基于遗传算法的概率准则组合证券模拟求解①

王燕青, 唐万生, 韩其恒
(天津大学管理学院系统工程研究所, 天津 300072)

摘要:针对概率准则意义下的组合证券投资模型,采用随机模拟技术和遗传算法相结合的思想, 设计出求解算法, 并用 Matlab 语言实现. 求解算法适用于证券收益率服从任意分布的情况, 甚至不考虑证券收益率分布, 用实际数据进行模拟和优化. 实例证明, 该算法有很好的收敛性及较高的计算效率.

关键词:概率准则; 组合证券投资; 遗传算法; 随机模拟; Matlab 语言
中图分类号:F830 文献标识码 A: 文章编号:1007 - 9807(2002)06 - 0029 - 05

0 引 言

组合证券投资, 是指为避免或分散大的风险, 投资者将资金分散投资到若干种证券, 以降低风险[1,6—10,13]. 一般讲, 投资者的预期收益越大,承担的风险也越大. 不同投资者, 或同一投资者在不同市场条件下, 对获利性的要求都不相同, 但无论其预期收益为何, 总是希望找到一组投资, 使其实现的可能性最大, 故提出了一种在概率准则意义下的最优组合投资模型[2,14,15], 即模型以实现预期收益的概率为目标函数, 在满足约束的条件下, 选择投资比例, 使目标函数达到最大.

随机模拟(Monte Carlo 模拟)是一种实现随机(或确定)系统抽样实验的技术. 虽然随机模拟是一种不很精确的技术, 它提供的仅仅是一个统计估计而非一个精确解, 但对那些无法用解析方法处理的模型, 随机模拟却是一种获得问题答案的好方法.

遗传算法(genetic algorithm)是一种基于自然选择和自然遗传机制的搜索算法[5,11,12]. 它吸取了自然界的自然选择、适者生存以及遗传、变异的思想, 从一组初始解群开始迭代, 逐步淘汰较差的解, 产生更好的解, 直到满足某种收敛指标为止. 大量试验表明, 利用遗传算法不仅可以获得较好的全局最优解, 而且算法简单, 收敛速度快, 计算时间短, 是求解组合优化问题的有效算法.

本文首先用传统方法求解正态性假设下概率准则的最优组合证券投资模型, 其后假定收益率不服从正态分布, 利用计算机, 根据随机模拟技术和遗传算法相结合的思想给出该模型的求解算法, 并用 Matlab 语言予以实现.

1 数学模型

选定 n 种证券, ξ_i 表示第 i 种证券持有期收益率, $E\xi_i = r_i$, $i = 1,2,\cdots,n$. 令 $\xi = (\xi_1,\xi_2,\cdots,\xi_n)^{\mathrm{T}}$, $\xi \sim \mathrm{N}(\mu, \mathrm{Q})$, 其中 $\mu = (r_1, r_2, \cdots, r_n)^{\mathrm{T}}$, Q 为随机向量 ξ 协方差阵, 通常以方差作为证券风险的度量指标. 假定各种证券投资比例为 x_i, $i = 1,2,\cdots,n$, 要求 $\sum_{i=1}^{n} x_i = 1$, 所以组合证券的收益率为 $Y = \sum_{i=1}^{n} x_i\xi_i$. 令 $x = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$, 则有 $Y \sim N(\mu^{\mathrm{T}}x, x^{\mathrm{T}}\mathrm{Q}x)$.

① 收稿日期:2001 - 01 - 29; 修订日期: 2001 - 11 - 21.
基金项目:国家自然科学基金资助项目(70171004); 天津市自然科学基金资助项目(013602611).
作者简介:王燕青(1974 —),女,河南汤阴人,博士生.

在证券市场中,为了维持稳定性,往往不允许卖空. 不允许卖空时的数学模型为

$$\max_x P(Y \geq R)$$

$$\text{s.t.}\begin{cases}\sum_{i=1}^{n} x_i = 1 \\ x_i \geq 0 \quad i = 1,2,\cdots,n\end{cases} \tag{1}$$

其中:R 为预期的收益率;$P(Y \geq R)$ 为实现组合证券收益率不低于预期水平的概率值.

