

基于 Python 的马科维茨投资组合理论的实证研究

摘要: 随着中国金融体系的逐步完善和大数据时代的到来,新技术在金融领域的应用不断深化。本文在马科维茨投资组合理论的基础上,借助 Python 工具,在 20 只来自不同行业的股票中选取 5 只进行了组合投资分析,通过实证得到夏普比率最大的最优投资组合及方差最小的最优投资组合,对它们的预期收益率、标准差及夏普比率进行对比分析,并给出资产组合的有效边界。通过实证分析,进一步说明马科维茨投资组合理论在金融风险管理中的重要意义。

关键词: 马科维茨模型 Python 夏普比率 有效边界 最优投资组合

● 孙丽泊

一、引言

(一) 相关理论

1. 马科维茨投资组合理论。1952 年,哈里·马科维茨的论文《Portfolio Selection》运用概率论和求解二次规划的方法解决投资组合的选择问题,这是现代资产组合管理理论诞生的标志。在投资者厌恶风险并寻求期望收益最大化的假设条件下,马科维茨投资组合理论包括均值-方差模型和有效边界理论两方面内容。该理论提出用均值和方差分别量化投资组合的期望收益和投资风险,投资决策的目标便是寻找相同收益水平下投资风险最小或同等风险水平下收益最高的投资组合,位于有效边界上的投资组合同时满足这两个条件。

(1) 均值-方差模型。

假设投资者在某单一投资期投资于 n 种风险资产构成的组合, r_i 表示第 i 个资产的预期收益率,则资产组合的期望收益率为:

$$E(r_p) = \sum_{i=1}^n x_i E(r_i) \quad (1)$$

其中 x_i 表示第 i 个资产所占的投资比重。

用 σ_i^2 表示第 i 个资产的方差,则由 n 项资产构成的投资组合的方差为:

$$\sigma_p^2 = \sum_{i=1}^n \sum_{j=1}^n x_i x_j \text{cov}(r_i, r_j) = \sum_{i=1}^n x_i^2 \sigma_i^2 + \sum_{i=1}^n \sum_{j \neq i}^n x_i x_j \rho_{ij} \sigma_i \sigma_j \quad (2)$$

其中, i, j 表示不同的资产, $\text{cov}(r_i, r_j)$ 是资产 i 和资产 j 之间的协方差,它是用来衡量两种资产收益率联动性的指标。 ρ_{ij} 为资产 i 与资产 j 之间的相关系数,可用来比较两种资产之间相关性的强弱。 σ_i 和 σ_j 分别表示资产 i 和资产 j 的标准差。根据公式 (2) 可以看出,投资组合的风险主要取决于各个资产的投资比重、不同证券之间的相关系数、以及各项资产的标准差。因此,我们应优先选择方差较小且两两之间相关系数较低的资产构建投资组合从而降低投资

风险。在实际应用中,通常用过去收益数据的样本平均值和样本方差来估计未来收益与风险。

(2) 资产组合的有效边界。

投资组合中各项资产的所有可能组合构成可行集,其形状类似图 1 所示的一个左凸的实心区域。MVP 点标注的为最小方差资产组合,以该点为中心可将可行集分成两部分,其中位于上半部分边缘线的投资组合同时满足既定收益水平下风险最小和同等风险水平下收益最高的条件,被称为资产组合的有效边界。

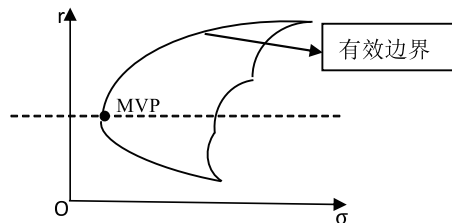


图 1 资产组合的可行集和有效边界

2. 夏普比率。1966 年,夏普提出用夏普比率 (Sharpe ratio) 作为衡量基金业绩的风险调整指标,夏普比率是投资组合的超额期望收益率与整体标准差之比,计算公式为:

