

Generalized Skew t Distribution of Log-Return Rate in Stock Market

Xin Yang

Department of Mathematics, Guilin University of Aerospace Technology, Guilin
Email: xinyang_emily@sina.cn

Received: Sep. 12th, 2014; revised: Oct. 11th, 2014; accepted: Oct. 20th, 2014

Copyright © 2014 by author and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The main aim of this paper is to study the distribution characteristics of log-return rate in stock market. The closing data of a year trading day of the 6 indexes (the Shanghai composite index, Shenzhen stock index, industrial index, real estate index, consumer services index, and food and beverage index), are done with the empirical analysis by using the statistical methods. Results show that the log-return rates of the stock indexes do not obey the normal distribution, with the characteristics of high peak, heavy tail and skew distribution. And results also show that the log-return rates are with high probability to be accepted as obeying the generalized skew t distribution. So the generalized skewed t distribution is a reasonable distribution to research the log-return rate of stock market.

Keywords

Stock Index, Log-Return Rate, Generalized Skewed t Distribution

证券市场对数收益率的 广义偏斜 t 分布

杨 昕

桂林航天工业学院, 数理部, 桂林
Email: xinyang_emily@sina.cn

收稿日期: 2014年9月12日; 修回日期: 2014年10月11日; 录用日期: 2014年10月20日

摘要

本文的主要问题是研究证券市场中对数收益率的分布特征。对上证指数、深证成指、工业指数、地产指数、消费服务和食品饮料等6个指数一个年度的交易日的收盘数据，利用统计检验方法进行实证分析，分析结果表明：证券指数的对数收益率不服从正态分布，具有尖峰、厚尾、偏斜等特征，它们均以较大的概率被接受为服从广义偏斜 t 分布，所以广义偏斜 t 分布是研究证券市场对数收益率的合理分布。

关键词

证券指数，对数收益率，广义偏斜 t 分布

1. 引言

对数收益率是研究证券市场中一个重要变量，用它可以研究证券投资组合和相应的投资风险管理。例如，马柯维茨的均值-方差投资组合理论，威廉·夏普(William F Sharpe)的资本资产定价模型 CAPM，都是建立在证券对数收益率的基础上。又如，在证券风险管理中利用证券投资组合的对数收益率的风险在值 VaR (Value at Risk)，条件风险在值 CVaR(Conditional Value at Risk)，都是用来刻画投资风险。

为了这些研究和相应理论的需要，人们通常假设对数收益率服从正态分布(Duffie and Pan [1]1997)。然而，有许多实际数据说明证券市场的对数收益率不服从正态分布(Mandelbrot [2] 1963, Fama [3]1965, Blattberg and Gonedes [4]1974, Cont [5]2001)，其分布通常具有尖峰、厚尾、偏斜和非对称等特点。因此，人们假设或检验对数收益率服从其它非正态分布，并以此研究证券市场的各种问题。可选择的分布有： t 分布(Seneta [6]2004, Lin et al. [7]2006, Ferguson and Platen [8]2006)，幂律分布(Wu [9]2006, Klass [10]2007, Plerou and Stanley [11] 2008)，Laplace 分布(刘建元和刘琼荪[12]2007，宋丽娟和杨虎[13]2008，陈晔等人[14]2010)，等等。

大家知道， t 分布是尖峰厚尾对称分布，幂律分布是厚尾非对称分布，Laplace 分布的厚尾性是处在正态分布与 t 分布之间的尖峰厚尾对称分布。我们注意到偏斜 t 分布同时具有尖峰、厚尾、偏斜和非对称等特点，因此本文重点研究偏斜 t 分布，在 Azzalini [15](1985)、Nadarajah and Kotz [16](2007)和 Nadarajah [17](2009)等人对标准偏斜 t 分布定义的基础上，我们考虑一种广义偏斜 t 分布(见文中定义 2)，对上证指数、深证成指、工业指数、地产指数、消费服务、食品饮料等 6 个指数的实际数据，利用极大似然方法估计分布参数，检验这些指数数据的分布，研究证实这 6 个指数的对数收益率都服从广义偏斜 t 分布。

