

重尾指数估计中阈值 k 的简便优化估计

刘维奇^{1,2}, 邢红卫²

(1. 山西大学 管理科学与工程研究所, 太原 030006; 2. 山西大学 数学科学学院, 太原 030006)

摘要 重尾分布在风险管理和金融领域有极为重要的应用价值. 提出了一种选取重尾指数估计阈值 k 的解析方法, 称为简便优化估计. 并且利用 Hill 估计和已知重尾分布进行 Monte-Carlo 模拟, 发现简便优化方法与 Sum-plot 方法和 Bootstrap 方法具有同样的精确性与稳健性, 三种方法都能得到令人满意的结果. 简便优化方法在精确性方面似乎略占优势, 计算方法简单, 且不受分布类型和重尾指数范围的影响, 在分布为 Pareto 型时基于简便优化方法的 Hill 估计依然具有相合性.

关键词 Hill 估计; Sum-plot 方法; Bootstrap 方法; 简便优化方法; 相合性

Simple and optimized choice of the threshold for heavy-tailed index estimation

LIU Wei-qi^{1,2}, XING Hong-wei²

(1. Institute of Management Science and Engineering, Shanxi University, Taiyuan 030006, China;
2. School of Mathematical Science, Shanxi University, Taiyuan 030006, China)

Abstract Heavy-tailed distribution models have revealed to be quite useful in fields such as risk management and finance. We provide an analytical method of choosing threshold for heavy-tailed index estimation, named Simple and Optimized Estimator(SOE). Then we proceed to an intensive computer simulation that enables us to find, through Monte-Carlo techniques, the SOE method is accurate and robust as well as Sum-plot method and Bootstrap method, the three methods can derive satisfactory results. The SOE method seems to have a slight advantage in terms of accuracy, the calculation method is simple, but not be affected by the types of the heavy-tailed distributions and the scope of the index, the Hill estimator based on the SOE method is still consistent whenever the distributions are Pareto-type models.

Keywords Hill estimator; Sum-plot method; Bootstrap method; simple and optimized method; consistency

1 引言

重尾分布在风险管理和金融等领域有极为重要的应用价值. 在极值理论中 F 是重尾分布, 如果 $\bar{F} := 1 - F \in RV_{-\alpha}$, 即

$$\bar{F}(x) := 1 - F(x) = P[X_1 > x] = x^{-\alpha}L(x) \quad (1)$$

定义分位数函数

$$b(t) = \left(\frac{1}{1-F}\right)^{\leftarrow}(t) = F^{\leftarrow}\left(1 - \frac{1}{t}\right) \quad (2)$$

1975 年 Hill^[1] 提出了基于条件极大似然估计的重尾指数 α 的估计方法,

$$\hat{H}_{n,k}^{-1} = \hat{\alpha}_{n,k}^{-1} = \hat{\gamma} = \frac{1}{k} \sum_{i=1}^k \log X_n^{(i)} - \log X_n^{(k+1)} \quad (3)$$

收稿日期: 2009-05-26

资助项目: 教育部人文社会科学研究项目 (07JA630027, 06JA630035); 山西省高校人文社科重点研究基地项目 (20083006)

作者简介: 刘维奇 (1963-), 男, 教授, 博士生导师, 主要从事金融工程和时间序列等领域的研究, E-mail: liuwq@sxu.edu.cn.

www.cnki.net

则以下三条等价:

- i) $\bar{F} \in RV_{-\alpha}, \alpha > 0$.
- ii) 存在序列 b_n 且 $b_n \rightarrow \infty$,

$$\lim_{n \rightarrow \infty} n\bar{F}(b_n x) = x^{-\alpha}, \quad x > 0.$$

- iii) 存在序列 b_n 且 $b_n \rightarrow \infty$, 则在 $M_+(0, \infty]$,

$$\mu_n(\cdot) := nP\left[\frac{X_1}{b_n} \in \cdot\right] \xrightarrow{v} \nu_\alpha(\cdot) \tag{8}$$

其中 $\nu_\alpha(x, \infty] = x^{-\alpha}, x > 0, \alpha > 0$.

