五一数学建模竞赛

承 诺 书

我们仔细阅读了五一数学建模竞赛的竞赛规则。

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与本队以外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的, 如果引用别人的成果或其它公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们愿意承担由此引起的一切后果。

我们授权五一数学建模竞赛组委会，可将我们的论文以任何形式进行公开展示（包括进行网上公示，在书籍、期刊和其他媒体进行正式或非正式发表等）。

参赛题号（从A/B/C中选择一项填写）： B

参赛队号：

参赛组别（研究生、本科、专科、高中）： 本科

所属学校（学校全称）： 江苏科技大学

参赛队员： 队员1姓名： 令狐岩松

队员2姓名： 夏佳伟

队员3姓名： 夏鸿鑫

联系方式： Email： 联系电话：

日期： 年 月 日

**（除本页外不允许出现学校及个人信息）**

**五 一 数 学 建 模 竞 赛**

****

**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-半差值模型、线性约束非线性二次规划、历史模拟法、一维搜索求优化**

**摘 要：**

**（前面两页必须使用模板格式，否则论文检测不通过）**

**此页为论文正文开始处**

**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：

本问题要求根据附件一的资产配置信息建立数学模型，并以此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。若想求得公司的资产配置策略。然而仅凭现有的资料我们很难得出规律，根据对问题的分析，我们认为可以从简单做起，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。在此基础上我们便可以建立起相关的数学模型进行求解。

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，则需要我们根据附件二的资料首先确定股票价格关于时间的数学模型以确定每只股票的价格变动，并在此基础上可以确定收益的问题。在此基础上则可以建立公司的收益模型，进而解决该问题。

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：

**三、模型的假设**

1、假设股票价格的变动只与时间有关

**四、名词解释与变量说明**

1、投资占比：由一个公司所持单只股票持股总市值与公司的投资总额相除得到，体现了该股票在公司投资中的比重

2、皮尔逊相关系数：在统计学中，皮尔逊相关系数( Pearson correlation coefficient），又称皮尔逊积矩相关系数（Pearson product-moment correlation coefficient，简称 PPMCC或PCCs），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。（百度百科）

3、对数收益率：对数收益率是两个时期资产价值取对数后的差额，即资产多个时期的对数收益率等于其各时期对数收益率之和。

4、置信区间：置信区间是指由[样本统计量](https://baike.baidu.com/item/%E6%A0%B7%E6%9C%AC%E7%BB%9F%E8%AE%A1%E9%87%8F/7378689" \t "https://baike.baidu.com/item/%E7%BD%AE%E4%BF%A1%E5%8C%BA%E9%97%B4/_blank)所构造的总体参数的估计区间。在统计学中，一个[概率](https://baike.baidu.com/item/%E6%A6%82%E7%8E%87/828845)样本的置信区间（Confidence interval）是对这个样本的某个总体参数的[区间估计](https://baike.baidu.com/item/%E5%8C%BA%E9%97%B4%E4%BC%B0%E8%AE%A1/6611490" \t "https://baike.baidu.com/item/%E7%BD%AE%E4%BF%A1%E5%8C%BA%E9%97%B4/_blank)。置信区间展现的是这个参数的真实值有一定概率落在测量结果的周围的程度，其给出的是[被测量](https://baike.baidu.com/item/%E8%A2%AB%E6%B5%8B%E9%87%8F/5141218)参数的测量值的可信程度，即前面所要求的“一个概率”。

