# ****2022年第三届“大湾区杯”粤港澳****

# ****金融数学建模竞赛****

**题目**：**基于宏观经济周期的大类资产配置策略构建**

****摘要****

大类资产配置是根据投资需求，将投资资金在不同资产类别之间进行分配，构造的一个包含多类资产的投资组合。全球资产配置之父Gary.P.Brinson的研究表明，从长远看，超过90%的投资收益来自于成功的大类资产配置。这得益于大类资产配置通过分散投资为投资者平滑了投资组合风险。但是，我们很难判断投资组合当中哪一类资产在未来表现会更好，这就意味着我们需要基于对未来宏观经济的判断以及大类资产的预测，来动态调整大类资产配置。本文基于DEA算法、熵权法以及LSTM算法，构建了一套更为合适的宏观经济与大类资产预测模型。

**针对问题一，**为了找到高频有效的宏观经济指标，首先，我们把每个指标的数据以取平均值的方式处理成以年为单位的数据，筛选出2001~2020年的数据，剔除掉缺失值占比过高的指标，并填充缺失值。其次，我们通过主成分分析(PCA)对宏观经济指标数据进行降维处理，减少主成分的偏差。接着，基于熵权法，对宏观经济指标进行权重分析，得到权重列表，取权重前5名作为高频指标。然后再基于数据包络分析法，评价模型的有效性。最后。我们根据选出来的指标，将中国的宏观经济运行状况划分为不同的经济状态。

**针对问题二，**本文采用LSTM算法来预测中国未来五年的经济状态。首先使用经过PCA处理的高频指标数据训练LSTM。随后，将近几年的数据输入LSTM并预测中国未来五年的高频指标趋势。根据问题一所建立的模型与预测的数据，划分出中国未来五年的宏观经济环境。

**针对问题三，**本文首先使用熵权法对每类资产下的指数进行权重分析。其次，根据分析的结果选取每类资产中权重最大的指数作为代表该类资产的指数。随后，通过计算夏普比率，得到不同指数的风险收益特征。根据问题一与问题二所划分出的经济状态，得到各种经济状态下的指数风险收益特征。最后，对所选指数进行了相关性分析，并得出指数之间有一定的相关性。

**针对问题四，**本文首先预测了中国未来五年所选指标的状态。随后计算问题三中所选指数的夏普比率。最后，本文采用序列最小二乘规划(SLSQP)算法优化投资组合优化找出使夏普率最大的最优投资组合，并预测所选投资组合的风险收益特征。

**关键词：PCA、熵权法、DEA、夏普比率、LSTM、SLSQP**

目录

[2022年第三届“大湾区杯”粤港澳 1](#_Toc28869)

[金融数学建模竞赛 1](#_Toc6972)

[摘要 1](#_Toc26108)

**[一、 问题背景与问题重述 3](#_Toc19519)**

[1.1. 背景与问题意义 3](#_Toc5918)

[1.2问题重述 3](#_Toc2553)

**[二．问题分析 4](#_Toc4532)**

[2.1问题一分析 4](#_Toc2449)

[2.2问题二分析 4](#_Toc9470)

[2.3问题三分析 4](#_Toc31000)

[2.4问题四分析 5](#_Toc15192)

**[三． 名词解释 5](#_Toc25566)**

**[四． 模型的建立与求解 5](#_Toc23467)**

**[4.1问题一： 6](#_Toc8858)**

[4.1.1问题一求解流程图 6](#_Toc31608)

[4.1.2数据处理 6](#_Toc32201)

[4.1.3 PCA降维模型建立 6](#_Toc20041)

[4.1.4熵权法模型的建立得到高频指标 6](#_Toc7136)

[4.1.5 DEA法模型处理得到高频有效指标 7](#_Toc25989)

[4.1.6 根据指标划分各时期的经济状态 8](#_Toc25167)

**[4.2问题二： 9](#_Toc19444)**

[4.2.1问题二求解流程图 9](#_Toc26474)

[4.2.2通过LSTM算法模型得到预测曲线 10](#_Toc25221)

[4.2.3模型预测 11](#_Toc17578)

**[4.3问题三： 12](#_Toc12300)**

[4.3.1问题三求解流程图 12](#_Toc21477)

[4.3.2数据处理 12](#_Toc9227)

[4.3.3熵权法模型的建立得到四个代表指数 13](#_Toc30581)

[4.3.4夏普比率模型建立 14](#_Toc3280)

[4.3.5划分经济各种经济状态下的风险收益特征 14](#_Toc19557)

[4.3.6各资产指数之间的相关性 17](#_Toc17255)

**[4.4问题四： 18](#_Toc24338)**

[4.4.1问题四求解流程图 18](#_Toc25647)

[4.4.2 LSTM算法预测未来五年各指数的投资组合 19](#_Toc32239)

[4.4.3 SLSQP算法得到最优投资组合 19](#_Toc24858)

[4.4.4 得到最优组合的夏普率 20](#_Toc25417)

**[五.模型的评价 20](#_Toc25306)**

[5.1 模型的优点 20](#_Toc30266)

[5.2模型的缺点 21](#_Toc22669)

**[六.参考文献 21](#_Toc9265)**

**[七.附录及支撑材料 23](#_Toc20894)**

[7.2.1引用程序及实现平台 24](#_Toc8118)

[7.2.2 PCA(主成分分析)代码 25](#_Toc23554)

[7.2.3 多变量LSTM模型 26](#_Toc30292)

[7.2.4 熵权法 29](#_Toc12529)

[7.2.5 夏普比率计算 30](#_Toc16078)