引理 2^{[11]79-80} 假设 $\{X_i > 0, i \geq 1\}$ 是分布 F 的 iid, 非负随机变量, F 满足 (1),

$$\frac{n}{k}P\left[\frac{X_1}{b(n/k)} \in \cdot\right] \xrightarrow{v} \nu_\alpha(\cdot) \tag{9}$$

当 $n \rightarrow \infty, k = k(n) \rightarrow \infty$ 且 $n/k \rightarrow \infty$. 则在 $M_+(0, \infty]$,

$$\nu_n \Rightarrow \nu_\alpha \tag{10}$$

其中 $\nu_\alpha(x, \infty] = x^{-\alpha}, x > 0, \alpha > 0$. 然而由于 $b(\cdot)$ 未知, $b(n/k)$ 可以由其相合估计 $\hat{b}(n/k)$ 代替,

$$\hat{\nu}_n := \frac{1}{k} \sum_{i=1}^n \epsilon_{X_i/\hat{b}(n/k)} \tag{11}$$

引理 3^{[11]81-85} 如果 (10) 成立, 当 $n \rightarrow \infty, k \rightarrow \infty$ 且 $n/k \rightarrow \infty$, 则

$$\hat{H}_{n,k}^{-1} \xrightarrow{P} \frac{1}{\alpha} \tag{12}$$

定理 1 如果 F 满足 (1), 且用于 Hill 估计的次序统计量个数由简便优化方法选取, 即 \hat{k} 由 (6) 得到. 则 $n \rightarrow \infty$,

$$\hat{H}_{n,\hat{k}}^{-1} \xrightarrow{P} \frac{1}{\alpha} \tag{13}$$

证明 由比较优化法的定义, 式 (4), (5), 和 (6) 可以分别改写为

$$k(n, X_n^{(k)}) = \sum_{i=1}^n 1_{[X_n^{(k)}, \infty)}(X_n^{(i)}) = k \tag{14}$$

$$\max_k Z = \frac{k}{n} - \int_{X_n^{(k)}}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \tag{15}$$

$$\hat{k} = \arg \max_k \left(\frac{k}{n} - \int_{X_n^{(k)}}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \right) \tag{16}$$

第一步, 证明当 $n \rightarrow \infty, \hat{k} \rightarrow \infty$ 且 $n/\hat{k} \rightarrow \infty$.

很明显, $\frac{k}{n} > 0, \int_{X_n^{(k)}}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx > 0$. 所以为了最大化 (15), 即最大化 $\frac{k}{n}$ 且同时最小化 $\int_{X_n^{(k)}}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$, 当 $n \rightarrow \infty, k \rightarrow \infty$ 且 $X_n^{(k)} \rightarrow \infty$ 即 $\hat{k} \rightarrow \infty$ 且 $n/\hat{k} \rightarrow \infty$.

第二步, 如果 F 满足 (1), 即 $\bar{F} \in RV_{-\alpha}, \alpha > 0$, 由分位数函数 $b(n/k)$ 定义及定理 1 即证

$$\frac{n}{\hat{k}}P\left[\frac{X_1}{b(n/\hat{k})} \in \cdot\right] \xrightarrow{v} \nu_\alpha(\cdot),$$

当 $n \rightarrow \infty, \hat{k} \rightarrow \infty$ 且 $n/\hat{k} \rightarrow \infty$, (10) 成立. 由定理 3, (13) 成立.

3 Monte-Carlo 模拟

为了研究简便优化方法的稳健性, 以已知的重尾分布和 Hill 估计对其进行 Monte-Carlo 模拟. 选取 Student- $t(\alpha)$ 分布, Fréchet(α) 分布, Weibull($\lambda, 0.5$) 分布 (当 $0 < \lambda < 1$, Weibull 分布是重尾分布, 假设 $\lambda = \alpha/4$), 逆 Ga($\alpha, 1$) 分布, Burr(a, b) 分布 ($a = 2, b = 2$, 重尾指数 $\alpha = ab$) 和 Pareto($1, \alpha$) 分布, $\alpha(0 < \alpha \leq 4)$ 以 0.2 为间隔 (见表 1).

利用 Student- t 分布, Fréchet 分布, Weibull 分布, 逆 Ga 分布, Burr 分布和 Pareto 分布, 取重尾指数 α 的范围从 0 至 4, 通过 Monte-Carlo 模拟, 发现简便优化方法似乎不受分布类型和重尾指数范围影响, 具有比较好的稳健性.

