

# Correlation Analysis of Shanghai and Shenzhen Stock Market Even Rising Yield—Based on Copula Connection Function

Wenqing Yang

Tianjin University of Finance and Economics Pearl River College, Tianjin, 301811, China

## Abstract

In this paper, we use non-reference kernel density to estimate the continuous yield as the edge distribution and the Copula function, and find that t-Copula can well describe the dependence structure between SSE and deep continuous yield sequence. The establishment of the t-Copula model is divided into two steps: ① Estimate the edge distribution function of the continuous yield sequence using the non-reference kernel density. ② Given the type of Copula function, parameter estimation is performed using the CML method, and then the optimal Copula model is selected based on the AIC criterion.

## Keywords

kernel density estimation; Copula function; CML method

## 沪深股市连涨收益率相关性分析——基于 Copula 连接函数

杨文青

天津财经大学珠江学院, 中国·天津 301811

## 摘要

论文利用非参核密度估计拟合连涨收益率, 以此作为边缘分布, 结合 Copula 函数, 发现 t-Copula 能很好地描述上证与深成连涨收益率序列之间相依性结构。t-Copula 模型的建立分为两步: ① 利用非参核密度估计连涨收益率序列的边缘分布函数。② 在给定 Copula 函数类型的前提下, 利用 CML 方法进行参数估计, 然后基于 AIC 准则选择最优的 Copula 模型。

## 关键词

核密度估计; Copula 函数; CML 方法

## 1 引言

Copula 函数是一种连接函数, 它将联合分布函数与其边缘分布函数连接起来, 在衡量变量相关结构方面具有优良的特性。目前很多学者将 Copula 模型应用在相关性方面, 中国学者张尧庭阐述了 Copula 技术在相关性指标选择上的优势。对股票对数收益率的文献有很多, 但对股指的连涨收益率的研究相对来说较少。雷鸣从生存分析的角度定义了连跌连涨收益率, 把连涨连跌收益率视为每次涨跌的“寿命”, 这样股指连续上涨与下跌的过程就可以看做是一个特殊的生存过程。胡心瀚等在对股票连涨和连跌收益率风险分析时引入 Copula-ACD 模型, 结果表明该方法拟合效果优于传统方法<sup>[1]</sup>。论文以上证指数(上证)的连涨收益率与深证成指(深成)的连涨收益率为研究对象, 结合 Copula 模型, 探究上证及深成的连涨收益率序列之间的相关性。

【作者简介】杨文青(1991-), 女, 中国河南周口人, 助教, 硕士, 从事经济统计研究。

## 2 数据的选取及分析

论文选取 2000 年 1 月 4 日至 2015 年 4 月 11 日的上证指数和深证成指的收盘价  $P_t$  为原始数据<sup>[2]</sup>, 样本容量为 3691 个, 数据来源于: 网易财经个股行情网站。对数收益率定

义为:  $R_t = P_t - P_{t-1}$ , 连涨收益率定义为:  $x_{m+1} = \sum_{k_{2m}}^{k_{2m+1}} R_t$  其中,  $k_0 = \inf\{t : P_{t+1} - P_t < 0\}$ ,  $k_{2m+1} = \inf\{t : t > k_{2m}, P_t - P_{t-1} < 0\}$ ,  $P_t$  是股票的当期收盘价。利用统计软件 R 对数据进行处理, 最终得到各个收益率对应的 718 组有效数据<sup>[3]</sup>。

对上证和深成连涨收益率序列进行描述性分析, 结果见图 1 和表 1。

从图 1 直观的可以看出量上证及深成连涨收益率不具有对称性, 再观察表 1 可以看出上证及深成连涨收益率的偏度均大于 0, 说明每种连涨收益率序列都是右偏的, 而且, 它们的峰度均大于 3。可以初步判断两种收益率序列存在尖峰厚尾的特征, 不满足正态分布对称的特征<sup>[4]</sup>。为了更进一步的说明两收益率序列不服从正态分布, 分别对收益率序

列做 k-stest、jbtest、lillietest 正态性检验，表 1 中的三种检验的 p 值均远小于 0.05，说明上证及深成连涨收益率均不服从正态分布<sup>[5]</sup>。

