.8%,而正态分布比历史数据低 14.1%;在偏离均值两个和三个标准差的区间, scaled-*t*分布比历史数据高 16.0%,而正态分布比历史数据低 91.0%;在偏离均值三个和四个标准差的区间,

scaled-*t*分布比历史数据低 3.3%、而正态分布比历史数据低 99.9%.

如果再做进一步计算的话,可以算出在  $[A + 5S, A + 6S]$  以后的所有区间里,按 scaled-*t*分布模拟的股指收益的偏离度均远远小于正态分布.

上述的结果显示,如果投资者认为 A 股股指收益服从正态分布,他将会明显低估市场风险.在分布的尾部这种低估尤其严重,也就说按照正态分布,投资者将大大低估高损失和高收益.正态分布下的收益落在偏离均值三个标准差以外的地方的概率几乎可以忽略不计的,但是在 scaled-*t*分布下,收益落在偏离均值三个标准差以外地方的机会要频繁得多,尤其在偏离均值较远的尾部,

scaled- $t$ 分布和实际数据是契合得最好的. 下面给出了 scaled- $t$ 分布拟合曲线、正态分布拟合曲线和历史数据的对照图. 上述的结论, 在图中有很好的显示.

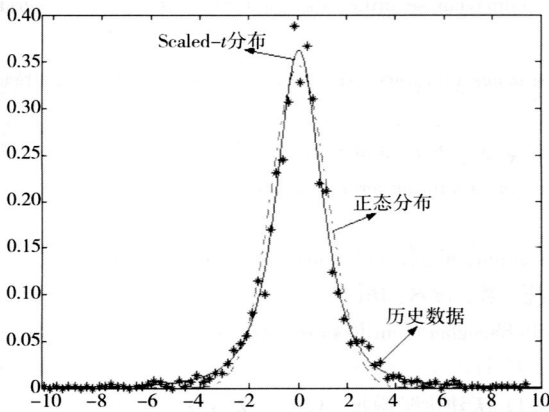


图 1 拟合曲线与历史数据的对照图

Fig 1 Comparison between fitted curves and historical data

注: 图 1 是用 scaled- $t$ 分布、正态分布拟合上证综指的效果对照, 图 2 是图 1 的尾部放大, 图 3 是图 1 的顶部放大, 散点 (·) 为历史数据, 实线为按照 scaled- $t$ 分布拟合的分布图, 虚线 (- -) 为按照正态分布拟合的分布图