表 1 简便优化方法的稳健性模拟结果

α	Student- $t(\alpha)$	Fréchet(α)	Weibull($\lambda, 0.5$)	IGa($\alpha, 1$)	Burr(2, b)	Pareto(1, α)
0.2	0.2128	0.2298	0.3747	0.2363	0.2987	0.2128
0.4	0.4130	0.3963	0.6378	0.4003	0.4675	0.2536
0.6	0.6519	0.5582	0.8489	0.5575	0.6119	0.4265
0.8	0.9056	0.7586	1.2812	0.7306	0.7686	0.5900
1.0	0.9937	0.8956	1.3571	0.9148	0.9045	0.2950
1.2	1.1760	1.1215	1.4773	1.1590	1.0919	1.0245
1.4	1.4147	1.2989	1.7851	1.3086	1.1850	1.1259
1.6	1.6044	1.4888	2.0939	1.5168	1.3885	1.4450
1.8	1.7989	1.7574	2.0927	1.7602	1.6067	1.7909
2.0	2.0009	1.9461	2.1636	1.9942	1.7181	2.1213
2.2	2.1840	2.1908	2.3025	2.1100	1.9566	2.4394
2.4	2.3465	2.3920	2.5296	2.2689	2.1562	2.1791
2.6	2.5500	2.6383	2.7728	2.4971	2.3295	2.5806
2.8	2.7022	2.8101	2.9841	2.6907	2.5198	3.2468
3.0	2.9209	3.0568	3.2518	2.9113	2.8565	3.0172
3.2	3.1718	3.1869	3.3604	3.1093	3.0293	3.3263
3.4	3.3420	3.4598	3.2717	3.3637	3.1373	3.4006
3.6	3.6112	3.6709	3.5017	3.5297	3.2492	3.5841
3.8	3.8398	3.8620	3.7992	3.7390	3.5150	3.4751
4.0	3.9355	4.0453	3.9323	3.9119	3.7944	3.8197

以 Sum-plot 方法和 Bootstrap 方法为参照, 对简便优化方法的精确性进行研究. 分别从 $t(4)$ 分布, Cauchy 分布, Fréchet(3) 分布, 逆 Ga(2,1) 分布, Burr(2,2) 分布和 Pareto(1,1.5) 分布抽取样本容量为 8000 的样本, 采用 Sum-plot 方法, Bootstrap 方法和简便优化方法选取 k 代入 Hill 估计 (见表 2). Sum-plot 图仅在开始的区域内呈线性, 一次作图效果不是很清晰, 二次作图非常必要. 为了更准确地确定图形的线性部分, 再次作呈线性部分的那些较大顺序统计量的 Sum-plot 图 (见图 1-图 6).

Sum-plot 方法不受分布类型和样本异常值影响, 但是由于 Sum-plot 方法需要观察以坐标 $\{(k, S_k), 1 \leq k \leq n\}$ 画成的散点图在哪一点偏离直线, 因此选择的 k 有一定的猜测性, 会对估计结果造成不可避免的误差.

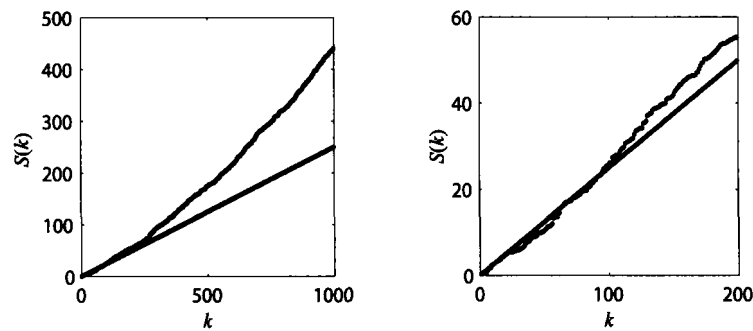
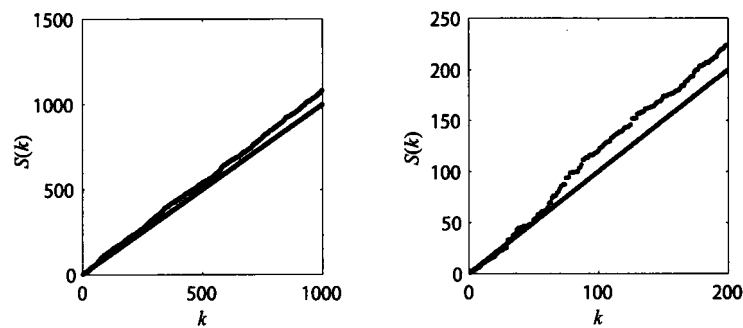
图 1 $t(4)$ 分布的两次 Sum-plot 图

