

文章编号: 1005-2542(2006)02-0149-09

基于 Copula-GARCH-EVT 的 资产组合选择模型及其混合遗传算法

刘 志 东

(中央财经大学, 北京 100081)

【摘要】在非正态分布的条件下, Markowitz 的均值-方差资产组合选择模型存在不足。为此, 以 VaR 和 CVaR 作为风险度量方法, EVT 反映收益率的尾部分布, GARCH 反映收益率的波动性, Copula 函数反映金融资产收益的相关性, 构建了基于 Copula 函数的资产组合选择模型。针对非正态分布条件下 VaR 非凸性和分布函数不连续性导致资产组合选择优化计算复杂、不精确的难题, 设计了基于单纯形和传统遗传算法的混合遗传算法。最后, 根据中国证券市场数据, 采用该混合遗传算法对建立的资产组合选择模型求解。

关键词: Copula 函数; VaR 和 CVaR; 极值分布; GARCH; 资产组合选择; 遗传算法; 单纯形

中图分类号: F 832.5 **文献标识码:** A

A Portfolio Selection Model on Copula-GARCH-EVT Based and Its Hybrid Genetic Algorithm

LIU Zhi-dong

(Center Univ. of Finance and Economics, Beijing 100081, China)

【Abstract】 There are some drawbacks in Markowitz's mean-variance portfolio selection model under the condition of no normal distributions. So the article constructed a portfolio selection model based on Copula-EVT-GARCH, which measures the risk by VaR and CVaR, reflects the tail distributions by EVT (extreme value theory), reflects the volatility by GARCH, reflects the dependence of financial assets returns by copula function. Then according to the fact VaR's no convexity and the discontinuity in distribution, which make the computation of the portfolio selection optimization very complex and inaccurate, the article designed a hybrid genetic quantitative algorithm based on Nelder-Mead simplex and traditional genetic algorithms. Finally, according to the data from China securities market, the article does empirical research for the portfolio selection model by the hybrid genetic quantitative algorithm.

Key words: Copula; VaR and CVaR; portfolio selection; extreme value theory; Garch; genetic algorithm, simplex search

在 Markowitz^[1,2] 的均值-方差资产组合选择模型中, 风险被定义为期望收益的可能变化, 一般用方差或标准差表示。Markowitz 的资产组合理论主要是规范分析, 告诉人们如何进行资产选择。其理论忽

略了金融市场的实证特征。用方差或标准差度量风险隐含的假设是投资者对负的损失和正的收益赋予相同的权重, 对待两者的态度是相同的。尽管 Markowitz 的资产组合理论思想非常直观, 但将收益的方差或标准差等同于风险只有在投资者具有二次效用函数或资产收益呈正态分布时才是可行的。在 Markowitz 的资产组合选择模型中, 采用 Pearson 的线性相关系数反映金融资产收益的相关性。

收稿日期: 2005-04-18 修订日期: 2005-07-25

作者简介: 刘志东 (1973-), 男, 讲师, 博士。主要从事投资组合管理与风险控制研究。

Pearson 的线性相关只适用于椭圆分布,要求金融资产风险程度适中,只能度量随机变量之间的线性关系。由于 Pearson 的线性相关不是根据随机变量联合分布度量随机变量相关性的方法,它具有一些缺陷,常常导致错误的结论。另外,Markowitz 的均值-方差资产组合选择方法的计算任务比较重,需要计算很大的方差和协方差矩阵。所以,Markowitz 的基于均值-方差的资产组合选择模型存在一定局限性。

现实中,投资者对资产组合的损失和盈利持不同的态度,通常人们更关心资产组合收益分布下侧,下侧代表损失,应该对下风险单独度量。VaR 和 CVaR 就是一种下侧风险度量方法。在用 VaR 和 CVaR 作为度量风险的工具时,基于均值-风险准则的资产组合模型面临以下几个难题:

(1) VaR 的非凸性。在实际中,金融资产收益率是非正态分布,VaR 是离散的、非连续和非凸的,不满足次级可加。在基于 VaR 风险度量方法进行资产组合选择优化时,可能存在很多局部 VaR 最小解。传统的优化算法可能不能达到使 VaR 最小的全局最优解或次优解。尽管 CVaR 满足凸性和次级可加要求,但由于其计算依赖于 VaR,在实际中,以其作为风险度量方法进行组合选择也很难找到最优解。VaR 的非凸性是 VaR 和 CVaR 难以在实际资产组合优化中应用的最大障碍。

(2) 分布函数尾部的非连续可微。由于 VaR 是基于分位数的风险度量方法,分布函数在尾部是离散的,不一定是连续和可微的,可能存在“跃迁”(Jump)的现象。在用历史模拟法和蒙特卡洛模拟计算 VaR,该现象更为明显。分布函数尾部的这种特性表明传统依靠梯度和导数求使 VaR 最小的资产组合几乎是不可行的。

(3) 如何反映资产收益率的相关性对资产组合选择的影响。现实中金融资产收益之间不一定符合线性相关。度量真实的相关性不仅对组合风险管理有重要意义,而且对资产组合选择也有重要的影响。相关性直接影响到分散化投资降低风险的效果。

鉴于以上事实,必须建立一种反映组合资产收益率实际分布和相关性的资产组合选择模型。并采用可行的方法,对模型求解,从而得到资产组合选择边界。本文首先建立基于 Copula 函数的资产组合选择模型,然后研究用混合遗传算法对所建立的资产组合选择模型求解。

