题 目：基于系统性风险角度的基金资产配置策略分析

关键词：聚类分析 均值-半方差模型 非线性二次规划 多项式逼近拟合 一维搜索求最优解算法

摘 要：

本文通过对数据进行分析和利用，度量了 2019 年不同基金公司的策略相似性，并设计出投资效用最大化的规划模型。在拟合大量数据后，配合风险价值模型设计出可以在投资效用最大化前提下风险价值最小的优化算法，为广大基金公司提供有效的投资建议。

针对问题一：通过对数据的宏观分析，得出数据在分布上的特点，研究基金公司的投资倾向，发现基金公司**投资分布均匀，更倾向于低股价股票**。接着对数据进一步进行皮尔逊相关系数计算和**聚类分析**，最终得出不同基金公司之间资产配置策略的相似性结论：**公司 5 和 10，8 和 9 的投资策略比较相似，公司 2，5，10 之间的投资策略也有着一定的相似性**。并且股票价格和基金公司投资占比之间是中等程度相关。

针对问题二：首先在分析分布直方图后选择对数收益率作为重要参数。接着通过均值-方差模型的延伸改进得到**均值-半方差模型**，并进一步修正普遍的效用模型得到投资效用最大化模型。为求解模型将其转化为线性条件约束的**非线性二次规划**模型，不断转化模型和提取系数矩阵，最终利用得到 matlab 中 quadprog() 函数成功求解最优的股票投资组合策略。

针对问题三：在使用 Eviews 对股票价格序列进行 ADF 检验后发现**大部分股票的价格序列都是不稳定的**。在拥有大量数据的前提下最终选用历史模拟法得出历史经验向量 VaR，通过简单的公式计算得出各基金公司 2020 年按老配置 95%置信水平下的首个交易日的风险价值，按照从高到低排序：**H、J、G、I、E、D、C、B、F、A**。

针对问题四：斟酌不同方法后采用**多项式逼近拟合**来预测 2020 年 45 天后、90 天后、180 天后的期望对数收益率。将问题二和问题三中的两种模型进行整合，衍生出**一维搜索求最优解算法**模型，套用不同的参数后分别得到不同情况下的投资组合策略和投资建议。最终得出结论：2020 年首个交易日可以得到最优组合，并且投资效用为 **1.2970** 和系统风险价值为 **1301301.1 万元**；45 天后首个交易日最优组合投资效用为 **0.8804**，系统风险价值为 **980831.7 万元**；90 天后首个交易日最优投资组合投资效用为 **0.7199**，系统风险价值为 **979967.9 万元**；180 天后首个交易日最优投资组合投资效用为：**1.1008**，系统风险价值为 **1047337.2 万元**。

综合模型可以得出结论，**基金公司可以根据自身经济情况和市场情况选取不同的参数，模型会给出短期最优组合和长期大体的推荐投资方向。**并且长期投资下会给出**不同风险下最佳投资区间和较大的风险能力调控范围**，方便基金公司使用。

一、问题重述

1.1 问题背景

近年来，随着改革开放程度的不断提高，我国经济运行中的各种风险逐渐暴露并集中传导和体现于金融领域。作为金融系统的重要组成部分之一，资产管理业务的风险状况同样是我国金融体制改革过程中值得关注的问题。企业在进行资产配置时常常会受到各种经济风险的影响，这使得企业不得不尝试去调整制定适合自身实际情况和战略定位的资产分配计划，以给自身带来一定的经济效益。因此如何平衡基金投资收益和系统性风险之间的关系值得深入探究。

1.2 问题提出

（1）问题一：根据材料所给的十家公募基金公司 2019 年所持有的股票种类以及数量、价值。比较不同基金公司之间资产配置战略的相似性。

（2）问题二：根据材料所给的 2019 年样本股票的价格统计数据，以所有基金的持股市值总和作为初始财富，努力追求企业投资效益的最大化，建立数学模型，确定最合适的股票投资组合策略。

（3）问题三：在 2020 年所有基金公司依旧采用 2019 年资产配置规划进行投资的前提下。根据材料所给数据，建立数学模型，度量每个公司 2020 年 95%置信水平下的风险价值，并比较排序。

（4）问题四：考虑 2020 年公司在同时满足投资效用最大化和风险价值最小化的基础上，建立数学模型，制定出最合适的股票投资组合策略，同时给出投资效用和风险价值。

二、问题分析

2.1、度量不同基金公司之间资产配置策略的相似性的分析：

本题要求根据已有的资产配置信息建立数学模型，并分析度量不同基金公司之间资产配置策略的相似性。想要知道各家公司的资产配置策略，则要对其所持股票进行分析。可以从宏观和微观两个方面入手。首先将每支股票的持股总市值除以持股总量得到平均股价，将每家公司所有股票的持股总市值相加可得每家公司的投资总额，随后将单只股票的持股总市值与投资总额相除便可以得到每只股票对应的投资占比。宏观上可以研究不同情况的分布，微观上我们可以建立起相关的数学模型进行分析和求解。

2.2、确定最优的股票投资组合策略以达到收益最大化分析：

本问题要求实现投资效用最大化，则需要我们根据已有得资料首先确定股票价格的收益率，然后采用正确的经济学模型对其进行构造规划模型，不断地转化分析已经构造好的模型进行求解，最终得到最优解。

2.3、度量每个基金公司 2020 年 95%置信水平下的风险价值的分析：

本问题我们需要计算风险价值。风险价值是指在一定的置信水平下，某一金融资产（或证券组合）在未来特定的一段时间内的最大可能损失。想要求得风险价值，则需要我们在股票价格变动模型的基础上来进一步确定投资结束时手上资产的最小值，我们也可以通过分析资产组合值的概率分布得出相同的结果。由于本题中所给的数据整体上较为全面,所以综合考虑下采用历史模拟法构建风险价值的数学模型。

2.4、确定最优的股票投资组合策略分析：

由题意可知这是双目标问题。首先要通过拟合得到 2020 年的一些重要参数，然后可以根据已有的模型进行组合构建新的模型，或者把不同的模型用新的算法进行

配合来求出最优解。

三、模型的假设

1) 投资者在考虑每一次投资选择时，其依据是某一持仓时间内的证券收益的概率分布。
2) 投资者是根据证券的期望收益率的方差或标准差估测证券组合的风险。
3) 投资者的决定仅仅是依据证券的风险和收益。
4) 在一定的风险水平上，投资者期望收益最大；相对应的是在一定的收益水平上，投资者希望风险最小。

四、变量说明

序号	符号	意义
1	R_i	证券预期收益率的期望值
2	ρ	总体相关系数
3	r	皮尔逊相关系数
4	σ^2	证券投资收益率的方差
5	R	选定的 n 种证券预期收益率的期望值向量
6	X	n 种证券投资比例系数向量
7	Q	n 种证券收益率的协方差矩阵
8	F	n 维单位列向量
9	x_i	证券的投资比例系数
10	R_0	证券组合的预期收益率
11	SV	证券投资收益率的半方差
12	U	效用值
13	A	投资者回避风险程度的指数