图 2 Cauchy 分布的两次 Sum-plot 图

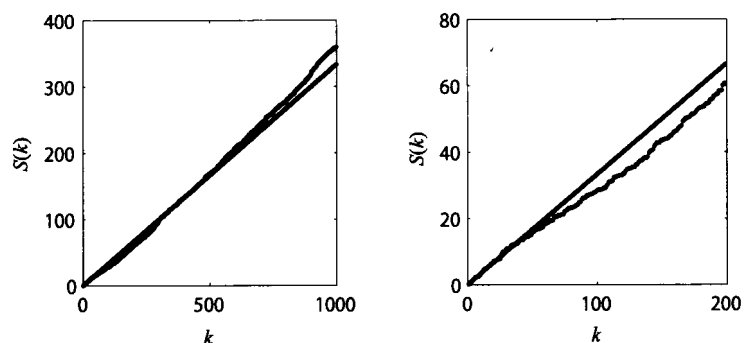


图 3 Fréchet(3) 分布的两次 Sum-plot 图

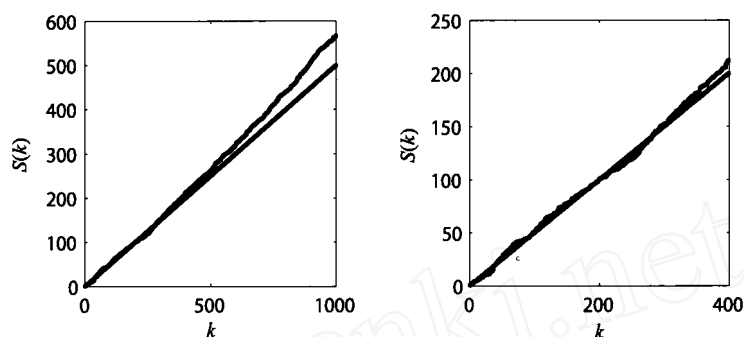


图 4 逆 Ga(2,1) 分布的两次 Sum-plot 图

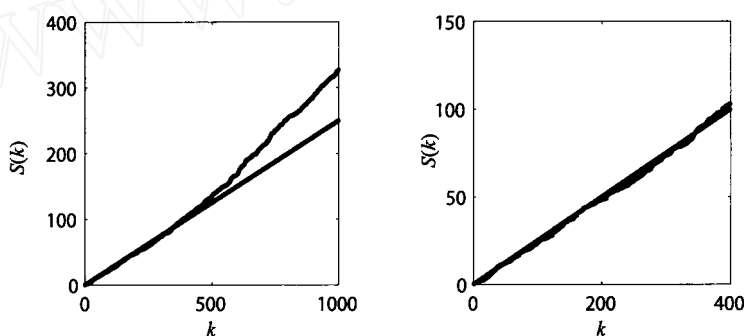


图 5 Burr(2,2) 分布的两次 Sum-plot 图

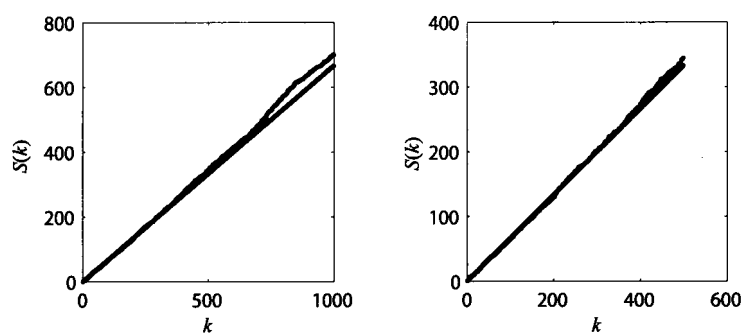


图 6 Pareto(1,1.5) 分布的两次 Sum-plot 图

应用 Sum-plot 方法, Bootstrap 方法和简便优化方法选取 k 都能得到令人满意的估计结果, 简便优化方法精确性似乎优于 Sum-plot 方法和 Bootstrap 方法 (见表 2).

4 实证

股指收益率作为金融资产价格波动的指标, 对投资者的理性投资起着很大的指导作用. 金融资产价格的巨幅波动, 如股票的暴涨暴跌, 反映在收益率这一指标上就是股票极端收益率情形, 而极端收益率发生情况由收益率分布的尾部特征来决定. 我们选取 1993 年 12 月 14 日到 2009 年 2 月 23 日上证综合指数每日收盘价 p_t (共 3695 个数据) 和 1999 年 3 月 31 日到 2009 年 4 月 1 日深证综合指数每日收盘价 p_t (共 2418 个数据)

作为基础数据, 股指收益 $R_t = \log p_t - \log p_{t-1}$ 作为样本数据. 利用简便优化方法, 计算结果表明选取上证指数股指收益率的极大次序统计量个数 $\hat{k} = 43$, 将 \hat{k} 代入 Hill 估计, 计算分布的右重尾指数 $\alpha = 2.664$. 选取上证指数股指收益率的极大次序统计量个数 $\hat{k} = 17$, 将 \hat{k} 代入 Hill 估计, 计算分布的右重尾指数 $\alpha = 3.905$. 通过比较精确地估计我国上证综指收益率和深证综指收益率分布的重尾指数, 对我国上证综指收益率和深证综指收益率的分布性质给出了基本判断, 为进一步研究我国股市的股指收益, 金融市场风险, 金融资产定价以及金融创新提供了依据.

