

基于改进的聚类和核密度估计的 分布鲁棒均值-CVaR投资组合优化

于疆¹, 李永明^{2*}

¹东华理工大学理学院, 江西 南昌

²上饶师范学院数学与计算科学学院, 江西 上饶

收稿日期: 2024年8月10日; 录用日期: 2024年9月2日; 发布日期: 2024年9月12日

摘要

在金融市场中, 如何构建最优投资组合来平衡风险和回报是当今研究者所面临的主要问题之一。为了构建最优投资组合, 研究者们通常使用的是VaR或CVaR模型。本研究通过综合运用聚类、核密度估计以及分布鲁棒均值-CVaR模型的方法, 从而达到提升股票投资组合的构建和风险管理能力的目的。本文考虑了包含100只股票日收益数据的实验数据集, 通过优化聚类方法, 利用核密度估计确定了K-means算法的最佳聚类中心和k值选取。随后, 将聚类后的数据输入核密度估计的分布鲁棒均值-CVaR模型中进行分析。通过窗口滚动实验, 比较了在有聚类条件下模型对投资组合收益率的影响。结果显示, 应用聚类方法后的模型具有更高的投资组合收益率, 有助于投资者更好地平衡风险与回报, 构建最优的投资组合。

关键词

K-Means, 核密度估计, 分布鲁棒优化, CVaR

Robust Mean-CVaR Portfolio Optimization Based on Enhanced Clustering and Kernel Density Estimation

Jiang Yu¹, Yongming Li^{2*}

¹College of Science, East China University of Technology, Nanchang Jiangxi

²School of Mathematics and Computer Science, Shangrao Normal University, Shangrao Jiangxi

Received: Aug. 10th, 2024; accepted: Sep. 2nd, 2024; published: Sep. 12th, 2024

Abstract

In financial markets, how to construct an optimal investment portfolio that balances risk and return

*通讯作者。

文章引用: 于疆, 李永明. 基于改进的聚类和核密度估计的分布鲁棒均值-CVaR 投资组合优化[J]. 应用数学进展, 2024, 13(9): 4099-4107. DOI: 10.12677/aam.2024.139391

is one of the main challenges faced by researchers today. To build an optimal portfolio, researchers typically use VaR or CVaR models. This study aims to enhance the construction of stock portfolios and risk management capabilities by comprehensively utilizing methods such as clustering, kernel density estimation, and distributionally robust mean-CVaR models. The paper utilized an experimental dataset containing daily returns of 100 stocks. By optimizing clustering methods and determining the optimal clustering centers and k values of the K-means algorithm using kernel density estimation, we then input the clustered data into the robust mean-CVaR model for analysis. By rolling window experiments, we compared the impact of the model on portfolio returns with and without clustering conditions. The results show that the model with clustering methods applied has higher portfolio returns, helping investors better balance risk and return to construct optimal portfolios.

Keywords

K-Means, Kernel Density Estimation, Distribution Robust Optimization, CVaR

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

近年来, 随着金融市场的不断发展和变化, 投资者对于构建最优的投资组合以平衡风险与回报的需求日益迫切。为了解决这一问题, 研究者们开始探索各种方法和模型来优化投资组合的构建过程。投资组合优化问题涉及在资产和风险之间进行权衡, 通过在不同资产上进行投资以实现利益最大化, 形成最佳的投资组合, 这是当前面临的核心挑战。

