

沪深股市收益率的Copula模型及参数估计

李晓康

陕西理工大学, 陕西 汉中

收稿日期: 2021年9月30日; 录用日期: 2021年10月15日; 发布日期: 2021年10月29日

摘要

沪深股市对数收益率具有较强的相关性。使用GPD分布描述其边际分布, 对其阈值采用Hill矩法进行估计, 对其参数使用极大似然估计给出, 并求其数值解。对其相关结构的描述采用二元正态Copula和二元t-Copula, 并将其相关参数与样本进行比较。结果表明: 二元t-Copula能够较好地描述其相关结构, 与样本数据拟合较好。

关键词

收益率, 相关性, GPD分布, Copula函数

Copula Model and Parameter Estimation about Shanghai and Shenzhen Stock Market Returns

Xiaokang Li

Shaanxi University of Technology, Hanzhong Shaanxi

Received: Sep. 30th, 2021; accepted: Oct. 15th, 2021; published: Oct. 29th, 2021

Abstract

There is a strong correlation between the logarithmic return of Shanghai and Shenzhen stock markets. The GPD distribution is used to describe its marginal distribution, the hill moment method is used to estimate its threshold, the maximum likelihood estimation is used to estimate its parameters, and the numerical solution is obtained. For the description of the correlation structure, the binary normal copula and binary t-copula are used, and compare the correlation parameters with the samples. The results show that the binary t-copula can describe its correlation structure well and fit the sample data well.

Keywords

Rate of Return, Correlation, GPD Distribution, Copula Function

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

1. 引言

股市是市场经济的产物,在资源配置(特别是资本配置)中发挥着关键性作用。随着改革开放和市场经济的发展,我国股市经历了从无到有、有小到大的发展。1990年12月19日上海证券交易所开始正式营业。1990年12月1日,深圳证券交易所试营业。三十年来,沪深股市对我国的改革开放事业和市场经济的发展发挥了非常重要的作用。但是,沪深股市也面临着许多问题,主要表现在:我国股市是一个典型的主力交易者占主导地位的市场;绝大多数投资者非理性;缺乏衍生市场。按照效率市场理论的观点,我国股市是一个无效市场,即没有达到弱势有效[1]。

股票市场的变化(股票价格、市场指数)受许多因素的影响,是一个随机过程。描述市场变化规律对指导投资者理性投资、促进市场健康有效发展具有十分重要的意义。股市的暴涨暴跌使投资者面临着极大的风险。极值理论是描述这类现象的有力工具,并且在金融、环境、气象等许多方面取得了成功应用[2] [3] [4] [5] [6]。

本文将应用极值分布对沪深股市的暴涨、暴跌进行建模,对极值分布参数进行估计,利用 Copula 函数对其进行相关性分析。

2. 极值分布简介

样本数据的选取通常有两种:

第一种是区间选取法,也称为块方法(block method),它是将样本数据按照时间顺序分为若干个区间,从每个区间选取一个极大、小值,从而得到极值样本。

第二种是超阈值法(peak over threshold),也称为 POT 法,它是按照某种方法确定一个阈值,将超过阈值的数据作为样本。

基于不同的样本数据选取,极值分布通常有两种:广义极值分布(GEV)与广义帕累托分布(GPD)。

2.1. GEV 分布

广义极值分布(GEV)的分布函数为:

$$H_{u,\xi,\sigma}(x) = \begin{cases} \exp\left\{-\left[1+\xi\frac{x-u}{\sigma}\right]^{-1/\xi}\right\}, \xi \neq 0 \\ \exp\left\{-\exp\left(-\frac{x-u}{\sigma}\right)\right\}, \xi = 0 \end{cases}$$

其中, u 为位置参数, σ 为尺度参数, ξ 为形状参数。

2.2. GPD 分布

广义帕累托分布(GPD)的分布函数为:

$$G_{u,\xi,\sigma}(x) = \begin{cases} 1 - \left[1 + \xi \frac{x-u}{\sigma} \right]^{-1/\xi}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{x-u}{\sigma}\right), & \xi = 0 \end{cases}$$

其中, u 为位置参数, σ 为尺度参数, ξ 为形状参数, 随形状参数 ξ 由小到大(由负值变为正值), 尾部逐渐变厚。

二者具有如下关系:

$$G(x) = 1 + \ln H(x)$$

即: GPD 分布的分布函数可以看做 GEV 分布的分布函数的一个单增变换。

本文采用超阈值法选取极值样本, 利用 GPD 分布描述极值样本的分布。

3. 参数估计

GPD 分布的参数估计方法主要有极大似然估计与分布估计。

3.1. 极大似然估计

按照极大似然原理, 对给定样本, 所求参数估计应使得似然函数或对数似然函数达到极大, 此即极大似然估计(MLE)。

GPD 分布的密度函数为:

$$g(x) = \frac{1}{\sigma} \left(1 + \xi \frac{x-u}{\sigma} \right)^{-1/\xi-1}, \xi \neq 0$$