1. **问题背景与问题重述**
   1. **背景与问题意义**

全球资产配置之父Gary P.Brinson 的研究表明，从长远看，超过 90% 的投资收益都是来自于成功的大类资产配置。了解什么是大类资产配置，首先需要了解什么是“大类资产”，什么是“配置”。我们所说的“大类资产”的概念，可以理解为具有相似性质资产的聚类，比如说，无论是投资级债券还是垃圾债券，都属于“债券”这一大类资产。“配置”是根据 投资需求将投资资金在不同资产类别之间进行分配。大类资产配置的根本思想出发点是不把鸡蛋放在一个篮子里的，构造一个包含多类资产的投资组合。相比于独立资产，大类资产配 置通过分散投资为投资者平滑了投资组合风险。从这个意义上讲，大类资产配置这一工具天 然具备风险控制的优势。全球范围来看，大类资产配置作为核心投资方法已是业界共识。头部资管公司希望通过 将投资能力建立在大类资产配置和科技平台上的方式，从而突破投资能力天花板，形成规模效应和网络效应，规模效应是指因规模增大带来的经济效益提高；网络效应是指一个产品或 服务的用户越多，创造的价值就越大；这就意味着大类资产配置使得投资能力随着 AUM（资产管理规模）增长而持续增强。目前全球资管公司竞争激烈但市场分散，未来的增长主要来 自份额提高，领先的一站式大类资产配置平台预计将成为终极赢家，先锋基金和贝莱德是最佳的实践案例。不过，由于我们很难判断投资组合当中哪一类资产在未来表现会更好，并且我们无法明 确哪一类资产的长期表现能够持续占优于其他资产，这意味着大类资产配置在实践过程中的 必然面临动态调整问题。常见的做法是，当我们有把握判断某一类资产优于其他资产的时候，自然会提高这一类资产在整体组合的比例。但是，大类资产配置动态调整的核心在于对未来宏观经济的判断以及大类资产的预测，只有在准确预测出未来经济状况和资产表现的情况之下，我们才能有效辨别出表现较佳的资产类别。先锋的 VCMM 全球资产定价模型是其开展资产配置业务的核心模型，它能够有效模 拟美国的宏观经济运行状态，并预测出在某个经济状态下各类资产的风险收益特征以及资产 之间的相关性。然而，针对中国的投资者与投资环境，我们需要有一套更为适合的宏观经济 与大类资产预测模型。

**1.2问题重述**

**问题一：**寻找出高频有效的宏观经济指标，将 2001 年-2021 年国内的宏观经济运行状况划分成不同的经济状态；（比如，美林时钟框架将宏观经济运行状况划分成衰退，复苏，过热及滞胀四个经济状态）要求宏观经济指标至少选择两个或以上

**问题二：**通过宏观经济模型或其它数学模型模拟中国未来五年的经济增长、通胀、利率（反映货币政策松紧程度）等宏观经济环境，并说明未来五年中国将面临的经济状态处于第 1 问划分出的经济状态中的什么经济状态

**问题三：**挑选出能够代表四类资产（股票、大宗商品、债券、现金及其等价物）的四个指数， 预测大类资产指数在第 1 问划分出的各种经济状态下的风险收益特征（期望收益，收益率标准差，夏普比率或其它），以及大类资产指数之间的相关性

**问题四：**基于你们的模型预测出的国内未来五年的经济状态，挑选出合适的大类资产指数构建投资组合，并预测投资组合的风险收益特征

**二．问题分析**

**2.1问题一分析**

问题一要求我们，在题目所给的宏观指标中，寻找出高频有效的宏观经济指标，并将2001~2021年国内宏观经济运行状况划分为不同的经济状态。首先，我们把每个指标的数据以取平均值的方式处理成以年为单位的数据，再筛选出2001~2020年的数据，剔除掉缺失值占比过高的指标，并填充缺失值，完成对宏观经济指标数据的整合。接着，我们通过主成分分析对宏观经济指标数据进行降维处理。然后基于熵权法。对宏观经济指标进行权重分析，提取权重前五名作为高频指标。最后，我们基于数据包络分析法，评价我们模型的有效性。我们根据选取的高频有效指标，将2001~2021年国内宏观经济运行状况划分为不同的经济状态。

**2.2问题二分析**

问题二要求我们建立模型，来模拟预测中国未来五年的宏观经济环境，然后根据问题，说明未来五年中国将面临什么经济状态。模拟预测中国未来五年的宏观经济环境的关键点在于拟合的准确性，因此，我们基于LSTM算法，使用经过PCA处理的高频指标数据训练LSTM，将近几年的数据进行训练，然后开始预测中国未来五年的经济状态。再根据问题一所建立的模型与预测数据，划分中国未来五年的宏观经济环境

**2.3问题三分析**

问题三要求我们挑选出能够代表四类资产的四个指数，分析四类指数在第一问划分的各种经济状态下的宏观指标风险收益特征，并分析大类资产指数之间的相关性。首先，我们使用熵权法对每类资产下的指数进行权重分析，选取每类资产中权重最大的指数作为代表该类资产的指数。其次，因为风险收益特征是指分析风险与收益的关系，所以我们通过夏普比率来计算得到风险和回报，使用RSR秩和比对不同的风险收益特征进行综合评价。根据问题一、问题二所划分的经济状态，得到各种经济状态下的指数风险收益特征。最后，对所选指数进行相关性分析。

**2.4问题四分析**

问题四可以拆解为两个小问题：①是要求我们预测国内未来五年的经济状态；②是挑选出合适的大类资产指数构建组合，并预测投资组合的风险收益特征。对于问题①，我们参考问题二所建立的模型，先预测指标，再划分经济状况。对于问题②，可以通过问题三先挑选出合适的大类资产指数，然后进行投资组合，得到组合权重，再通过问题三所建立的宏观指标风险收益特征模型，对风险收益特征进行预测。