最初, 风险的衡量方法由 Markowitz (1952) [1]提出的方差理论开始, 许多学者都在该理论上展开了深入研究。然而, 方差理论只考虑了资产收益的波动性, 不能全面考虑投资者对不同类型风险的偏好和敏感度, 因此没有充分实现投资者最佳投资组合的目标。之后, Konno 等(1993) [2]提出了均值 - 绝对偏差模型, 以绝对偏差而非方差来衡量风险。然而, 这种方法对尾部风险的度量能力有限。于是, Markowitz (1959) [3]提出了半方差来度量风险。随后, VaR (Value at Risk)模型和 CvaR (Conditional Value at Risk)模型成为常用的风险度量方法之一。Gaivoronski & Pflug (1999) [4]将 VaR 引入投资组合中进行建模求解, 然而, VaR 有一些固有的局限性。它只关注损失超过某个临界值的部分, 而不考虑超过该临界值后的损失情况, 导致一些实际情况并不能考虑充分。同时, VaR 在极端市场情况下的表现不够稳健, 因为它通常假设损失的分布是固定的。所以后来, Rockafellar & Uryasev (2000) [5] [6]研究了 CVaR, 并指出 VaR 不满足次可加性的性质。但它的计算较为复杂, 尤其是当涉及到大规模投资组合或高维数据时。这种复杂性导致计算和优化难度增加, 从而在实际应用中变得不够高效。同时, 它的准确性高度依赖于输入数据的质量和准确性。如果数据存在噪声或不完全, 会影响 CVaR 的计算结果, 从而影响投资决策的有效性。在国内, 牛昂(1997) [7]研究了 VaR 的计算方法, 并对其进行了比较, 然而, 这种计算方法依赖于对收益分布的特定假设, 但在实际的金融市场中, 资产收益的分布可能呈现更复杂的特征, 如厚尾分布或非对称分布等, 这些都可能影响 VaR 的准确性。随后, 姚刚(1998) [8]、刘宇飞(1999) [9]、王春风(2000) [10]等文献提出了改进 VaR 计算方法的方向, 虽然 VaR 的计算结果更为精确, 但这些方法还是没有解决对收益分布的特定假设问题。

为了进一步提高投资组合模型在实际生活中的应用, 准确估计参数成为其中的核心问题之一。许多

研究者尝试采用鲁棒优化的方法来解决参数难以估计的挑战。Lobo 等(1998) [11]提出了鲁棒均值 - 方差模型, 该模型在考虑投资损失最严重情况下, 追求最小偏差。虽然考虑了模型的不确定性, 但仍然需要对不确定性进行某些假设或简化, 这些假设不完全符合实际金融市场条件, 从而影响了模型的实际效果, 同时在面对极端市场情况或极端事件时, 模型的表现仍然不够理想。市场的非线性特征和突发事件超出了模型的处理能力。Zhu 和 Fukushima (2009) [12]则提出了鲁棒-CVaR 投资组合优化模型, 并将问题转化为半定规划问题, 但构建这个模型需要一个确定的不确定性集来描述输入数据的可能变化范围, 这个不确定性集难以构造, 从而导致模型过于保守或不够精确, 而且随着金融市场的不断变化, 这个不确定性集也需要不断变化。另外, Lei 等(2009) [13]提出了基于均值-CVaR 的鲁棒优化模型, 将问题最终转化为二次规划问题, 并广泛应用于投资组合领域。模型考虑了尾部风险的同时提升了投资组合的稳健性, 但对于高维数据或大规模投资组合二次规划的求解是非常复杂的, 这就限制了模型的实际应用。

除了参数估计外, 实际金融数据通常是高维且规模庞大的数据集, 这也会影响投资组合模型的输出结果。为了解决这个问题, 研究者们尝试使用 K-means 聚类对数据进行降维处理。然而, 传统的 K-means 聚类方法在处理大规模复杂数据集时效果不佳。为了改进这一问题, 杨娟等(2017) [14]提出了改进的 K-means 聚类算法, 虽然聚类的速度得到了优化, 但初始聚类中心没有明确给出, 而是随机选取; 杜洪波等(2018) [15]则通过改进的密度峰值算法对 K-means 算法进行了优化, 虽然通过密度峰值确定了聚类中心, 但一开始的聚类数目没有根据数据特征进行选取而是随机选取, 容易导致聚类效果不佳; 而熊开玲等(2017) [16]则通过核密度估计的方法对 K-means 聚类进行了优化, 通过确定半径阈值和密度阈值来确定了初始聚类数目和初始聚类中心。

综上所述, 为了更好地构建投资组合模型, 本文采用了窗口滚动实验方法。选取了包含 100 只股票的日收益数据作为实验数据集, 并利用核密度估计分析了数据集的高斯核密度分布。随后, 根据密度分布设定了半径阈值和密度阈值, 以确定 K-means 聚类的初始聚类中心和聚类簇数, 从而对数据集进行了聚类降维处理, 让投资组合能够适应高维数据。将处理后的数据集代入投资组合模型中, 实验结果表明, 经过聚类处理的投资组合模型具有更高的收益。这些研究结果对于投资者和机构有效地管理投资组合风险并提高投资绩效具有重要的学术和实践价值。