为便于进一步讨论,将模型(1) 中的目标函数作如下变换. 因为

$$Y = \sum_{i=1}^{n} x_i \xi_i \sim N(\mu^{\mathrm{T}} x, x^{\mathrm{T}} Q x)$$

令 $U = \frac{Y - \mu^{\mathrm{T}} x}{(x^{\mathrm{T}} Q x)^{1/2}}$, 则 $U \sim N(0,1)$.

故

$$P(Y \geq R) = P\left(\frac{Y - \mu^{\mathrm{T}} x}{(x^{\mathrm{T}} Q x)^{1/2}} \geq \frac{R - \mu^{\mathrm{T}} x}{(x^{\mathrm{T}} Q x)^{1/2}}\right) = P\left(U \geq \frac{R - \mu^{\mathrm{T}} x}{(x^{\mathrm{T}} Q x)^{1/2}}\right) = \frac{1}{\sqrt{2\pi}} \int_{\frac{R - \mu^{\mathrm{T}} x}{(x^{\mathrm{T}} Q x)^{1/2}}}^{+\infty} \exp\left(-\frac{1}{2} t^2\right) \mathrm{d}t$$

则目标函数变形为 $\min_x \frac{R - \mu^{\mathrm{T}} x}{(x^{\mathrm{T}} Q x)^{1/2}}$.

所以,模型(1) 可表示为下面的形式

$$\min_x f(x) = \min_x \frac{R - \mu^{\mathrm{T}} x}{(x^{\mathrm{T}} Q x)^{1/2}}$$

$$\text{s.t.}\begin{cases}\sum_{i=1}^{n} x_i = 1 \\ x_i \geq 0 \quad i = 1,2,\cdots,n\end{cases} \tag{2}$$

2 随机模拟

模型(2) 是在假定 ξ 服从正态分布的前提下进行讨论的. 但由于证券的收益率会随着时间的推移而经常无规律地变化,可能不服从正态分布,设 $\xi \sim F(\cdot)$,此时的 $P(Y \geq R)$ 不一定有相应的解析表达式,故很难写出其确定型等价类模型,计算起来也很困难. 所以对于这类模型,需采用随机模拟来解决.

本文模型(1) 中的概率值 $P(Y \geq R)$ 可由随机模拟(Monte Carlo 模拟) 方法近似计算,具体操作过程如下:

步骤 1 对固定的 x,置 $N' = 0$;

步骤 2 从概率分布 $F(\cdot)$ 中生成随机向量 $\xi^{(i)}$,$(i = 1,2,\cdots,N)$;

步骤 3 若 $Y = x^{\mathrm{T}} \xi^{(i)} \geq R$,$i = 1,2,\cdots,N$,则 $N' = N' + 1$;

步骤 4 重复步骤2 和步骤 3,共 N 次;

步骤 5 根据大数定律,则有 $P(Y \geq R) \approx \frac{N'}{N}$.

3 遗传算法

本文用遗传算法求解模型(1) 的基本思想是:随机产生一种群,然后反复进行遗传操作(交叉,变异等) 并进行评价,从而得到近似最优解. 具体算法如下:

1) 初始种群的产生

一般地,种群越大,搜索范围越广,越容易获取全局最优解. 然而数目越大,每次迭代时间越长,所以通常定义种群大小 10 到 160 之间. 本文选取种群为 20 条染色体. 具体编码和操作如下:

随机产生 $\left\{x^{(i)}\right\}_{i=1}^{\text{pop\_size}}$,其中:$x^{(i)} = (x_1^{(i)}, x_2^{(i)}, \cdots, x_j^{(i)}, \cdots, x_n^{(i)})^{\mathrm{T}}$,并要求 $\sum_{j=1}^{n} x_j^{(i)} = 1$ 且 $x_j^{(i)} \geq 0$,其中 $x_j^{(i)}$ 表示第 i 条染色体上第 j 种证券的投资比例.

2) 适值函数的选取

在一般的遗传算法中,用适应度函数来评价染色体的优劣,并在此基础上进行各种遗传操作. 因此适应度函数的选择是否适当对算法的性能好坏影响很大,应根据实际问题的特性来具体确定.