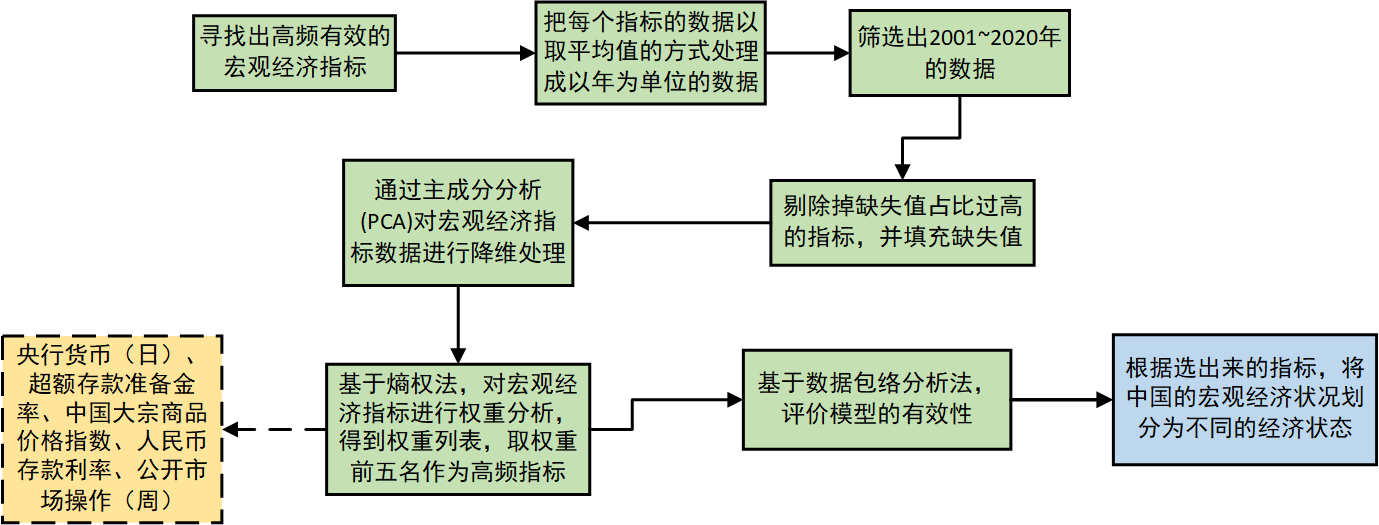
1. **名词解释**

|  |  |
| --- | --- |
| **名词** | **解释** |
| **PCA** | 主成分分析（PCA）是一种降维方法，通常用于通过将数量很多的变量转换为仍包含集合中大部分信息的较少变量来降低数据集的维数。因为较小的数据集更易于探索和可视化，并且使机器学习算法更容易和更快地分析数据，而无需处理无关的变量。 |
| **熵权法** | 在信息论中,熵是对不确定性的一种度量。不确定性越大,熵越大,包含的信息量越多;不确定性越小,熵越小,包含的信息量就越小。根据熵的特性,可以通过计算熵值来判断一个事件的随机性及无序程度,也可以用熵值来判断某个指标的离散程度,指标离散程度越大,该指标对综合评价的影响(权重)越大。 |
| **DEA** | 数据包络分析(DEA)是根据多项投入和多项产出指标，利用线性规划的方法，对具有可比性的同类型单位进行相对有效性评价的一种数量分析方法。 |
| **夏普比率** | 计算投资组合每承受一单位总风险，会产生多少的超额报酬。 |
| **LSTM** | 长短期记忆网络（LSTM，Long Short-Term Memory）是一种时间循环神经网络，是为了解决一般的RNN（循环神经网络）存在的长期依赖问题而专门设计出来的。LSTM的表现通常比时间递归神经网络及隐马尔科夫模型（HMM）更好，在时间序列任务上常常有着优秀的表现。 |

1. ****模型的建立与求解****

**4.1问题一：**

**4.1.1问题一求解流程图**



**4.1.2数据处理**

由于各个指标数据里并不都是统一以一个时间段为自变量来统计且有些指标是存在某一段时间缺失数据的，对于这两个问题，本文先用取平均值的方法将所有的指标数据统一成以年为单位的数据，再通过Python里面的pandas模块中的fillna函数进行缺失值填充，得到一个完整的数据集。