1 基于 Copula-GARCH-EVT 的资产组合选择模型

1.1 Copula 函数及其在反映随机变量相关性上的优势

Copula 一词原意是交换、连接的意思。在数学中,它是指把多个变量的联合分布与它们的边缘分布连接在一起的函数。如果 F_1, F_2, \dots, F_d 是一元分布函数, $u_i = F_i(x_i), i = 1, 2, \dots, d$, 则 $C(F_1(x_1), F_2(x_2), \dots, F_d(x_d))$ 是具有边缘分布函数 F_1, F_2, \dots, F_d 的多元分布函数。 d 维 Copula 函数 C 是把多个随机变量 Y_1, Y_2, \dots, Y_d 的联合分布与它们各自的边缘分布连接在一起的函数。Copula 函数对于构造和模拟多元分布函数具有重要的意义。根据关于 Copula 函数最重要的 Sklar 定理^[3],令 F 是具有边缘分布函数 F_1, F_2, \dots, F_d 的 d 维分布函数,若边缘分布函数 F_1, F_2, \dots, F_d 连续,则存在一个唯一满足 $F(x_1, x_2, \dots, x_d) = C(F_1(x_1), F_2(x_2), \dots, F_d(x_d))$ 关系的连接函数 C 。对于多元连续分布函数,一元边缘分布函数和多元分布函数相关结构能够被分离,多元变量之间的相关结构可以用适当的 Copula 函数表示。Copula 函数与多元分布函数一样,包含随机变量之间的所有相关信息。

为了有效地度量金融资产收益之间的真实相关性,需要多元分布函数理论。多元分布函数是描述随机变量相关性的最根本的方法。但传统的多元分布函数在实际应用中存在一些缺陷,即在变量较多时解析式很难处理,并且存在一系列约束条件,不仅要求各个边缘分布函数类型与多元分布函数类型一样,而且各个边缘分布必须完全相同。资产组合尤其是含有不同种类资产的资产组合(股票和外汇),各种金融资产边缘分布函数通常不符合同一类型的分布函数,这种情况使得多元分布函数很难在资产组合管理中得到应用。而通过 Copula 函数技术可以构造灵活的多元分布函数,掌握资产组合内各金融资产收益真实相关关系,在此基础上,运用 VaR 和 CVaR 等指标度量资产组合风险,基于均值-风险准则的资产组合选择模型才有意义。

本文采用以下几个步骤来构造资产组合选择模型:(1)根据 Copula 函数构建反映金融资产收益率相关性的联合分布函数。(2)根据构建的联合分布函数,用蒙特卡罗技术模拟资产组合中各金融资产在不同情景下的收益率。(3)根据模拟的金融资产收益率,用 VaR 和 CVaR 计算资产组合风险,构建基于均值-风险准则的资产组合选择模型。

1.2 根据 Copula 函数构建反映金融资产收益率相关性的联合分布函数

通过 Copula 函数技术构建多元分布函数时需要两个步骤: ① 构建各个变量的边缘分布函数; ② 选择合适的 Copula 函数。下面根据 Copula 构建反映组合内金融资产收益率相关性的联合分布函数的具体步骤。

1.2.1 资产组合中各资产收益率随机扰动项边缘分布函数的构建 通常情况下, 金融资产收益率不服从正态分布, 呈现出一定的“尖峰”和“厚尾”特征。传统的极值理论虽然直接研究金融资产收益率分布的尾部, 但它忽略了金融资产收益率分布是时变的, 假设资产收益率是独立同分布的。传统的 ARMA 和 GARCH 模型虽然可以描述时间序列的条件均值和条件方差, 但其假设随机扰动项服从正态分布和 t 分布, 并关注整个分布, 而不是直接对风险管理所关心分布的尾部进行建模。

由于金融资产收益率一般呈非对称分布, 具有“杠杆效应”。正态分布和 t 分布假设与此特征不符, 不能预测金融资产收益率的极端变化情况。McNeil 和 Frey^[4]、Diebold 等^[5]也开始探讨了把极值理论和 GARCH 进行组合的可能性。本文在遵循上述学者的研究思路基础上, 把 POT 极值理论和 ARMA-GARCH 模型进行适当的组合。文献研究表明, 可以用 AR(1) 度量金融资产收益率的条件均值, GARCH(1,1) 度量金融资产收益率的条件方差。

假设资产组合中有 d 种金融资产, 对于金融资产 i ($i = 1, 2, \dots, d$), 直接根据最近 n 期历史收益率数据 $(R_{i,t-n+1}, \dots, R_{i,t-1}, R_{i,t})$ 运用 AR(1) 和 GARCH(1,1) 模型建模。通常在残差符合条件正态分布的假设下, 可以利用 ARMA-GARCH 模型的似然函数, 给出模型参数的估计值。即使在金融资产收益率序列残差不满足条件正态分布的情况下, 使用正态极大似然估计, 仍然可以得到参数的一致渐进正态非最小方差估计。但是这样得到的随机扰动项 z_t 将有很大的误差, 而 z_t 是本文下一步进行 EVT 尾部估计的输入变量, 它的有效性将直接影响整个估计结果, 为此必须寻找一个更有效的参数估计方法。GMM 估计恰好可以满足其要求, 它不需要假设 z_t 符合任何分布, 只需要 z_t 的一阶和二阶条件矩, 就可以对模型参数进行估计。Skoglund 给出了采用 GMM 对 GARCH 模型的估计方法和计算过程, 以及参数的收敛情况。另外, 要进行 GMM 估计还需要一个对模型参数和随机扰动项 z_t 的三阶矩和四阶矩的初始估计值, 这些初始值可以通过对

ARMA-GARCH 模型随机扰动项符合正态分布的情况进行最大似然估计得到。这样就得到模型有效的参数估计值和随机扰动项序列 z_t , 进而对随机扰动项进行极值分析。

在采用 GMM 方法估计出模型参数基础上, 可以得到最近 n 期的条件均值 $(\hat{z}_{i,t-n+1}, \dots, \hat{z}_{i,t-1}, \hat{z}_{i,t})$ 和条件方差 $(\hat{z}_{i,t-n+1}^2, \dots, \hat{z}_{i,t-1}^2, \hat{z}_{i,t}^2)$, 则最近 n 期的随机扰动项 z 为

$$(Z_{i,t-n+1}, \dots, Z_{i,t}) = \left(\frac{R_{i,t-n+1} - \hat{z}_{i,t-n+1}}{\hat{z}_{i,t-n+1}}, \dots, \frac{R_{i,t} - \hat{z}_{i,t}}{\hat{z}_{i,t}} \right) \quad (1)$$