似然函数(样本联合密度)为:

$$\begin{aligned} L(x_1, \dots, x_n) &= \prod_{i=1}^n g(x_i) \\ &= \prod_{i=1}^n \frac{1}{\sigma} \left(1 + \xi \frac{x_i-u}{\sigma} \right)^{-1/\xi-1} \\ &= \left(\frac{1}{\sigma} \right)^n \prod_{i=1}^n \left(1 + \xi \frac{x_i-u}{\sigma} \right)^{-1/\xi-1} \end{aligned}$$

对数似然函数为:

$$l(x_1, \dots, x_n) = -n \ln \sigma - \left(1 + \frac{1}{\xi} \right) \sum_{i=1}^n \ln \left(1 + \xi \frac{x_i-u}{\sigma} \right)$$

由极值原理, 所求极大似然估计应满足如下正规方程组:

$$\begin{cases} \frac{\partial l}{\partial u} = 0 \\ \frac{\partial l}{\partial \xi} = 0 \\ \frac{\partial l}{\partial \sigma} = 0 \end{cases}$$

即:

$$\begin{cases} \frac{\partial l}{\partial u} = (1+\xi) \sum_{i=1}^n \frac{1}{\sigma + \xi(x_i - u)} = 0 \\ \frac{\partial l}{\partial \xi} = \frac{1}{\xi^2} \sum_{i=1}^n \ln \left(1 + \xi \frac{x_i - u}{\sigma} \right) - (1+\xi) \sum_{i=1}^n \frac{x_i - u}{\sigma + \xi(x_i - u)} = 0 \\ \frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + (1+\xi) \sum_{i=1}^n \frac{x_i - u}{\sigma^2 \left(1 + \xi \frac{x_i - u}{\sigma} \right)} = 0 \end{cases}$$

可求上述方程组的解析解，即得参数极大似然估计，在实际应用中，可使用数值方法求上述方程组的数值解。

3.2. 分布估计

对 GPD 分布的参数估计可以分两步进行：

第一步：估计位置参数 u (见下节)；

第二步：按极大似然法估计形状参数与尺度参数，即选取 $\hat{\xi}, \hat{\sigma}$ ，使：

$$\hat{\xi}, \hat{\sigma} = \arg \max l(X_1, \dots, X_n, \xi, \sigma)$$

3.3. 位置参数的估计

位置参数(阈值)的估计有两种方法：

第一种是样本超阈值函数法。

样本超阈值函数定义为：

$$MEF(u) = \frac{\sum_{i=1}^n (X_i - u)}{\sum_{i=1}^n I_{X_i > u}}$$

其中： $I_{X_i > u}(\cdot)$ 为示性函数，当 $X_i > u$ 时，其值为 1，否则为 0。

MEF 的含义为所有大于阈值的量的平均值。阈值的选取为样本超阈值函数图中由水平向正斜率变化的 u 的值。

第二种是 Hill 矩法。

令 $X_{(1)} > X_{(2)} > \dots > X_{(n)}$ 为独立同分布降序统计量。Hill 统计量定义如下：

$$H_{k,n} = \frac{1}{k} \sum_{l=1}^k \ln \frac{X_{(l)}}{X_{(k)}}$$

Hill 图为点 $\{(k, H_{k,n}^{-1}), 1 \leq k \leq n-1\}$ 的集合。阈值选取为 Hill 图中稳定区域起始点的横坐标 k 对应的 X_k 。

4. Copula 函数

Copula 函数是联系多维随机变量联合分布与各自边缘分布的桥梁，能够很好地描述多维随机变量的相关结构[7] [8] [9] [10] [11]。Copula 函数的定义如下：

定义 1 若多维随机变量 X_1, \dots, X_p 的联合分布为 $F(x_1, \dots, x_p)$ ，各自的边缘分布为 $F_1(x_1), \dots, F_p(x_p)$ ，存在多元函数 $C(\cdot)$ 满足： $F(x_1, \dots, x_p) = C(F_1(x_1), \dots, F_p(x_p))$ ，称 $C(\cdot)$ 为 Copula 函数。

本文采用二元 Copula 函数描述沪、深股市收益率极端情况的相关结构。

常用的 Copula 函数有以下几种:

1) 正态 Copula 函数

正态 Copula 分布函数形式如下:

$$C(u_1, \dots, u_N; \rho) = \Phi_\rho(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \dots, \Phi^{-1}(u_N))$$

其中: ρ 为 N 元标准正态分布的相关系数阵, $\Phi_\rho(\cdot)$ 表示相关系数阵为 ρ 的 N 元标准正态分布分布函数, Φ^{-1} 表示标准正态分布的逆函数。即正态 Copula 是由标准正态分布的逆函数和多元正态分布函数组合而成。

对二元情形, 若变量间线性相关系数为 ρ , 则二元正态 Copula 函数为:

$$C(u, v; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \exp\left[-\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)}\right] ds dt$$