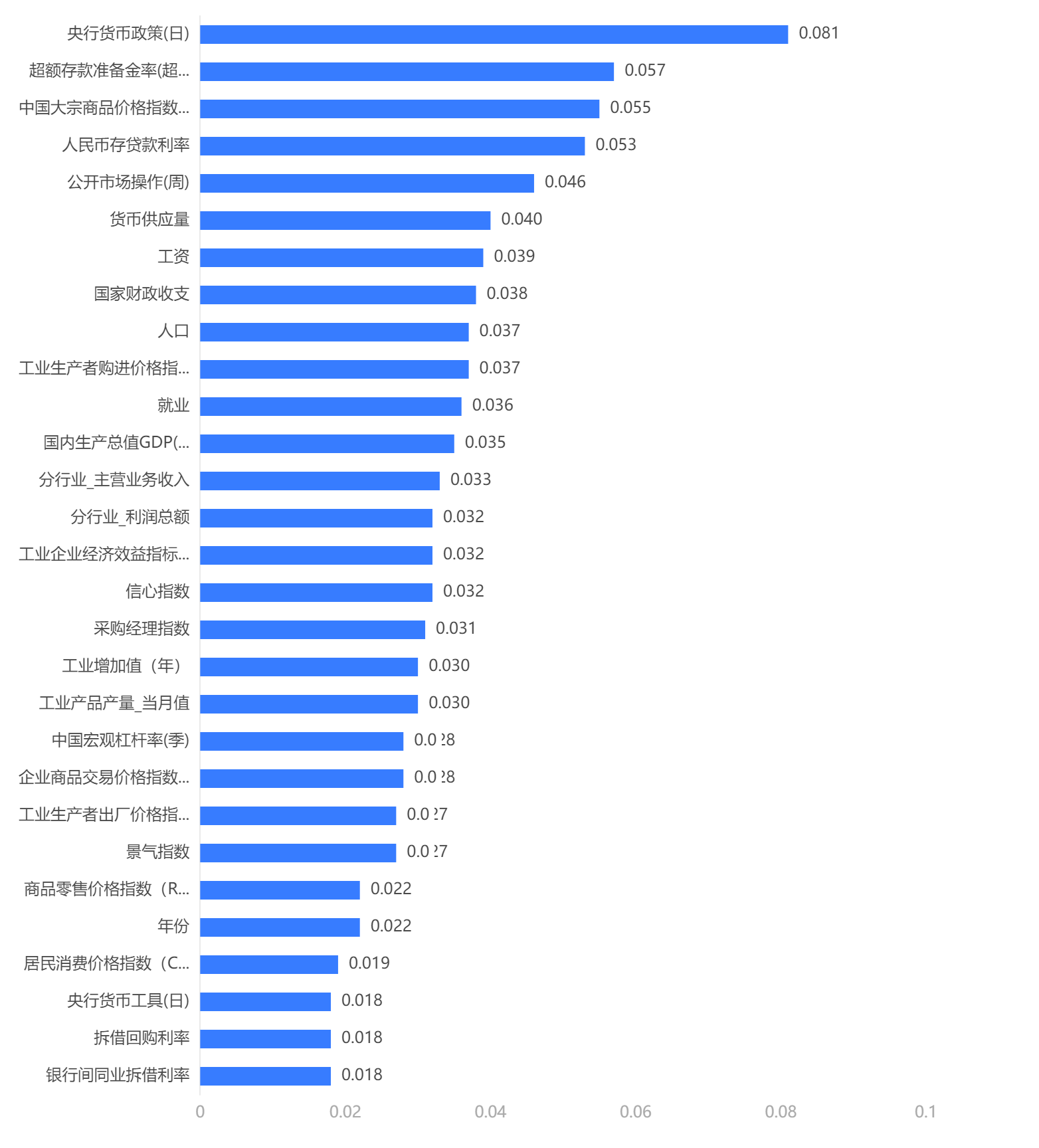
**4.1.3 PCA降维模型建立**

由于各数据集维度不一样，难以将各数据放一起进行比较处理，通过PCA方法可以减少数据集维数，并保留尽可能多的信息，所以使用PCA可以很好的解决该问题。

主成分分析（PCA）属于一个降维方法，通常是在将数量很多的变量转换为集合中大部分信息的少变量来降低数据集维数情况下使用。这是因为较小的数据集更容易去探索和处理，并且也可以使机器学习算法更轻松和速度更快地分析数据，且无需去处理没有关系的变量。

**4.1.4熵权法模型的建立得到高频指标**

PCA方法处理后，我们需要从处理后的数据中找出可以量化的指标，并从中找出最合适的几个特征指标，使用熵权法可以很好的解决这个问题。

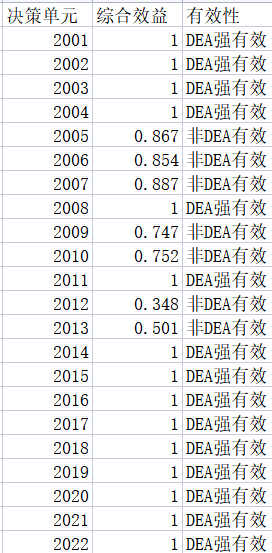
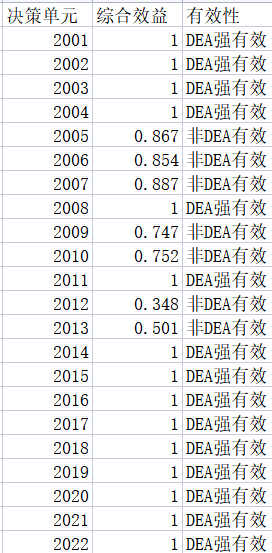
熵权法是能客观赋权的方法，它能反映出指标的有效信息量，熵值越大，指标的有效信息量会越小，所占权重也愈小，因此，熵权法解释性更具道理。因此，我们采用熵权法选出高频指标。 本文在spsspro数据平台里以年份为负向指标定量，其他参数为正向指标定量，得到权重图。

**4.1.5 DEA法模型处理得到高频有效指标**

本文从权重图里选出前五的指标作为高频指标，为验证其有效性，本文将其与低频有效指标进行DEA方法处理，判断其是否可以有效反映国家经济。

数据包络分析(DEA)是根据多项投入和多项产出指标，通过线性规划的方法，对具有可比性的同类型单位用相对有效性评价的数量分析方法，若评分为DEA强有效，可说明投入和投出相关性强，反之则弱。