2. 研究方法

2.1. 核密度估计

核密度估计是一种用于估计随机变量概率密度函数的非参数方法, 能够通过数据样本来推断数据的概率密度分布, 从而更好地理解数据分布的特征, 为后续的数据分析和建模提供基础, 因此被广泛应用于各个领域。其定义如下所示:

假设 x_n 是属于 R 的独立分布随机变量, 分布密度函数为 $f(x)$, 则

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

其中 $\hat{f}_h(x)$ 是 $f(x)$ 的核密度估计, h 为窗宽, $K(\cdot)$ 为核函数。

令 $K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right)$, 则(1)式可写成

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K_h(x_i - x), x \in R. \quad (2)$$

根据以上式子我们可以发现核密度估计不仅与数据集本身有关, 还和核函数与窗宽的选取有关。

通过熊开玲等[16]使用核密度估计对K-means聚类的优化, 我们可以解决多维数据的核密度估计问题。假设 $x_1, x_2, \dots, x_n \in \mathbb{R}^d$, $x_i = [v_1, v_2, \dots, v_d]$, 其中 x_i 是一个 d 维向量。令 $f_j(v), j=1, 2, \dots, d$ 是第 j 维概率密度函数。则 $f_d(x) = \prod_{j=1}^d f_j(v)$ 。 d 维的高斯核函数为

$$K_{d,\delta}(x) = (\sqrt{2\pi}\delta)^{-d} \prod_{j=1}^d e^{-\frac{v_j - x_j}{2\delta^2}}. \tag{3}$$

高斯核密度估计图如下图 1 所示。

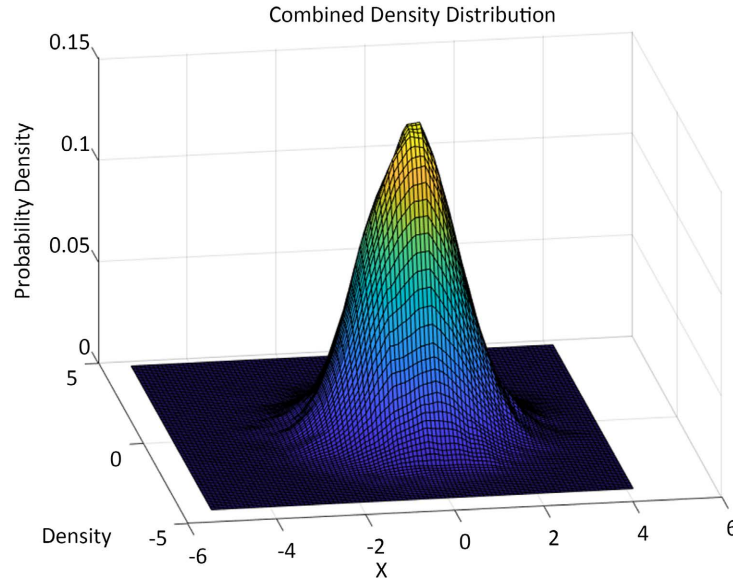


Figure 1. Gaussian kernel density estimation plot
图 1. 高斯核密度估计图

由图 1 所示, 我们可以通过密度分布设定密度阈值和半径阈值, 从而确定 k-means 聚类的初始聚类中心和聚类簇数 k 。具体做法是将小于半径阈值的数据点归为一类, 并且将大于密度阈值的极大值点作为初始聚类中心。

2.2. 均值-CVaR 模型

在传统形式下均值-CVaR 模型可以表示为

$$\min_{\omega \in W, \eta \in R} -u^T \omega + \lambda CVaR_r(X), \tag{4}$$

其中 u 代表的是期望收益, ω 代表的是投资权重比, λ 表示厌恶系数, r 表示置信水平。通过 Rockafellar & Uryasev [17] [18]对条件风险价值的定义 $CVaR_r(X)$ 可以表示为:

$$CVaR_r(X) = \min_{\alpha \in R} \left\{ \alpha + \frac{1}{1-r} E_p \left[[X - \alpha]_+ \right] \right\}. \tag{5}$$

假设损失 $X = -\xi^T \omega$, 我们把(5)式代入(4)式可以得到均值-CVaR 模型可以变形为

$$\min_{\omega \in W, \alpha \in R} -u^T \omega + \lambda \left\{ \alpha + \frac{1}{1-r} E_p \left[[-\xi^T \omega - \alpha]_+ \right] \right\}. \tag{6}$$