序号	符号	意义
14	w	资产组合的初始价值
15	R	持有期末的期望收益
16	μ	持有期末的期望收益的数学期望
17	σ	持有期末的期望收益的标准差
18	α	标准正态分布相应的分位数
19	c	给定的置信区间
20	W	资产组合的价值
21	S_i	第 i 天股票价格
22	R'	2019 年期望收益率矩阵
23	VaR	风险价值的经验向量

表 1 符号说明

五、问题一模型的建立与求解

5.1、资产配置策略相似性的宏观分析

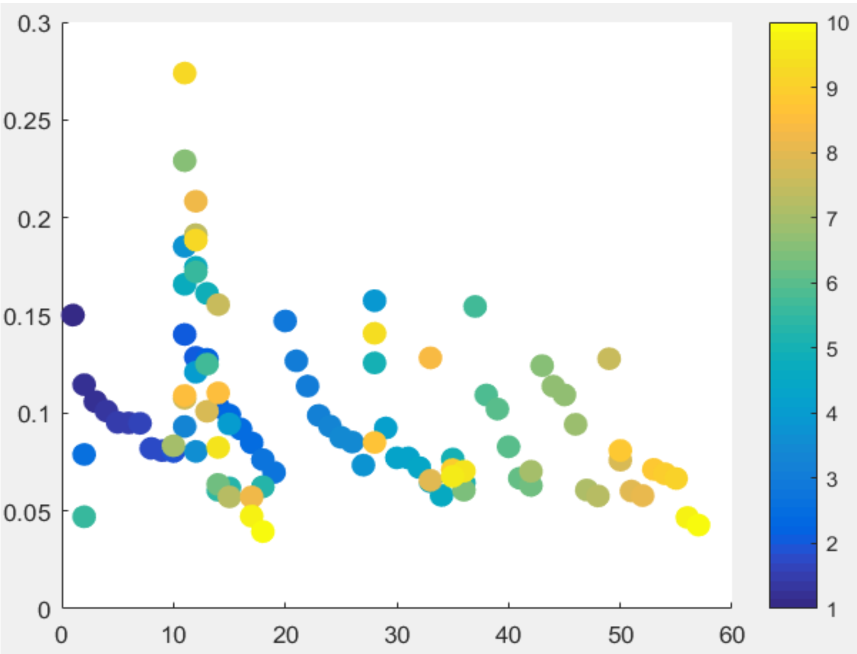
我们将已知的数据进行简单的处理：

1）将公司所持股票的持股总市值（万元）相加可得公司的投资总额（万元）。

2）公司投股总市值（万元）/投股总量（万股）=平均股价（元）

3）投资占比=单个股票投股总市值（万元）/投资总额（万元）

根据处理后的数据绘制股票种类和投资占比的散点泡沫图，如下：

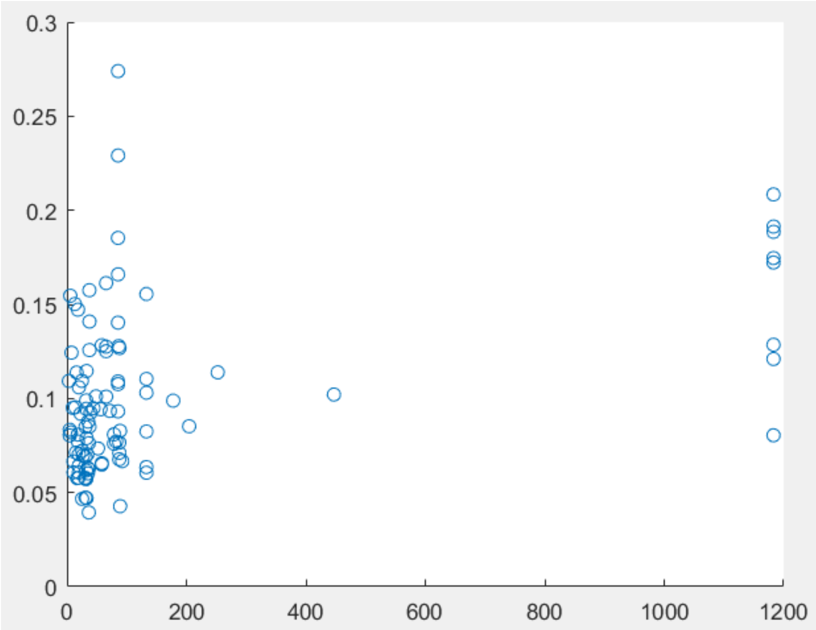


（图一：股票种类和投资占比的散点泡沫图）

此图中以横坐标表示股票的编号，纵坐标表示该股票的投资占比，图中 1——10 来对应资金公司的编号 A——J。

由此图我们可以看出各家基金公司在投资组合分布上还是比较平均的， 投资重复度上看相对的更青睐于股票 11。

换了一个角度继续进行分析，构建平均股价和投资占比的散点泡沫图以确定公司投资和股票价格的关系，如下：



（图二：平均股价和投资占比的散点泡沫图）

图中纵坐标表示股票的投资占比，横坐标表示平均股价，每一个点都代表了一只股票。

由上图可以得出结论为基金公司更青睐于平均股价较低的股票。

5.2、相关性和聚类分析

5.2.1、皮尔逊相关系数分析

在统计学中，皮尔逊相关系数(Pearson correlation coefficient）是用于度量两个变量 X 和 Y 之间的相关（线性相关)，其值介于-1 与 1 之间。两个变量之间的皮尔逊相关系数定义为两个变量之间的协方差和标准差的商：

$$\rho_{X,Y}=\frac{\operatorname{cov}(X,Y)}{\sigma_X\sigma_Y}=\frac{E\left[(X-\mu x)(Y-\mu y)\right]}{\sigma_X\sigma_Y}$$