2. 广义偏斜 t 分布的定义

Azzalini [15](1985)首先提出偏斜正态分布(skew-normal distribution)的概念，其理论依据是如下引理。

引理 1 设 $f(x)$ 是关于原点对称的密度函数， $G(x)$ 是绝对连续分布函数，其导数 $G'(x)$ 关于原点对称，则对于任意的实数 λ ，函数

$$p_{\lambda}(x) = 2f(x)G(\lambda x), \quad -\infty < x < \infty \quad (1)$$

是一个密度函数。

利用引理 1 可以定义各种偏斜分布(Azzalini [15]1985, Nadarajah 和 Kotz [16]2007, Nadarajah [17]2009)，如偏斜正态分布、偏斜 t 分布、偏斜柯西分布、偏斜 Logistic 分布、偏斜 Laplace 分布，等等。其中偏斜 t 分布定义如下：

定义 1 设 $f(x)$ 是自由度为 m 的 t 分布的密度函数, 即

$$f(x) = \frac{\Gamma((m+1)/2)}{\sqrt{m\pi}\Gamma(m/2)} \left(1 + \frac{x^2}{m}\right)^{-(m+1)/2}, \quad (2)$$

$G(x)$ 是 $f(x)$ 相应的分布函数, 则由(1)式确定的密度函数的分布称为偏斜 t 分布(skew t distribution), 其自由度为 m , 偏斜参数为 λ , 并记为 $ST(m, \lambda)$ 。

Nadarajah 和 Kotz [16](2007)研究了偏斜 t 分布 $ST(m, \lambda)$ 的密度函数的表示, 以及矩的计算公式。

在偏斜 t 分布 $ST(m, \lambda)$ 中, 自由度参数 m 是反映分布的厚尾特征, 自由度 m 愈小分布的尾部愈厚, 自由度 m 愈大分布的尾部愈轻; 参数 λ 是反映分布的偏斜方向和偏斜程度, 当 $\lambda > 0$ 时分布重心偏右, 当 $\lambda < 0$ 时分布重心偏左, λ 的绝对值愈大偏斜程度愈大。显然, 这个分布还缺少两个参数, 一个是位置参数 μ , 另一个尺度参数 σ , 这会影响它在实际应用中的适应性。因此, 本文考虑使用如下广义的偏斜 t 分布。

定义 2 设随机变量 X 服从定义 1 所定义的偏斜 t 分布 $ST(m, \lambda)$, 令 $Y = \mu + \sigma X$, 称随机变量 Y 服从广义偏斜 t 分布, 记为 $GST(\mu, \sigma, m, \lambda)$, 其中位置参数 $-\infty < \mu < \infty$, 尺度参数 $\sigma > 0$ 。

广义偏斜 t 分布 $GST(\mu, \sigma, m, \lambda)$ 的密度函数为

$$g(x; \mu, \sigma, m, \lambda) = \frac{2}{\sigma} f\left(\frac{x - \mu}{\sigma}\right) G\left(\frac{\lambda(x - \mu)}{\sigma}\right), \quad -\infty < x < \infty \quad (3)$$

其中 $f(x)$ 和 $G(x)$ 分别是自由度为 m 的 t 分布 $t(m)$ 的密度函数和分布函数。显然, 广义偏斜 t 分布具有如下基本性质。

性质 1 $GST(0, 1, m, 0) = t(m)$ 。

性质 2 $GST(0, 1, m, \lambda) = ST(m, \lambda)$ 。

3. 广义偏斜 t 分布的参数估计

广义偏斜 t 分布 $GST(\mu, \sigma, m, \lambda)$ 有四个参数, 可以利用极大似然估计方法估计, 其似然函数为

$$L(\mu, \sigma, m, \lambda) = \frac{2^n}{\sigma^n} \prod_{i=1}^n f\left(\frac{x_i - \mu}{\sigma}\right) G\left(\frac{\lambda(x_i - \mu)}{\sigma}\right), \quad (4)$$