$$S_p = \frac{E(r_p) - R_f}{\sigma_p} \quad (3)$$

其中, S_p 表示夏普比率, σ_p 为投资组合整体标准差, R_f 是无风险利率。

S_p 表示资产组合每多承担一单位风险时,能有多少超额收益。 S_p 综合考虑了收益与风险,本文的实证研究部分也将使用这一指标作为判断投资组合优劣的标准。

(二) 研究内容

本文研究的主要问题是,利用 Python 语言说明马科维茨投资组合理论在股市最优投资组合选择中的一个应用实例。Python 语言具有完整的数据分析套件和能够调取股价等金融数据的程序包,为解决金融问题提供了很大



的方便。利用 Python 语句求解最优投资组合可有效解决各项资产之间的协方差和相关系数计算复杂等问题。研究表明在 Python 中能够很快计算出最优投资组合，新技术对于马科维茨的投资组合理论在中国金融市场中的应用有重要价值。

二、实证研究

本文选取了市场上 20 只来自不同行业比较有代表性的股票，首先利用过去三个月的历史数据计算出这 20 只股票的收益率均值、股票相互之间的协方差与相关系数，进而从中筛选出收益率较高且相互之间相关系数较小的 5 只股票；然后通过夏普比率最大化和方差最小化这两种优化方法来找到最优的投资组合权重参数，并计算出这两种最优投资组合的期望收益率、预期波动率和夏普比率；最后对投资组合的有效边界进行可视化。

(一) 样本股票的选择及数据获取

首先新建一个 Python3 文件并将要使用的软件包引入，代码如下：

```
import tushare as ts
import pandas as pd
import numpy as np
import statsmodels.api as sm
import scipy.stats as scs
import matplotlib.pyplot as plt
```

选取市场中 20 只来自不同行业的比较有代表性的股票，利用 tushare 库中的 get_hist_data 函数，获取它们自 2019 年 1 月 1 日至 2019 年 12 月 31 日之间一共 246 个交易日的每日收盘价历史数据，分别用 mean 函数、cov 函数和 corr 函数估计出每只股票的预期年化收益率、相互之间的协方差和相关系数。表 1 列出了 20 只股票的股票代码和预期年化收益。

表 1 20 只股票的预期年化收益 (%)

股票代码	600519	601398	000002	601628	000538	601633	600009
年化收益	74.5	13.69	28.52	62.19	22.33	42.78	52.68
股票代码	603866	600030	600053	600764	600755	600601	600808
年化收益	-7.93	47.06	12.45	20.89	5.29	35.82	-14.74
股票代码	600090	600150	603899	603897	603898	600050	
年化收益	-31.76	54.11	63.96	-21.21	5.77	13.22	

从以上 20 只股票中筛选出收益率较高且相互之间相关系数较小的 5 只股票，分别是：贵州茅台（600519）、万科 A 股（000002）、方正科技（600601）、中国船舶（600150）和晨光文具（603899）（本文后面投资组合的权重向量各股排序与此处相同）。

(二) 5 只股票的预期收益率、相互之间的协方差与相关系数

重新计算所选 5 只股票的预期年化收益率、各只股票相互之间的协方差与相关系数，结果如表 2 所示：

根据表 2 可见，5 只股票的预期收益率均大于 0，各股票之间的协方差和相关系数均较小，适合构造投资组合。

表 2 5 只股票的预期收益率、协方差矩阵和相关系数矩阵

预期年化收益率	股票代码	600519	000002	600601	600150	603899
预期收益率 (%)		72.17	31.54	39.57	52.52	55.57
协方差矩阵		600519	000002	600601	600150	603899
	600519	0.0005	0.0001	0.0001	0.0002	0.0001
	000002	0.0001	0.0004	0.0001	0.0001	0.0000
	600601	0.0001	0.0001	0.0006	0.0002	0.0001
	600150	0.0002	0.0001	0.0002	0.0011	0.0001
	603899	0.0001	0.0000	0.0001	0.0001	0.0004
相关系数矩阵		600519	000002	600601	600150	603899
	600519	1.0000	0.3661	0.1169	0.3080	0.2606
	000002	0.3661	1.0000	0.2012	0.2274	0.0337
	600601	0.1169	0.2012	1.0000	0.1975	0.1342
	600150	0.3080	0.2274	0.1975	1.0000	0.1492
	603899	0.2606	0.0337	0.1342	0.1492	1.0000