通常随机扰动项 z 是独立同分布的, 但并不一定符合正态分布或 t 分布的假设条件。由于随机扰动项 z 满足独立同分布的假设条件, 可以采用一元极值理论中的 GPD (帕累托) 分布函数对随机扰动项 z 上下尾部分布进行建模。由于极值分布只描述样本的上下尾部分布, 用到很小一部分样本信息, 为使所有样本反映的信息得到充分运用, 对于处于上下尾部阈值之间的随机扰动项 z , 采用正态分布 (或经验分布) 方法估计分布函数。这样资产组合中每种金融资产收益率随机扰动项 z 的边缘分布为

$$F_i(Z_i) = \begin{cases} \frac{N_{u_i}^L}{N} \left[1 + \Upsilon \frac{|Z_i - u_i^L|}{\Upsilon} \right] & Z_i < u_i^L \\ H_i(x) & u_i^L \leq Z_i \leq u_i^R \\ 1 - \frac{N_{u_i}^R}{N} \left[1 + \Upsilon \frac{(Z_i - u_i^R)}{\Upsilon} \right] & Z_i \geq u_i^R \end{cases} \quad i = 1, 2, \dots, d \quad (2)$$

式中: $H_i(x)$ 为标准正态分布函数; u_i^R 、 u_i^L 分别表示随机扰动项 Z 的上下尾部阈值; N 表示随机扰动项 Z 的样本数; $N_{u_i}^R$ 表示高于上尾部阈值 u_i^R 的随机扰动项个数; $N_{u_i}^L$ 表示低于下尾部阈值 u_i^L 的随机扰动项个数。

在运用广义帕累托分布极值理论时, 样本阈值 u 的选取至关重要。如果 u 选取的太大, 超过 u 的样本数会很少, 参数估计的误差较大。如果 u 选取的太小, $F_u(y) \approx G^w(y)$ 的关系不成立, 导致参数估计是有偏的。在实证研究中, 选择 u^R 和 u^L 的标准是: $H(x) < 1 - \frac{N_{u_i}^R}{N}$, $H(x) > \frac{N_{u_i}^L}{N}$, 这样可以保证式 (2) 中的 $F_i(Z)$ 是连续的。

以上这种分段求分布函数的方法, 既能通过极值分布考虑到样本分布的厚尾特性, 又能通过正态分布使处于上下尾部之间的样本数据得到充分的运用, 反映真实的信息。

1.2.2 选择合适的 Copula 函数度量金融资产收益的相关性 在现实金融市场中,金融资产收益率的联合分布中存在 2 种非对称现象:①非对称现象表现为单个股票收益率具有非对称分布,偏度不等于零。②非对称是金融资产收益率之间相关的非对称,这种非对称相关表现为在市场处于下降的趋势时(熊市),尤其是极端下降时,金融资产收益率之间的相关性比正常时或上升时(牛市)的相关性大。最近 Ang 等^[6-8]报道了股票间这种非对称相关现象。刘志东等^[9,10]通过对中国股票收益相关性的研究,发现中国股票收益存在尾部极值相关。由于存在非对称相关性和尾部极值相关,在熊市时,分散化投资降低资产组合风险的效果就会减弱,资产组合的风险将会增加。金融资产收益率在尾部具有更强的相关性,并且这种相关性的与金融资产收益率的频率有关,高频数据比低频数据具有更强的相关性。忽略金融资产收益率的尾部相关性将会导致在市场趋于下降时过高估计资产组合分散化投资降低风险的作用。度量金融资产收益率的非对称相关和尾部极值相关性对于资产组合风险度量和选择具有重要的意义。

根据以上对现实金融资产收益率相关性的分析,本文采用 Copula 函数来度量资产组合资产收益率的相关性。Copula 函数的类型很多,总体可以分为椭圆类分布连接函数和阿基米德连接函数,而每类又分为许多具体的连接函数。在资产组合风险管理中,选择哪种连接函数要考虑两方面的影响因素:①该 Copula 函数的特征是否与现实金融市场中金融资产收益率之间的相关性相符合;②该 Copula 函数在实际应用中的可行性,是否存在计算技术上的难题。

现实的金融市场中,各种金融资产的收益率并不符合正态分布的假设条件,通常表现为“尖峰”和“厚尾”的特征。与此同时,各种金融资产的收益率之间也不符合多元正态分布的假设,呈现出尾部极值相关性。符合此特征的分布函数主要有 t -Copula 连接函数和阿基米德连接函数中的 Clayton 类连接函数。在 Clayton 类连接函数中,一般的 Clayton 连接函数只能度量单侧极值相关,只有 Joe-Clayton 连接函数在分布的上下尾部均具有相关性,而且这种相关性是非对称的。从理论上讲它比 t -Copula 连接函数更完美。但是 Joe-Clayton 连接函数在维数增加时,其计算任务是复杂和繁琐的,实际中很难运用。因此,综合考虑到以上两种约束条件,本文只研究用高斯连接函数、 t -Copula 连接函数度量金融资产收

益的相关性。 t 连接函数能够反映尾部相关性,而高斯连接函数不能反映尾部相关性。当自由度趋近于正无穷时,除非在相关系数等于 1,否则尾部相关系数将变为零,此时, t 连接函数与 Gauss 连接函数相同。因此, t 连接函数比 Gauss 连接函数应用更广泛。