分布鲁棒优化模型是考虑在最坏分布的情况下投资者达到最佳收益, 传统的优化模型为

$$\min_{x \in X} \sup_{P \in \phi} E_P [h(x, \xi)]. \quad (7)$$

2.3. 核密度估计的分布鲁棒均值-CVaR 模型

通过构建概率空间, (6)式可以变化为

$$\min_{\omega \in W, \alpha \in R} \left\{ F_{\beta}^{\kappa_1}(\omega, \alpha) := \lambda \alpha + E_P \left[-\xi^T \omega \right] + \frac{\lambda}{1-r} E_P \left[\left(-\xi^T \omega - \alpha \right)_+ \right] \right\}, \quad (8)$$

其中 $\kappa_1 = \{r, \lambda\}$, r 和 λ 分别表示置信水平和厌恶系数, $[x]_+ := \max\{x, 0\}$ 。 P 是资产收益 ξ 的概率分布函数。如果 ξ 独立同步来自函数 P , 则(8)式近似于

$$\min_{\omega \in W, \alpha \in R} \left\{ F_{\beta}^{\kappa_2}(\omega, \alpha) := \lambda \alpha + \frac{1}{T} \sum_{i=1}^T \left(-r_i^T \omega + \frac{\lambda}{1-r} \left(-r_i^T \omega - \alpha \right)_+ \right) \right\}. \quad (9)$$

其中 $\kappa_2 = \{r, \lambda, T\}$, 由于收益金融数据容易出现维数灾难, 所以我们通过核密度估计方法, 将它转化为有限维优化问题, 假设随机变量 v , 它的概率密度函数 $\rho(v)$ 的核密度估计可以表示为:

$$\hat{\rho}_{\delta}(v) = \frac{1}{h} \sum_{i=1}^T \delta_i K \left(\frac{v - v_i}{h} \right). \quad (10)$$

其中核函数 $K(\cdot) > 0$ 并且满足 $\int_{-\infty}^{\infty} K(v) dv = 1$, $K(-v) = K(v)$, $\int_{-\infty}^{\infty} v^2 K(v) dv < \infty$, $h = h(T)$, 为带宽, $v_i, i = 1, 2, \dots, T$ 为数据集, δ 表示权向量。

由(10)式可得(8)式中 $E_P \left[\left(-\xi^T \omega - \alpha \right)_+ \right]$ 的估计形式为:

$$\begin{aligned} & E_{\hat{\rho}_{\delta}} \left[\left(-\xi^T \omega - \alpha \right)_+ \right] \\ &= \int_{-\infty}^{+\infty} [x - a]_+ \hat{\rho}_{\delta}(x) dx \\ &= \int_{-\infty}^{+\infty} [x - a]_+ \frac{1}{h} \sum_{i=1}^T \delta_i K \left(\frac{x + r_i^T \omega}{h} \right) dx \\ &= \sum_{i=1}^T \frac{\delta_i}{h} \int_{-\infty}^{+\infty} (x - a) K \left(\frac{x + r_i^T \omega}{h} \right) dx \\ &= \sum_{i=1}^T \delta_i h \int_{(r_i^T \omega + \alpha)/h}^{+\infty} \left(\frac{-r_i^T \omega - \alpha}{h} + t \right) K(t) dt \\ &= \sum_{i=1}^T \delta_i h \left(\frac{-r_i^T \omega - \alpha}{h} G_k \left(\frac{-r_i^T \omega - \alpha}{h} \right) - \tilde{G}_k \right) \left(\frac{-r_i^T \omega - \alpha}{h} \right), \end{aligned} \quad (11)$$

对于上面公式的表达式我们可以做出如下假设 $G_k(u) = \int_{-\infty}^u K(x) dx$, $\tilde{G}_k(u) = \int_{-\infty}^u x K(x) dx$ 。如果 $\varsigma_k(u) := u G_k(u) - \tilde{G}_k(u)$, 则 $\tau_k(u, h) := h \varsigma_k(h^{-1}u)$, 并且它们满足以下条件

- 1) $\nabla \tau_k(u, h) \geq 0$,
- 2) $\tau_k(u, h)$ 是联合凸函数,
- 3) $[u]_+$ 是 $\varsigma_k(u)$ 的衰退函数 ($\varsigma_k 0_+(u)$), 同时 $\tau_k(u, h) = [u]_+ + o(h)$ 。