2) t-Copula 函数

t-Copula 分布函数形式如下:

$$C(u_1, \dots, u_N; \rho, k) = t_{\rho, k}(t_k^{-1}(u_1), t_k^{-1}(u_2), \dots, t_k^{-1}(u_N))$$

其中: ρ 为自由度为 k 的 N 元标准 t 分布的相关系数阵, $t_{\rho, k}$ 表示相关系数阵为 ρ 、自由度为 k 的 N 元标准 t 分布的分布函数, t_k^{-1} 表示自由度为 k 的一元标准 t 分布的逆函数。即 t-Copula 是由标准 t 分布的逆函数和多元 t 分布分布函数组合而成。

对二元情形, 若变量间线性相关系数为 ρ 、自由度为 k , 则二元 t-Copula 函数为:

$$C(u, v; \rho, k) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{t_k^{-1}(u)} \int_{-\infty}^{t_k^{-1}(v)} \left[1 + \frac{s^2 - 2\rho st + t^2}{k(1-\rho^2)}\right]^{-(k+2)/2} ds dt$$

5. 实证分析

本文选取上证指数与深证成指每日收盘价历史数据(2010.1.4~2018.12.28), 对数据采用对数变换, 得对数收益率, 即:

$$r_{i+1} = \ln X_{i+1} - \ln X_i = \ln \frac{X_{i+1}}{X_i}, i = 1, \dots, n-1$$

对沪深股市对数收益率的极端值(暴涨、暴跌)利用极值分布(GPD 分布)进行建模, 以此为边际分布, 利用二元 Copula 函数描述其相关结构, 并对相关参数进行估计, 以此为根据, 对沪深股市相关性进行定量分析。

5.1. 数据的描述性统计分析

对以上数据进行描述性统计分析, 可得以下结果:

Table 1. Descriptive statistics of logarithmic returns about Shanghai and Shenzhen stock markets

表 1. 沪深股市对数收益率描述性统计表

上证指数对数收益率		深证成指对数收益率	
平均	-0.000120259	平均	-0.00029
标准误差	0.000296113	标准误差	0.000352

Continued

中位数	0.000505808	中位数	-5.4E-05
标准差	0.013844673	标准差	0.016461
方差	0.000191675	方差	0.000271
峰度	5.997824969	峰度	3.656566
偏度	-0.95008146	偏度	-0.70689