1.2.3 Copula 函数参数估计 在用 Copula 函数模拟金融资产收益率时,采用什么样的参数估计至关重要。通常采用极大似然估计 (ML) 方法。高斯连接函数中的参数是相关系数矩阵 R , t -Copula 函数的参数包括相关系数矩阵 R 和自由度。对于 t -Copula 函数理想的参数估计方法应该是根据 EML (exact likelihood method) 对 R 和自由度 v 进行同时估计。但这种同时估计几个参数的方法在计算上相当困难。为此必须采用一种新的方法估计 R 。对于“厚尾”分布的数据,传统的线性相关估计方法缺乏稳定性,可以用 Kendall 的 τ 相关系数矩阵估计线性相关矩阵 R 。对于椭圆分布有:

$$R = \sin \left[\frac{\pi}{2} \right] \quad (3)$$

方法的最大优点是基于秩相关,克服了传统的线性相关的不足,在估计矩阵 R 时不涉及到自由度的信息。在估计出 R 后,在采用极大似然方法估计自由度 v 。对于本文的问题,整个算法如下:

(1) 用经验分布或上面的边缘分布函数对金融资产收益率的随机扰动项进行概率转化

$$F(Z_{ij}) = u_{ij} \quad (4)$$

$$i = 1, 2, \dots, n; \quad j = 1, 2, \dots, d$$

需要特别指出的是,本文并不直接求金融资产收益率的 Copula 函数,而是求其随机扰动项的 Copula 函数,根据文献中关于 Copula 函数的推论 2 可知二者相等。但是通过求其随机扰动项的 Copula 函数,可以在边缘分布函数中利用 GARCH 模型反映资产收益率的动态变化。

(2) 根据 $R = \sin \left[\frac{\pi}{2} \right]$, 利用 Kendall 的 τ 估计随机扰动项的相关系数矩阵

$$\hat{R}_{ij} = \sin \left[\frac{\pi}{2} \hat{\tau}_{ij} \right] \quad (5)$$

(3) 运用极大似然方法,估计自由度参数 v :

$$\hat{v} = \arg \max_{v \in [2, \infty)} \left[\sum_{i=1}^n \log(c(\hat{U}_i; v, \hat{R})) \right] \quad (6)$$

式中:

$$\hat{U}_i = (u_{i1}, u_{i2}, \dots, u_{id})$$

$$c(u_1, u_2, \dots, u_d; R; v) =$$

$$|R|^{-\frac{1}{2}} \frac{\Gamma\left(\frac{v+d}{2}\right) \left(\frac{\Gamma\left(\frac{v}{2}\right)}{\Gamma\left(\frac{v+d}{2}\right)}\right)^d \left(1 + \frac{1}{v} Y^T R^{-1} Y\right)^{-\frac{v+d}{2}}}{\prod_{j=1}^d \left(1 + \frac{Y_j^2}{v}\right)^{-\frac{v+1}{2}}} \\ \Upsilon \sim T_v^{-1}(u_1), \dots, T_v^{-1}(u_d)$$

1.3 用蒙特卡罗方法模拟资产组合资产的收益率

(1) 根据 t -Copula 函数模拟随机扰动项 t -

Copula 函数模拟随机扰动项计算步骤为: ① 计算随机扰动项的相关系数矩阵 R 的 Cholesky 分解矩阵 A 如果 R 是正定的, 则存在 $n \times n$ 矩阵, 使得 $R = AA^T$ ② 根据标准正态分布, 模拟 d 个相互独立的随机变量 $y = (y_1, \dots, y_d)^T$ ③ 根据 i^2 分布, 模拟独立于 y 的随机变量 s ④ 令 $w = Ay, X = \frac{-v}{s} w$ ⑤ 令

$u_i = t_v(X_i), i = 1, 2, \dots, d$, 则 $(u_1, u_2, \dots, u_d) \sim C_{v,R}^t$

⑥ 根据 $(Z_1, Z_2, \dots, Z_d) = (F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))$, 得到联合分布为 $F(Z_1, Z_2, \dots, Z_d)$, 连接函数为 $C_{v,R}^t$ 的 d 维随机扰动项 (Z_1, Z_2, \dots, Z_d)

(2) 资产组合资产收益率的模拟。在 (1) 的基础上, 根据随机波动方程 $R_t = \mu_t + \epsilon_{i,t} Z_i$ 对金融资产收益率的分析。用前面基于 GMM 估计的 AR(1) 模型模拟金融资产收益率的条件均值 $\mu_{i,t}, i = 1, 2, \dots, d$, 基于 GMM 估计的 GARCH(1, 1) 模拟金融资产收益率的条件方差 $\epsilon_{i,t}, i = 1, 2, \dots, d$ 。然后根据随机波动方程 $R_{i,t} = \mu_{i,t} + \epsilon_{i,t} Z_i, i = 1, 2, \dots, d$ 得到资产组合资产的收益率向量 $(R_{1,t}, \dots, R_{d,t})$ 。这种模拟资产组合资产收益率的方法既考虑了金融资产收益的相关性, 同时也反映了金融资产收益率的动态变化特性。

1.4 采用 VaR 和 CVaR 度量风险, 构建基于均值-风险准则的资产组合选择模型

假设在单一期限投资中, 有一包含 d 种金融资产的资产组合 $x = (x_1, x_2, \dots, x_d)' \in \mathbf{R}^d$, 表示资产组合选择的决策向量, x_i 是期初对第 i 种资产投资的权重, $\sum_{i=1}^d x_i = 1$ (预算约束)。 $\tilde{r} = (\tilde{r}_1, \tilde{r}_2, \dots, \tilde{r}_m)'$ 表示影响资产组合收益的不确定性随机向量, 用 r_i 表示。如: 利率、汇率、股票指数、价格风险因子等, $\tilde{r} \in \mathbf{R}^m$ 。对于任意给定的决策变量 x , 在持有期期末, 资产组合收益率 $\tilde{r}_p(x, \tilde{r}) = x' \cdot \tilde{r} = \sum_{i=1}^d x_i \tilde{r}_i$, 为一随机变量, 由期初决策变量 x 和随机向量 \tilde{r} 决定。分布函数是对随机变量最完整的描述, 在持有期期末, 资产组合的收益的分布函数为: $F(x, u) = P\{r(x, \tilde{r}) \leq u\}$ 。可以用一确定的离散情景集合 $K =$