如果知道参数 $\kappa_3 = \{r, \lambda, T, K(\cdot), h\}$, 并且给出权向量 δ , 我们可以知道核密度估计均值-CVaR 模型, 如下形式

$$\min_{\omega \in W, \alpha \in R} \left\{ F_{\beta}^{\kappa_3}(\omega, \alpha, \delta) := \lambda \alpha + \sum_{i=1}^T \delta_i \left(-r_i^T \omega + \frac{\lambda}{1-r} \tau_k \left(-r_i^T \omega - \alpha, h \right) \right) \right\}. \quad (12)$$

由于权向量 δ 不能够准确预估, 所以我们需要分布鲁棒优化, 即在给定的最坏分布下, 我们希望达到的投资损失最小, 模型公式如下所示:

$$\min_{\omega \in W, \alpha \in R} \max_{p \in I} \left\{ F_{\beta}^{\kappa_1}(\omega, \alpha, p) := E_p[-\xi^T \omega] + \lambda \alpha + \frac{1}{1-r} E_p[(-\xi^T \omega - \alpha)_+] \right\}. \quad (13)$$

由于(13)式涉及到维数问题, 因此我们采用核密度估计的方法, 并结合 ϕ 散度, 将问题转化为凸优化问题。我们假设样本 $X_i, i=1, 2, \dots, T$ 独立同分布于 $p(\cdot)$, 则不确定集为

$$I_{KDE-\phi}^c = \left\{ p_{\delta}(x) = \delta_i h^{-1} K\left(\frac{x - X_i}{h}\right), \delta \in \Gamma_{\phi}^c \subseteq R^T \right\}, \quad (14)$$

其中权向量定义为 $\Gamma_{\phi}^c = \left\{ \delta : \delta \in \Gamma, \theta^{\phi}(\delta, \delta^0) \leq c \right\}$, $\theta^{\phi}(\delta, \delta^0) := \sum_{i=1}^T \delta_i^0 \phi(\delta_i, \delta_i^0)$ 。

如果我们知道参数 $\kappa_4 \{r, \lambda, R, K(\cdot), h, \phi(\cdot), c\}$, 并且把(13)式的不确定集 I 换成(14)式, 则可以得到核密度估计的分布鲁棒均值-CVaR 模型

$$\min_{\omega \in W, \alpha \in R} \max_{p \in I_{KDE-\phi}^c} F_{\beta}^{\kappa_3}(\omega, \alpha, p). \quad (15)$$

由(12)式可得, (15)式可以变化为如下形式

$$\min_{\omega \in W, \alpha \in R} \max_{\delta \in I_{KDE-\phi}^c} \left\{ F_{\beta}^{\kappa_4}(\omega, \alpha, \delta) := \lambda \alpha + \sum_{i=1}^T \delta_i \left(-r_i^T \omega + \frac{\lambda}{1-r} \tau_k(-r_i^T \omega - \alpha, h) \right) \right\}. \quad (16)$$

下面我们将把(16)式转化为凸优化问题。其中(16)式的最大化问题的拉格朗日方程为

$$L(\delta, u, \partial) = \lambda \alpha + \sum_{i=1}^T \delta_i \left(-r_i^T \omega + \frac{\lambda}{1-r} \tau_k(-r_i^T \omega - \alpha, h) \right) + u \left(c - \frac{1}{T} \sum_{i=1}^T \phi(\delta_i T) \right) + \partial (1 - 1^T \delta). \quad (17)$$

假设 $g(\omega, \alpha; u, \partial) = \max_{\delta \geq 0} L(\delta, u, \partial)$, 由于 $\delta^0 = T^{-1} \mathbf{1} \in \Gamma_{\phi}^c$, 所以 Γ_{ϕ}^c 表示的意义是 $1^T \delta^0 = 1$ 并且 $I_{\phi}(\delta^0, \delta^0) = 0 < c$ 。根据 Γ_{ϕ}^c 的规律性, 具有强对偶性成立, 即 $\max_{\delta \in \Gamma_{\phi}^c} F_{\beta}^{\kappa_4}(\omega, \alpha, \delta) = \min_{u \geq 0, \partial} g(\omega, \alpha; u, \partial)$ 。则对偶目标函数满足