本文采用一种基于指数适值的适值函数[11]. 首先,将所有的染色体按目标函数值的升序进行排序;然后,定义三个偏好参数 ρ_1, ρ_0, ρ_2 $(0 < \rho_1 < \rho_0 < \rho_2 < 1)$,分别用来确定 3 个标准染色体 u_1, u_0, u_2,使得 $u_1 = \lceil \rho_1 \cdot \text{pop\_size} \rceil$,$u_0 = \lceil \rho_0 \text{ pop\_size} \rceil$ 和 $u_2 = \lceil \rho_2 \cdot \text{pop\_size} \rceil$,符号 $\lceil x \rceil$ 表示大于 x 的最小整数. 其中 ρ_1 和 ρ_2 是设计来删除具有极端值的染色体的. 则第 i 条染色体的适值函数为

$$\mathrm{eval}(i) = \begin{cases}\exp\left[-\frac{i - u_0}{u_1 - u_0}\right] & i < u_0 \\ 2 - \exp\left[-\frac{i - u_0}{u_2 - u_0}\right] & i \geq u_0\end{cases}$$

3) 选择运算

计算第 i 条染色体的选择概率 p_i 和累积概率

q_k,随机产生一小数 $r \in [0,1]$,若 $q_{k-1} < r < q_k$,则选择第 k 条染色体,重复 pop_size 次,这样可以得到 pop_size 个复制的染色体.

4) 交叉运算

从经选择运算后得到的种群中随机产生 2 条染色体 $x^{(r_1)}$ 和 $x^{(r_2)}$,且随机产生一小数 $c \in [0,1]$,按

$$x^{(r_1)} = cx^{(r_1)} + (1 - c)x^{(r_2)}$$

$$x^{(r_2)} = cx^{(r_2)} + (1 - c)x^{(r_1)}$$

进行交叉操作,产生两个新的染色体 $x^{(r_1)}$ 和 $x^{(r_2)}$.

重复操作 pop_size/2 次.

5) 变异运算

从区间 $[0,1]$ 中产生随机数 r,若 $r < p_m$(变异率),则按自由方向 d 变异染色体,即 $x^{(k)} = x^{(k)} + m \cdot d$,其中 m 为一个大的正数,d 为 n 维空间的一随机方向.

如果 $x^{(k)}$ 不可行,即 $x^{(k)}$ 中的各分量之和不为 1 时,则置 m 为 0 和 M 之间的随机数,直至可行为止,其中 M 是初始化过程定义的一个足够大的数.如果在预先定义的迭代次数之内没有找到可行解,则置 $M = 0$,无论 M 为何值,总用 $x^{(k)} + m \cdot d$ 代替 $x^{(k)}$.

重复操作次 pop_size 次.

6) 下一代染色体的产生

从父代及经选择、交叉、变异产生的后代中按目标函数值的升序进行排序,选择最好的染色体,组成新的种群.

转至 3),直至 gen = max_gen.

7) 结束准则

遗传算法是一种反复迭代的搜索方法.它通过多次进化逐渐逼近最优解,而不是恰好等于最优解,因而需要确定终止条件.最常用的终止方法是规定遗传(迭代)的代次.刚开始时,迭代次数小一些,如规定 100 次,然后视情况逐渐增加次数,可达到上千次.

所以,当迭代次数大于要求的最大的迭代次数时,求解结束.

对于上述的算法,作者已使用 Matlab 语言实现.在实际使用中,只需要输入相应的参数即可很快求得相应的解,但要注意所得到的解可能只是近似最优解.

4　模型求解与实例应用

4.1　模型求解

当 ξ 服从正态分布时,模型(1) 的确定性等价类模型为模型(2).对模型(2) 可以按照传统的方法求解.