$\{s: s = 1, 2, \dots, S\}$ (为了使离散分布更接近连续分布, S 的取值很大, 本文取 $S = 15\,000$) 代表未来的不确定性^[11]。某一特定情景 $s \in K$ 发生的概率为 $p_s >$

0, 满足 $\sum_{s=1}^S p_s = 1$ 。在此情景下的资产收益率 $r_s = (r_{1s}, \dots, r_{ds})'$, 资产组合收益率 $r_p(x, r_s) = x' \cdot r_s = \sum_{i=1}^d x_i r_{is}$, 资产组合收益率的期望值 $\bar{r} = \sum_{s=1}^S p_s r_{s0}$ 。

根据资产组合中各种金融资产在未来各种情景下收益率, 对于指定的决策变量 $x = (x_1, x_2, \dots, x_d)$, 可以计算资产组合收益率在未来各种情景下的分布函数, 进而能够计算出资产组合收益率的标准差、VaR 和 CVaR 等风险指标。资产组合选择就是在影响资产组合价值的风险因子分布确定, 满足风险与期望收益等约束条件, 通过事前 (ex ante) 决策对资产组合内各种金融资产进行有效配置, 实现事后 (ex post) 收益最大, 损失最小。

根据前面的分析, 可以利用 Copula 函数构造多元分布函数, 对未来 S 种情景下的金融资产的收益率进行模拟, 可以得到 S 个 $r_p(x, r)$ 。则资产组合的 VaR- τ 、CVaR- τ 和标准差 W 分别为:

$$\text{VaR}_{1-\tau}(x) = -\inf\{R | P[\tilde{r}_p(x, r) < R] > \alpha\} \quad (7)$$

$$\text{CVaR}_{1-\tau}(x) = E[-\tilde{r}_p(x, r) | -\tilde{r}_p(x, r) \geq \text{VaR}_{1-\tau}(x)] = \text{VaR}_{1-\tau}(x) + \int_{\text{VaR}_{1-\tau}(x)}^{\infty} [-\tilde{r}_p(x, r) - \text{VaR}_{1-\tau}(x)] p(y) dy \quad (8)$$

对于离散分布, CVaR- τ 可用下式表示:

$$\text{CVaR}_{1-\tau}(x) = \text{VaR}_{1-\tau}(x) + \frac{1}{1-\tau} \sum_{k=1}^S [-r(x, r_k) - \text{VaR}_{1-\tau}(x)] \quad (9)$$

$$W = \sqrt{E[(\tilde{r}(x, \tilde{r}) - E(\tilde{r}(x, \tilde{r})))^2]} \quad (10)$$

若用 $\text{risk}(x, \tilde{r})$ 统一表示上述 VaR- τ 、CVaR- τ 和标准差 W , $E(x, \tilde{r})$ 表示资产组合的期望损失, 则资产组合选择模型的一般表示为:

$$\left. \begin{aligned} \min \text{risk}(x, \tilde{r}) \\ E(x, \tilde{r}) \geq u \\ \text{s. t. } x' \cdot e = 1 \\ x_i > 0 \end{aligned} \right\} \quad (11)$$

式中: u 为常数; $e = (1, \dots, 1)^T$; $x_i > 0$, 表示不允许卖空。根据前面的分析, 在非正态分布条件下, 对式 (11) 数学规划不能用传统的方法求得最优解, 本论文将设计混合遗传算法, 对式 (11) 的数学规划求解。

需要指出的是: 当用方差或标准差度量风险时, 式 (11) 的资产组合选择模型也不是马克维茨的均值

方差模型 马克维茨的资产组合选择模型用协方差矩阵表示资产收益率的线性相关关系,然后根据二次规划求解 本文的均值 方差模型是基于 Copula 函数的资产组合选择模型 首先根据 Copula 函数构造多元分布函数,反映组合资产收益的相关关系,然后采用混合遗传算法求解,不需要计算协方差矩阵 这样既可以反映随机变量之间的真实相关关系,能够适应复杂的情况,又能够克服马克维茨的资产组合选择模型在计算中对协方差矩阵的限制 因此,本文的均值 方差资产组合选择模型是一种修正的均值 方差资产组合选择模型

2 基于混合遗传算法的资产组合选择计算流程设计

2.1 基于 GA 和单纯形算法的混合遗传算法

针对 GA 和单纯形算法 (Nelder-mead) 的不足^[12-15],综合两种方法的优势,本文按照以下思路构建混合 GA 首先利用 GA 的全局搜索 快速收敛特性,搜索到包含最优解的解区域 通常,传统 GA 只能搜索到的全局最优解附近 然后,以 GA 搜索的最优解为初始解,在其周围的解空间里构造解初始单纯形,利用单纯形算法进行局部搜索,直到达到一定的收敛准则为止,此时的最优解近似为全局最优解 混合 GA 算子的选择及参数的设置如下:

(1) 初始种群的生成 指定初始种群规模 (Pop size) 在种群的初始化阶段,为了使初始种群在解空间的分布较均匀,在不增加太多计算时间的前提下,应使初始种群的规模大一些 种群数目的选择与所解决问题变量的多少有关系,随着变量的增加,初始种群的规模也应该大一些 另外,初始种群数目的适当增大可以减少用单纯形算法局部搜索的计算时间 由于资产组合中资产数目较多,本文经过多次实验,证明在实证中初始种群规模设为 300 时的效果较好

(2) GA 算子选择与设计 本文采用基于序 (rank) 的选择算子 它的特点是个体被选择的概率与目标函数的具体值无关 将种群中的所有个体按其目标函数值的大小进行由大到小的降序排列 定义每个基于序的 i 个体被选择的概率为 p_i $p_i = q' (1 - q')^{r-1}$, q' 表示选择最好的个体的概率;通常 $q' = 0.08$; r 表示个体的列顺序, 1 表示最好; $q' = \frac{q}{1 - (1 - q)^p}$; p 表示种群的规模 由于该选择算子只与目标函数的排序有关,可以使对目标函数贡献大的个体被以较大的概率选中