$$\begin{aligned} g(\omega, \alpha; u, \partial) &= \lambda \alpha + \partial + cu + \max_{\delta \geq 0} \sum_{i=1}^T \left\{ \delta_i \left(-r_i^T \omega + \frac{\lambda}{1-r} \tau_k(-r_i^T \omega - \alpha, h) - \partial \right) - \frac{u}{T \phi(\delta_i T)} \right\} \\ &= \lambda \alpha + \partial + cu + \sum_{i=1}^T \max_{\delta_i \geq 0} \left\{ \delta_i \left(-r_i^T \omega + \frac{\lambda}{1-r} \tau_k(-r_i^T \omega - \alpha, h) - \partial \right) - \frac{u}{T \phi(\delta_i T)} \right\} \\ &= \lambda \alpha + \partial + cu + \frac{u}{T} \sum_{i=1}^T \max_{t \geq 0} \left\{ \frac{t}{u} \left(-r_i^T \omega + \frac{\lambda}{1-r} \tau_k(-r_i^T \omega - \alpha, h) - \partial \right) - \phi(t) \right\} \\ &= \lambda \alpha + \partial + cu + \frac{u}{T} \sum_{i=1}^T \phi^* \left(\frac{-r_i^T \omega + \frac{\lambda}{1-r} \tau_k(-r_i^T \omega - \alpha, h) - \partial}{u} \right). \end{aligned} \quad (18)$$

其中函数 ϕ 的共轭函数 $\phi^*(\cdot)$ 满足这些形式: $R \rightarrow R$, $\phi^*(s) := \sup_{t \geq 0} \{st - \phi(t)\}$ 。

由公式(18)模型(16)最后可以转化为

$$\min_{\omega \in W, \alpha, \partial, u \geq 0} \lambda \alpha + \partial + cu + \frac{u}{T} \sum_{i=1}^T \phi^* \left(\frac{-r_i^T \omega + \frac{\lambda}{1-r} \tau_k(-r_i^T \omega - \alpha, h) - \partial}{u} \right). \quad (19)$$

3. 数值实验

根据对我国股票市场实际情况及上述模型理论的综合考量, 我们选取了恒生指数 100 的股票收益率数据作为研究对象, 目的是运用该理论模型进行深入分析。我们利用 Matlab 软件编写了相关程序, 对这些股票的历史数据进行了全面的分析和优化。通过比较在有聚类和无聚类条件下, 模型对投资组合收益率的影响, 通过分析在有聚类和无聚类条件下, 实验得到的夏普比率、平均换手率以及模型运行时间等数据结果, 我们得出了一些相关结论。

3.1. 实验准备

在这个实验中, 采用了滚动窗口的方法, 选取了恒生指数 100 的股票收益率数据作为研究对象, 共有 $D = 2250$ 个数据集。其中, 我们设置了窗口宽度 $T = 2000$ 天, 从 $t = T + 1$ 天起, 用前 $T = 2000$ 天的数据计算最佳投资权重, 然后利用得到的投资权重计算第 $t = T + 1$ 天的收益率。每次滚动时, 我们增加一个新的数据并去除一个旧的数据, 总共进行了 $D - T$ 次滚动。其中实验参数为: $r = 0.95$; $\delta^0 = \frac{1}{T}1$; $\lambda = 10$; $c = 1$ 。

对经过核密度估计优化的聚类结果进行了细致评估, 并将聚类后的数据集代入到投资组合模型中。通过采用滚动窗口实验的方法, 比较了计算得出的收益率曲线、夏普比率、平均换手率以及模型运行时间等数值结果, 以此来深入探索聚类对投资组合模型的影响。