上式定义了总体相关系数，常用希腊小写字母 ρ 作为代表符号。估算样本的协方差和标准差，可得到皮尔逊相关系数，常用英文小写字母 r 代表：

$$r=\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2}}$$

我们根据上述方法对价格和投资进行了分析，最终求出价格和投资占比的皮尔

逊相关系数为 0.4182，二者中等程度相关。

5.2.2、聚类分析

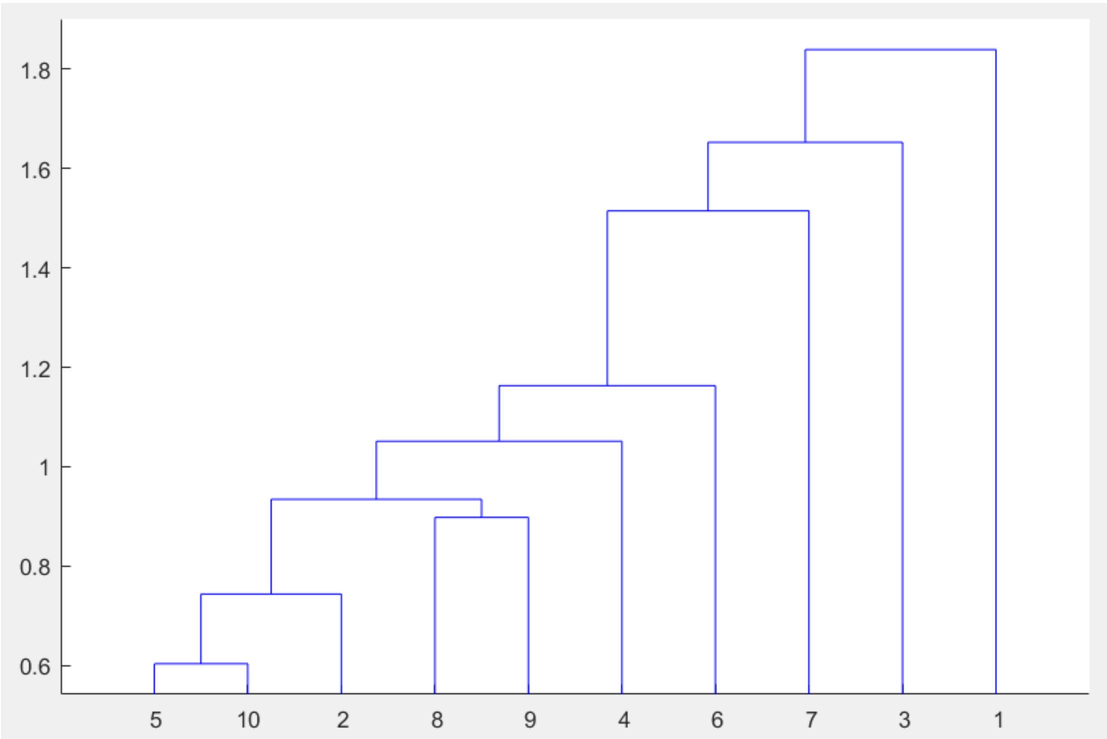
聚类分析指将物理或抽象对象的集合分组为由类似的对象组成的多个类的分析过程。

聚类分析是一组将研究对象分为相对同质的群组的统计分析技术，其聚类方法包括：层次聚类（如：合并法、分解法、树状图），非层次聚类（如：划分聚类、谱聚类）。

我们对股票种类和投资占比进行了聚类分析。输入股票种类和投资占比则将得到一个 10 × 57 的距离矩阵：

$$\begin{pmatrix}
0 & 1.8422 & 2.0000 & 2.0000 & 1.9059 & 2.0000 & 1.8393 & 2.0000 & 2.0000 & 2.0000 \\
1.8422 & 0 & 1.6528 & 1.2885 & 0.7439 & 1.3656 & 1.6051 & 1.0052 & 1.3186 & 1.1238 \\
2.0000 & 1.6528 & 0 & 1.6528 & 1.6528 & 1.8391 & 1.8136 & 1.6528 & 1.6528 & 1.6528 \\
2.0000 & 1.2885 & 1.6528 & 0 & 1.0511 & 1.7580 & 1.5150 & 1.4130 & 1.2401 & 1.1057 \\
1.9059 & 0.7439 & 1.6528 & 1.0511 & 0 & 1.1632 & 1.5538 & 1.1125 & 0.9993 & 0.6037 \\
2.0000 & 1.3656 & 1.8391 & 1.7580 & 1.1632 & 0 & 1.8739 & 1.3265 & 1.5284 & 1.4072 \\
1.8393 & 1.6051 & 1.8136 & 1.5150 & 1.5538 & 1.8739 & 0 & 1.7848 & 1.7819 & 1.5417 \\
2.0000 & 1.0052 & 1.6528 & 1.4130 & 1.1125 & 1.3265 & 1.7848 & 0 & 0.8982 & 1.1481 \\
2.0000 & 1.3186 & 1.6528 & 1.2401 & 0.9993 & 1.5284 & 1.7819 & 0.8982 & 0 & 0.9345 \\
2.0000 & 1.1238 & 1.6528 & 1.1057 & 0.6037 & 1.4072 & 1.5417 & 1.1481 & 0.9345 & 0
\end{pmatrix}$$

其中 10 指的是十家公司，57 则是 57 种股票所对应的投资占比。这样这个矩阵便能在一定程度上体现各家公司的投资策略。由此矩阵进行聚类分析则可得到一个聚类图，如下：



（图三：聚类分析图）

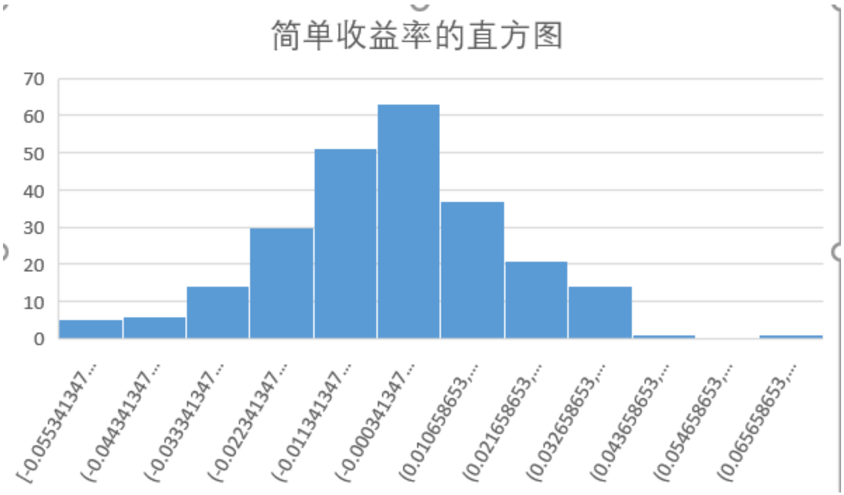
5.2.3、模型结果分析

根据先前的定义以皮尔逊相对系数法和聚类分析法对已有数据进行了详细的分析，由此我们则可以得出结论：公司 5 和 10，8 和 9 的投资策略比较相似，公司 2，5，10 之间的投资策略也有着一定的相似性。整体上基金公司都倾向于低股价的股票。