由上表 1 可以看出, 沪深股市对数收益率具有明显的尖峰(峰度大于 3)、偏态(偏度小于 0)分布特征, 为非正态分布。

由图 1、图 2 可以看出, 深证指数收盘价和对数收益率具有明显的变化趋势及阶段特征, 呈现出明显的波动性和聚集性。

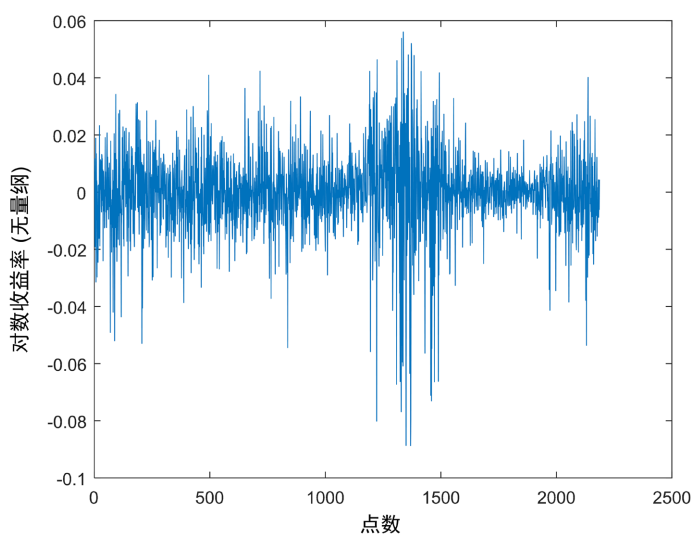


Figure 1. Logarithmic yield of Shanghai stock index

图 1. 上证指数对数收益率

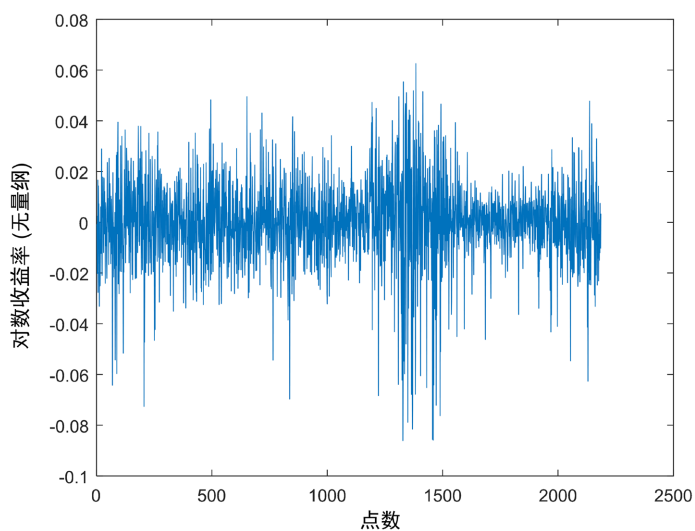


Figure 2. Logarithmic yield of Shenzhen composite index

图 2. 深证成指对数收益率

由图 3、图 4 可以看出，沪深股市对数收益率具有明显的非正态分布、尖峰、厚尾特征。

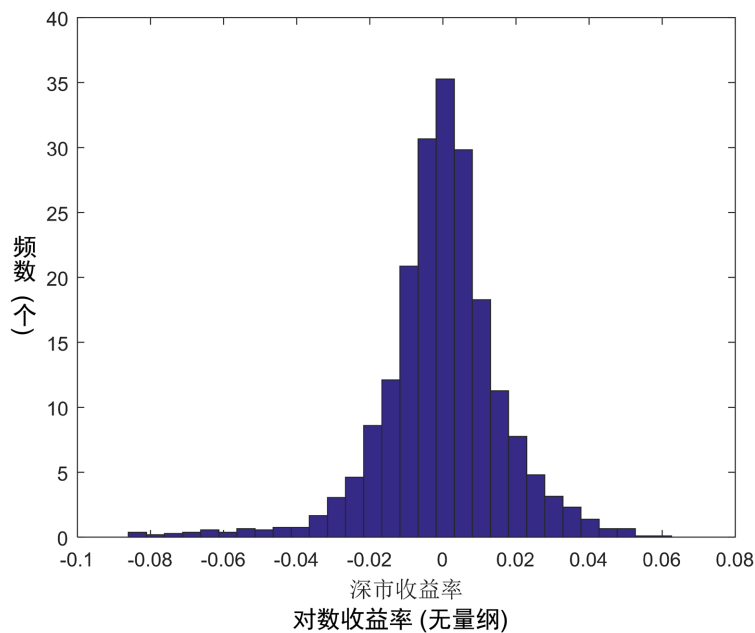


Figure 3. Log return histogram of Shanghai stock index
图 3. 上证指数对数收益率直方图

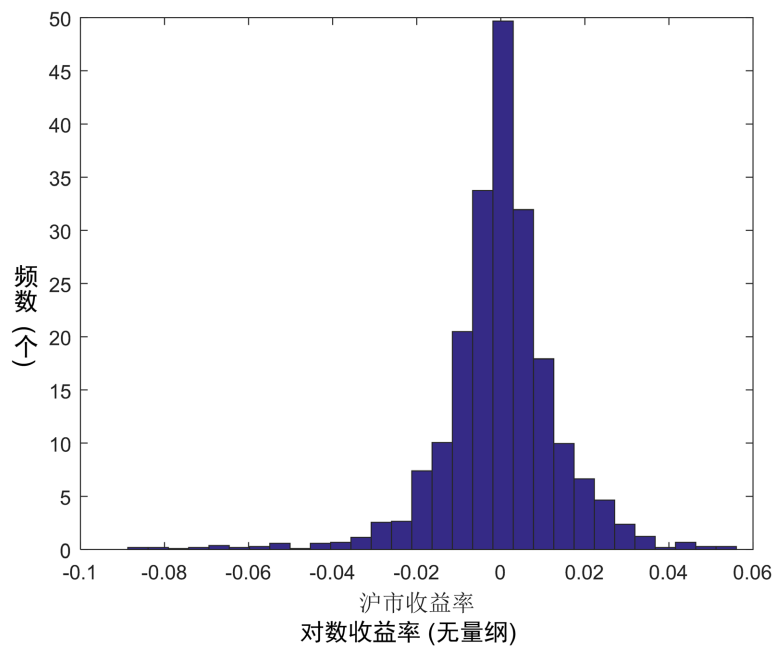


Figure 4. Log return histogram of Shenzhen stock index
图 4. 深证成指对数收益率直方图

对沪深股市对数收益率数据分别进行正态性检验，结果如图 5、图 6。

由图 5、图 6 可以看出，沪深股市对数收益率不服从正态分布。

为考查沪深股市对数收益率的相关性，计算其相关系数，结果如表 2。

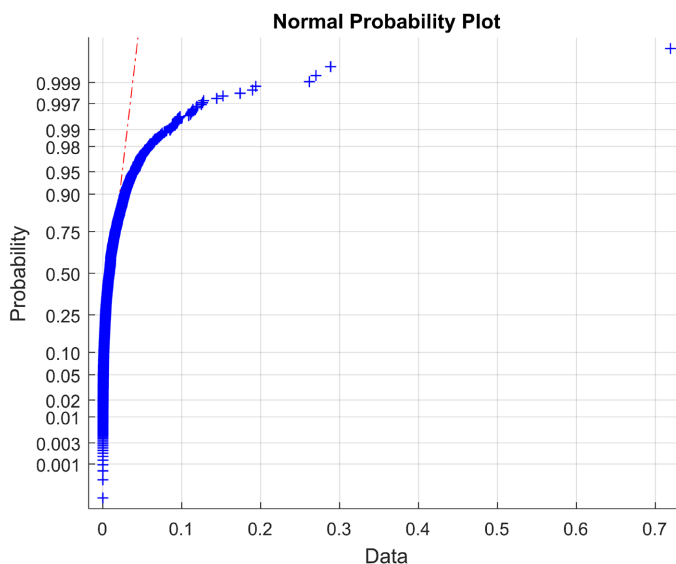


Figure 5. Test chart for normality of log return of Shanghai stock index
图 5. 上证指数对数收益率正态性检验图