交叉算子是 GA 区别于其他进化算法的重要特征,在算法中起关键作用,是产生新个体的主要方法 交叉算子的设计和实现与所研究的问题密切相关,一般要求既不过多破坏群体中的优良个体,又要能够较好地产生一些新个体 本文采用算术交叉算子,该算子由两个个体的线性组合产生两个新的个体 算术交叉算子专门针对采用浮点编码的种群进行操作 首先依据交叉概率 p_c 随机确定进行交叉操作的父代,并两两随机配对,其中任意一对 (x_i, x_j) 按下式交叉产生两个新个体: $x_i^{t+1} = U x_i^t + (1 - U) x_j^t$ 和 $x_j^{t+1} = U x_j^t + (1 - U) x_i^t$ $U \in (0, 1)$, 为进化代数决定的变量; t 为进化代数

变异操作的主要目的是改善算法的局部搜索能力,并维持种群的多样性,防止发生早熟现象 均匀变异算子包括单点和多点非均匀变异算子 单点非均匀变异算子指按一定变异概率 p_m 随机产生变异操作的父代 x_i ,并按下面方式进行变异:

$$x_i' = \begin{cases} x_i + (b - x_i)f(G) & r_1 < 0.5 \\ x_i - (b - a_i)f(G) & r_1 \geq 0.5 \\ x_i & \text{其他} \end{cases} \quad (12)$$

式中: $f(G) = \left[r_2 \left(1 - \frac{G}{G_{\max}} \right) \right]^b$; r_1, r_2 为属于 $U(0, 1)$ 均匀分布的随机变量; G 为当前进化代数; G_{\max} 为允许的最大进化代数; b 为系统参数 a, b 分别为上下限

与单点非均匀变异算子不同,多点非均匀变异算子的变异范围较大,对所有父代个体都进行单点非均匀变异运算 这样可以使在进化初期,搜索在较大空间中进行,有利于维持种群的多样性,有利于提高算法的全局搜索能力 进化后期,使得最优解的搜索集中在最有希望的重点局部区域,在一定程度上可以有效地防止早熟现象的发生,有利于提高算法的局部搜索能力 为此,本文选用多点非均匀变异算子

(3) GA 的终止和单纯形算法的转入 GA 通过选择、交叉、变异操作,完成了一次进化,得到了一个新的种群,重复上述过程,直到满足停止准则条件 通常最为常用的停止准则是指定一个最大的进化代数 当 GA 进化的次数满足指定的进化代数时,GA 收敛到一个单一的解 在混合 GA 中,选择适当的进化代数对于整个算法有着重要的影响 进化代数太少,很难收敛到含最优解的区域,GA 的全局寻优能力没有得到充分的发挥 而进化代数太大,会发生早熟现象,导致程序计算效率较低 因此,应在 GA 发生早熟的情况下,将 GA 搜索的最优解输出,以其为初始解,转入单纯形算法,进行局部搜索 这里关键是早熟的判断 早熟的判断一般有两种方法:① 种群

的最优个体的适应值连续几代不发生变化;②种群中最优个体的适应值小于一个事先给定的较小的正数。根据本文所研究的问题,当初始种群为 200~300 时,通常在进化到 350 代左右时,其后 20~30 代的适应值基本不发生变化或变化不大,发生早熟现象,因此,本文设定的最大进化代数 400。

(4) 初始单纯形的构建及用单纯形算法进行局部搜索。由于 GA 最终只输出一个解,而用单纯形算法解决含有 n 个变量的优化问题,需要对由 $n+1$ 个解构成的初始单纯形进行操作。因此,首先以上述 GA 得到的最优解为基础,按照以下方式构建初始单纯形。

给定 $X^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})^T$, 令 $X^{(i)} = X^{(0)}$, 对于 $j = 1, 2, \dots, n$, 如果 $x_j^{(i)} \neq 0$, 则 $x_j^{(i)} = (1 - W_1)x_j^{(i)}$; 否则, 如果 $x_j^{(i)} = 0$, 则 $x_j^{(i)} = W_2$ (其中: $W_1 = 0.05, W_2 = 0.00025$)。然后令 $X^{(i+1)} = X^{(i)}$, 重复上述计算 n 次。这样就在 X_0 的周围构建了由 $n+1$ 个解构成的初始单纯形。然后,按照单纯形算法进行局部搜索,得到的最优解为整个混合 GA 的最优解。

2.2 用经典函数测试混合 GA

为了验证本文所设计算法的有效性,选择几个具有代表性比较复杂的优化测试函数,对没有单纯形的 GA 和带有单纯形的混合 GA 进行测试,并比较它们的性能。测试函数包括: De Jong 函数、Rosenbrock 函数、Griewangk 函数、Ackley 函数^[16]。通常评价 GA 计算性能的指标包括计算效率指标和计算质量指标。为了避免初始种群对实验结果的影响,分别对上述几个函数采用 GA 进行各 50 次优化实验。并以在指定最大进化代数时,目标函数成功收敛到最优解析解的次数(比率),目标函数最小值平均值和方差等指标来评价算法的性能。从各个指标的对比中可以看到,本文设计的混合 GA 的效率较高。这里省略实验的结果统计。

3 实证研究

3.1 资产组合选择模型适应值函数的设计

利用第 2 节中设计的混合 GA,对式(11)的资产组合选择模型求解。在利用 GA 计算过程中,为了评价个体,应该结合所要解决的问题设计合适的适应值函数。在具有等式或不等式的线性或非线性规划中,由于对染色体做遗传运算通常获得不可行的后代,故运用 GA 解线性或非线性规划的核心是如何满足约束的问题。近年来已经提出了几种用 GA 满足约束的技术,这些技术大致主要包括拒绝策略、修复策略、改进遗传算子策略和惩罚策略。其中前 3