将这五个指标放入投入指标变量中，将常用于衡量国家宏观经济的五个有效低频指标（GDP、CPI、人口、就业、工资）设为产出指标变量，年份设为决策单元，得到下图



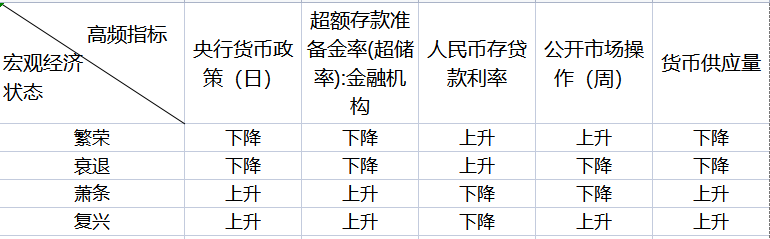
由图可以说明这五个高频指标除了在少数年份外并不能很好地作用于国家经济，多数情况下还是可以有效反映的。

因此我们选择的高频有效指标为 **1.央行货币政策（日） 2.超额存款准备金率(超储率):金融机构 3.中国大宗商品价格指数:总指数 4.人民币存贷款利率 5.公开市场操作（周）**

**4.1.6 根据指标划分各时期的经济状态**

本文引用了《宏观经济指标体系运行的实证分析》这篇论文对宏观经济的定义，将宏观经济状况分为四种状态，分别是**繁荣、衰退、萧条、复苏**

本文选用公开市场操作（月）这一指标来划分，由金融的理论知识可以知道，在理论情况下，各高频指标变化对应的经济状态如下图



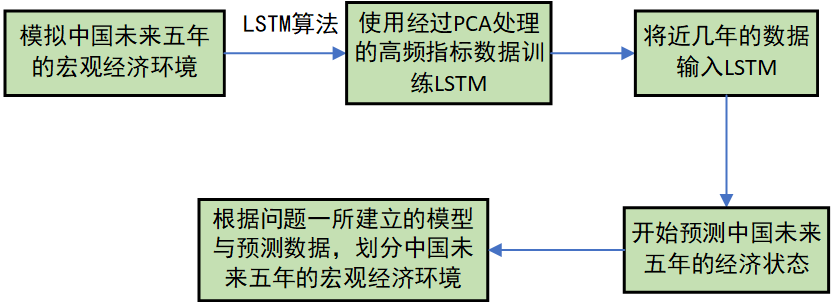
然后将公开市场操作（周）与年份的对应图画出，得到下图



分析，由引用论文可知，在世纪初的2000年，正处于萧条时期，一直持续到2002年开始进入复苏,之后根据公开市场操作（周）的数据，04-06年公开市场操作指数斜率大，可视为进入繁荣时期，06-07年出现跌落，但斜率并不高，可视为进入衰退时期，07-08年末重新进入复苏繁荣时期，08年末-11年，可能金融危机的影响，国家经济进入衰退期，11年-12年，经济出现复兴迹象，但12-14年又衰退，14-15经济出现复苏，15年经济开始繁荣，一直直到15年末经济衰退，16-17年经济衰退，17-20年经济进入萧条期，一直直到20年开始复苏，但因疫情原因，只复苏不到一年的时间，经济再次下跌进入萧条期；

**4.2问题二：**

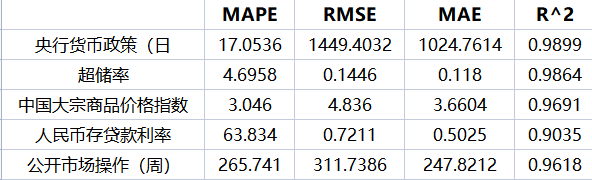
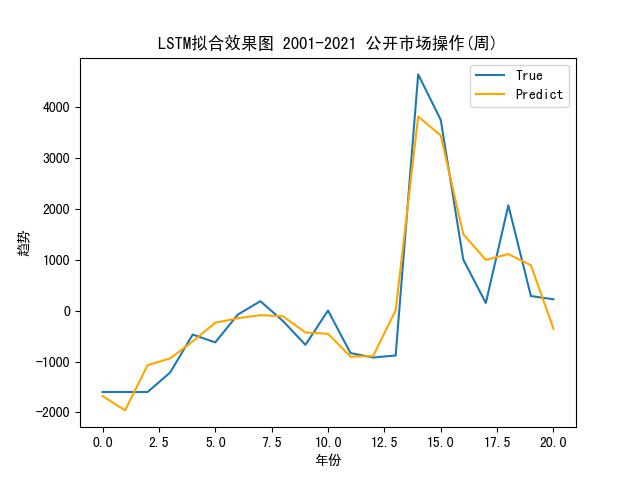
**4.2.1问题二求解流程图**