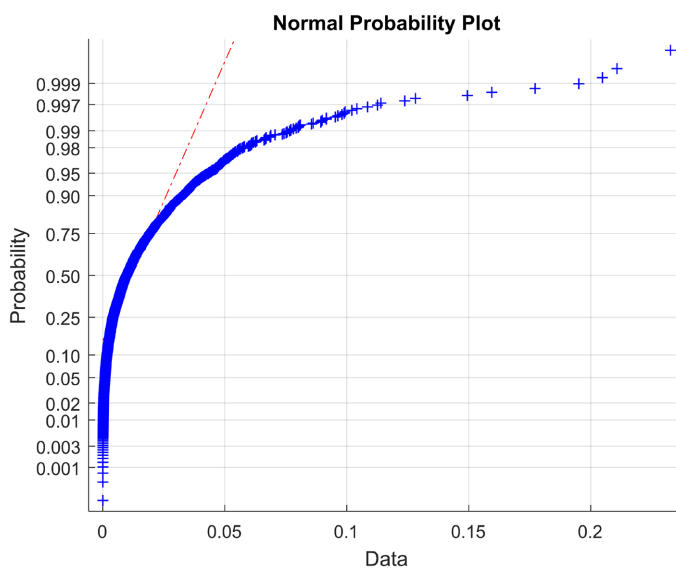


Figure 6. Test chart for normality of log return of Shenzhen stock index
图 6. 深证成指对数收益率正态性检验图

Table 2. Table of linear correlation coefficient of log return in Shanghai and Shenzhen stock markets
表 2. 沪深股市对数收益率线性相关系数表

	上证指数对数收益率	深证成指对数收益率
上证指数对数收益率	1	0.9282
深证成指对数收益率	0.9282	1

由上表可以看出，沪深股市对数收益率具有高度的相关性，可采用合适的 Copula 函数描述其相关性。

由表 2 及图 7 可以看出，沪深股市对数收益率具有高度相关性(线性相关系数为 0.928238)，呈现出明显的线性相关性。

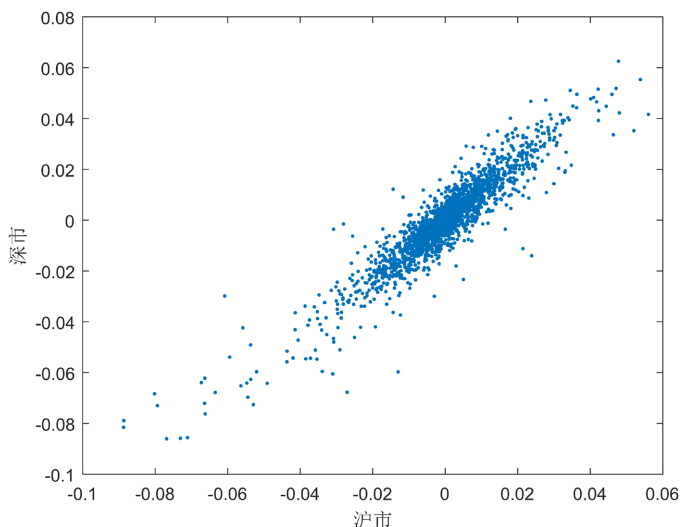


Figure 7. Relationship between log return of Shenzhen composite index and log return of Shanghai composite index
图 7. 深证成指对数收益率与上证指数对数收益率关系图

基于以上特征及分析，下面采用 GPD 分布对沪深股市对数收益率的极端情况进行建模，描述其边际分布。

5.2. GPD 分布阈值的确定

按照 2.3 节介绍的方法，使用 Hill 统计量：

$$H_{k,n} = \frac{1}{k} \sum_{l=1}^k \ln \frac{X_{(l)}}{X_{(k)}}$$

确定阈值。

为此，选取沪深股市对数收益率大于 0 (上涨)的数据，将其由大到小排序，计算 Hill 统计量值，并绘制 Hill 图(点 $\{(k, H_{k,n}), 1 \leq k \leq n-1\}$ 的集合)。选取 Hill 图中稳定区域起始点的横坐标 k 对应的 X_k 为相应 GPD 分布的阈值。Hill 图如下：