首先对该模型做恒等变换,得到以下形式

$$\min_x f(x) = \min_x \frac{R - \mu^{\mathrm{T}} x}{(x^{\mathrm{T}} Qx)^{1/2}} \tag{3}$$

$$\text{s.t.}\begin{cases} h(x) = \sum_{i=1}^{n} x_i - 1 = 0 \\ g_i(x) = x_i \geq 0 \quad i = 1,2,\cdots,n \end{cases}$$

令 $F = (1,1,\cdots,1)^{\mathrm{T}} \in R^n$,因为 $\nabla h(x) = F$,$\nabla g_i(x)$ 均为单位向量,并且 $g_i(x) = x_i$,$(i = 1,2,\cdots,n)$ 不全是起作用的约束,所以起作用的约束的梯度是线性无关的.本模型的最优解满足 Kuhn-Tucker 条件

存在 γ 和 $\Lambda = (\lambda_1,\lambda_2,\cdots,\lambda_n)^{\mathrm{T}}$ 使得模型最优解满足下列条件

$$\begin{cases} \nabla f(x^*) - \gamma \nabla h(x^*) - \sum_{i=1}^{n} \lambda_i \nabla g_i(x^*) = 0 \\ \lambda_k g_k(x) = 0 \qquad k = 1,2,\cdots,n \\ \lambda_k \geq 0 \qquad k = 1,2,\cdots,n \end{cases}$$

即

$$\begin{cases} \nabla f(x^*) - \gamma F - \Lambda = 0 \\ \lambda_k g_k(x) = 0 \qquad k = 1,2,\cdots,n \\ \lambda_k \geq 0 \qquad k = 1,2,\cdots n \end{cases}$$

当 ξ 不服从正态分布时,模型(1) 无法转化成上述的确定性等价类问题,从而也很难用传统方法求解,但可用本文给出的算法求解.

4.2　实例应用

例 1　证券的收益率服从正态分布 $N(\mu, Q)$

利用文[8] 给出的一个 7 种证券组合模型进行计算仿真.其中,收益率协方差阵为

$$Q=\begin{bmatrix}0.142 & -0.188 & 0.168 & 0.189 & -0.185 & -0.193 & 0.162\\ -0.188 & 0.266 & -0.219 & -0.247 & 0.254 & 0.256 & -0.215\\ 0.168 & -0.219 & 0.225 & 0.239 & -0.216 & -0.237 & 0.210\\ 0.189 & -0.247 & 0.239 & 0.275 & -0.246 & -0.259 & 0.221\\ -0.185 & 0.254 & -0.216 & -0.246 & 0.256 & 0.257 & -0.216\\ -0.193 & 0.256 & -0.237 & -0.259 & 0.257 & 0.293 & -0.229\\ 0.162 & -0.215 & 0.210 & 0.221 & -0.216 & -0.229 & 0.210\end{bmatrix}$$

各种证券的期望收益率为 $\mu=(0.12\quad 0.09\quad 0.10\quad 0.09\quad 0.09\quad 0.11\quad 0.11)^T$,预期组合证券收益率 $R=0.1$

1) 利用传统方法求解时,计算结果如下:

组合证券的最佳投资比例为

$x^*=(0.5846\quad 0.1696\quad 0.0000\quad 0.0000\quad 0.0184\quad 0.2274\quad 0.0000)^T$

模型(1) 的概率最优值为

$P^*(Y\geq 0.1)=0.5980$

模型(2) 的目标函数最优值为

$f^*=f(x^*)=-0.251$

2) 利用随机模拟和遗传算法求解时,结果如下:

输入　交叉率 $p_c=0.5$

　　　变异率 $p_m=0.1$

　　　迭代次数 max. gen = 150

组合证券的最优投资比例为

$x^*=(0.5410\quad 0.0222\quad 0.0297\quad 0.0001\quad 0.1949\quad 0.2052\quad 0.0069)^T$

模型(1) 的概率最优值为

$P^*(Y\geq 0.1)=0.6390$

图 1 表示在不允许卖空的情况下,迭代次数与目标函数值的关系.