种策略的共同特点是不会产生不可行解,缺点是无法考虑可行域外的点。对于约束严的问题,不可行解在种群中的比例很大,这样将搜索限制在可行域内很难找到可行解。惩罚策略就是这类在遗传搜索中考虑不可行解的技术。惩罚策略本质上是通过惩罚不可行解将约束问题转为无约束问题。在约束算法中,惩罚技术用来在每一代的种群中保持部分不可行解,使遗传搜索可以从可行域和不可行域两边来达到最优解。惩罚策略的关键是如何设计一个惩罚函数 $p(x)$,从而能有效地引导遗传搜索达到解空间的最好区域。

对于 1.4 节中的资产组合选择问题,综合考虑所有约束条件,可以构造以下惩罚函数:

$$p(x, \tilde{r}) = [\max\{0, -(E(x, \tilde{r}) - u)\}]^2 + (x' \cdot e - 1)^2 + \sum_i^n [\max\{0, -x_i\}]^2 \quad (13)$$

并构建辅助适应值函数:

$$F(x, \tilde{r}) = \text{risk}(x, \tilde{r}) + W \cdot p(x, \tilde{r}) \quad (14)$$

当 x 为可行点时, $p(x, \tilde{r}) = 0$; 当 x 为不可行点时, $p(x, \tilde{r}) > 0$ 。对于极小化问题, W 取一个很大的正实数; 对于极大化问题, W 取一个很小的负实数。通过这样设计辅助适应值函数,就可以把具有约束的资产组合选择问题转化成无约束的资产组合选择问题。在实际的计算中,惩罚因子 W 的选择十分重要。如果 W 过大,则给惩罚函数的极小化增加计算上的困难; 如果 W 太小,则惩罚函数的极小点远离约束问题的最优解,计算效率很差。因此,一般的策略是取一个趋向无穷大的严格递增正数列 $\{W_k\}$ 。通常取 $W_{k+1} = W_k, k = 0, 1, \dots, +\infty, W_0 > 0$ 为选取的初始惩罚因子,递增系数 $c = 5 \sim 10$ 。从某个 W_1 开始,对于每个 k ,求解

$$\min \text{risk}[(x, \tilde{r}) + W \cdot p(x, \tilde{r})] \quad (15)$$

从而得到一个极小点的序列 $\{\tilde{x}_k\}$,在适当的条件下,这个序列将收敛到约束问题的最优解^[17]。合理的 W 应满足 $W_p(x_k) < \epsilon$ (ϵ 非常小)。根据计算实验,对于资产组合选择问题,可选 $W = 150$,此时,求得最优解满足 $W \cdot p(x) < 10^{-2}, F(x, \tilde{r}) \approx \text{risk}(x, \tilde{r})$ 。

3.2 收益率非正态分布条件下资产组合选择

本文利用中国证券市场的实际数据对建立的资产组合选择模型和算法进行验证。根据一些学者研究的结果和经验,在中国股票市场的发展中,大盘股

* 数据来源: 北京大学色诺芬金融数据服务公司,并且采用负对数收益率。资产组合资产的构成: 真空电子、原水股份、华晨集团、上海石化、济南轻骑、悦达投资、马钢股份、华北制药、东方明珠、四川长虹。

与小盘股所表现出来的风险和收益特性不同 因此, 本文在构建资产组合时, 考虑的影响因素有各只股票上市的时间和流通股本的大小, 资产组合流通股本相对较大 由于中国证券市场在 1997 年以后实行涨跌停制度, 故样本从 1997 年开始 这样使样本数据具有可比性, 减少对实证结果的影响 根据顾岚等学者对中国资产组合规模的研究结果, 资产组合规模在 5~ 10 比较好. 本文实证研究中的资产组合由 10 只股票组成*, 数据样本期为 1997-01-02~ 2004-05-28 首先在 2004-05-28, 置信水平为 97. 5% 时, 根据 1. 3 中方法对资产组合内各种金融资产未来一期的收益率模拟 15 000 次 并假设未来一期资产组合的期望收益率为 0. 045%. 则可以根据上述混合 GA 求出使 VaR 最小时, 对每种资产的投资比重 图 1 表示 GA 的迭代过程

从图 1 可以明显看到, 在迭代到 350 代左右时, 适应值 (VaR) 基本不发生变化, 说明早熟现象发生 此时最小 VaR 值等于 4. 214 如果采用本文的混合 GA, 最小 VaR 值等于 3. 788, 最小 VaR 值降低了 10% 左右 同理可以变化期望收益率, 根据上述算法求出在收益率非正态分布条件下, 基于均值-标准差、均值-VaR 均值-CVaR 资产组合选择边界, 如图 2 所示 由于组合内有 10 种资产, 要求得到资产组合选择边界, 计算量相当大. 为此, 本文采用 MATLAB6. 5 编程模拟资产组合选择边界曲线和对各种股票投资的权重.

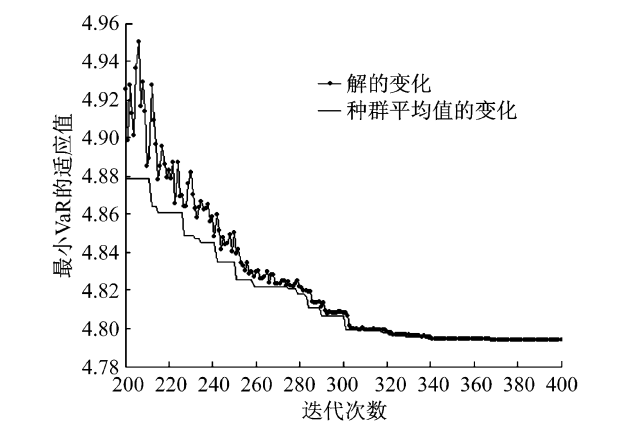


图 1 遗传算法的寻优性能

由于这里假设存在卖空约束条件, 故资产组合的边界是一条连续的曲线段. 由于 GA 是一种随机优化方法, 本文也对每一条边界曲线进行多次重复模拟, 每次得到的边界基本重合, 说明本文提出的资产组合选择模型和设计的混合 GA 比较可信