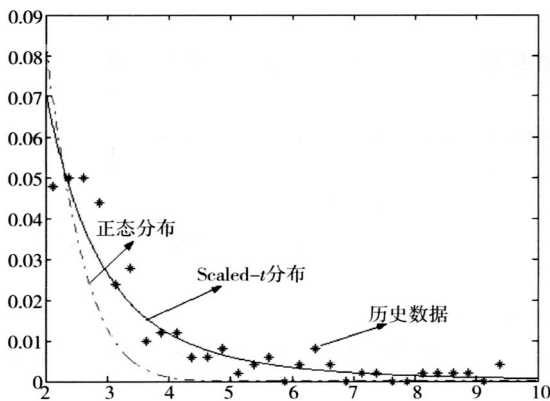


图 2 拟合曲线与历史数据的尾部对照图

Fig 2 Comparison of tails between fitted curves and historical data

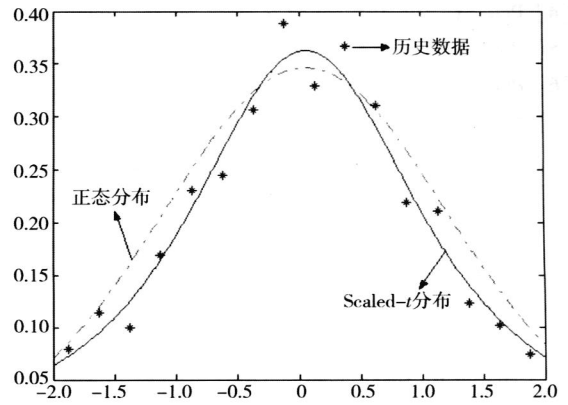


图 3 拟合曲线与历史数据的顶部对照图

Fig 3 Comparison of peaks between fitted curves and historical data

## 6 结 论

本文从实证分析出发, 多角度考察了上证综指和深证综指收益的分布特性: 股指收益数据基本统计量的正态性检验很明确的拒绝了正态性假设; 而使用 scaled- $t$ 分布、逻辑斯谛分布、指数幂分布、混合正态分布、ARCH-M 模型、GARCH-M 模型对股指收益数据分别拟合以后, 发现 scaled- $t$ 分布和历史数据的分布最为接近, 而且 scaled- $t$ 分布拟合下不同置信水平的 VaR 值也和历史数据有着最好的契合.

股票价格是投资者行为的体现, 而投资者行为一直处在变化之中, 股票价格的表现模式也会发生着变化. 因此对股票收益分布的实证研究, 是一个没有止境的过程. 对新的历史数据、不同行业的数据、不同时段的数据进行实证分析, 是一件很有意义的工作.

## 参 考 文 献:

[1] Peters E. Chaos and order in the capital markets: A new view of cycles, prices and market volatility[J]. New York: John Wiley and Sons, 1991.

[2] Smith J. The Probability Distribution of Market Returns: A Logistic Hypothesis[D]. Salt Lake City: University of Utah, 1981.

[3] Gray B, French D. Empirical comparisons of distributional models for stock index returns[J]. Journal of Business, Finance & Accounting, 1990, 17: 451—459.

为使图形不显得杂乱, 未画出按指数幂分布、逻辑斯谛分布、ARCH-M 模型以及 GARCH-M 模型拟合的收益曲线.