(三) 模拟大量随机投资组合

为了寻找满足条件的最优投资组合以及有效边界，首先通过一次蒙特卡罗模拟得到 50000 组随机权重向量，并通过数组记录各组合的期望收益、标准差和夏普比率，这里主要用到 for 循环语句、append 函数和 numpy 库中的 random 函数和 array 函数，其中组合期望收益为 5 只股票预期收益的加权和，计算夏普比率时无风险利率设定为 4%，另外由于中国的股市不允许卖空，还需将各股权重数值限定在 0 和 1 之间。在图 2 中可以看到这些随机组合的分布，它们构成了资产组合的可行集。

(四) 找到夏普比率最大的最优投资组合

首先定义一个 stat 函数记录投资组合的预期收益、标准差和夏普比率，然后引入 scipy 库中的 optimize 模块，在权重和等于 1 的约束下，利用 minimize 函数最小化夏普比率的负值，起始参数列表采用均匀分布，即 5 只股票的权重均为 0.2，最终求得夏普比率最大的投资组合的权重向量为 [0.435, 0.057, 0.127, 0.021, 0.36]，预期收益率、标准差、夏普比率构成的向量为 [0.593, 0.226, 2.443]。

(五) 找到方差最小的最优投资组合

用 minimize 函数得到方差最小的最优投资组合权重向量为 [0.134, 0.336, 0.162, 0.022, 0.346]，该组合预期收益率、标准差、夏普比率向量为：[0.471, 0.200, 2.155]。

(六) 最优投资组合的比较和可视化

将以上得到的夏普比率最大的投资组合和方差最小的投资组合结果整理为表 3，以便投资者更直观地对它们进行比较，选择适合自己的最优投资组合。

表 3 两种最优投资组合的比较

资产组合	各股权重	预期收益率 (%)	标准差	夏普比率
夏普比率最大的最优组合	[0.435, 0.057, 0.127, 0.021, 0.36]	59.3	0.226	2.443
方差最小的最优组合	[0.134, 0.336, 0.162, 0.022, 0.346]	47.1	0.200	2.155

(下转第 50 页)

[2] 蒋向利. 金融机构科技转型有了“定心丸” 专家解读《金融科技 (Fin Tech) 发展规划 (2019—2021 年)》[J]. 中国科技产业, 2019 (11): 57-58.

[3] 金 B. 银行 4.0 [M]. 广州: 广东经济出版社, 2018: 268-289.

[4] 刘勇, 李达. 开放银行: 服务无界与未来银行 [M]. 北京: 中信出版社, 2019: 118-119.

[5] 王蕊, 颜大为. 开放银行生态圈的理论基础、经验探索与发展路径 [J]. 西南金融, 2019 (11): 70-79.

[6] 吴朝平. API 开放银行: 金融科技背景下商业银行转型升级的重要方向 [J]. 金融理论与实践, 2020 (01): 67-72.

[7] Competition and Markets Authority. Retail Banking Market Investigation. Final Report [J]. 2016.

[8] Gozman D, Hedman J, Olsen K S. Open

Banking: Emergent Roles, Risks & Opportunities [J]. 2018.

[9] Omarini A E. Banks and fintechs: how to develop a digital open banking approach for the bank's future [J]. 2018.

[10] Premchand A, Choudhry A. Open Banking & APIs for Transformation in Banking [C] // 2018 International Conference on Communication, Computing and Internet of Things (IC3IoT). IEEE, 2018: 25-29.

[11] Treasury H M. Data sharing and open data for banks [J]. 2014.

