

基于Copula函数的汽车与钢铁股票价格相关性分析

严彦文, 李瑞宁

中国石油大学(北京)理学院, 北京

收稿日期: 2024年3月25日; 录用日期: 2024年4月30日; 发布日期: 2024年5月28日

摘要

本文基于Copula函数理论, 研究了我国汽车与钢铁股票价格之间的相关性。在边缘分布的确定上, 采取非参数核密度法估算汽车与钢铁股票的边缘分布函数; 在Copula函数的参数估计上, 通过极大似然法对引入的五种不同Copula函数进行参数估计; 在函数模型的选择上采用欧式平方距离法则和AIC、OLS最小法则相结合的方法进行优选。研究表明t-Copula函数模型能够更好地描绘汽车-钢铁之间的相关结构, 两者存在中等程度的正相关性, 且上下尾相关系数相等。

关键词

Copula模型, 核密度估计, 统计推断, 相关性

Correlation Analysis of Automobile and Steel Stocks Prices Based on Copula Function

Yanwen Yan, Ruining Li

College of Science, China University of Petroleum, Beijing

Received: Mar. 25th, 2024; accepted: Apr. 30th, 2024; published: May 28th, 2024

Abstract

The study is based on Copula function theory and investigates the correlation between automobile and steel stock prices in China. Determining the marginal distribution using non-parametric kernel density estimation for automobile and steel stock returns; Estimating parameters for five different Copula functions introduced through maximum likelihood estimation; Selecting the functional model using a combination of Euclidean distance method and AIC, OLS minimization method.

The research indicates that the t-Copula function model can better depict the correlation structure between automobiles and steel, showing a moderate level of positive correlation between the two, with equal upper and lower tail correlation coefficients.

Keywords

Copula Model, Kernel Density Estimation, Statistical Inference, Correlation

Copyright © 2024 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

随机变量相依性的度量问题一直是统计的热点问题,多年来广泛应用于各个领域。在传统的概率统计学中,相关性度量多采用皮尔逊相关系数 ρ 、肯德尔秩相关系数 τ 和斯皮尔曼秩相关系数 ρ_s 。但这要求随机变量的方差保持在特定的区间内,所描绘相关关系的类型也只适用于线性关系或单调关系。针对上述方法的不足, Copula 函数理论逐渐发展完善,它可以精准地描绘变量间的相依结构,并且可以用于各种类型的变量,包括连续型、离散型和混合型变量。随着经济一体化全球化的发展,资产结构也日益复杂,一个行业市场的波动都可能会引起另一个或多个市场的波动。投资回报和投资选择是每一位投资者密切且直接关注的重点,研究金融资产的收益率显得尤为重要。但金融收益率的序列通常不满足正态假设,各资产收益率的相关性也并不是线性的,彼此之间的复杂相关性并不能简单地由相关系数度量,而 Copula 函数恰能解决这类问题。通过深入地研究并把握各行业股票之间的相关性以及尾部相关性,以规避风险,合理投资,减少损失。

近年来在社会不断发展的同时,大量国民开始涌入金融市场,而现实中金融市场正是糅合了非对称性、非线性、时变性等多特征的复杂系统。恰好 Copula 函数理论能够更完美地刻画资产收益率的真实分布,有效减少因为强行捕捉多个序列之间的相关关系而造成的谬论。所以随着金融市场的不断扩大, Copula 函数广泛应用在金融领域。而汽车和钢铁作为我国重要的两支关乎民生和经济发展的股票,其相关结构和相关性的研究较少。本文构建 Copula 模型对汽车与钢铁股票进行实证分析,通过非参数核密度法确定边缘分布、对不同的 Copula 模型进行参数估计,比较五种 Copula 模型的拟合效果,进行拟合优度检验。最后依据所选择的拟合效果最优的 Copula 模型计算相关性系数以及尾部相关系数。研究汽车与钢铁股票价格之间的相依性,从相依性结构,整体相依性以及尾部相依性进行了定量分析。实证相关 Copula 函数理论的同时,研究汽车与钢铁行业股票间的相依性,根据结果进行分析,为我国两支重要股票给出合理的投资建议,以减少损失和风险,获得最大的收益。

2. Copula 函数模型的理论基础

2.1. Sklar 定理

Sklar 定理[1]: 假设 F 为 n 维联合分布函数, F_1, F_2, \dots, F_n 分别为 n 个边缘分布函数,则存在一个 n 元的连接函数 $C(u_1, u_2, \dots, u_n)$, 这个连接函数叫做 Copula 函数,使得:

$$F(x_1, x_2, \dots, x_n) = C[F_1(x_1), F_2(x_2), \dots, F_n(x_n)] \quad (1)$$

由上式也可看出, Copula 函数是将 $F(x_1, x_2, \dots, x_n)$ 剔除边缘分布信息之后, 余留下来的重要信息, 涵盖了变量之间的联系, 通常被称为变量间的相依结构。且分布函数 F 的密度函数满足下式:

$$f(x_1, x_2, \dots, x_n) = c[F_1(x_1), F_2(x_2), \dots, F_n(x_n)] \prod_{i=1}^n f_i(x_i) \quad (2)$$