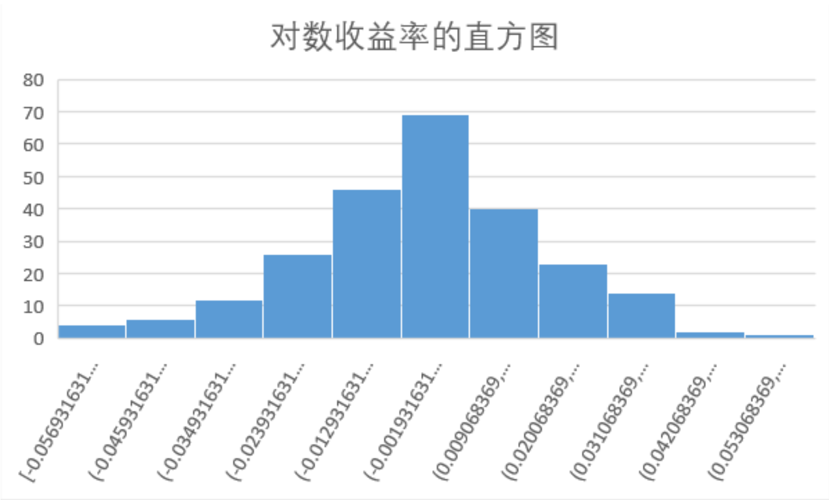
六、问题二模型的建立与求解：

6.1 数据选取和处理：

为了确定最优的股票投资组合策略，在衡量采用哪一种收益率的问题上，对股价数据进行两种不同的计算得到简单收益率和对数收益率。我们可以统计它们的分布图：



（图四：简单收益率的直方图）



（图五：对数收益率的直方图）

显然对数收益率更符合正态分布，因此我们选用对数收益率。

在决定使用对数收益率后，利用 matlab 软件对附件 2 中的数据进行处理，对每一种股票的对数收益率进行统计，将平均值作为期望对数收益率进行分析。

6.2 模型建立：

6.2.1 均值-半方差模型：

Markowitz 的均值-方差模型：

均值-方差模型是由哈里·马科维茨 (H. M. Markowitz) 在 1952 年提出的风险投资模型。[6]马科维茨把风险定义为收益率的波动率，首次将数理统计的方法应用到投资组合选择的研究中。这种方法使收益与风险的多目标优化达到最佳的平衡效果。马科维兹证券组合投资模型可以表述为以下形式：

$$\min \sigma^2 = X^{'}QX$$

$$s.t.\begin{cases} R^{'}X = R_0 \\ F^{'}X = 1 \end{cases} \qquad (1)$$

其中 $R=(R_1,R_2,\cdots,R_n)^{'}$ 为选定的 n 种证券预期收益率的期望值向量，σ^2 为证券投资收益率的方差，$X=(x_1,x_2,\cdots,x_n)^{'}$ 为 n 种证券投资比例系数的向量，$Q=(\sigma_{ij})_{m\times n}$ 表示 n 种证券收益率的协方差矩阵，R_0 为证券组合的预期收益率，$F=(1,1,\cdots,1)^{'}$ 是 n 维单位列向量。

6.2.2 对均值-方差模型的改进：

虽然 Markowitz 的均值-方差模型是解决这类经济问题的基础方法，但是由于其计算过于繁琐，限制条件比较苛刻，同时其本身也存在着一定的局限性，这使其与现实之间存在着一定的脱节。[2]再加上其基于历史数据的客观实际，随着各种变量的不断变化，这种估计的有效性也存在着很大的争议，在实际应用中反而会受到很大的限制和约束。

而均值-半方差模型不同点在于：它将位于投资期望线之上和之下的收益率分开计算，在规避了风险的同时还确保投资者的最大收益不会被均值一方差模型过滤掉可能的超额收益机会。

这种方法与均值-方差模型相比显然更加完善，因此我们选择建立均值-半方差模型进行求解。

均值-半方差组合投资模型的形式为：

$$\min SV = \frac{1}{n}\sum_{i=1}^{n}(u_i)^2 \qquad (2)$$

$$s.t.\begin{cases} R^{'}X \geq R_0 \\ F^{'}X = 1 \\ u_i = (R_0 - R_i)^{+} \end{cases}$$

其中 SV 表示证券投资收益率的半方差。

均值一方差模型将投资风险定义为投资收益的不确定性，而半方差模型则将投资风险定义为可能的损失。其思想是用半方差替换方差来减小潜在的误差，在此基础上更好地反映投资者的偏爱。[7]因为均值一方差模型将风险定义为投资，然而这和投资者的想法往往不同，投资者一般将收益率低于预期收益率视为风险，即实际收益率在预期收益率下方；而把预期收益率上方看作为超额收益。均值一半方差模型作为均值一方差模型的改进，以收益率作为度量风险的标准，低于预期收益率的

则表示风险，这样更加接近现实中投资者的感受，可以为投资者提供更好的指导。

6.2.3 普遍的效用最大化模型：

效用函数使用均值-方差来表现风险-预期收益率相互替换的大小和形式，其一般形式是：

$$U = R_0 - \frac{1}{2}A\sigma^2 \tag{3}$$

其中U为效用值，A是反应投资者回避风险程度的指数，由（3）式可以看出：效用会随着预期收益率的增大而增大，随着方差的减小而增大。

效用最大化模型的形式如下：

$$\max U = R_0 - \frac{1}{2}A\sigma^2 \tag{4}$$

$$s.t.\begin{cases} R_0 = R^{'}X \\ F^{'}X = 1 \\ \sigma^2 = X^{'}QX \end{cases}$$

6.2.4 基于均值-半方差模型的修正：

由于均值-半方差模型采用了半方差代替了方差，所以在形式上和原模型有所区别，并且在投资风险上定义也不同，因此有必要进行调整。在综合考虑各种因素并加上不能卖空的限制条件之后，修正后的形式为：

$$\max U = R_0 - \frac{1}{2}A\sigma^2 \tag{5}$$

$$s.t.\begin{cases} R_0 = R^{'}X \\ F^{'}X = 1 \\ \sigma^2 = X^{'}QX \end{cases}$$

6.3 模型求解：

6.3.1 转化为非线性二次规划模型：

由于（5）式的目标函数中存在$-\frac{1}{2}A\times\frac{1}{n}\sum_{i=1}^{n}(u_i)^2$这一多项式，因此该目标函数不是凸的，初步处理后模型如下：

$$\min -U = -R_0 + \frac{1}{2}A\times\frac{1}{n}\sum_{i=1}^{n}(u_i)^2$$

$$s.t.\begin{cases} R_0 = R^{'}X \\ F^{'}X = 1 \\ x_i \ge 0, i = 1, 2, \cdots, n \\ u_i = (R_0 - R_i)^{+} \end{cases} \tag{6}$$

由于$u_i=(R_0-R_i)^+$这一条件很难在实际中应用，所以我们将其转化为等价形式$\begin{cases}u_i\ge R_0-R_i\\ u_i\ge 0\end{cases}$，很容易就可以证明这两项是等价的。此外为了计算方便，应该把未知数都存放在一个向量里面，即令$x_{n+i}=u_i$， $i=1,2,\cdots,n$