**4.2.2通过LSTM算法模型得到预测曲线**

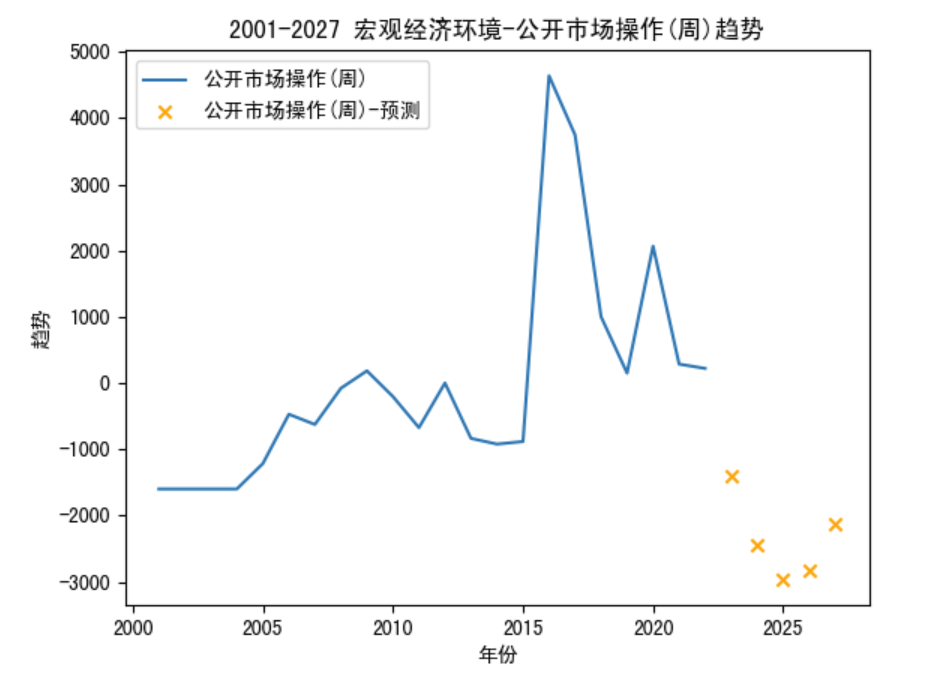
用数学模型模拟未来五年经济状况，实质上就是一个典型的时间序列预测问题，而LSTM倾向于在具有更多固定成分的不稳定时间序列上做得更好，且由于LSTM算法是经典的时间序列预测算法模型，其适用于处理和预测时间序列中间隔和延迟非常长的重要事件，此外，宏观经济状态具有许多影响因素，其随时间的变化也是非常不稳定的，所以本文采用LSTM算法模型去推得预测曲线。

在python3.8环境下分别对所选的五个高频指标调用keras.models的LSTM进行模型训练，得到以下图形， 当LSTM模型测试集R^2达到0.90时，说明拟合度是较高的，可以采用用该训练好的LSTM模型进行未来五年指数的预测。