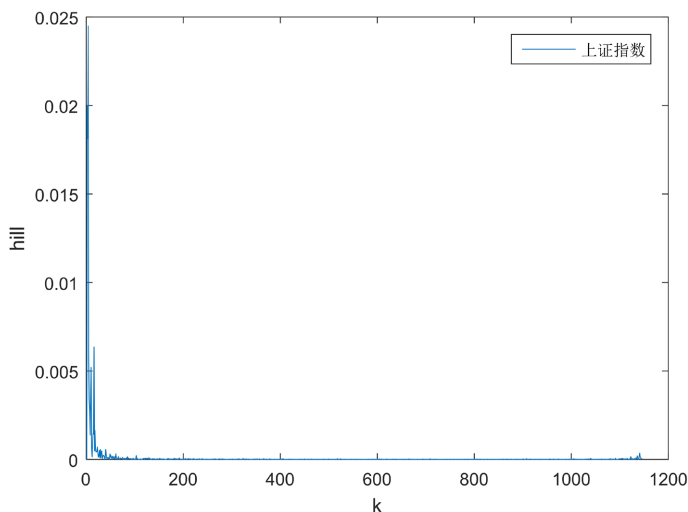


Figure 8. Hill chart of log return of Shanghai stock index
图 8. 上证指数对数收益率 Hill 图

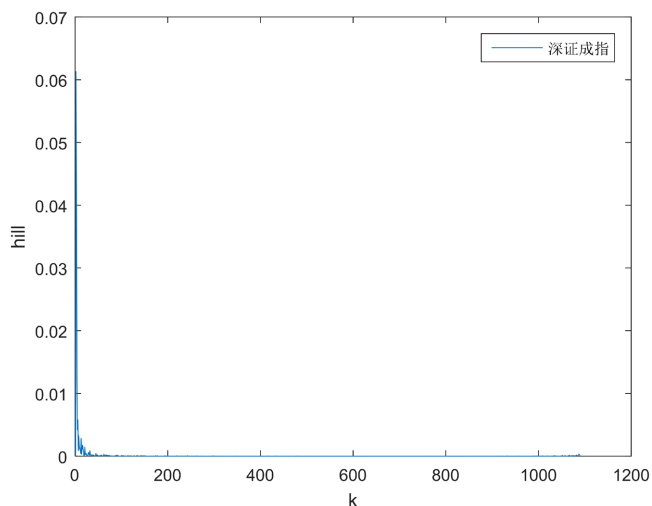


Figure 9. Hill chart of log return of Shenzhen stock index
图 9. 深证成指对数收益率 Hill 图

由上图 8、图 9 可以看出：对上证指数对数收益率大于 0 的样本数据，当 $u = 0.0002$ 时，Hill 图基本稳定；对深证成指对数收益率大于 0 的样本数据，当 $u = 0.0006$ 时，Hill 图基本稳定。故选取阈值如下表 3：

Table 3. GPD distribution threshold table of log return in Shanghai and Shenzhen stock markets
表 3. 沪深股市对数收益率 GPD 分布阈值表

	阈值
上证指数对数收益率	0.0002
深证成指对数收益率	0.0006

5.3. 沪深股市对数收益率的 GPD 分布及参数估计

选取样本数据中对数收益率为正(上涨)的数据，利用 matlab 统计工具箱中的函数 ecdf 绘制经验分布函数图形如图 10、图 11：

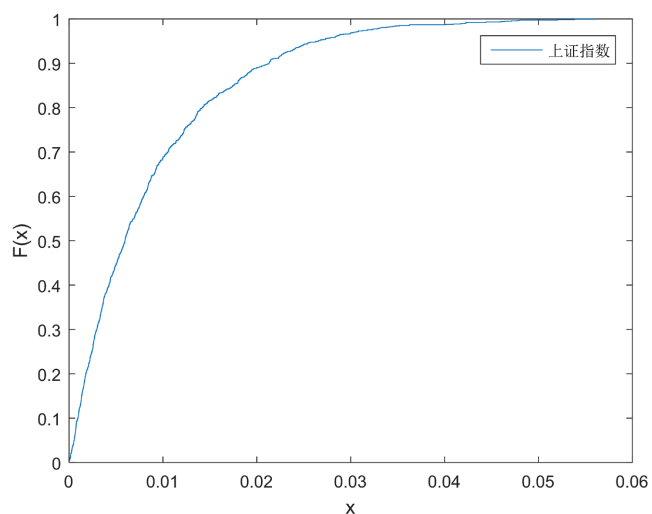


Figure 10. Empirical distribution function of log return about Shanghai stock index
图 10. 上证指数对数收益率经验分布函数图