表 2 三种方法用于 t , Cauchy, Fréchet, 逆 Ga, Burr 和 Pareto 的结果

分布 方法	$t(4)$		Cauchy		Fréchet(3)		逆 Ga(2,1)		Burr(2,2)		Pareto(1,1.5)	
	\hat{k}	$\hat{\alpha}$	\hat{k}	$\hat{\alpha}$	\hat{k}	$\hat{\alpha}$	\hat{k}	$\hat{\alpha}$	\hat{k}	$\hat{\alpha}$	\hat{k}	$\hat{\alpha}$
Sum-plot 方法	104	3.7754	62	0.9099	50	3.2110	320	1.9488	350	3.9376	400	1.4610
Bootstrap 方法	40	4.5978	132	0.8392	240	3.3214	16	2.4581	116	4.3764	236	1.4763
比较优化法	84	4.0997	13	1.0735	243	3.1276	94	1.9645	311	4.0274	34	1.5441

5 结论

利用 Monte-Carlo 模拟对 Sum-plot 方法、Bootstrap 方法和本文提出的简便优化方法进行比较, 得到以下结论:

- 1) 由于简便优化方法只是解决简单的优化问题, 因此其计算量非常小;
- 2) 通过对多种已知重尾分布 Monte-Carlo 模拟, 简便优化方法也不受重尾分布类型和重尾指数范围的影响, 具有良好的稳健性;
- 3) 简便优化方法精确性优于 Sum-plot 方法和 Bootstrap 方法. Bootstrap 方法受异常值和样本容量影响较大, Sum-plot 方法需要观察散点图在哪一点偏离直线, 因此选择的 k 有一定的猜测性, 因而估计结果不稳定, 而简便优化方法得到的是 k 的解析值;
- 4) 在分布函数为 Pareto 型时, 由简便优化方法选取次序统计量的 Hill 估计依然具有相合性.

同时, 对我国上证综指收益率和深证成指收益率的分布性质给出了基本判断, 比较精确地估计了分布的重尾指数, 为进一步研究我国股市的股指收益, 金融市场风险, 金融资产定价以及金融创新提供了依据, 本文对提出的简便优化方法的精确性, 以及在分布类型和重尾指数范围方面的稳健性进行了实证分析, 样本容量以及样本相关性对简便优化方法的影响将是我们以后研究的内容.

参考文献

- [1] Hill B. A simple general approach to inference about the tail of a distribution[J]. *Annals of Statistics*, 1975, 3: 1163-1174.
- [2] Resnick S, Starica C. Tail index estimation for dependent data[J]. *The Annals of Applied Probability*, 1998, 4: 1156-1183.
- [3] Kratz M, Resnick S. The qq-estimator and heavy tails[J]. *Stochastic Models*, 1996, 12(4): 699-724.
- [4] Beirlant J, Vynckier P, Teugels J L. Tail index estimation, pareto quantile plots, and regression diagnostics[J]. *Journal of the American Statistical Association*, 1996, 436: 1659-1667.
- [5] Resnick S, Starica C. Smoothing the Hill estimator[J]. *Advances in Applied Probability*, 1997, 29: 271-293.
- [6] Drees H, Haan L D, Resnick S. How to make a Hill plot[J]. *Annals of Statistics*, 2000, 28: 254-274.
- [7] Sousa B. A contribution to the estimation of the tail index of heavy-tailed distributions[D]. The University of Michigan, 2002.
- [8] Hall P. Using the bootstrap to estimate mean square error and select smoothing parameters in non-parametric problems[J]. *Journal of Multivariate Analysis*, 1990, 32: 177-203.
- [9] Danielsson J. Using a bootstrap method choose the sample fraction in tail index estimation[J]. *Journal of Multivariate Analysis*, 2001, 76: 226-248.
- [10] 赫英迪. 重尾分布的尾部指数估计及其沪深股市实证分析 [D]. 太原: 山西大学, 2007.
He Y D. Tail index estimation of heavy-tailed distribution and empirical analysis of China's stock markets[D]. Taiyuan: Shanxi University, 2007.
- [11] Resnick S. Heavy-Tail Phenomena: Probabilistic and Statistical Modeling[M]. New York, USA: Springer Science+Business Media LLC, 2007.