5、持股时间：持股时间是指股票买入时间与卖出时间的持有期限

**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

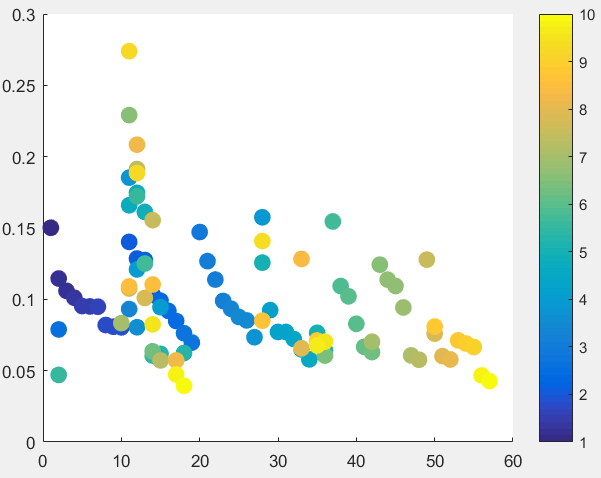
我们将已知的数据进行简单的处理：

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）。

公司投股总市值（万元）/投股总量（万股）=平均股价（元）

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：

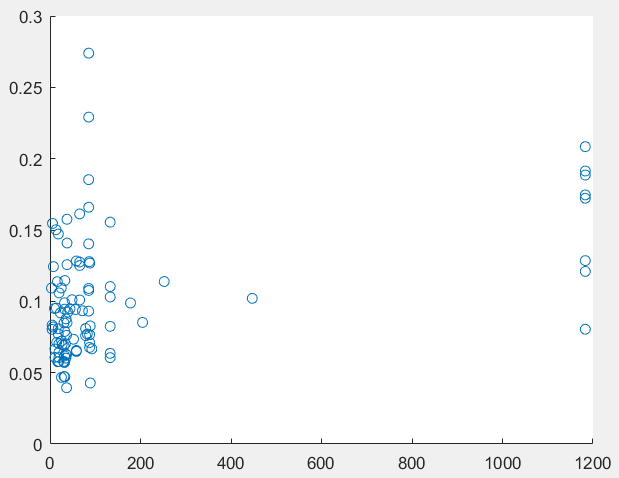


（图一：股票种类和投资占比的散点泡沫图）

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](https://baike.baidu.com/item/%E7%BB%9F%E8%AE%A1%E5%AD%A6/1175" \t "_blank)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](https://baike.baidu.com/item/%E7%9B%B8%E5%85%B3/9882881" \t "_blank)（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：



上式定义了总体相关系数，常用希腊小写字母作为代表符号。估算[样本](https://baike.baidu.com/item/%E6%A0%B7%E6%9C%AC/19974592)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：



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。

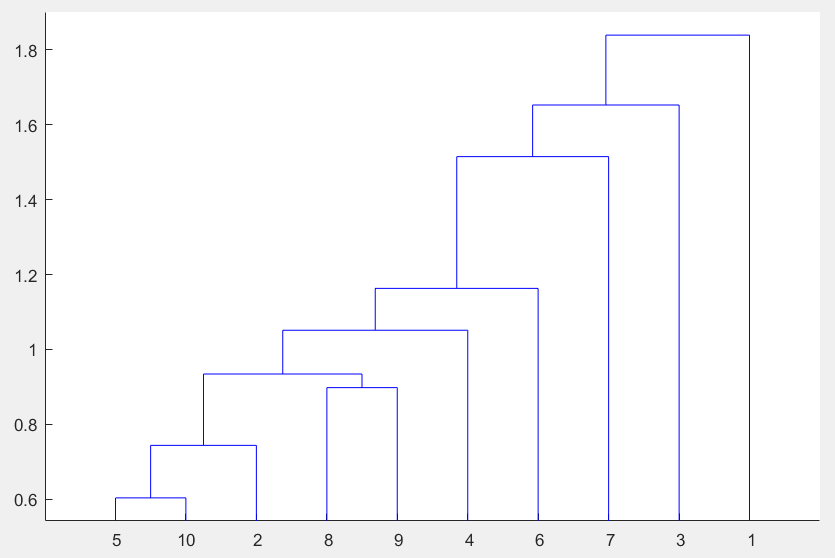
5.2.2、聚类分析

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](https://baike.baidu.com/item/%E5%88%86%E8%A7%A3%E6%B3%95)、[树状图](https://baike.baidu.com/item/%E6%A0%91%E7%8A%B6%E5%9B%BE)），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：