其中 x_1, x_2, \dots, x_n 是样本数据。相应的对数似然函数为

$$\ln L(\mu, \sigma, m, \lambda) = n \ln 2 - n \ln \sigma + \sum_{i=1}^n \ln f\left(\frac{x_i - \mu}{\sigma}\right) + \sum_{i=1}^n \ln G\left(\frac{\lambda(x_i - \mu)}{\sigma}\right). \quad (5)$$

通过极大化对数似然函数(5)式可以计算出参数的估计值, 在具体计算中我们是利用 **R** 软件中的非线性极小化 **nlm** 函数(Non-Linear Minimization)。注意对数似然函数极大化等价于对负的对数似然函数极小化, 所以这个函数 **nlm** 可以对极小化和极大化进行求解。

4. 证券指数对数收益率的分布检验

4.1. 证券指数的样本数据

本文选择上证指数、深证成指、工业指数、地产指数、消费服务、食品饮料等 6 个指数作为研究对象, 对这些指数收集从 2012 年 7 月至 2013 年 6 月完整一年时间共 239 个交易日的收盘价数据。设 S_t 表示某证券指数在第 t 日的收盘指数, 则 $R_t = \ln(S_t/S_{t-1})$ 就是该指数在第 t 日的对数收益率, 下面将讨论这

些对数收益率的分布特征。

4.2. 对数收益率的正态性检验

正态性检验有多种方法，这里考虑利用偏度(Skewness)检验、峰度(Kurtosis)检验和夏皮洛-威尔克(Shapiro-Wilk)检验，计算结果见表 1。从偏度检验看，深证成指、工业指数和地产指数等 3 个指数的偏度检验 P 值都小于 0.05，说明有显著的偏度；而上证指数、消费服务和食品饮料等 3 个指数的偏度检验 P 值都大于 0.05，说明没有显著的偏度。从峰度检验看，6 个指数的峰度检验 P 值都远远小于 0.01，注意表 1 中的峰度值(kurtosis)是减了 3 的，所以这些指数的对数收益率的分布的峰值显著高于正态分布。图 1 是正态分布拟合曲线图，它更直观地看出这些指数的对数收益率的直方图分布的峰值显著高于正态分布。另外，从 Shapiro-Wilk 正态性检验看，所有指数的检验 P 值也都远远小于 0.01，这说明它们与正态分布有非常显著的差异。这些统计检验事实表明：这 6 个证券指数的对数收益率在研究时间段内不服从正态分布。

Table 1. Skewness, kurtosis and Shapiro-Wilk normality test
表 1. 偏度检验、峰度检验与 Shapiro-Wilk 正态性检验

指标名称	偏度检验		峰度检验		Shapiro-Wilk 检验	
	Skewness	P 值	Kurtosis	P 值	统计量	P 值
上证指数	-0.177797	6.570e-02	2.957307	3.113e-21	0.9564757	1.333e-06
深证成指	-0.223255	3.992e-02	2.578106	1.179e-16	0.9667588	2.357e-05
工业指数	-0.287442	1.756e-02	1.893313	6.222e-10	0.9769529	6.425e-04
地产指数	-0.681269	4.452e-06	3.375882	5.351e-27	0.9507866	3.178e-07
消费服务	0.030143	2.124e-01	1.681642	2.965e-08	0.9779888	9.273e-04
食品饮料	-0.082256	1.511e-01	1.029707	2.961e-04	0.9819077	3.936e-03