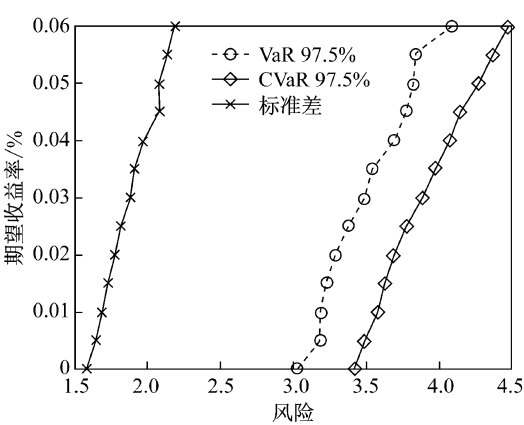


图 2 基于不同风险度量方法的资产组合选择边界

4 结 语

本文针对传统资产组合选择模型采用方差或标准差度量风险的不足, 以及不能真实有效地反映组合内资产收益相关性, 采用 VaR 和 CVaR 度量风险, Copula 函数反映组合内资产收益相关性, 构建了基于 Copula 函数的资产组合选择模型 针对 VaR 非凸性和分布函数不连续性导致资产组合选择优化计算复杂、不精确的难题, 设计了基于单纯形和传统 GA 的混合 GA, 并采用该混合 GA 对本文建立的资产组合选择模型求解. 实证研究证明该混合遗传算法在资产组合选择应用中具有可靠性

参考文献:

[1] Markowitz H. Portfolio selection [J]. Journal of Finance, 1952(7): 71- 93.

[2] Markowitz H. Portfolio selections Efficient Diversification of Investment [M]. New York: John Wiley & Sons, 1959.

[3] Nelsen R. An Introduction to Copulas [M]. Springer: Lecture Notes in Statistics, 1999.

[4] McNeil A J, Frey R. Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach [J]. Journal of Empirical Finance, 2000(7): 271- 300.

[5] Diebold F, Schuermann T, Stroughair J. Pitfalls and opportunities in the use of extreme value theory in risk management [J]. Journal of Risk Finance, 2000, 1(Winter): 30- 36.

[6] Ang A, Chen J. Asymmetric correlation of equity portfolio [J]. Journal of Financial Economics, 2002, 63(3): 443- 494.

[7] Erb Claude B, Harvey Campbell R, Viskanta, et al. Forecasting international equity correlation [J]. Financial Analysis Journal, 1994, 50: 32- 45.

[8] Login F, Solnik B. Extreme correlation of international equity markets [J]. Journal of Finance,

2001, 56(2): 649– 676

[9] 刘志东.资产组合风险度量与选择研究 [D].北京: 中国矿业大学 (北京)管理学院, 2003.

[10] 韦艳华,张世英,郭 焱.金融市场相关程度与相关模式的研究 [J].系统工程学报, 2004, 19(4): 355– 362.

[11] Jamshidian F, Zhu Y. Scenario simulation theory and methodology [J]. Finance Stochastic, 1997, 1: 43– 67.

[12] Nelder J A, Mead A. A simplex method for function minimization [J]. The Computer Journal, 1965, 7: 308– 313.

[13] Durand N, Alliot J M. A combined Nelder-Mead simplex and genetic algorithm [A]. Proceedings of the Genetic and Evolutionary Computation Conference GECCO 99 [C]. Morgan Kaufman, Orlando, FL, USA, 1999. 1– 7.

[14] Lagarias J C, Reed J A, Wright M H ,*et al.* Convergence properties of the Nelder-Mead simplex algorithm in low dimensions [J]. SIAM Journal Optimization, 1998, 9: 112– 147.

[15] McKinnon K. Convergence of the Nelder-Mead to a no stationary point [J]. SIAM Journal Optimization, 1998, 9: 148– 158.

[16] 飞思科技产品研发中心.辅助优化计算与设计 [M].北京: 电子工业出版社, 2003. 158– 170.

[17] 陈宝林.最优化理论与算法 [M].北京: 清华大学出版社, 2003. 480– 490.

(上接第 148页)

参考文献:

[1] 朱宝宪,何治国. U值和帐面市值比与股票收益的实证研究 [J].金融研究, 2002, (4): 71– 79.

[2] 汪 炜,周 宇.中国股市“规模效应”和“时间效应”的实证分析——以上海股票市场为例 [J].经济研究, 2002, (10): 16– 30.

[3] Fama E F, French K R. Common risk factors in the returns on stocks and bonds [J]. Journal of Financial Economics, 1993, (33): 3– 56

[4] Daniel Kent, Titman Sheridan, Wei K C John. Explaining the cross-section of stock returns in Japan: Factors or characteristics? [J]. Journal of Finance, 2001, (56).

[5] Fama Eugene F, French K R. Multifactor explanations of asset pricing anomalies [J]. Journal of Finance, 1996, (53): 55– 84.

下期发表论文摘要预报

一种新型的 VaR计算方法: $g-h$ VaR法

潘志斌¹, 田 澎², 朱海霞³

(1.华东师范大学 商学院,上海 200062; 2 上海交通大学 安泰经济与管理学院,上海 200052
3.上海海运学院,上海 200135)

摘 要: 根据 $g-h$ 分布的统计特性,提出了基于金融资产损失的 VaR计算方法—— $g-h$ VaR法.该方法结合了分析方法、历史模拟方法和极值理论方法的优点,能够很好地处理组合回报的不对称现象和厚尾现象.实证研究表明,该方法优于现有的德尔塔-正态方法.

ArcGIS在市场交通污染分析中的应用

侯立文¹, 谭家美², 蒋 馥¹

(1.上海交通大学 安泰经济与管理学院,上海 200052
2 上海海事大学 交通运输学院,上海 200135)

摘 要: 简要介绍了 ArcGIS在城市交通污染分析中的作用,然后根据机动车尾气排放模型和扩散模型,构建了基于 ArcGIS的城市交通大学污染分析系统框架.最后,指出了 ArcGIS在城市环境污染监测和管理体系中的重要地位.