其中, $c(u_1, u_2, \dots, u_n) = \frac{\partial C(u_1, u_2, \dots, u_n)}{\partial u_1 \partial u_2 \dots \partial u_n}$, f_n 是边缘分布 F_n 的密度函数。

Sklar 定理的提出具有十分重要的理论意义, 为统计领域提供了一种新的求取联合分布的方法。并且对于连续的边缘分布, 只能得到唯一的连接函数。

2.2. Copula 函数

从定义来看, Copula 函数具有种类丰富的不同形式。本文研究其两个分布族: 椭圆分布族、Archimedean 分布族[2] [3]。并针对这两个分布族, 我们选择了其中的五个 Copula 函数作为建立模型的基础。

2.2.1. 椭圆 Copula 函数族

(1) Gaussian Copula 函数

高斯 Copula 是边缘分布为正态分布的连接函数, 对应的分布函数和密度函数表达式为:

$$C(u_1, \dots, u_n, \rho) = \Phi_\rho(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n)) \quad (3)$$

$$c(u_1, \dots, u_n; \rho) = |\rho|^{-1/2} \exp\left(-\frac{1}{2} \zeta'(\rho^{-1} - I) \zeta\right) \quad (4)$$

其中 ρ 为 n 阶对称正定矩阵, Φ_ρ 表示相关系数矩阵为 ρ 的 n 元标准正态分布的分布函数; I 表示单位矩阵, $\zeta' = (\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n))$ 。

设相关系数为 ρ , 则二元函数如下:

$$C^{Ga}(u, v, \rho) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)}\right) ds dt \quad (5)$$

正态 Copula 函数取样方便, 参数数量较少, 函数分布性质易掌握模拟。

(2) t-Student Copula 函数

t-Copula 指多元 t-Student 分布的连接函数, 若 $u_i = t_k(X_i \sqrt{k} / \sqrt{Y})$ 服从 t 分布, 则自由度为 k 的多维 t-Copula 函数分布函数和密度函数为:

$$C(u_1, \dots, u_n, \rho, k) = t_{\rho, k}(t_k^{-1}(u_1), \dots, t_k^{-1}(u_n)) \quad (6)$$

$$c(u_1, \dots, u_n; \rho, k) = |\rho|^{-1/2} \frac{\Gamma\left(\frac{k+n}{2}\right) \left[\Gamma\left(\frac{k}{2}\right)\right]^{n-1} \left(1 + \frac{1}{k} \zeta' \rho^{-1} \zeta\right)^{\frac{k+n}{2}}}{\left[\Gamma\left(\frac{k+1}{2}\right)\right]^n \prod_{i=1}^n \left(1 + \frac{\zeta_i^2}{k}\right)^{\frac{k+1}{2}}} \quad (7)$$

其中 $t_{\rho, k}$ 表示相关系数矩阵为 ρ , 自由度为 k 的 n 元标准 t 分布的分布函数, $\zeta' = (t_k^{-1}(u_1), \dots, t_k^{-1}(u_n))$ 。

对于二元情况, 存在如下函数形式, 其中 ρ 为变量间线性相关系数。

$$C^t(u, v, \rho, k) = \int_{-\infty}^{t_k^{-1}(u)} \int_{-\infty}^{t_k^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{s^2 - 2\rho st + t^2}{k(1-\rho^2)}\right)^{-(k+2)/2} ds dt \quad (8)$$

t-Copula 具有较厚的对称的尾部, 这使得我们能够通过该函数模型捕捉到尾部的变化, 且尾部相关性随着自由度 ρ 的增大而增强。

2.2.2. Archimedean Copula 函数

Archimedean Copula 函数具有分布类型众多、方便计算、可结合性的优良性质, 蕴含了良好的统计特性, 也因此获得了广泛的应用, 常被用在研究金融资产的厚尾性, 可结合性等。阿基米德 Copula 函数的定义如下:

$$C(u_1, u_2, \dots, u_n) = \begin{cases} \varphi^{-1}(\varphi(u_1), \dots, \varphi(u_n)), & \sum_{i=1}^n \varphi(u_i) \leq \varphi(0) \\ 0, & \text{其他} \end{cases} \quad (9)$$

其中, 函数 $\varphi(u)$ 称为阿基米德 Copula 函数的生成元, 满足 $\varphi(1) = 0$, 且对于所有属于 $[0, 1]$ 的 u , 有 $\varphi'(t) < 0, \varphi''(t) > 0$, 即生成元 $\varphi(u)$ 是一个凸的减函数。

下面引入本文应用到的二元阿基米德 Copula 函数。

(1) Gumbel Copula 函数

生成元函数 $\varphi(t) = (-\ln t)^\alpha$, 其分布函数为:

$$C_G(u, v, \alpha) = \exp\left\{-\left((-\ln u)^\alpha + (-\ln v)^\alpha\right)^{1/\alpha}\right\} \quad (10)$$

其中 α 表示该 Copula 函数中的参数, 当 α 趋于无穷大时, u, v 完全相依, 即 $C_G(u, v, \alpha) = \min(u, v)$ 。