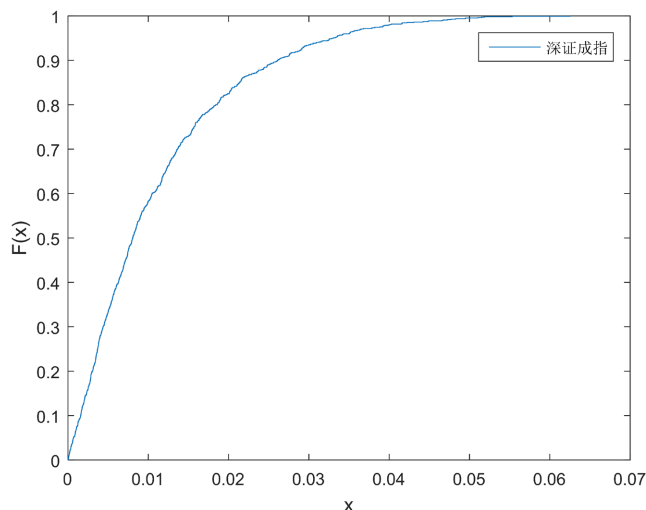


Figure 11. Empirical distribution function of log return about Shenzhen stock index
图 11. 深证成指对数收益率经验分布函数图

利用 GPD 分布对沪深股市对数收益率数据进行描述，使用 matlab 统计工具箱的函数 gpfir 进行参数估计，得如下结果(表 4):

Table 4. GPD distribution parameters of log return in Shanghai and Shenzhen stock markets
表 4. 沪深股市对数收益率 GPD 分布参数表

	尺度参数 σ	形状参数 ξ
上证指数对数收益率	0.0104	0.1827
深证成指对数收益率	0.0133	0.0817

上证指数对数收益率分布如下:

$$G_{hushi}(x) = 1 - \left[1 + 0.1827 \frac{x - 0.0002}{0.0104} \right]^{-1/0.1827}$$

深证成指对数收益率分布如下:

$$G_{shengshi}(x) = 1 - \left[1 + 0.0817 \frac{x - 0.0006}{0.0133} \right]^{-1/0.0817}$$

5.4. 沪深股市对数收益率的二元 Copula 模型

利用二元 Copula 函数描述沪深股市收益率的相关性，即在二元 Copula 函数族中寻找一个合适的 Copula 函数，使其能够很好的描述沪深股市的相关性。

首先，根据沪深股市对数收益率样本数据绘制二元频率直方图如图 12。

由图 12 可以看出，其尾部基本对称，故可选取二元正态 Copula 或二元 t-Copula 函数来描述其尾部及相关结构。

下面利用二元正态 Copula 和二元 t-Copula 对沪深股市对数收益率进行描述。

首先对二元 Copula 模型的参数进行估计。利用 MATLAB 统计工具箱中的 copulafit、copulastat、copulaparam、copulapdf、copulacdf 等函数对二元正态 Copula 模型的相关参数、Kendall 秩相关系数、Spearman 秩相关系数、密度函数值、分布函数值等进行计算，结果如下(表 5):

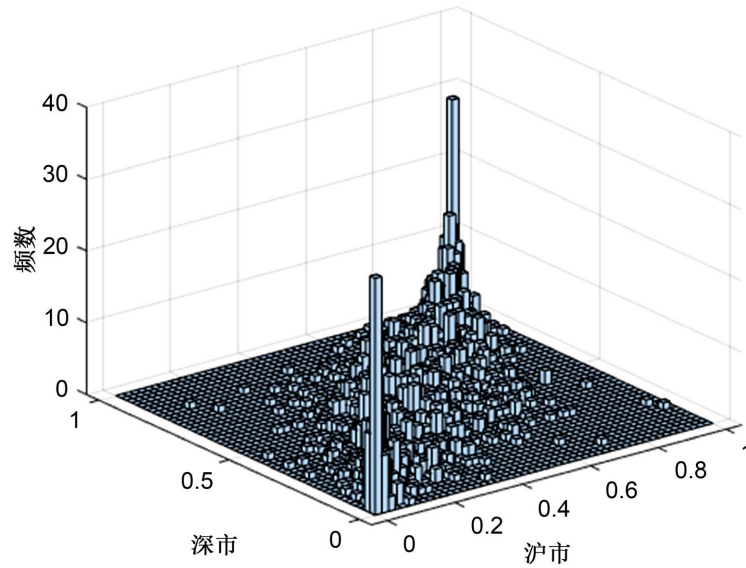


Figure 12. Binary frequency histogram of log return in Shanghai and Shenzhen stock markets
图 12. 沪深股市对数收益率二元频率直方图

Table 5. Correlation coefficient table of Copula function of logarithmic rate of return in Shanghai and Shenzhen stock markets
表 5. 沪深股市对数收益率 Copula 函数相关系数表

	二元正态 Copula	二元 t-Copula	样本
线性相关系数	0.9234	0.9289	0.9282
Kendall 秩相关系数	0.7492	0.7584	0.7525
Spearman 秩相关系数	0.9166	0.7584	0.9085
自由度		3.57	

由以上结果可以得出，对应的二元正态 Copula 为：

$$C(u, v; \rho) = \frac{1}{2\pi\sqrt{1-0.9234^2}} \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \exp\left[-\frac{s^2 - 2 \times 0.9234st + t^2}{2(1-0.9234^2)}\right] dsdt$$