3.2. 数值结果

根据我们改进的聚类方法, 可以确定选定数据集的聚类中心, 如图 2 所示。

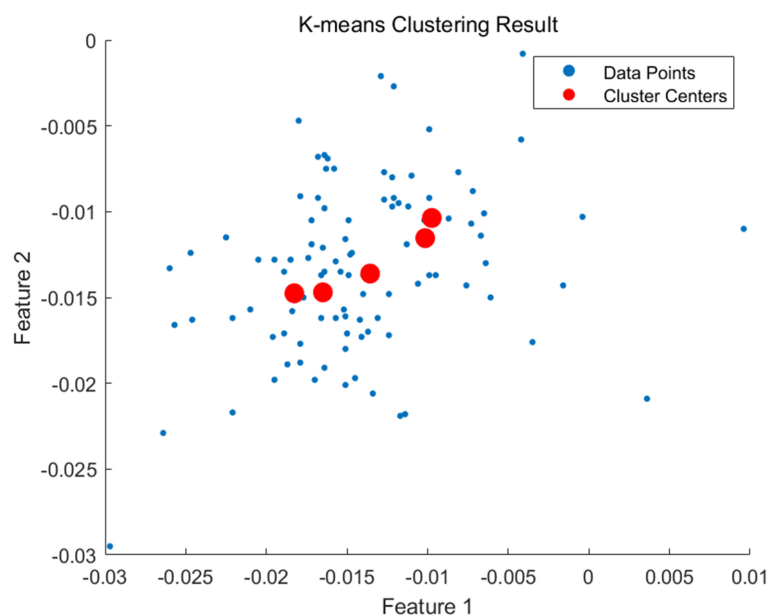


Figure 2. Cluster centroids plot
图 2. 聚类中心图

从图中我们可以看出聚类中心为 5 个点, 所以我们可以确定聚类簇数 k 为 5。

我们将经过聚类处理的数据集输入到投资组合模型中, 并绘制了具有和没有聚类的投资组合模型的收益率曲线, 如图 3 所示和实验的数值结果, 如表 1 所示。

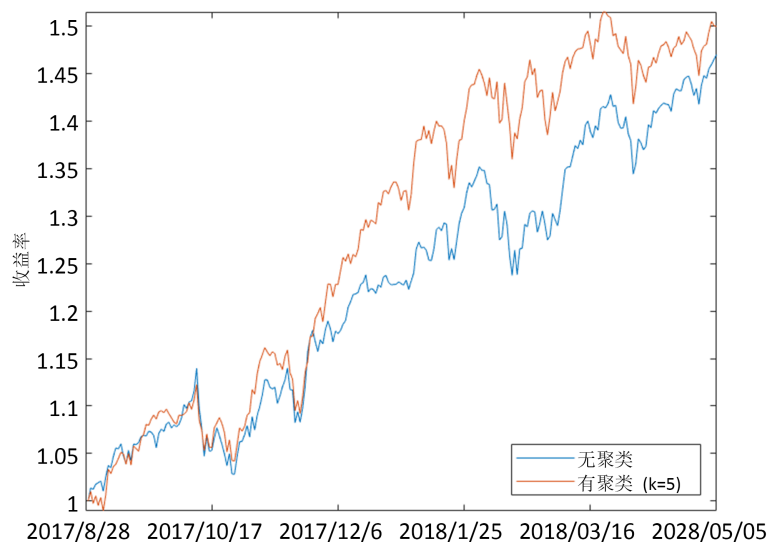


Figure 3. Yield curve plot

图 3. 收益率曲线图

Table 1. Numerical result

表 1. 数值结果

评价指标	无聚类	有聚类
平均收益率	0.0016	0.0017
收益率标准差	0.0106	0.0111
夏普比率	0.1502	0.152
平均换手率	0.0028	0.0000
模型运行时间	2183 秒	55.21 秒

根据图 3 和表 1 的结果, 我们可以观察到经过聚类后的模型具有更高的收益率, 并且夏普比率也更高, 这意味着可以实现更高的投资回报。与此同时, 聚类后的模型平均换手率较低, 这表明投资者可以减少高额的交易费用, 并且显著缩短模型的运行时间。因此, 我们可以得出结论, 具有聚类的模型在投资效果上更为出色。

4. 结论

本文探索了综合运用聚类、核密度估计以及分布鲁棒均值-CVaR 模型的方法, 以提升股票投资组合的构建和风险管理能力。通过对股票的风险度量、收益率曲线以及最优权重的计算进行分析, 得出以下结论:

- 1) 在确定聚类数目和选择聚类中心的过程中, 我们设定了初始聚类中心和聚类数目, 这样实验结果不会因为聚类而随意改变模型的投资实验结果。
- 2) 通过实验结果我们可以得出, 在投资组合模型中使用聚类能够产生更高的收益, 相比于没有使用聚类的模型, 能够提高夏普比率, 并且降低平均换手率, 同时也提高了模型的运算速率。
- 3) 结合聚类的投资组合模型能够构建更出色的投资组合, 在重新平衡投资风险的同时, 可以让投资者获得更高的回报。这表明基于改进的聚类和核密度估计的分布鲁棒均值-CVaR 投资组合模型具有更广泛的适用性。