表 1 上证及深成连涨收益率描述性统计

		深证成指	上证指数
均值		0.027200	0.023300
最小值		0.000077	0.000177
最大值		0.196000	0.165000
标准差		0.028535	0.02385
偏度		2.2904	2.1767
峰度		6.9167	6.0137
正态性 检验 (p 值)	kstest	7.8595e-19	9.5549e-18
	jbtest	1.0000e-03	1.0000e-03
	lillietest	1.0000e-03	1.0000e-03

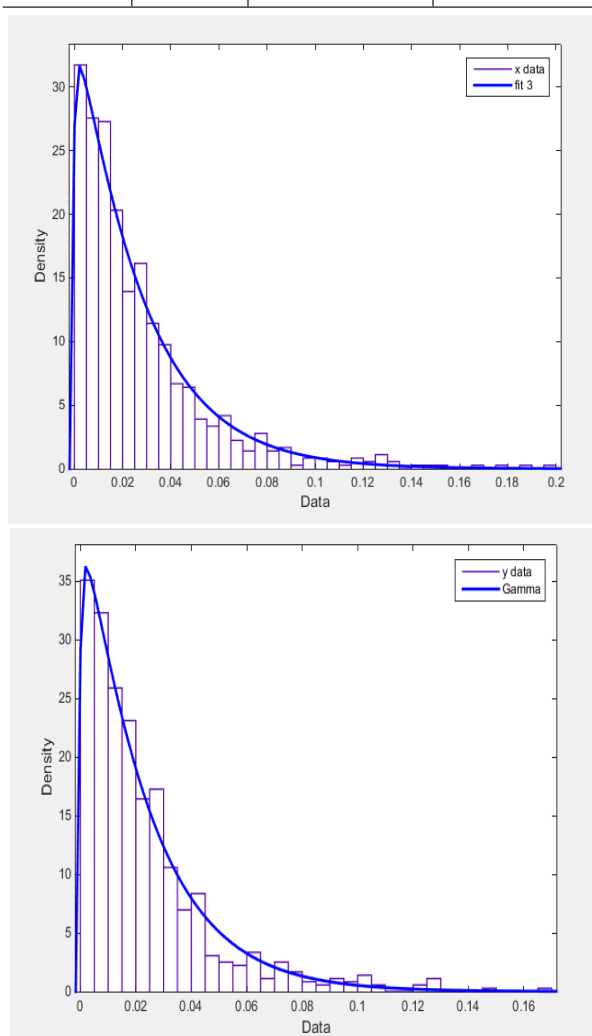


图 1 上证及深成连涨收益率直方图

### 3 Copula 相依结构的构建

基于原始数据  $(x_i, y_i)$  的核分布估计值，画出深成连涨收

益率序列与上证连涨收益率序列的联合边缘直方图，如图 2 所示，两者的相关性具有对称性。结合 Copula 函数的性质，最终我们选取 Gaussian Copula、t-Copula、Gumbel Copula、Frank Copula 来描述两连跌收益率序列的相关性结构。

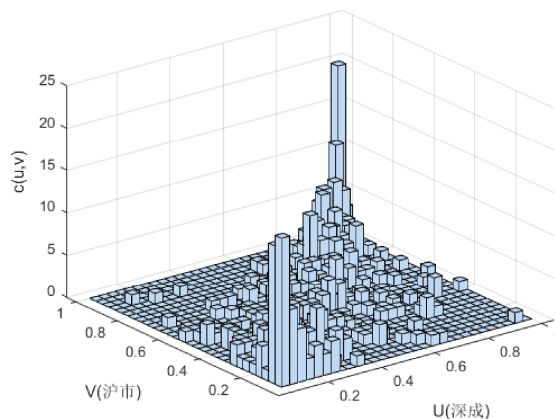


图 2 深成与上证指数收益率的联合边缘分布直方图

论文是利用 CML 方法求得 Copula 的参数，然后利用 AIC 准则判定 Copula 拟合的好坏<sup>[6]</sup>。在实证当中，基于 CML 方法估计未知参数，利用估计参数分别绘制出基于估计参数的 Copula 密度函数图与基于实际观测值的联合边缘分布图进行比较，如图 3 所示，图 3 的右边图形是基于实际观测值绘制出的联合分布图，右边是基于估计参数画出的 Copula 密度函数图。