经过上述的处理，则可以将（6）式化作

$$\min -U=-R_0+\frac{1}{2}A\times\frac{1}{n}\sum_{i=1}^{n}(x_{n+i})^2$$

$$s.t.\begin{cases}R_0=R^{'}X\\ F^{'}X=1\\ x_i\ge 0,i=1,2,\cdots,2n\\ x_{n+i}\ge R_0-R_i\end{cases}$$

至此将改进后的效用最大化模型转化为线性约束的非线性二次规划模型。由解向量中的$x_1,x_2,...,x_{57}$所组成的一维子向量就是我们要求的投资方案。

6.3.2 运用 quadprog() 函数求解二次规划模型：

对于二次规划模型有很多的解法，比如拉格朗日方法等。由于我们的求解目标是线性约束的非线性二次规划模型，所以选择使用 matlab 的 quadprog() 函数来求解这个问题。

$$\min_x \frac{1}{2}x^T Hx+f^T x \; such\,that \begin{cases}A\cdot x\le b,\\ Aeq\cdot x=beq,\\ lb\le x\le ub.\end{cases}$$

首先我们将目标函数转化成如下的形式：

$$f(x)=\frac{1}{2}x^T Hx+f^T x$$

显然 R’向量也要拓展到 114 个元素，从第 58 个元素开始用 0 填充。

$$R_0=R^{'}X=R_1x_1+R_2x_2+\cdots+R_{57}x_{57}$$

则对应的$f^T=-R^{'}=\begin{bmatrix}-R_1\\ -R_2\\ \vdots\\ -R_{57}\\ 0\\ \vdots\\ 0\end{bmatrix}$。

参数中矩阵 H 的元素是二次型中矩阵元素的两倍，即：

$$H_{(i,j)}=\begin{pmatrix} 0 & 0 & & \cdots & & \cdots & 0 \\ 0 & 0 & & & & & \\ & & \ddots & & & & \vdots \\ \vdots & & & A & & & \\ & & & & \ddots & & \vdots \\ \vdots & & & & & \frac{A}{n-1} & 0 \\ 0 & & \cdots & & \cdots & 0 & \frac{A}{n} \end{pmatrix}$$

其约束条件要转化为$\begin{cases} A\cdot x\le b \\ A_{nq}\cdot x=b_{nq} \\ lb\le x\le ub \end{cases}$的形式。

由于 $F^{'}X=1$为线性拘束，其中$F^{'}=\begin{bmatrix}1\\1\\\vdots\\1\\0\\0\\\vdots\\0\end{bmatrix}$，$F^{'}$的前 57 个元素为 1，其余为 0。

因此有$A_{eq}=F^{'}\quad b_{eq}=1$。

在二次规划模型中的约束条件$x_{n+i}-R_0\ge R_i$可以转化为$R_0-x_{n+i}\le -R_i$，其展开后为

$$\begin{pmatrix} R_1 & R_2 & \cdots & R_{57} & -1 & 0 & 0 & \cdots & 0 \\ R_1 & R_2 & \cdots & R_{57} & 0 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & & & \vdots & \vdots & \vdots & & \vdots \\ R_1 & R_2 & \cdots & R_{57} & 0 & 0 & 0 & \cdots & -1 \end{pmatrix}X\le -R^{'}，$$

所以 $A=\begin{pmatrix} R_1 & R_2 & \cdots & R_{57} & -1 & 0 & 0 & \cdots & 0 \\ R_1 & R_2 & \cdots & R_{57} & 0 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & & & \vdots & \vdots & \vdots & & \vdots \\ R_1 & R_2 & \cdots & R_{57} & 0 & 0 & 0 & \cdots & -1 \end{pmatrix} X \le -R^{'}$，$B=-R^{'}$。

至此所有的参数都已经转化完毕。

6.4 模型结果分析

6.4.1 A 值分析

输入 $R^{'}$ 和 A 参数后带入 quadprog()函数后得出如下表格：

A 值	最大效用	是否满足投资所有股份
……	……	……
17.2	0.0026	否
17.3	0.0025	否
17.4	0.0025	是
17.5	0.0025	是
……	……	……
22.3	0.0016	是
22.4	0.0016	否
22.5	0.0016	否
……		

（表二：A 值与最大效用之间的关系）

在此处 A 值代表了投资者回避风险的程度，当 A 越大时投资者越倾向于波动较小的投资方案，随着 A 的增加，方差较大的股票的份额则会逐渐变小，方差较小的股票所占份额会逐渐变大。

在本次的分析中，当 A 值选取 17.3~22.3 时对结果影响显著，超过此区间对结果影响不显著。随着 A 的增大，预期收益率也会逐渐减小。所以投资者应该首先找到 A 值变化影响显著的取值区间，并在此基础上仔细地考虑自己应该以何种组合方式进行投资。

6.4.2 最优解

根据先前结果可以看出当 A 值选取 17.3 时配置方案没有卖空股票的前提下实现效用最大。

此时分配方案如下：

股票代码	投资占比	股票代码	投资占比	股票代码	投资占比
股票 1	0.01%	股票 21	0.04%	股票 41	98.67%
股票 2	0.01%	股票 22	0.16%	股票 42	0.01%
股票 3	0.02%	股票 23	0.02%	股票 43	0.01%
股票 4	0.02%	股票 24	0.03%	股票 44	0.01%
股票 5	0.02%	股票 25	0.01%	股票 45	0.01%
股票 6	0.02%	股票 26	0.03%	股票 46	0.02%
股票 7	0.01%	股票 27	0.02%	股票 47	0.01%
股票 8	0.01%	股票 28	0.02%	股票 48	0.03%
股票 9	0.01%	股票 29	0.01%	股票 49	0.03%
股票 10	0.02%	股票 30	0.01%	股票 50	0.02%
股票 11	0.02%	股票 31	0.01%	股票 51	0.01%
股票 12	0.03%	股票 32	0.01%	股票 52	0.01%
股票 13	0.02%	股票 33	0.02%	股票 53	0.01%
股票 14	0.11%	股票 34	0.01%	股票 54	0.02%
股票 15	0.01%	股票 35	0.02%	股票 55	0.01%
股票 16	0.01%	股票 36	0.01%	股票 56	0.01%
股票 17	0.01%	股票 37	0.05%	股票 57	0.01%
股票 18	0.02%	股票 38	0.01%		
股票 19	0.06%	股票 39	0.14%		
股票 20	0.01%	股票 40	0.04%		

（表三：最优解时的分配方案）

七、问题三模型的建立与求解

7.1 模型建立：

7.1.1VaR 模型：

VaR 按字面的解释就是 value-at-risk,即“处于风险状态的价值”，是在一定置信水平和一定持有期内，某一金融工具或其组合在未来资产价格波动下所面临的最大损失额。