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**：

6.2.1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：

手机屏幕截图

描述已自动生成

手机屏幕截图

描述已自动生成

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。

6.1模型建立：

6.1.1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](https://baike.baidu.com/item/%E5%93%88%E9%87%8C%C2%B7%E9%A9%AC%E7%A7%91%E7%BB%B4%E8%8C%A8/2695688) (H. M. Markowitz) 在 1952 年提出的风险投资模型。马科维茨把风险定义为[收益率](https://baike.baidu.com/item/%E6%94%B6%E7%9B%8A%E7%8E%87/4533890)的[波动率](https://baike.baidu.com/item/%E6%B3%A2%E5%8A%A8%E7%8E%87/8350659)，首次将数理统计的方法应用到[投资组合](https://baike.baidu.com/item/%E6%8A%95%E8%B5%84%E7%BB%84%E5%90%88/3520475)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：



 （1）

其中为选定的n种证券预期收益率的期望值向量，为证券投资收益率的方差，为n种证券投资比例系数的向量，表示n种证券收益率的协方差矩阵，为证券组合的预期收益率，是n维单位列向量。

6.1.2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。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－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：

 （2）



其中表示证券投资收益率的半方差。

均值－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。因为均值－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－半方差模型作为均值－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。

6.1.3基于均值-半方差模型的效用最大化模型：

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，其一般形式为： （3）

其中为效用值，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：

 （4）



6.1.4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：

 （5）



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：



 （6）

由于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，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，



经过上述的处理，则可以将（6）式化作





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。



首先我们将目标函数转化成如下的形式：



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。



则对应的。

参数中矩阵H的元素是二次型中矩阵元素的两倍，即：



其约束条件要转化为的形式。

由于为线性拘束，其中，的前57个元素为1，其余为 0。

因此有。

在二次规划模型中的约束条件可以转化为，其展开后为，

所以，。

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 值分析

输入和参数后带入quadprog()函数后得出如下表格：

|  |  |  |
| --- | --- | --- |
| 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 |
| …… |  |  |

在此处值代表了投资者回避风险的程度，当越大时投资者越倾向于波动较小的投资方案，随着的增加，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 | 分配权重 | 股票代码 | 分配权重 | 股票代码 | 分配权重 |
| 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% |  |  |

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](https://baike.baidu.com/item/%E7%BD%AE%E4%BF%A1%E6%B0%B4%E5%B9%B3)和一定持有期内，某一金融工具或其组合在未来[资产价格](https://baike.baidu.com/item/%E8%B5%84%E4%BA%A7%E4%BB%B7%E6%A0%BC)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，持有期末的期望收益为，的数学期望和标准差依次为和，在给定的置信区间下，期末资产组合的最低值为，其中为相对应的最低收益率（一般情况下为负值），则：

 （1）

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。

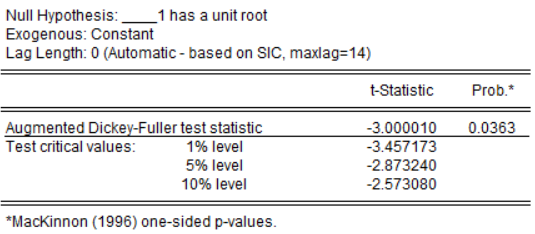
从投资者的角度来说，资产组合的持有期应由资产组合自身的特点来决定。 资产的流动性越强，相应的持有期越短；反之，流动性越差，持有期则越长。根据已有数据特点，我们选择持有期为1天。

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。



（图四：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 |

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](https://baike.baidu.com/item/%E5%8E%86%E5%8F%B2)[资料](https://baike.baidu.com/item/%E8%B5%84%E6%96%99)，不需要加诸[资产](https://baike.baidu.com/item/%E8%B5%84%E4%BA%A7)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。

则一个公司的总风险价值，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 |

(公司风险价值降序排列表)

八、参考资料