Gumbel Copula 函数上尾高于下尾, 不具有对称性。此函数模型说明上尾部存在较强的相关性, 下尾部渐进独立。

(2) Clayton Copula 函数

生成元为 $\varphi(t) = (t^{-\alpha} - 1)/\alpha$, 分布函数、密度函数如下:

$$C_C(u, v, \alpha) = \max\left(\left(u^{-\alpha} + v^{-\alpha} - 1\right)^{-1/\alpha}, 0\right) \quad (11)$$

$$c(u, v, \alpha) = (1 + \alpha)(uv)^{-\alpha-1} \left(u^{-\alpha} + v^{-\alpha} - 1\right)^{-2-1/\alpha} \quad (12)$$

其中 α 表示函数中的参数, $\alpha \in (-1, \infty), \alpha \neq 0$ 。当 θ 趋于零时, 随机变量 u, v 趋向于独立, 即 $C_C(u, v, \alpha) = uv$, 当 α 趋向于正无穷大时, u, v 趋向于完全相关, $C_C(u, v, \alpha) = \max(u, v)$ 。

Clayton Copula 函数下尾高, 上尾低。能够捕捉到下尾相关的变化。所以在实际中, 可以较为完美地反映具有明显下尾相关关系的资产变量。

(3) Frank Copula 函数

生成元 $\varphi(t) = -\ln \frac{e^{-\alpha t} - 1}{e^{-\alpha} - 1}$, 其分布函数为:

$$C_F(u, v, \alpha) = -\frac{1}{\alpha} \ln \left(1 + \frac{(e^{-\alpha u} - 1)(e^{-\alpha v} - 1)}{e^{-\alpha} - 1} \right) \quad (13)$$

其中, α 作为参数, 取值范围为 $\alpha \in (-\infty, \infty), \alpha \neq 0$ 。当 $\alpha > 0$, 随机变量 u, v 正相关, 反之, 存在负相关; 若参数无限趋向于零, 即 $C_F(u, v, \alpha) = uv$ 。

Frank Copula 函数具备对称相关模式, 且对上下尾部均不敏感。难以捕捉非对称二元变量的相依性和尾部相关性。该模型适用于具有对称尾部且尾部渐进独立的二元随机变量。

2.3. 边缘分布的拟合和 Copula 函数的参数估计

构建 Copula 模型的步骤为确定边缘分布和估计 Copula 函数的参数。在确定随机变量分布的研究中, 一般分为参数法和非参数法。参数法具有一定的主观性, 可能存在较大误差。非参数法引入经验分布函数和核分布, 这种方法适用于边缘分布难以确定的类型。

2.3.1. 非参数核密度估计法

设 x_1, x_2, \dots, x_n 是总体 X 的样本, 其总体密度函数 $f(x)$ 的核估计为[3] [4]:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \quad (14)$$

其中 $K(\cdot)$ 称为核函数, h 称为窗宽, 核密度估计的核心就是通过确立核函数和最优窗宽来对变量的样本值的密度函数进行估计。事实上, 根据文献调研和学者的大量研究已经发现核函数的选择在样本值足够大时, 影响微乎其微; 但是窗宽 h 会影响 $\hat{f}_h(x)$ 的光滑程度。在确定最优窗宽时, 依据积分均方误差 MISE:

$$MISE(\hat{f}_h) = E\left[\int(\hat{f}_h(x) - f(x))^2 dx\right] \quad (15)$$

核函数选定时, 选择使 MISE 值最小时的 h 作为最优窗宽。一般实际应用中, 时常将高斯函数作为核函数, 即 $K(x) = (1/\sqrt{2\pi})e^{-\frac{x^2}{2}}$, 有最佳窗宽为[3]:

$$\hat{h} = (4/3)^{1/5} \sigma n^{-1/5} \quad (16)$$

假设 $f(x), g(y)$ 分别是汽车和钢铁收益率的密度函数, $F(x), G(y)$ 为其边缘分布函数。其中 σ 由样本的标准差 S 来代替, n 为数据长度。由此得到 $f(x), g(y)$ 的非参数核密度估计为:

$$\hat{f}(x) = \frac{1}{n\hat{h}_x} \sum_{i=1}^n K\left(\frac{x-x_i}{\hat{h}_x}\right), \hat{g}(y) = \frac{1}{n\hat{h}_y} \sum_{i=1}^n K\left(\frac{y-y_i}{\hat{h}_y}\right) \quad (17)$$

收益率序列 (x, y) 在某一时刻的分布函数值为:

$$\hat{u} = \hat{F}_h(x_t) = \int_{-\infty}^{x_t} \hat{f}(x) dx, \hat{v} = \hat{G}_h(y_t) = \int_{-\infty}^{y_t} \hat{g}(y) dy \quad (18)$$

2.3.2. 基于极大似然估计对 Copula 函数的参数进行估计[5]