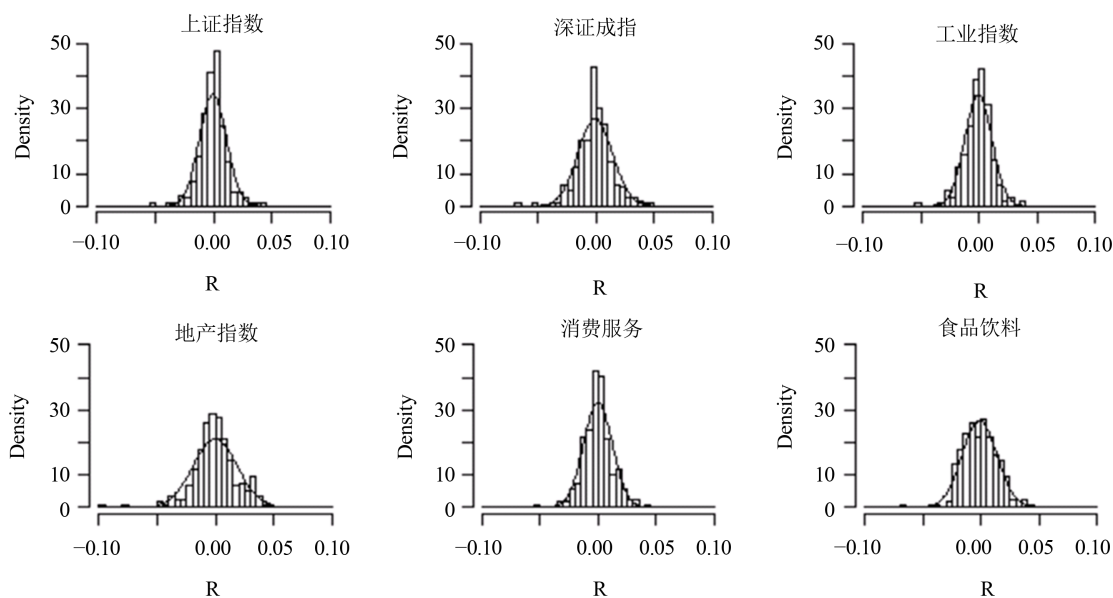


Figure 1. Fitting curve of normal distribution
图 1. 正态分布拟合曲线

4.3. 对数收益率的广义偏斜 t 分布检验

从前面的检验结果和图 1 知，证券指数的对数收益率具有尖峰、偏斜的特点，所以我们考虑检验它们是否与广义偏斜 t 分布有较好的拟合，为此做广义偏斜 t 分布的 K-S 检验。在做检验前，首先要利用极大似然方法估计广义偏斜 t 分布的四个参数，然后再计算相应的 K-S 检验，具体结果见表 2。

从广义偏斜 t 分布的 K-S 检验的结果(表 2)看，我们有如下结论：

1) P 值：所有指数的 K-S 检验的 P 值都非常大，前面 5 个指数(上证指数、深证成指、工业指数、地产指数和消费服务)的 P 值都大于 0.8，最后一个指数食品饮料的 P 值也接近 0.7，这远远大于 0.05 的概率水平，所以这些指数的对数收益率的分布与广义偏斜 t 分布没有显著差异，并以较大的概率水平可以接受它们是服从广义偏斜 t 分布。

2) μ 值：位置参数 μ 的值都比较小，几乎为零，说明对数收益率是以 0 为中心左右波动。

3) σ 值：注意到只有当 $\mu=0$ 和 $\sigma=1$ 时，广义偏斜 t 分布才等于标准偏斜 t 分布，即 $GST(\mu, \sigma, m, \lambda) = ST(m, \lambda)$ 。由于表 2 中尺度参数 σ 的值远远小于 1，所以说明只能用广义偏斜 t 分布 $GST(\mu, \sigma, m, \lambda)$ ，而不能用标准偏斜 t 分布 $ST(m, \lambda)$ 来拟合这些指数的分布。

4) m 值：前面 5 个指数(上证指数、深证成指、工业指数、地产指数和消费服务)的自由度都在 4 左右，自由度比较小，说明它们确实具有比较厚尾的特征。最后一个指数食品饮料的自由度为 13.266127，这个自由度稍微大一些，说明该指数的尾部分布与正态分布没有显著差异，并且其峰度值 1.029707(见表 1)也是这 6 个指数中最小一个，所以该指数的分布稍微接近正态分布。

5) λ 值：上证指数、工业指数和地产指数的偏斜参数是负值，它们的分布向左偏斜，向左偏斜最大的是工业指数的分布。深证成指、消费服务和食品饮料的偏斜参数是正值，它们的分布向右偏斜，其中消费服务和食品饮料的分布向右偏斜较大。