基金项目

国家自然科学基金项目(12161075), 江西省自然科学基金重点项目(20212ACB201006)。

参考文献

- [1] Markowitz, H. (1952) Portfolio Selection. *The Journal of Finance*, **7**, 77-91. <https://doi.org/10.2307/2975974>
- [2] Konno, H., Shirakawa, H. and Yamazaki, H. (1993) A Mean-Absolute Deviation-Skewness Portfolio Optimization Model. *Annals of Operations Research*, **45**, 205-220. <https://doi.org/10.1007/bf02282050>
- [3] Markowitz, H. (1959) Portfolio Selection: Efficient Diversification of Investments. Wiley, Hoboken.
- [4] Gaivoronski, A.A. and Pflug, G. (1991) Finding Optimal Portfolios with Constraints on Value at Risk. *The III Stockholm Seminar on Risk Behavior and Risk Management*, Stockholm, 10 January 1991, 1.
- [5] Rockafellar, R.T. and Uryasev, S. (2000) Optimization of Conditional Value-at-Risk. *The Journal of Risk*, **2**, 21-41. <https://doi.org/10.21314/jor.2000.038>
- [6] Rockafellar, R.T. and Uryasev, S. (2002) Conditional Value-At-Risk for General Loss Distributions. *Journal of Banking & Finance*, **26**, 1443-1471. [https://doi.org/10.1016/s0378-4266\(02\)00271-6](https://doi.org/10.1016/s0378-4266(02)00271-6)
- [7] 牛昂. Value at Risk: 银行风险管理的新方法[J]. 国际金融研究, 1997, 27(4): 61-65.
- [8] 姚刚. 风险值测定法浅析[J]. 经济科学, 1998, 19(1): 56-51.
- [9] 刘宇飞. VAR 模型及其在金融监管中的应用[J]. 经济科学, 1999, 20(1): 39-50.
- [10] 王春风, 万海晖, 张维. 金融市场风险测量模型-VAR [J]. 系统工程学报, 2000, 15(1): 67-76.
- [11] Lobo, M.S., Vandenbergh, L., Boyd, S. and Lebet, H. (1998) Applications of Second-Order Cone Programming: Interior-Point Methods and Engineering Applications. *Linear Algebra and Its Applications*, **284**, 193-228. [https://doi.org/10.1016/s0024-3795\(98\)10032-0](https://doi.org/10.1016/s0024-3795(98)10032-0)
- [12] Zhu, S. and Fukushima, M. (2009) Worst-Case Conditional Value-at-Risk with Application to Robust Portfolio Management. *Operations Research*, **57**, 1155-1168. <https://doi.org/10.1287/opre.1080.0684>
- [13] Zhu, L., Coleman, T. and Li, Y. (2009) Min-Max Robust and Cvar Robust Mean-Variance Portfolios. *The Journal of Risk*, **11**, 55-85. <https://doi.org/10.21314/jor.2009.191>
- [14] 杨娟, 屈传慧. 改进 K 均值聚类算法[J]. 舰船电子对抗, 2017, 40(6): 91-93.
- [15] 杜洪波, 白阿珍, 朱立军. 基于改进的密度峰值算法的 K-means 算法[J]. 统计与决策, 2018, 34(18): 20-24.
- [16] 熊开玲, 彭俊杰, 杨晓飞, 黄俊. 基于核密度估计的 K-means 聚类优化[J]. 计算机技术与发展, 2017, 27(2): 1-5.
- [17] Rockafellar, R.T. and Uryasev, S. (2000) Optimization of Conditional Value-at-Risk. *The Journal of Risk*, **2**, 21-41. <https://doi.org/10.21314/jor.2000.038>
- [18] Rockafellar, R.T. and Uryasev, S. (2002) Conditional Value-at-Risk for General Loss Distributions. *Journal of Banking & Finance*, **26**, 1443-1471. [https://doi.org/10.1016/s0378-4266\(02\)00271-6](https://doi.org/10.1016/s0378-4266(02)00271-6)