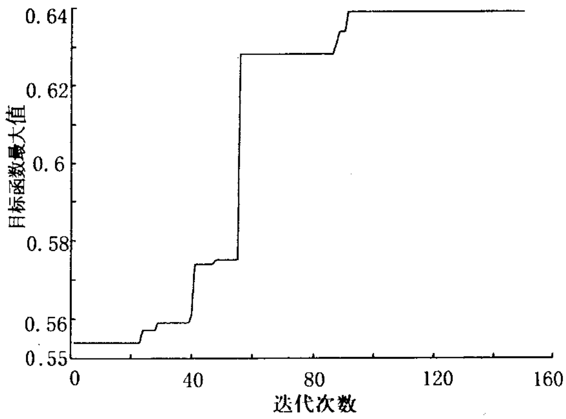

图 1　迭代次数与目标函数的关系

从计算结果和图表可以看出,该算法经过 92 次迭代后开始收敛,并且得到了该问题的近似最优解.

例 2　证券的收益率不服从正态分布

设证券的收益率服从混合正态分布. $\xi\sim f(x)$,其中,$f(x)=cf_1(x)+(1-c)f_2(x)$ 为混合正态分布的密度函数,$f_1(x)$ 是正态分布 $N(\mu_1,Q_1)$ 的密度函数,$f_2(x)$ 是正态分布 $N(\mu_2,Q_2)$ 的密度函数,常数 $c\in[0,1]$,其中,$c=0.3$.

$\mu_1=(0.0230, 0.0763, 0.3383, 0.7617, 0.4117, -0.2637)^T$

$$Q_1=\begin{bmatrix}1.6967 & 0.5052 & -0.4144 & -0.9764 & 0.1327 & 0.5611\\ 0.5052 & 1.7830 & -0.2389 & -0.0043 & 0.5711 & -0.0165\\ -0.4144 & -0.2389 & 7.4756 & 0.4553 & -0.5256 & -1.0731\\ -0.9764 & -0.0043 & 0.4553 & 4.2709 & 0.0454 & -0.3337\\ 0.1327 & 0.5711 & -0.5256 & 0.0454 & 5.5701 & -1.3716\\ 0.5611 & -0.0165 & -1.0731 & -0.3337 & -1.3716 & 2.2662\end{bmatrix}$$

$\mu_2=(0.0796, 0.3671, 0.3234, 0.1603, 0.0940, 0.5556)^T$

$$Q_2=\begin{bmatrix}2.1354 & 0.2848 & -0.2641 & 0.0597 & 0.2083 & 0.3237\\ 0.2848 & 6.6053 & 0.1319 & 0.2235 & 0.4514 & 1.1873\\ -0.2641 & 0.1319 & 6.9752 & -0.1091 & 1.3533 & -0.7517\\ 0.0597 & 0.2235 & -0.1091 & 2.2668 & -0.4064 & 0.3043\\ 0.2083 & 0.4514 & 1.3533 & -0.4064 & 3.3663 & -0.1656\\ 0.3237 & 1.1873 & -0.7517 & 0.3043 & -0.1656 & 2.9665\end{bmatrix}$$

利用随机模拟和遗传算法求解时:

输入　交叉率 $p_c=0.5$

　　　变异率 $p_m=0.1$

　　　迭代次数 max. gen = 100

　　　预期组合证券收益率 $R=0.18$

组合证券的最优投资比例为

$$x^* = (0.0034, 0.1768, 0.1932, 0.3354, 0.1687, 0.1224)^T$$

模型(1) 的概率最优值为

$$P^*(Y \geq 0.18) = 0.5920$$

图 2 表示迭代次数与目标函数值的关系.

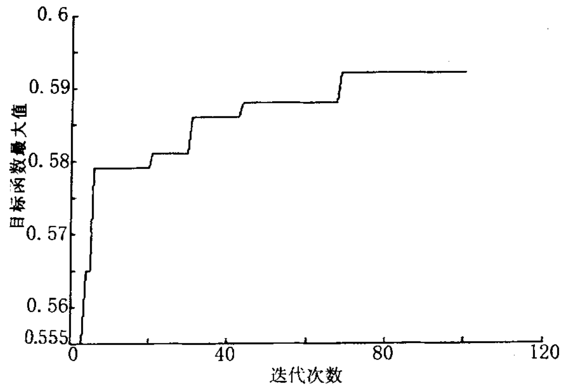

图 2　迭代次数与目标函数的关系

分析上面两种解法所得到的结果,基于遗传算法的随机模拟在求解概率准则组合证券模型时可行,且得到的解是全局最优解或近似最优解,而传统解法易陷入局部最优解.