我们将相应的广义偏斜 t 分布的密度展示在图 2 中，图中直观显示了广义偏斜 t 分布比正态曲线更好拟合数据的直方图分布。

我们注意到：除了偏斜 t 分布外，常见的偏斜分布还有偏斜正态分布、偏斜 Logistic 分布、偏斜 Laplace 分布、Pareto 分布、指数分布等。关于偏斜正态分布、偏斜 Logistic 分布和偏斜 Laplace 分布，这三个分布的两边尾部都是以不低于指数衰减速度($e^{-|x|}$)趋于 0，均属于轻尾分布，所以这三个分布不太适合于反映对数收益率的尖峰厚尾特征。Pareto 分布和指数分布都是属于下方受限分布(即分布的左边尾部是被截尾的分布)，这些与对数收益率的两边尾部无限是不相符，也不太适合描述对数收益率的分布。

广义偏斜 t 分布具有尖峰、厚尾、偏斜等特征，其尾部是以幂率速度衰减，并且能通过自由度 m 的

Table 2. K-S test for generalized skewed t distribution

表 2. 广义偏斜 t 分布的 K-S 检验

指标名称	位置参数	尺度参数	自由度	偏斜参数	K-S 检验	
	μ	σ	m	λ	统计量	P 值
上证指数	-0.000299	0.007784	3.267387	-0.015349	0.038770	0.866777
深证成指	-0.001187	0.010835	4.007825	0.017963	0.035389	0.926820
工业指数	0.001361	0.009149	4.862127	-0.182230	0.040275	0.834841
地产指数	-0.000199	0.013700	3.980028	-0.003006	0.038125	0.879560
消费服务	-0.002040	0.009664	4.692233	0.158127	0.034746	0.936142
食品饮料	-0.002874	0.013792	13.266127	0.133190	0.046230	0.689154

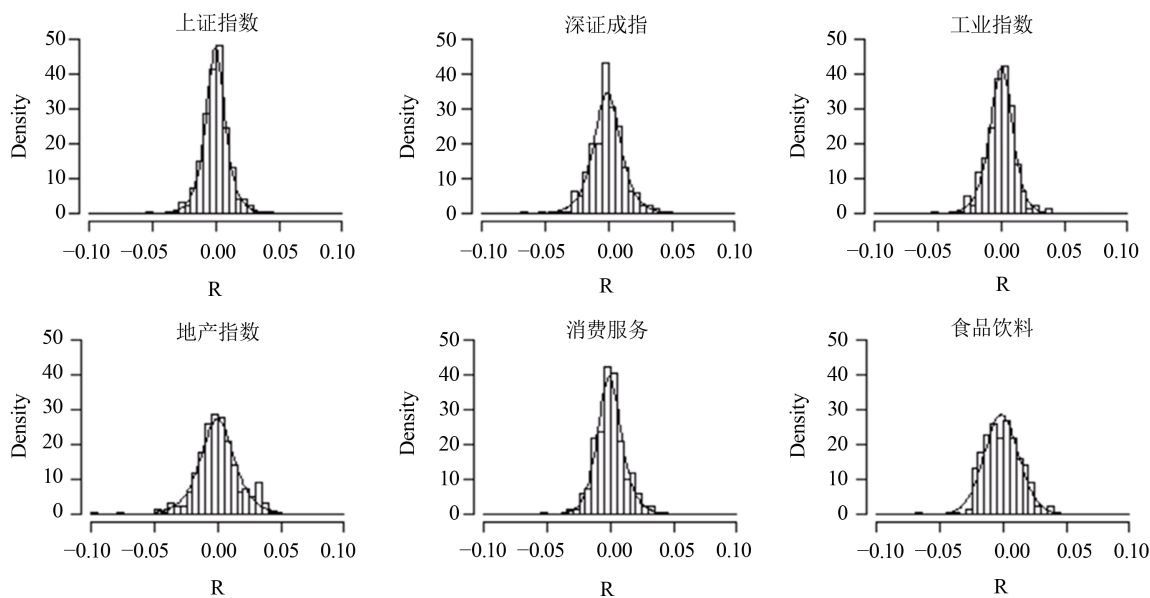


Figure 2. Fitting curve of generalized skewed t distribution
图 2. 广义偏斜 t 分布拟合曲线

大小调节尾部的厚薄程度。另外，通过广义偏斜 t 分布的参数 μ, σ, λ 可以调节分布的位置、尺度、偏度，从而使其适应更广泛的分布。在实际中，不同的指数可能有不同的分布，广义偏斜 t 分布能更好地适应这种变化。