构建 Copula 模型时主要分为两步, 包括拟合边际分布和 Copula 函数参数估计。基于 2.1 中所介绍的 Copula 函数形式 $C(u, v, \alpha)$, $u = F(x), v = G(y)$ 。设 (X, Y) 的联合分布函数为:

$$H(x, y; \alpha) = C(F(x; \theta_1), G(y; \theta_2); \alpha) \quad (19)$$

其中参数 α 为 Copula 函数中的待估计参数, θ_1, θ_2 为边缘分布的参数。联合密度函数为:

$$h(x, y; \alpha) = c(F(x; \theta_1), G(y; \theta_2); \alpha) f(x; \theta_1) g(y; \theta_2) \quad (20)$$

$(X_i, Y_i), i = 1, 2, \dots, n$ 的似然函数为:

$$L(\alpha) = \prod_{i=1}^n h(x_i, y_i; \alpha) \quad (21)$$

求解得到各参数的估计值:

$$\hat{\alpha} = \arg \max \ln L(\alpha) \quad (22)$$

实际情况中边缘分布有时难以确定, 且估计分布函数中的参数计算量较大。而基于非参数核密度法确定边缘分布函数, 随后再采用极大似然法估计 Copula 函数中的参数这种方法因计算过程较为简便, 且能够充分利用样本值而受到青睐。

此时采用核函数 $F_h(x)$, $G_h(y)$ 来代替边缘分布函数 $F(x, \theta_1)$ 、 $G(y, \theta_2)$, 只需估计 Copula 函数中参数即可[6], 代入伪似然函数如下即可得到各 Copula 函数的参数:

$$\hat{\alpha} = \arg \max \sum_{i=1}^n \ln c(F_h(x_i), G_h(y_i); \alpha) \quad (23)$$

2.4. Copula 函数的模型检验与优选

基于随机变量的样本值我们得到了各 Copula 模型的参数估计值, 那么对模型优劣的评价就要引入经验 Copula。

首先假定随机变量 X, Y 分布函数为 $F_n(x)$ 、 $G_n(y)$, 则经验 Copula 如下[7]:

$$\hat{C}_n(u, v) = \frac{1}{n} \sum_{i=1}^n I_{[F_n(x_i) \leq u]} I_{[G_n(y_i) \leq v]}, \quad u, v \in [0, 1] \quad (24)$$

其中 I 为示性函数, $F_n(x_i) \leq u, I_{[F_n(x_i) \leq u]} = 1$, 否则等于 0。

2.4.1. 欧式平方距离最小法

以二元经验 Copula 函数为参照, 计算估计得到的 Copula 函数模型值与经验 Copula 函数值的欧式平方距离。欧式平方距离 d^2 越小, 所选的 Copula 模型拟合数据效果越好, 所选模型能够更好地反应随机变量间的相关结构。欧式距离公式如下:

$$d_w^2 = \sum_{i=1}^n \left| \hat{C}_n(u_i, v_i) - \hat{C}_w(u_i, v_i) \right| \quad (25)$$

其中 W 代表所估计的 Copula 函数类型, u_i, v_i 代表各边缘分布函数的值。

2.4.2. AIC、OLS 最小准则[7]

以其中 m 为参数数量, n 为样本数量。AIC 值通常小于零, 当 AIC 和 OLS 值均最小时, 拟合优度最佳。

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n \left(\hat{C}_n(u_i, v_i) - \hat{C}_w(u_i, v_i) \right)^2 \quad (26)$$

$$\text{AIC} = n \log(\text{MSE}) + 2m \quad (27)$$

$$\text{OLS} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\hat{C}_n(u_i, v_i) - \hat{C}_w(u_i, v_i) \right)^2} \quad (28)$$

2.5. Copula 函数与相关性测度之间的关系

Copula 函数可以用于描述多元随机变量之间的依赖关系, 且与 Spearman 相关系数、Kendall 秩相关系数以及尾部相关系数具有一定的解析关系。此外, 相关性测度的使用通常局限于较小的样本和某些特定类型的序列, 但 Copula 函数可以适用于更广泛的数据类型和样本大小。下面着重介绍 Copula 函数与相关性测度之间的关系, 各相关测度的定义不做赘述, 详见参考文献[8]。

Kendall 秩相关系数度量了变量间变化的一致程度, 若有连续二维随机变量 (X, Y) , 设其存在 $C(u, v)$, 则满足[9]:

$$\tau = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1 \tag{29}$$

Spearman 相关系数将变量的数值进行排序, 衡量了两个变量的等级顺序之间的相关性。若连续二维随机向量 (X, Y) , 设其存在 $C(u, v)$, 则满足[9]:

$$\rho_s = 12 \int_0^1 \int_0^1 C(u, v) dudv - 3 \tag{30}$$

尾部相关性刻画了一个随机变量取极值时, 对另一个变量的取值产生影响的概率, 这对资产组合和分配既有重要意义。对于随机向量 (X, Y) , 相对应的 Copula 函数为 $C(u, v)$, 则上、下尾部相关系数满足[2] [9]:

$$\lambda_{up}(X, Y) = \lim_{u \rightarrow 1^-} \frac{1 - 2u + C(u, u)}{1 - u} \tag{31}$$