5　结论

本文提出了一种在概率准则意义下组合投资模型,介绍了其传统求解方法,并按照随机模拟和遗传算法相结合的思想,给出了新的求解算法,还利用 Matlab 语言编制出该算法程序. 经实例仿真,验证了该算法的可行性. 本文算法突破了传统的各种证券收益率必须服从正态分布的假定,可对任意分布形式的模型求解,且可不考虑证券收益率分布,直接利用历史数据模拟及优化.

参 考 文 献:

[1]Markowitz H M. Portfolio Selection: Efficient Diversification of Investment[M]. 2nd ed . Cambridge: Basil Blackwell ,1991

[2]梁建峰,唐万生. 有交易费用的组合证券投资的概率准则模型[J]. 系统工程,2001,19(2): 5 —10

[3]Liu B D, Ku C F. Probability criterion in inventory system with varying stochastic demands[J]. J Systems Science and Math Science, 1994,14(9):211 —217

[4]Fogel D B. An introduction to simulated evolutionary optimization[J]. IEEE Tran Neural Networks,1994,(5):3 —14

[5]Gen M, Liu B D. Evolution program for deterministic and stochastic optimization[J]. European Journal of Operational Research,1996, 94(3): 618 —625

[6]Bielecki T R, Pliska S R. Risk-sensitive dynamic asset management[J]. Applied Mathematics & Optimization,1999,39(3): 337 —360

[7]唐小我,傅 庚,曹长修. 非负约束条件下组合证券投资决策方法研究[J]. 系统工程,1994,12(6):23 —29

[8]王 竹,唐小我,曹长修. 加权行和指标下组合证券投资风险最小化迭代算法[J]. 系统工程,1994;12(5):53 —58

[9]马永开,唐小我. 不允许卖空的多因素证券组合投资决策模型[J]. 系统工程理论与实践,2000,20(2):37 —43

[10]荣喜民,张喜彬,张世英. 组合证券投资模型研究[J]. 系统工程学报,1998,13(1):81 —88

[11][日]玄光男,程润伟. 遗传算法与工程设计[M]. 北京:科学出版社,2000

[12]刘宝碇,赵瑞清. 随机规划与模糊规划[M]. 北京:清华大学出版社,1998

[13]黄登仕. 金融市场的标度理论[J]. 管理科学学报,2000,3(2):27 —33

[14]Tang Wansheng, Han Qiheng, Li Guangquan. Cash management decision with probability criterion[J]. Proc IEEE/SMC,2001: 2670 —2673

[15]Tang Wansheng, Han Qiheng, Li Guangquan. The portfolio selection problems with chance-constrained[J]. Proc IEEE/SMC,2001: 2674 —2679

(下转第 52 页)

Continuous time revenue management with maximal & minimal reservation prices

WEI Yi-hua , HU Qi-ying
School of Economics & Management , Xi 'an Electronic Science & Technology University , Xi 'an 710071 ,China

Abstract : This paper discusses a continuous time yield management model with the following characters : customers arrive according to a nonhomogeneous Poisson process , each arriving customer has maximal & minimal reservation prices , the distribution of which depends on the time of arrival. We find that all results of the general yield management model still hold.

Key words : pricing ; yield management ; nonhomogeneous Poisson process ; optimal value function ; optimal policy

(上接第33页)

Stochastic simulation based genetic algorithm for solving portfolio problem with probability criterion

WANG Yan-qing , TANG Wan-sheng , HAN Qi-heng
Institute of Systems Engineering , School of Management , Tianjin University , Tianjin 300072 , China

Abstract : A stochastic simulation base genetic algorithm is devised for solving portfolio investment model with probability criterion. The algorithm is programmed by using Matlab. It can solve all kinds of problems of any distribution form ,even not considering the distribution of rate of the return. Examples show that this algorithm is convergent and efficient.

Key words : probability criterion ; portfolio investment ; genetic algorithm ; stochastic simulation ; Matlab