综合上述讨论，我们认为广义偏斜 t 分布是拟合这些指数对数收益率分布的合理分布。

5. 结论

本文对上证指数、深证成指、工业指数、地产指数、消费服务和食品饮料等 6 个指数进行实证分析，利用偏度(Skewness)检验、峰度(Kurtosis)检验和夏皮洛-威尔克(Shapiro-Wilk)检验等方法进行正态性检验，结果表明：证券指数的对数收益率不服从正态分布，具有尖峰、厚尾、偏斜等特征。基于这些特点，我们选择广义偏斜 t 分布 $GST(\mu, \sigma, m, \lambda)$ 进行拟合对数收益率的分布，通过极大似然估计法和 K-S 检验法对证券指数对数收益率分布进行拟合与检验，结果表明：这些指数的对数收益率都服从广义偏斜 t 分布，并且接受的概率都很大，这说明广义偏斜 t 分布是研究证券市场对数收益率的合理分布。

致 谢

非常感谢审稿人和编辑提出修改意见，这些修改意见使本文的内容更加完整和充实。

基金项目

国家自然科学基金项目(11461009)，广西自然科学基金项目(2011GXNSFA018133)。

参考文献 (References)

- [1] Duffie, D. and Pan, J. (1997) An overview of value-at-risk. *The Journal of Derivatives*, **7**, 7-49.
- [2] Mandelbrot, B. (1963) The variation of certain speculative pricings. *Journal of Business*, **36**, 394-419.
- [3] Fama, E.F. (1965) The behavior of stock market prices. *Journal of Business*, **38**, 34-105.
- [4] Blattberg, R. and Gonedes, N. (1974) A comparison of stable and student distributions as statistical models for stock prices. *Journal of Business*, **47**, 244-280.

- [5] Cont, R. (2001) Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, **1**, 223-236.
- [6] Seneta, E. (2004) Fitting the variance-gamma model to financial data. *Journal of Applied Probability*, **1A**, 177-187.
- [7] Lin, S.K., Wang, R.H. and Fuh, C.D. (2006) Risk management for linear and non-linear assets: A Bootstrap method with importance resampling to evaluate Value-at-Risk. *Asia-Pacific Finan Markets*, **3**, 261-295.
- [8] Fergusson, K. and Platen, E. (2006) On the distributional characterization of daily log returns of a world stock index. *Applied Mathematical Finance*, **13**, 19-38.
- [9] Wu, L. (2006) Dampened power law: Reconciling the tail behavior of financial security returns. *The Journal of Business*, **79**, 1445-1473.
- [10] Klass1, O.S., Biham, O., Levy, M., Malcai, O. and Solomon1, S. (2007) The Forbes 400, the Pareto power-law and efficient markets. *The European Physical Journal B*, **55**, 143-147.
- [11] Plerou, V. and Stanley, H.E. (2008) Stock return distributions: tests of scaling and universality from three distinct stock markets. *Physical Review E*, **77**, 037101.1-037101.4.
- [12] 刘建元, 刘琼菀 (2007) 基于非对称 Laplace 分布研究 VaR. *统计与决策*, **18**, 33-35.
- [13] 宋丽娟, 杨虎 (2008) 基于 APARCH-Laplace 模型的 VaR 和 CVaR 方法. *统计与决策*, **14**, 16-19.
- [14] 陈晔, 晏小兵, 王跃恒 (2010) 股票投资组合的实证研究. *数学理论与应用*, **1**, 110-115.
- [15] Azzalini, A. (1985) A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171-178.
- [16] Nadarajah, S. and Kotz, S. (2007) Skew models I. *Acta Applicandae Mathematicae*, **98**, 1-28.
- [17] Nadarajah, S. (2009) The skew logistic distribution. *AStA—Advances in Statistical Analysis*, **93**, 187-203.