$$\lambda_{low}(X, Y) = \lim_{u \rightarrow 0^+} \frac{C(u, u)}{u} \tag{32}$$

3. 实证分析

3.1. 数据选取与处理

基于申万行业分类标准, 将钢铁行业、汽车整车行业的收盘指数 P_t 作为数据来源, 整理编制出对数日收益率 R_t 的相关数据:

$$R_t = 100(\ln P_t - \ln P_{t-1})$$

共选取了 2020/1/1~2022/12/31 的收盘指数数据, 编制得到 718 组有效数据。对所编制的收益率序列进行描述性分析(表 1), 并进行正态性检验。频率直方图(图 1)以及偏度和峰度反应出的信息可以初步判断收益率序列的分布总体呈现出尖峰厚尾的特点。通过 QQ 图(图 2)也可看出有相当一部分的样本点未分布在直线附近。进一步通 K-S, J-B 进行正态性检验, p 值小于 0.05, 说明两个收益率序列均不服从正态分布。

3.2. 基于非参数核密度法确定序列边缘分布函数

基于非参数核密度法估计两收益率序列的边际密度函数。计算得到基于正态核函数, 各序列的最优窗宽 h_x 、 h_y 。基于式(18)由 MATLAB 计算得到的经验分布值与核分布值。其值虽并不完全相同, 但汽车、钢铁日收益率的经验分布函数图与核分布函数图几乎完全重合(图 3)。由此可见核密度估计法估计总体分布的优良性与可行性。

Table 1. Descriptive statistics
表 1. 描述性统计

股票	均值	最小值	最大值	峰度	偏度	K-S 检验	J-B 检验
汽车	0.0407	-9.6910	5.3782	4.5095	-0.3213	$h = 1;$ $p = 0.0462$	$h = 1;$ $p = 1.00e-03$
钢铁	0.0085	-10.2979	7.3227	5.7676	-0.3023	$h = 1;$ $p = 0.00098$	$h = 1;$ $p = 1.00e-03$