为了能更透彻地理解 VaR 的相关概念，下面我们将推导其数学表达式。设资产组合的初始价值为w，持有期末的期望收益为R，R的数学期望和标准差依次为μ和σ，在给定的置信区间c下，期末资产组合的最低值为$w^* = W(1+R^*)$，其中R^*为相对应的最低收益率（一般情况下为负值），则：

$$Value\ at\ risk = E(W) - W^* = -W(R^* - \mu)$$

7.1.2 参数选取

VaR 模型具有两个重要的参数：资产组合的持有期 和 置信水平。

从投资者的角度来说，资产组合的持有期应由资产组合自身的特点来决定。资产的流动性越强，相应的持有期越短；反之，流动性越差，持有期则越长。根据已有数据特点，我们选择持有期为 1 天。

置信水平的选取反映了投资主体对风险的厌恶程度，置信水平越高，厌恶风险的程度越大。根据题意置信水平为 95%。

7.2 模型解法选取：

7.2.2、Eviews 分析股票价格序列 ADF 检验。

我们在此使用 Eviews 分析股票的价格序列，并使用 ADF 检验来对股票价格的稳定性进行评估。例如股票 1 的价格序列的 ADF 检验值小于 5%的显著水平，所以它是相对平稳的。

Null Hypothesis: ____1 has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=14)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.000010	0.0363
Test critical values:	1% level	-3.457173	
	5% level	-2.873240	
	10% level	-2.573080	

*MacKinnon (1996) one-sided p-values.

（图六：对股票 1 进行 ADF 检验的结果图）

同样对 57 种股票价格序列进行 ADF 检测，结果如下。

股票编号	稳定程度
1	稳定
11	较稳定
17	较稳定
28	稳定
33	较稳定
36	较稳定
45	稳定
55	稳定
56	较稳定

（表四：对股票进行 ADF 检测的结果）

除此以外其余股票全部不稳定

7.2.2 历史模拟法：

历史模拟法(Back/Historic Simulation Approach)是一个简单的、非理论的方法，有些金融商品不易取得完整的历史交易资料，此时可以借由搜集此金融商品之风险因子计算过去一段时间内的资产组合风险收益的频率分布,通过找到历史资料求出其报酬率，然后搭配持有资产的投资组合部位，则可以重新建构资产价值的历史损益分配，然后对资料期间之每一交易日重复分析步骤，如果历史变化重复时，则可以重新建构资产组合未来报酬的损益分配。

使用历史模拟法的好处有：利用历史资料，不需要加诸资产报酬的假设，不须对资产报酬的波动性、相关性做统计分配的假设。因此避免了许多估计所可能产生

的误差。

因为数据整体上较为全面，所以在综合考虑下采用历史模拟法更加方便。

7.3 模型求解：

通过 excel 表格功能对简单收益率进行升序排列，并按照历史模拟法选取向量 A。

这样就可以得出 $\mathrm{Var} = Wi \times Ai$。

则一个公司的总风险价值 $Var_{总} = \sum_{i} Var(i)$，其中 i 代表此公司所持的股票编号

计算后将其按照风险价值降序排列则可以得出各家公司风险价值排序表。

基金公司	公司风险价值
H	140070.6
J	117265.3
G	93162
I	85716.86
E	66219.51
D	60335
C	59540.86
B	41670.97
F	40396.07
A	34616.71

(表五：公司风险价值降序排列表)

八、问题四模型的建立与求解

8.1 数据准备

根据已有的数据，若将所有样本基金公司组成一个系统，则可以得到 10 个基金公司的投资总额为 24675114.49 万元。同样因为对数收益率的分布更符合正态分布，所以选取对数收益率作为模型的重要参数。

8.2 多项式逼近拟合

结合 ADF 检验的结果可以看出大部分股价序列是不稳定序列。根据均值-半方差模型以及延伸出来的非线性二次规划模型的参数要求，我们需要一些有说服力的新的参数矩阵 R'，以此来更好地给出 2020 年投资策略和指导意见。

对于有关于 2020 年的参数矩阵 R' 的选取自然来自于对 2019 年对数收益率数据的拟合预测值。对于不平稳序列，如果只需求下一个交易日或者短期预测，可以采用 Smoothing Spline 方法以此来达到最大拟合度，但显然用于长期预测会出现较大偏差。这里我们采用微分方程的拟合思想，选取多形式逼近拟合（Polynomial）方法来得到未来的长期预测值。

由对数收益率的可加性可知，如果我们要计算 45 天后的期望对数收益率，应该用 45 天后的对数收益率减去首日对数收益率求平均值，而对数收益率的计算满足：

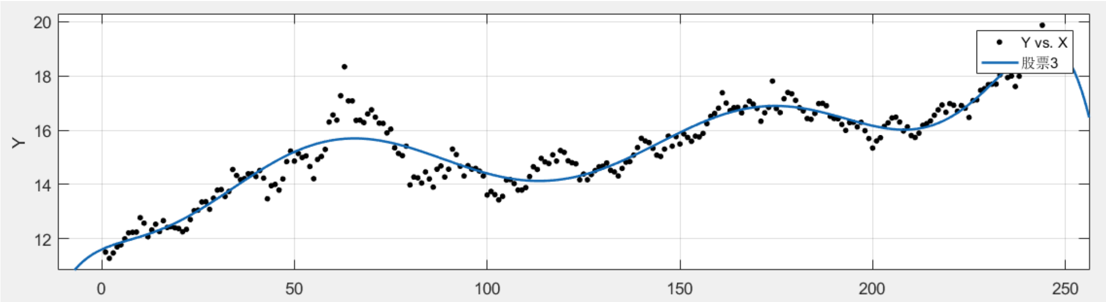
$$\ln\frac{S_2}{S_1}+\ln\frac{S_3}{S_2}+\cdots+\ln\frac{S_n}{S_{n-1}}=\ln\frac{S_n}{S_1}$$