对应的二元 t-Copula 为：

$$C(u, v; \rho, k) = \frac{1}{2\pi\sqrt{1-0.9289^2}} \int_{-\infty}^{t_k^{-1}(u)} \int_{-\infty}^{t_k^{-1}(v)} \left[1 + \frac{s^2 - 2 \times 0.9289st + t^2}{4(1-0.9289^2)}\right]^{-(4+2)/2} dsdt$$

由上表可以看出，线性相关系数为 0.9289、自由度为 4 的二元 t-Copula 更能反映沪深股市对数收益率间的尾部相关性。其尾部相关系数由下式计算：

$$\hat{\lambda}_{up} = \hat{\lambda}_{lo} = 2 - 2t_{k+1}\left(\frac{\sqrt{k+1}\sqrt{1-\rho}}{\sqrt{1+\rho}}\right) = 0.6856$$

以上结果说明，沪深股市对数收益率具有正的上、下尾相关性，即在上涨和下跌阶段具有同步变化的关系。

由上表的秩相关系数可以看出，二元正态 Copula 的秩相关系数为 0.7492，与样本秩相关系数更加接近，能更好反映沪深股市对数收益率间的秩相关性。

5.5. 模型评价

为评价两个模型的优劣，下面定义经验 Copula，利用经验 copula 与模型的距离来进行度量。

经验 Copula 若 $(x_i, y_i)(i=1, 2, \dots, n)$ 为来自总体 (X, Y) 的样本， X 的经验分布函数为 $F_n(x)$ ， Y 的经验分布函数为 $G_n(y)$ ，称

$$\hat{C}_n(u, v) = \frac{1}{n} \sum_{i=1}^n I_{F_n(x_i) \leq u} I_{G_n(y_i) \leq v}, u, v \in [0, 1]$$

为经验 Copula。其中： $I_{F_n(x_i) \leq u}$ 为示性函数，当 $F_n(x_i) \leq u$ 时， $I_{F_n(x_i) \leq u} = 1$ ，否则 $I_{F_n(x_i) \leq u} = 0$ 。

平方欧式距离 称所有样本的经验 Copula 和模型的差的平方和为平方欧式距离，定义如下：

$$d^2 = \hat{C}_n(u, v) = \sum_{i=1}^n [\hat{C}_n(u_i, v_i) - C(u_i, v_i)]^2$$

其中： $u_i = F_n(x_i), v_i = G_n(y_i)$ 。

利用如上的平方欧式距离可以度量模型与样本的拟合程度。对二元正态 Copula 和二元 t-Copula，平方欧式距离分别为 0.023 和 0.014。由此可见，二元 t-Copula 能够更好拟合沪深股市对数收益率观测数据。

6. 结论

沪深股市收益率变化具有较强的相关关系和长期系统的一致变化趋势。通过以上分析可以看出，二元正态 Copula 和二元 t-Copula 都从不同侧面描述了沪深股市对数收益率的相关关系。二元 t-Copula 的线性相关系数与样本相关系数更加接近，更能反映沪深股市对数收益率间的尾部相关性。二元正态 Copula 的秩相关系数与样本秩相关系数更加接近，能更好反映沪深股市对数收益率间的秩相关性。从平方欧式距离来看，二元 t-Copula 更小，能够更好拟合沪深股市对数收益率观测数据。

基金项目

本课题由陕西省教育厅自然科学基金项目(18JK0141)资助。

参考文献

- [1] 张元萍. 数理金融[M]. 北京: 中国金融出版社, 2004: 244-248.
- [2] 余笑荷, 艾蔚, 袁芳英, 徐吟川. 基于尾部相依视角的证券业系统性风险度量[J]. 统计与决策, 2019, 35(17): 162-165.
- [3] 王娜, 任燕燕. 基于固定效应面板数据的 Copula 分位数回归及模拟[J]. 统计与决策, 2019(19): 82-86.
- [4] 韦艳华. Copula 理论及其在金融分析上的应用[M]. 北京: 清华大学出版社, 2008.
- [5] 李强, 周孝华. 基于 Copula 的我国台湾和韩国股票市场相关性研究[J]. 管理工程学报, 2014(2): 100-107.
- [6] 李强. 基于 Copula 理论和 GPD 模型的金融市场风险测度研究[D]: [博士学位论文]. 重庆: 重庆大学, 2012.
- [7] 华祖林, 汪靓, 顾莉, 褚克坚. 基于门限极值理论的湖泊水质参照状态的确定[J]. 中国环境科学, 2004(12): 3215-3222.
- [8] 陈思柳. 基于 MRS Copula 模型的沪港股市相依关系研究[J]. 浙江金融, 2018(11): 43-49.
- [9] 张建伟. 基于极值理论的股票市场风险度量[D]: [硕士学位论文]. 南京: 南京财经大学, 2018.
- [10] 蒋涛, 李政育. 我国银行业系统性风险结构度量——基于尾部依赖视角[J]. 合肥学院学报(综合版), 2016(4): 17-23.
- [11] 马锋, 魏宇, 黄登仕. 基于 vine copula 方法的股市组合动态 VaR 测度及预测模型研究[J]. 系统工程理论与实践, 2015, 35(1): 26-36.