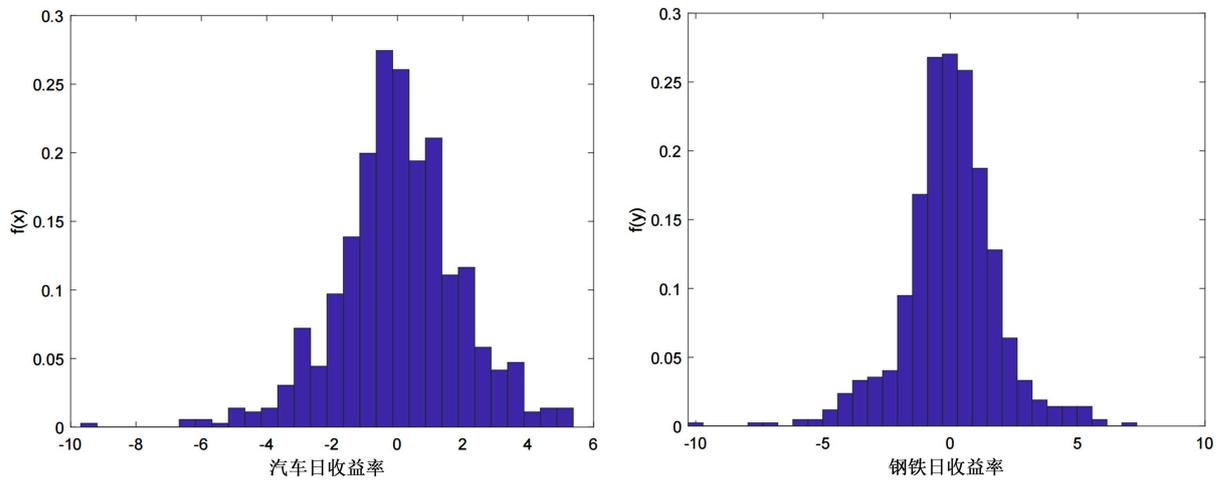


Figure 1. Frequency histogram

图 1. 频率直方图

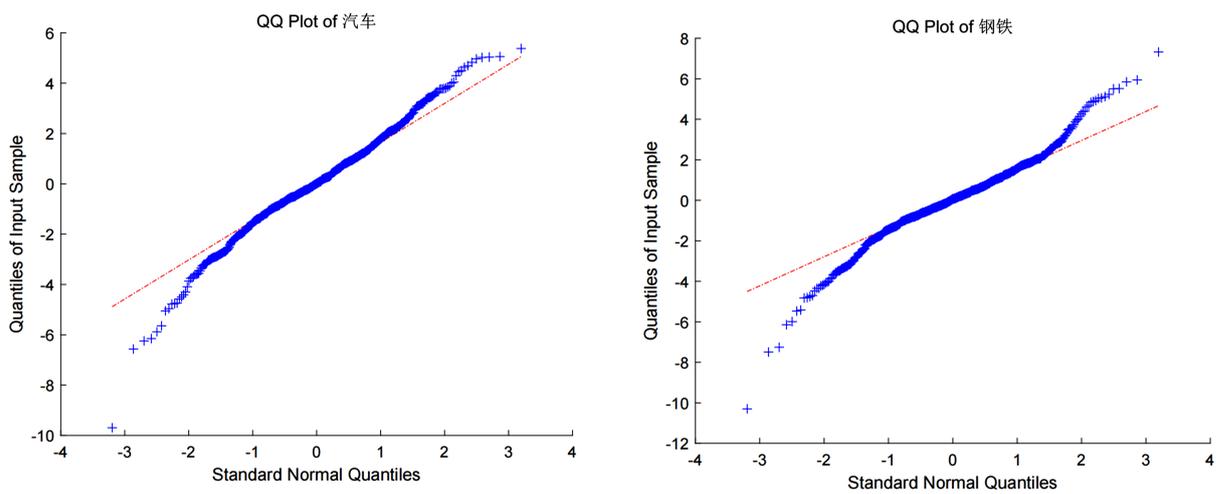


Figure 2. QQ diagram

图 2. QQ 图

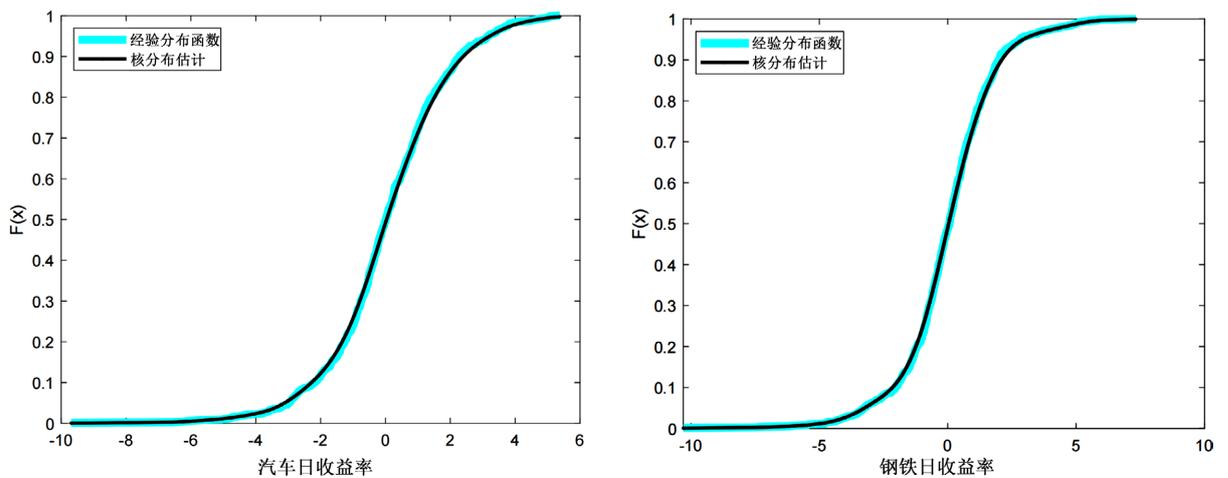


Figure 3. Kernel distribution estimate and empirical distribution chart of daily return rate of automobile, steel

图 3. 汽车和钢铁产日收益率核分布估计图和经验分布图

在确定汽车和钢铁收益率的边缘分布后, 得到新的分布序列 (U_x, V_y) , 对新的序列进行 K-S 检验, 检验结果均为 $h = 0$, 且 p 值小于 0.05, 则无法拒绝原假设即服从均匀分布, 满足用 Copula 函数连接成联合分布的要求。以非参数核密度估计得到各收益率序列的边缘分布, 避免了全参数估计法中主观判断样本分布类型的不准确性, 同时又使得建模过程完全基于样本数据本身。

3.3. Copula 模型求解与检验

3.3.1. Copula 函数模型的参数估计与相关性测度计算

基于核密度法得到的分布函数值, 采用极大似然法估计五种 Copula 函数中的未知参数; 并依据 Copula 函数与相关性测度直接的关系得到各测度值, 结果如下表 2 所示。

Table 2. Copula function parameter estimates and correlation measures

表 2. Copula 函数参数估计值及相关性测度值

	Gaussian	t	Gumbel	Clayton	Frank
参数	0.4653	$\rho = 0.4879, \nu = 5.3955$	1.4121	0.8193	3.2882
Kendall	0.3081	0.3245	0.2918	0.2906	0.3319
Spearman	0.1495	0.4614	0.421	0.4213	0.4824
λ^{up}	0	0.1884	0.3662	0	0
λ^{low}	0	0.1884	0	0.4291	0

从各 Copula 函数得到的相关性测度可以初步看出汽车与房地产之间存在中等强度的正相关性。从尾部相关系数也可以看出高斯 Copula、t-Copula 以及 Frank Copula 函数具备对称相关模式, 且高斯 Copula 和 Frank Copula 对上下尾部均不敏感。难以捕捉非对称二元变量的相依性和尾部相关性。Gumbel Copula 函数说明上尾部存在较强的相关性即当房地产股票较高时, 汽车的股票价格有 37% 的概率也较高, 但下尾部渐进独立。Clayton Copula 函数模型则说明两组股票之间存在较强的下尾相关性, 即当房地产股票较低时, 汽车的股票价格有 43% 的概率也较低, 但上尾渐进独立。