由下图可知，五个指数的拟合程度均是较高的。详细预测结果见附件。

**4.2.3模型预测**

该阶段继续采用公开市场操作（周）这一指标进行预测，得到的预测部分为下图的黄色部分

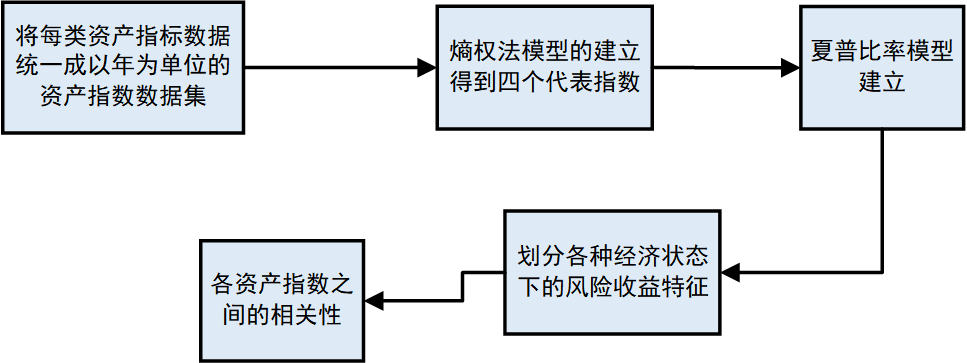


可见，2022-2024年该指标指数将继续大幅下降，对应的经济状态为萧条，之后2024-2026有一小段平缓并回升，此时经济状态为复苏。

**综上，未来五年中国的经济状态将处于从萧条到复苏的时期。**

**4.3问题三：**

**4.3.1问题三求解流程图**

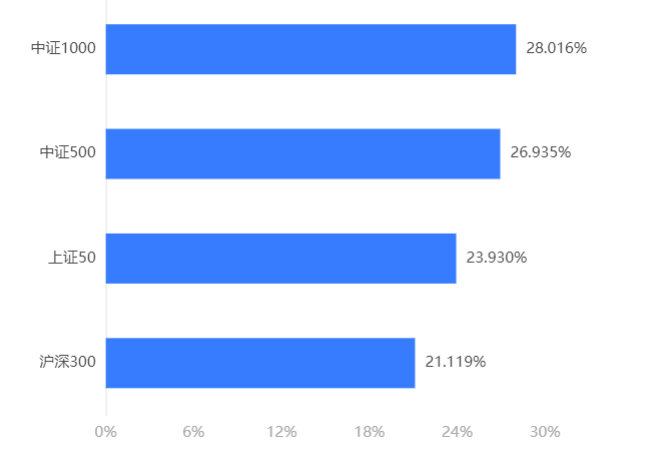


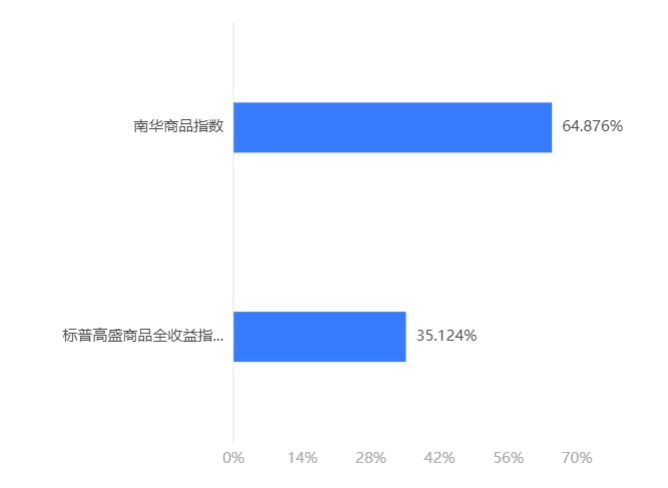
**4.3.2数据处理**

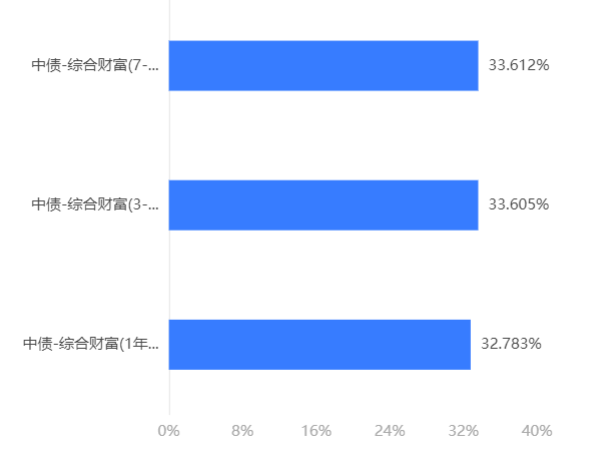
与第一问的数据处理方法类似，本文采用取平均值的方法将每类资产指数数据统一成以年为单位的资产指数数据集

**4.3.3熵权法模型的建立得到四个代表指数**

问题要求我们找出合适的四个代表指数，同样，与第一问解决方法一样，采用熵权法可以很好地解决该问题。 对每类资产分别经过熵权法处理后，得到以下图







选出每类资产权重最大的指数，即**中证1000、南华商品指数、中债-综合财富(7-10年)指数、货币基金**

**4.3.4夏普比率模型建立**

夏普比率是有诺贝尔奖得主威廉·夏普提出的，用以帮助投资者比较投资组合的回报和风险。理性的投资者一般是固定所能承受的风险，追求最大的回报；或者在固定预期回报，追求最小的风险。所以夏普比率计算的是，每承受一单位的总风险所产生的超额回报。计算公式如下。