- [4] Peir ÓA. The distribution of stock returns: International evidence[J]. Applied Financial Economics, 1994, 4: 431—439.
- [5] Press J. A compound events model for security prices[J]. Journal of Business, 1967, 40: 317—335.
- [6] Praetz P. The distribution of share price changes[J]. Journal of Business, 1972, 45: 49—55.
- [7] Felipe A, Javier E. Empirical distributions of stock returns: Scandinavian securities markets 1990—1995 [R]. Madrid: Carbs III University Working Paper, 1997.
- [8] Engle R F. Autoregressive conditional heteroscedasticity with estimates of United Kingdom inflation [J]. Econometrica, 1982, 50: 987—1007.
- [9] Bollerslev T. Generalized autoregressive conditional heteroskedasticity [J]. Journal of Econometrics, 1986, 31: 307—327.
- [10] Engle R F, David M L, Robins R P. Estimating time varying risk premia in the term structure: The ARCH-M model [J]. Econometrica, 1987, 55: 391—407.
- [11] Nelson B. Conditional heteroscedasticity in asset returns: A new approach [J]. Econometrica, 1991, 59: 347—370.
- [12] 闫冀楠, 张 维. 关于上海股市收益分布的实证研究 [J]. 系统工程, 1998, 16(1): 21—25.  
Yan Jinan, Zhang Wei. Empirical study on distribution of return in Shanghai securities exchange [J]. System Engineering, 1998, 16(1): 21—25. (in Chinese)
- [13] 闫冀楠, 张 维. 股市投资收益与风险直接关系的定量研究 [J]. 天津大学学报, 1999, 32(4): 454—458  
Yan Jinan, Zhang Wei. Empirical study on the relation of stock returns and risk [J]. Journal of Tianjin University, 1999, 32(4): 454—458. (in Chinese)
- [14] 陶亚民, 蔡明超, 杨朝军. 上海股票市场收益率分布特征的研究 [J]. 预测, 1999, 2: 57—58  
Tao Yam in, Cai Mingchao, Yang Chaojun. The Study of return distribution of Shanghai stock market [J]. Forecasting, 1999, 2: 57—58. (in Chinese)
- [15] 陈启欢. 中国股票市场收益率分布曲线的实证 [J]. 数理统计与管理, 2002, 9: 9—11.  
Chen Qi huan. The curve of stock market yield in China [J]. Application of Statistics and Managenent, 2002, 9: 9—11. (in Chinese)
- [16] 林美艳, 薛宏刚, 赵凤群. 上海证券市场收益率的正态性检验 [J]. 纺织高校基础科学学报, 2003, 16(3): 246—248  
Lin Meiyan, Xue Honggang, Zhao Fengqun. The normal test of the return ratio in Shanghai stock market [J]. Basic Sciences Journal of Textile Universities, 2003, 16(3): 246—248. (in Chinese)
- [17] 封建强. 上海证券市场收益率分布的对称性研究 [J]. 统计研究, 2001, 7: 29—33.  
Feng Jianqiang. Symmetry of yield distribution of Shanghai stock market [J]. Statistical Research, 2001, 7: 29—33. (in Chinese)
- [18] 张 维, 黄 兴. 沪深股市的 R/S 实证分析 [J]. 系统工程, 2001, 19(1): 1—5.  
Zhang Wei, Huang Xin. Empirical study on the Shanghai stock exchange and Shenzhen stock exchange with R/S [J]. System Engineering, 2001, 19(1): 1—5. (in Chinese)
- [19] 李亚静, 朱宏泉. 沪深股市收益率分布的时变性 [J]. 数学的实践与认识, 2002, 32(2): 228—233.  
Li Yajing, Zhu Hongquan. The time-varying property of the return s of Shanghai and Shenzhen stock markets [J]. Mathematics in Practice and Theory, 2002, 32(2): 228—233.
- [20] 马玉林, 施红俊, 陈伟忠. 沪深股市周收益率分布的实证研究 [J]. 经济数学, 2003, 20(4): 52—57.  
Ma Yulin, Shi Hongjun, Chen Weizhong. An empirical study on return distribution of portfolio in China security market [J]. Mathematics in Economics, 2003, 20(4): 52—57. (in Chinese)
- [21] 封建强, 王福新. 中国股市收益率分布函数研究 [J]. 中国管理科学, 2003, 11(1): 14—21.  
Feng Jianqiang, Wang Fuxin. A research on return distribution function of Chinese stock market [J]. Chinese Journal of Management Science, 2003, 11(1): 14—21. (in Chinese)
- [22] Benninga S, Wiener Z. Value-at-risk [J]. Mathematica in Education and Research, 1998, 4: 1—8
- [23] 王春峰, 万海辉, 李 刚. 基于 MCMC 的金融风险 VaR 的估计 [J]. 管理科学学报, 2000, 3(2): 54—61.  
Wang Chunfeng, Wan Haihui, LI Gang. Estimation of value-at-risk using MCMC [J]. Journal of Management Sciences, 2000, 3(2): 54—61. (in Chinese)

## Empirical study on distributions of stock index returns in China 's securities market

HUANG De-long, YANG Xiao-guang

Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Laboratory of Management Decision and Information System, Chinese Academy of Sciences, Beijing 100080, China

**Abstract:** Using daily data from 1996 to 2004 of Shanghai and Shenzhen composite indices, we investigate the empirical distributions of China 's stock index returns. First we find that the hypothesis of Normal Distribution is refused. Then we further test scaled- $t$  distribution, Logistic distribution, exponential power distribution, mixtures of two normal distributions, ARCH-M model, GARCH-M model from various view, e.g. goodness-of-fit test, VaR deviation, and probability deviation over different intervals. We find that scaled- $t$  Distribution is the best to describe China 's stock index returns.

**Key words:** stock index returns; distribution functions; goodness-of-fit test; VaR deviation

---

## 2008年商业智能和金融工程国际会议通知

由长沙理工大学、中国系统工程学会以及中国决策科学共同主办的“2008年商业智能与金融工程国际会议”将于今年10月28—30日在湖南长沙召开，中国科学院汪寿阳教授与长沙理工大学王耀中教授担任大会主席。本次会议竭力为从事商业智能和金融工程研究领域的专家、业界专业人员以及企业高级管理人员提供一个国际性的学术交流平台。

会议论文集将由法国 Atlantis Press 出版（论文有永久的 DOI 号，由 EI 和 ISTP 检索）。优秀论文将发表在 Journal of Applied Computational Intelligence 和 International Journal of Computational Sciences 的 Special issue 上。

有关会议具体情况，请浏览 <http://www.bife2008.net>