3.3.2. 最优 Copula 函数的选择

首先通过绘制二维分布直方图初步筛选可能拟合的 Copula 函数, 通过图 4 可以看出房地产与汽车股票价格之间存在尾部对称性, 则可以初步筛选出高斯 Copula、t-Copula 以及 Frank Copula 函数三个模型; 进一步通过二维分布直方图可观察到尾部具有一定厚度, 存在尾部相关性, 则 t-Copula 可能为最佳模型。

但二维分布直方图只能进行定性分析, 并不能定量地抉择出最优的 Copula 函数模型。所以我们以经验 Copula 为参照, 采取欧式距离最小法和 AIC、OLS 最小准则定量计算出对房地产与钢铁股票价格之间联合分布拟合效果最好的 Copula 模型(表 3)。

欧式平方距离和 AIC、OLS 反映了二元 Copula 模型对样本数据拟合情况的优劣, 欧式平方距离和 AIC、OLS 值越小, 拟合情况越好, 该 Copula 模型越能描述房地产与汽车股票收益率之间的联合分布, 即相关结构。由表 3 可以看出二元 t-Copula 模型的欧式平方距离和 AIC、OLS 值均最小, 能够更好地拟合汽车 - 钢铁收益率的样本数据。

此时我们通过定性定量分析, 对拟合效果优劣进行检验, 均发现对于所选取的收益率序列, t-Copula 函数是表征汽车 - 钢铁股票收益率相关结构最好的函数模型。

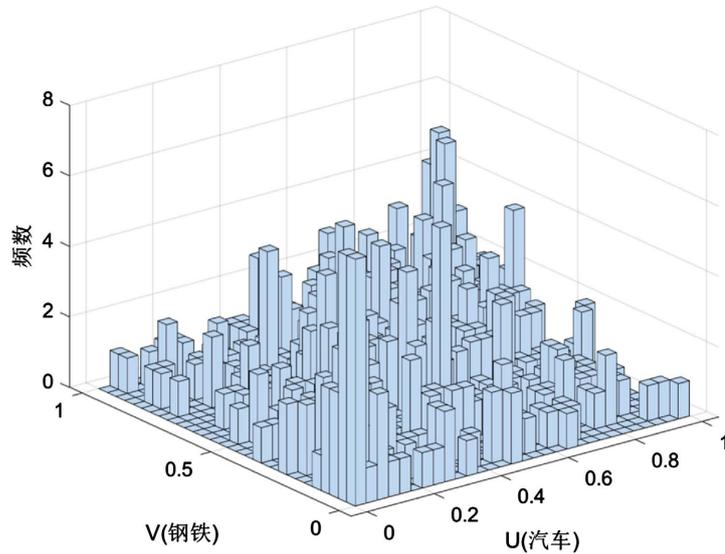


Figure 4. The histogram of daily returns for automobile and steel production
图 4. 汽车和钢铁产日收益率直方图

Table 3. The parameter estimator of Copula function and the test value
表 3. Copula 函数参数估计值及检验值

	Gaussian	t	Gumbel	Clayton	Frank
参数	0.4653	$\rho = 0.4879, \nu = 5.3955$	1.4121	0.8193	3.2882
d^2	0.0487	0.0419	0.1319	0.0459	0.0792
AIC	-6.8898e+03	-6.9957e+03	-6.1744e+03	-6.9323e+03	-6.5406e+03
OLS	0.0082	0.0076	0.0136	0.0080	0.0105

3.3.3. 相关性分析

将 t-Copula 模型的参数代入式(8)可以得到汽车 - 钢铁收益率之间的相关结构为:

$$\hat{C}_t(u, v) = \int_{-\infty}^{t_5^{-1}(\hat{u})} \int_{-\infty}^{t_5^{-1}(\hat{v})} \frac{1}{2\pi\sqrt{1-0.4879^2}} \left(1 + \frac{s^2 - 2*0.4879st + t^2}{5*(1-0.4879^2)} \right)^{-(5+2)/2} dsdr$$

对应的二元 t-Copula 函数模型的密度函数图和分布函数图如下。

图 5 表征刻画了两支收益率之间的相关结构、整体相关性。接下来进一步分析由 t-Copula 函数得到以下相关性测度的含义。

根据相关性测度(表 4)和上图得到的汽车 - 钢铁的联合概率密度以及联合分布函数图像可以看出, 汽车与钢铁之间存在中等程度的正相关性。同时汽车与钢铁收益率之间存在对称的尾部相关性, 即当房地产股票发生极端上涨或下跌时, 汽车股票也有将近 20% 的可能性发生上涨或下跌。

4. 结论

自由度为 5, 参数为 0.4879 的 t-Copula 函数模型能够更好地描绘汽车 - 钢铁收益率的相关结构, 且根据相关系数, 发现两者呈现中等强度的正相关性。上尾、下尾相关系数均为 0.1884, 即汽车股票收益率的剧烈上涨或下跌也会很可能引发钢铁股票收益的上涨或下跌。投资者可以同时参考消费者信心, 税