张力为云南财经大学商学院在读硕士研究生; 柳晓青就职于盐城农村商业银行, 为云南财经大学商学院在读硕士研究生

(上接第 47 页)

从表 3 可以看出, 投资组合的夏普比率最大可以达到 2.443, 对应的标准差为 0.226, 对那些希望每多承担一单位风险所获得的超额收益达到最大的投资者, 应该选择夏普比率最大的投资组合, 即对贵州茅台、万科 A 股、方正科技、中国船舶和晨光文具的投资比重分别为 43.5%、5.7%、12.7%、2.1% 和 36%, 该组合的预期收益率是 59.3%。对于追求投资风险最小的投资者, 应选择方差最小的最优投资组合, 即对贵州茅台、万科 A 股、方正科技、中国船舶和晨光文具的投资比重分别为 13.4%、33.6%、16.2%、2.2% 和 34.6%。该组合的标准差为 0.2, 相应的预期收益率为 47.1%。

利用 matplotlib.pyplot 中的相关函数对以上最优投资组合在可行集中的位置进行可视化, 得到图 2。

图 2 中, 五角星标记的是夏普比率最大的投资组合 (即风险收益均衡点), 正六边形标记的是最小方差投资组合。与图 1 类似, 以最小方差投资组合为界, 将可行集分为上下两个部分, 上半部分的边缘即为有效边界。

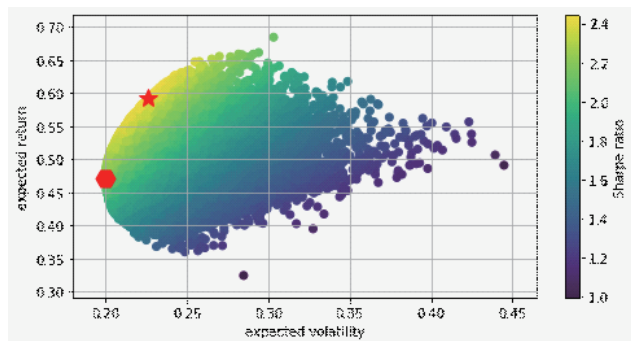


图 2 最优投资组合的可视化

从图 2 中可以看出, 所有投资组合的夏普比率均为正值, 有效边界上的投资组合达到了相同收益水平下投资风险最小和同等风险水平下收益最高的效果。此外, 有效边界上

的投资组合风险和收益一定是相对应的, 投资者若想达到得较高的收益, 相应的就必须承担更高的风险。

三、结论

实验结果表明, 利用马科维茨投资组合理论, 能够找到由多个资产组成的投资组合中风险最小或夏普比率最大的投资组合以及有效边界。投资者可以根据自身实际能力以及风险偏好进行理性投资。Python 在很大程度上方便了马科维茨投资组合理论中预期收益率、投资组合方差的计算, 并能快速找出最优投资组合, 对于马科维茨投资组合理论在中国金融市场中的应用有重要价值。将传统的金融理论和新兴程序设计语言结合起来, 更加快速、高效地解决金融问题, 应该受到金融分析人员和新兴技术人才的重视。

参考文献:

[1] Markowitz H.M..Portfolio Selection [J]. Journal of Finance.1952 (7).

[2] 朱顺泉. 投资学及其 Python 应用 [M]. 北京: 清华大学出版社, 2019.

[3] 张贺清. 均值和方差变动的马科维茨投资组合模型研究 [D]. 哈尔滨: 哈尔滨工业大学, 2015.

[4] 李洋, 余丽霞. 基于马科维茨理论的最优证券组合分析 [J]. 财会月刊. 2013 (22).

[5] 杨芷. 投资组合理论发展演变与模型构建 [J]. 经济研究导刊. 2018 (36).

[6] 刘科弟. 马科维茨理论构造投资组合 [J]. 现代商业. 2018 (36).

[7] 黎爽. 基于 Python 科学计算包的金融应用实现 [D]. 南昌: 江西财经大学, 2017.

作者单位: 苏州科技大学