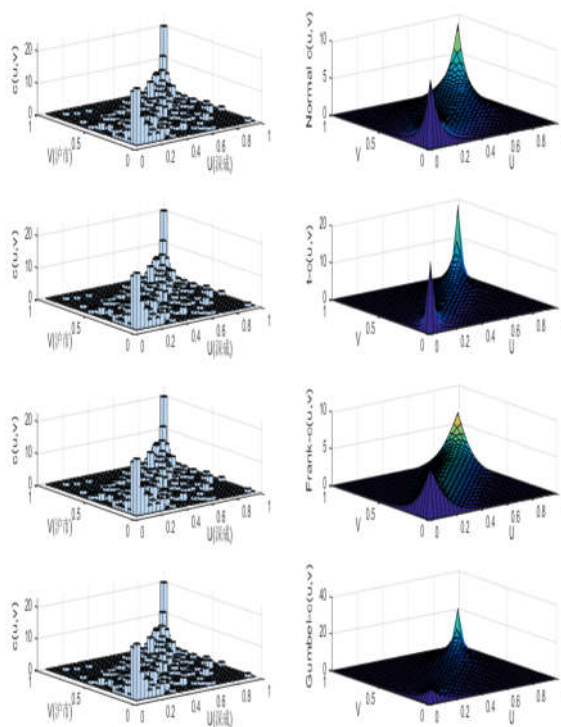


图 3 联合边缘分布频率直方图和指定的 Copula 密度函数图的比较

观察图 3, 直观上可以看出二元 t-Copula 能较好地反应两收益率的相关性结构<sup>[7]</sup>。为了进一步说明二元 t-Copula 能很好的拟合两者之间的相依性结构, 比较各种 Copula 的极大似然函数值, 结果如表 2 所示。

表 2 Copula 函数参数估计表

	参数估计	似然函数值	AIC
Gaussian Copula	0.8552	472.28	-1.3128
t-Copula	(0.8924, 2)	547.1225	-1.5212
Gumbel Copula	2.6989	502.6602	-1.3974
Frank Copula	9.7125	486.7479	-1.3531

从表 2 可知, t-Copula 的 AIC 值显然地小于其他类型 Copula 的值, 所以我们选择 t-Copula 来描述两收益率之间的相关性是合理的。得到的二元 t-Copula 的表达式如下:

$$\hat{C}(u, v) = \int_{-\infty}^{\xi^{-1}(u)} \int_{-\infty}^{\xi^{-1}(v)} \frac{1}{2\pi\sqrt{1-0.8924^2}} \left[ 1 + \frac{s^2 - 2 \times 0.8924st + t^2}{2 \times (1 - 0.8924^2)} \right]^{-2} ds dt$$

#### 4 结语

将 Copula 连接技术应用在上证指数的连涨收益率与深证成指的连涨收益率的相关性方面, 并用核密度估计拟合上证及深成连涨收益率, 以此作为 Copula 连接函数的边缘分布, 减少了估计未知参数的个数。然后采用 CML 方法来估

计给定 Copula 类型的未知相依参数, 根据 AIC 准则选取恰当的二元 t-Copula 来描述上证和深成连涨收益率的相关性结构<sup>[8]</sup>。这样, 当一个股票市场发生变化时, 投资者就可以根据相关性结构, 预测另外一个股票市场的变化概率。这在金融风险管理中有着重要的意义。

#### 参考文献

- [1] 郭名媛, 张世英. 基于高频数据的沪深股票市场的相关性研究 [J]. 系统工程学报, 2009(5):293-298.
- [2] 张尧庭. 我们应该选用什么样的相关性指标 [J]. 统计研究, 2002(9):41-44.
- [3] 雷鸣, 谭永春, 缪柏其. 运用生存分析与变点理论对上证指数的研究 [J]. 中国管理科学, 2007(15):1-8.
- [4] 胡心瀚, 叶五一, 缪柏其. 基于 COPULA-ACD 模型的股票连涨和连跌收益率风险分析 [J]. 系统工程理论与实践, 2010(30):298-305.
- [5] 谢中华. MATLAB 统计分析与应用:40 个案例分析 [M]. 北京: 北京航空航天大学出版社, 2010.
- [6] 雷鸣, 谭永春. 运用生存分析与变点理论对上证成指的研究 [J]. 中国管理科学, 2007(5):1-8.
- [7] 任仙玲, 张世英. 基于非参数核密度估计的 Copula 函数选择原理 [J]. 系统工程学报, 2010(25):36-42.
- [8] 张连增, 胡祥. Copula 的参数估计与半参数估计方法的比较 [J]. 统计研究, 2014(31):91-95.