所以我们只需采用首日股价 S_1 和 n 天后的股价 S_n，即可求出该种股票的期望对数收益率的未来预测值。

经过筛选发现，多形式逼近拟合中选取八次方多项式形式作为拟合表达式最为符合需求。

$$f(x)=p_1*x^8+p_2*x^7+p_3*x^6+p_4*x^5+p_5*x^4+p_6*x^3+p_7*x^2+p_8*x+p_9$$

以股票 3 的股价为例，采用八次方多项式形式拟合：



（图七：股票三拟合价格变动曲线）

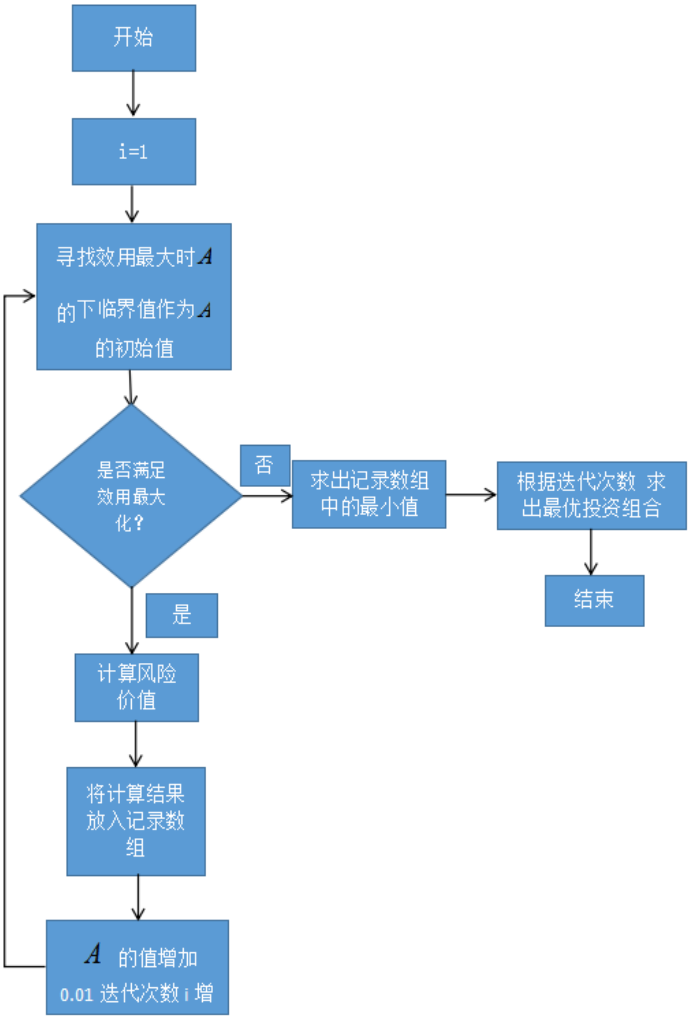
可以看出相比于其他形式和方法的拟合程度更大，且一定程度上适用于长期预测。并且八次方多项式形式拟合的预测值计算出的对数收益率期望在形式上更贴近于来自于 2019 年得出的对数收益率期望。

8.3 模型建立——一维搜索求最优解模型

在得到几个参数矩阵 R' 后，我们需要在满足投资效用最大的前提下找到风险价值最低的最优解。而我们采用的求投资效用最大的模型解法是非线性二次规划，如果强行加上一个新的目标组成多目标二次规划模型，很难用通常方法得到最优解。

为此我们决定用积极集法的思想，从 A 的下临界值开始进行一维搜索求最优解。因为风险价值的经验向量 VaR 是对 2019 的对数收益率数据采用历史模拟法总结而来的，而非线性二次规划中的解并不是线性的，所以只能采用一维搜索的遍历法。在已经求得效用最大时 A 的下临界值的前提下，通过不断对 A 值进行迭代，统计直到不满足投资效用最大条件前所有分配情况下的系统风险价值总和。最后进行排序求出系统风险价值最小值，并根据迭代次数得出最优解时 A 的值和最佳投资组合。

根据上述设计，此算法的流程图如下：



（图八：算法流程图）

8.4 模型求解结果

8.4.1 首个交易日

采用 2019 年经验向量 VaR 和 2019 年的期望收益率矩阵作为参数 R'，解得最优解：

把下面做成表格：

股票编号	投资占比	股票编号	投资占比	股票编号	投资占比
1	0.02%	21	0.07%	41	97.60%
2	0.02%	22	0.26%	42	0.01%
3	0.03%	23	0.04%	43	0.02%
4	0.04%	24	0.06%	44	0.02%
5	0.04%	25	0.03%	45	0.02%
6	0.03%	26	0.06%	46	0.05%

7	0.02%	27	0.03%	47	0.01%
8	0.02%	28	0.03%	48	0.06%
9	0.02%	29	0.01%	49	0.05%
10	0.03%	30	0.01%	50	0.03%
11	0.03%	31	0.02%	51	0.01%
12	0.06%	32	0.03%	52	0.03%
13	0.04%	33	0.03%	53	0.01%
14	0.17%	34	0.02%	54	0.03%
15	0.02%	35	0.03%	55	0.02%
16	0.01%	36	0.03%	56	0.02%
17	0.03%	37	0.08%	57	0.03%
18	0.05%	38	0.01%		
19	0.11%	39	0.23%		
20	0.02%	40	0.07%		

（表六：最优解情况下的结果）

由此可以看出，模型更推荐投资股票 41。此时整个系统的风险价值总和为 1301301.1 万元。

8.4.2 基于拟合预测 45 天、90 天、180 天后的策略

将多项式逼近拟合得出的数据进行计算，得出新的期望对数收益率作为 $R^{'}$，求得最优解，分以下三种情况

（1）无论 A 的取值，45 天后的策略中除了股票 9 和 42，其他的策略均不超过 0.0001%，因此可以忽略不计，可得出以下表格：

A 值	股票 9 投资占比	股票 42 投资占比
……	……	……
268.2	0.02%	99.98%
275	97.79%	2.21%
……	……	……

（表七：45 天后情况下 A 值与配置策略的关系）

结果显示 A 值取 268.2~275 之间时推荐方案变化明显。模型推荐不同的基金公司根据自身情况选取的 A 值不同，给出了两个大体的推荐投资方向。

模型取得最优解时投资效用为 0.8804,整个系统的风险价值总和为 980831.7 万元。

（2）无论 A 的取值，90 天后的策略中除了股票 9、股票 42 和股票 43，其他的策略均不超过 0.0001%，因此可以忽略不计，可得出以下表格：

A 值	股票 9 投资占比	股票 42 投资占比	股票 43 投资占比
……	……	……	…
155	0.14%	99.77%	0.09%
160	98.77%	0.03%	1.21%

300	0%	0%	100%
……	……	……	…

（表八：90 天后情况下 A 值与配置策略的关系）

当A取（155,160）时，最佳投资方案逐渐转移。当A值超过 160 越大，更偏向于股票 43。模型推荐不同的基金公司根据自身情况选取的A值不同，给出了三个大体的推荐投资方向区间。

模型取得最优解时投资效用为 0.7199,整个系统的风险价值总和为 979967.9 万元。

（3）无论A的取值，90 天后的策略中除了股票 21、股票 38 和股票 43，其他的策略均不超过 0.0001%，因此可以忽略不计，可得出以下表格：

A 值	股票 21 投资占比	股票 38 投资占比	股票 43 投资占比
……	……	……	…
212.5	99.99%	0.01%	0.00%
250	0.01%	99.99%	0.00%
512.5	0%	0.01%	99.99%
……	……	……	…