收政策, 宏观经济环境, 交通建设、城市化进程等影响因素, 在形势较好的情况下同时购入这两支股票, 以实现利益最大化。

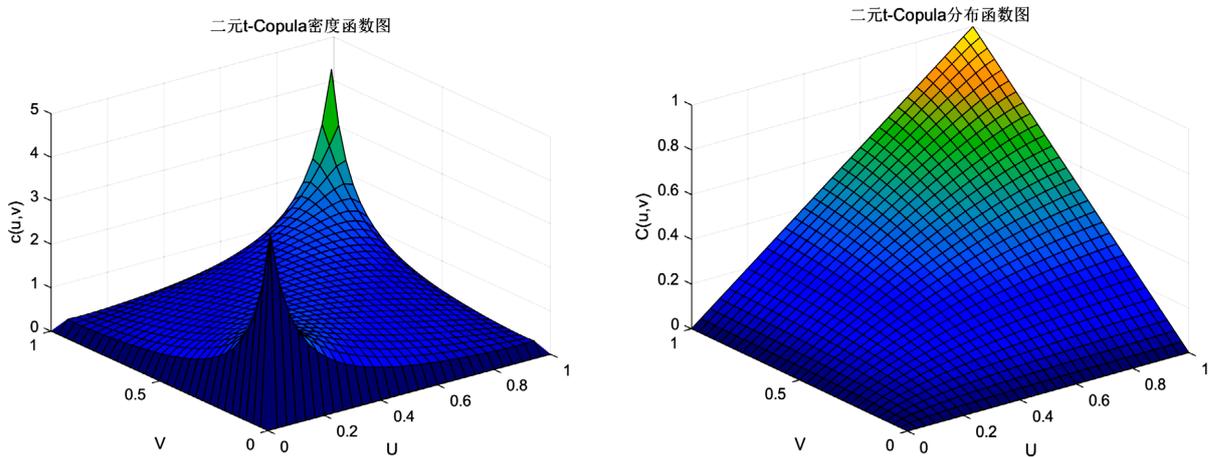


Figure 5. Binary t-Copula density function and distribution function diagram

图 5. 二元 t-Copula 密度函数和分布函数图

Table 4. t-Copula correlation measure

表 4. t-Copula 的相关性测度

	最佳 Copula 函数	Kendall	Spearman	λ^{up}	λ^{low}
汽车 - 钢铁	t-Copula	0.3245	0.4614	0.1884	0.1884

总之 Copula 函数能够描述和捕捉各个维度之间的非线性相关关系。其优势主要体现在：对数据分布没有假设限制；能够处理极端事件和崩盘情况，从而更准确地评估风险。其次在 Copula 模型的边缘分布的确定和参数估计中，假定分布类型进行参数估计存在一定的主观性，所得到的分布拟合效果很可能不尽如人意。那么对于边缘分布难以确定的情况下，基于核密度估计的估计法则更具有优势，主要体现在：更加灵活，不需要假设其边缘分布的类型；避免了过度拟合，减少了参数数量，简便了运算。但是 Copula 函数模型也存在一定的不足，对 Copula 函数种类的选取只采取了常见的五种函数形式，人为地限制了样本数据间存在的相关结构类型，比如正态 Copula 函数以及 Frank Copula 函数具有对称性，但尾部渐进独立，难以捕捉非对称二元变量的相依性和尾部相关性。Gumbel Copula 函数上尾高于下尾，不具有对称性。此函数模型说明上尾部存在较强的相关性，但下尾部渐进独立。Clayton Copula 函数下尾高，上尾低。能够捕捉到下尾相关的变化，上尾相关变化难以获取。所以面对更复杂的数据结构，可能存在更加合适的 Copula 函数类型，比如复合 Copula 函数。

基金项目

中国石油大学(北京)“1158 工程项目”。

参考文献

- [1] Nelsen, R.B. (1998) An Introduction to Copulas. Springer, New York, 7-28.
- [2] 荀鑫垚. 世界主要汽车市场间的风险相依性及风险溢出效应研究[D]: [硕士学位论文]. 大连: 东北财经大学, 2022.
- [3] 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: 179-181.

- [4] 陈希镇, 胡兆红. Copula 函数的非参数核密度估计方法[J]. 统计与决策, 2010(14): 27-28.
- [5] 邱小霞, 刘次华, 吴娟. Copula 函数中参数极大似然估计的性质[J]. 经济数学, 2008, 25(2): 210-215.
- [6] 陈希镇. 现代统计分析方法的理论和应用[M]. 北京: 国防工业出版社, 2016: 91-100.
- [7] 郭家骥. 基于 Copula 理论承德市降水与温度相关性量化研究[J]. 水科学与工程, 2020(2): 5-10.
- [8] 董智前, 李星野. Copula 函数在金融市场中的应用[J]. 数学理论与应用, 2016, 36(4): 106-115.
- [9] 吴娟. Copula 理论与相关性分析[D]: [博士学位论文]. 武汉: 华中科技大学, 2009.