（表九：180 天后 A 值与投资策略的关系）

当A取（212.5，250）时，最佳投资方案逐渐从股票 21 转移向股票 38。当A取（220，512.5）时，最佳投资方案逐渐从股票 38 转移向股票 43。当模型推荐不同的基金公司根据自身情况选取的A值不同，给出了四个大体的推荐投资方向区间。

模型取得最优解时投资效用为 1.1008，整个系统的风险价值总和为 1047337.2 万元。

8.4.3 模型结果分析

基于 2019 年庞大的数据支撑，首个交易日的最佳投资组合方案很明确。

由于部分股价数据缺失和 ADF 检验结果，拟合的结果不能给出具体的最优投资组合方案，但是能给出投资者和投资公司有实际意义的投资建议。根据基金公司的经济实力，选取不同的A值回避风险，模型会给出大体的推荐投资方向。

九、模型的评价与推广

9.1 模型的优点

1）创新性地将不同的小模型进行结合，采用积极集法的思想设计一维搜索优化算法，将原本非常难解决的包含二次规划的双目标非线性规划问题转化为优化算法求最优解。对数据进行了较为全面的分析和利用，包括所有的方法和参数的选取都是在多重考虑下决定的，因此在数据运用上有一定优越性。

2）模型的基础是均值-半方差模型，并且效用函数以及后面利用 quadprog()函数求解的参数转化都有详细的数学推导，所以模型在结构上一定的严谨性，弥补了一些均值-方差模型的不足之处。

3）对数据进行了较为全面的分析和利用，包括所有的方法和参数的选取都是在多重考虑下决定的，因此在数据运用上有一定优越性。

4）模型的参数数据简单且易获取，并且不同的公司可以自由根据自己的经济情况和回避风险的能力自由地选取参数。

9.2 模型的缺点

1）某些股票的数据样本实在太少，在数据拟合和对数收益率计算上会出现较大误差。

2）在寻找最大效用值的参数的下临界值时可能会因为人为因素出现误差。

3）虽然模型在使用了所有拟合类型后选择了多项式逼近的八次多项式形式的拟合，但是进行长期预测难免会出现误差，尤其是 ADF 检验后不平稳的序列。

9.3 模型的推广

1）可以为不同的基金公司进行简单的投资组合分析，并且可以给出短期最佳投资组合和长期投资的投资方向建议。

2）未来可以在以非线性二次规划为核心的算法上继续优化或者组合新的模型，在更多考虑更多因素的情况下求出最优解。

十、参考文献

［1］屠新曙，王春峰，巴曙松．投资组合效用问题的研究［J］．数 量 经 济技术经济研究，2002，19（5）：37－40

［2］武可栋，韦增欣．基于效用最大化的投资组合模型［J］．China Management Informationization，2017, 20(1)：119-122

［3］钱艳英，李建新．效用最大化投资模型的进一步分析［J］．湖南工程学院学报，2006，16（2）：59-60,79

［4］理查德．米肖．有效资产管理一股票组合优化与资产配置实用指南，上海财经大学出版社，2006 年：第 18 页

［5］屠新曙．证券市场风险管理［M］．北京：科学出版社，2008.

［6］H Markowitz．Portfolio Selection ［J］．The Journal of Finance，1952，7（1）： 77－91

［7］卫 海 英，徐 广 伟．半方差风险计量模型的实证比较及改进研究［J］．暨南学报：哲学社会科学版，2004，26（5）：24－30.

［8］ 雷涛.证券投资基金资产配置策略研究[D].天津财经大学,2009

附 录

Test_1.m

```
%一维搜索求最优解算法
ii = 1;%迭代次数
AA = 17.3;%效用最大时 A 值的下临界值
%根据 R 给 A 赋值
A = zeros(57,114);
for i = 1:57
    for j = 1:57
        A(i,j) = data_R(1,j);
    end
end
for i = 1:57
    A(i+57,i) = -1;
end
%二次规划求解
data_dR = data_R';
f = fuF .* data_dR;
H = zeros(114,114);
for i = 58:114
    H(i,i) = AA*(1/(i-57));
end
Aeq = F';
Beq = 1;
lb = zeros(114,1);
B = fuF(1:57,1) .* data_dR(1:57,1);
[xx,fval,exitflag,output,lambda] = quadprog(H,f,A,B,Aeq,Beq,lb);
xxx = xx';
result = zeros(1,100);
up_fval = fval + 0.0001;%规定区间上界
while fval <= up_fval
    data_dR = data_R';
    f = fuF .* data_dR;
    H = zeros(114,114);
    for i = 58:114
        H(i,i) = AA*(1/(i-57));
    end
    Aeq = F';
    Beq = 1;
    lb = zeros(114,1);
    B = fuF(1:57,1) .* data_dR(1:57,1);
    [xx,fval,exitflag,output,lambda] = quadprog(H,f,A,B,Aeq,Beq,lb);
    xxx = xx';
```

```
    vvalue = 24675114.49 * xxx(1,1:57);%分配的钱
    result(1,ii) = vvalue * vvar';
    ii = ii + 1;
    AA = AA + 0.001;
end
iii = 2;
min_var = result(1);
min_number = 1;
%风险价值最小的值和迭代次数
while result(1,iii) ~= 0
    if result(1,iii) < result(1,iii - 1)
        min_var = result(1,iii);
        min_number = iii;
        iii = iii + 1;
    end
end
%运行完毕后 min_var 有最小风险价值
AA = AA + 0.01*(min_number - 1);
%求出此时的最优分配方案
data_dR = data_R';
f = fuF .* data_dR;
H = zeros(114,114);
for i = 58:114
    H(i,i) = AA*(1/(i-57));
end
Aeq = F';
Beq = 1;
lb = zeros(114,1);
B = fuF(1:57,1) .* data_dR(1:57,1);
[xx,fval,exitflag,output,lambda] = quadprog(H,f,A,B,Aeq,Beq,lb);
best_solve = xx';%此时里面有最优投资组合
```

Test_2.m

```
%拟合预测的函数
lastday = 244;%最后的天数
lastpride =13.13;%最后的价格
R = zeros(3,1);
R(1,1) = log(solve(lastday+45)/lastpride)/45;
R(2,1) = log(solve(lastday+90)/lastpride)/90;
R(3,1) = log(solve(lastday+180)/lastpride)/180;
R %输出不同天数后的预测值
%用 cftool 拟合后将参数信息写入 fx 中
function fx = solve(x)
```

```
        p1 =  -9.722e-16;
        p2 =   9.075e-13;
        p3 =  -3.346e-10;
        p4 =   6.139e-08;
        p5 =   -5.76e-06;
        p6 =   0.0002503;
        p7 =   -0.003844;
        p8 =     0.06599;
        p9 =        11.6;

    fx = p1*x^8 + p2*x^7 + p3*x^6 + p4*x^5 + p5*x^4 + p6*x^3 + p7*x^2 +
p8*x + p9;
end
```