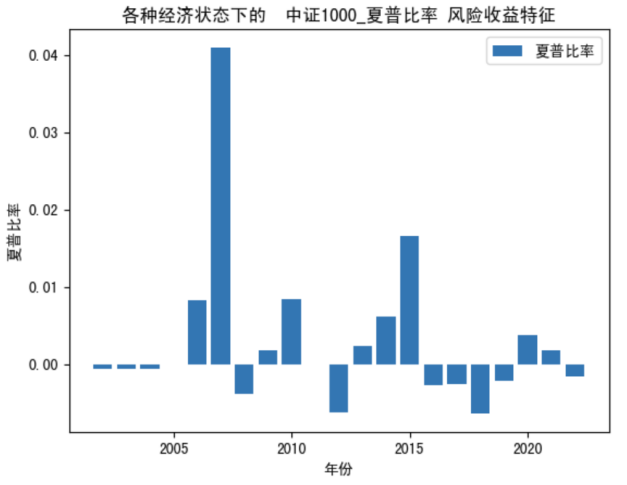
夏普比率=

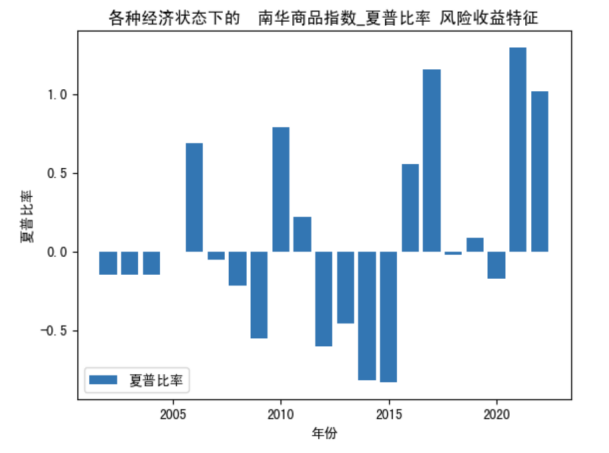
其中是预期回报率，是无风险利率，是超额收益的标准差，

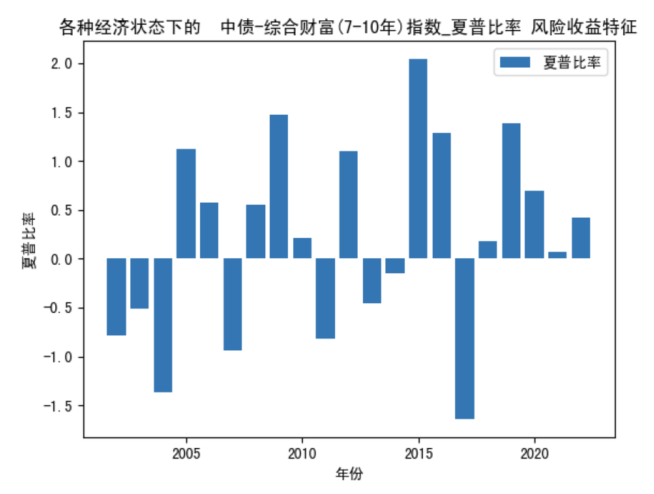
为i年的资产指数

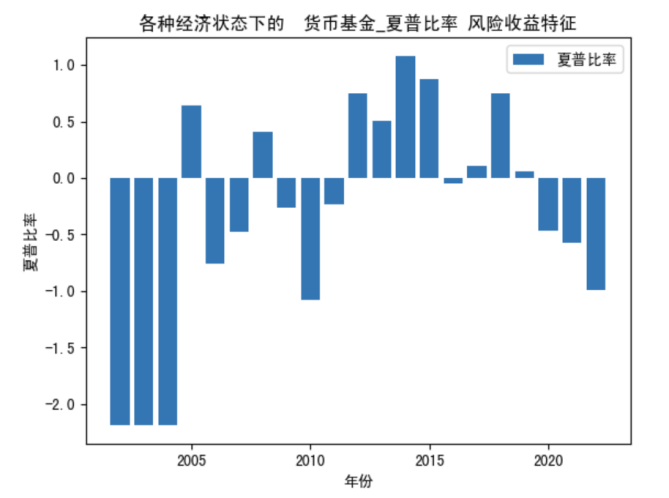
计算各指标当年风险收益；

### **4.3.5划分经济各种经济状态下的风险收益特征**

作出每一年的每一个指数夏普比率的柱状图



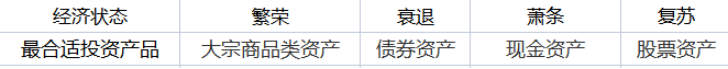




由夏普指数公式可知，分子代表的是回报，指某项投资的回报与代表整个投资类别的基准之差，分母的标准差则代表该投资的波动率，对应着风险，波动越大预示着风险就越高。 因此夏普比率越大，则可代表若分母固定的情况，即风险固定的情况下，回报越大或者若分子固定的情况下，即回报固定的情况下，风险越小。所以，越高的夏普比率就越好，越值得选.将第一题划分的经济状态图与之对比分析

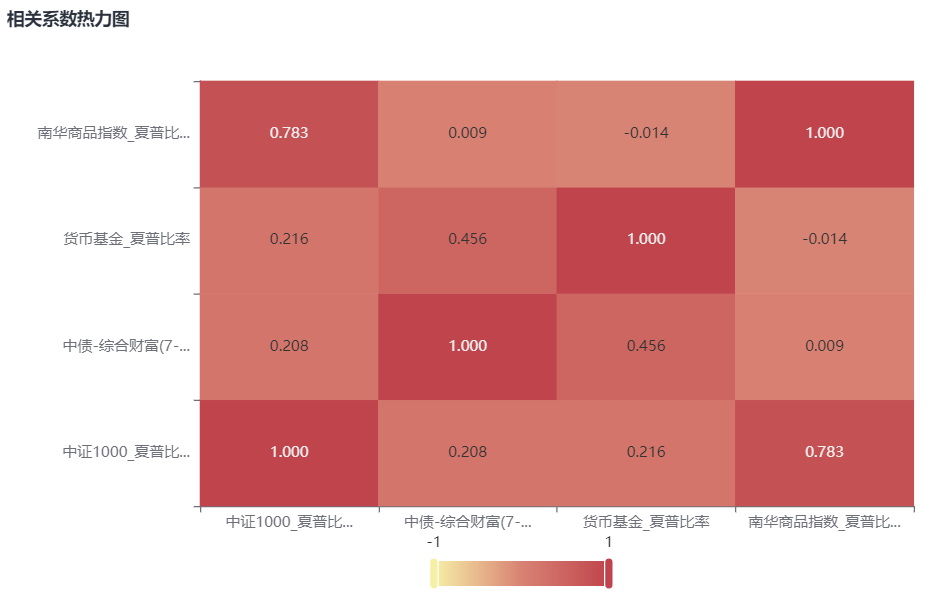


由图对比可知，繁荣时期南华商品的夏普比率最高，复苏时期中证1000的夏普比率最高，衰退时期中债指数夏普比率最高，萧条时期货币基金夏普比率最高；相当于繁荣时期优先投资大宗商品类资产，复苏时期优先投资股票资产，衰退时期优先投资债券资产，萧条期优先投资现金资产。



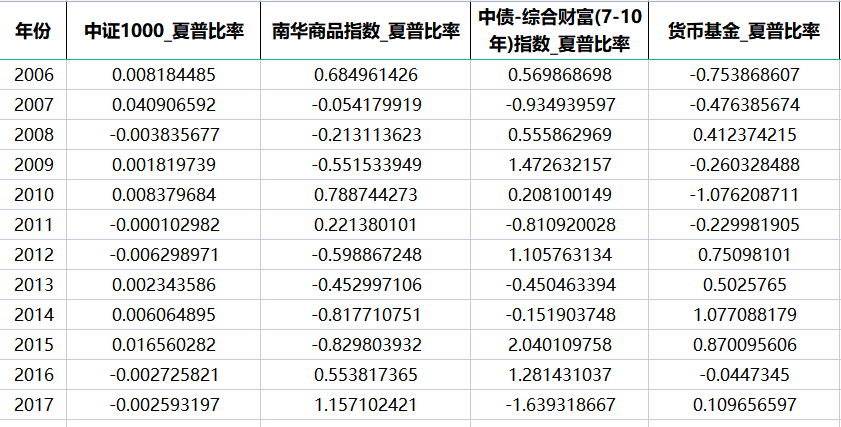
**4.3.6各资产指数之间的相关性**

将四个夏普比率指数表格放入spsspro进行相关性分析，得到以下图形。



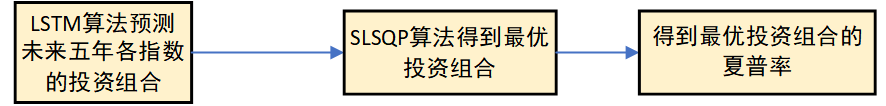
对图分析，可见中证1000指数与南华商品指数的相关性强，相当于股票与大宗商品的指数相关性强。而对于其他指数之间的相关性，就有些偏弱。

所以，得到四个指数对应的每年的夏普比率，如下图。



**4.4问题四：**

**4.4.1问题四求解流程图**



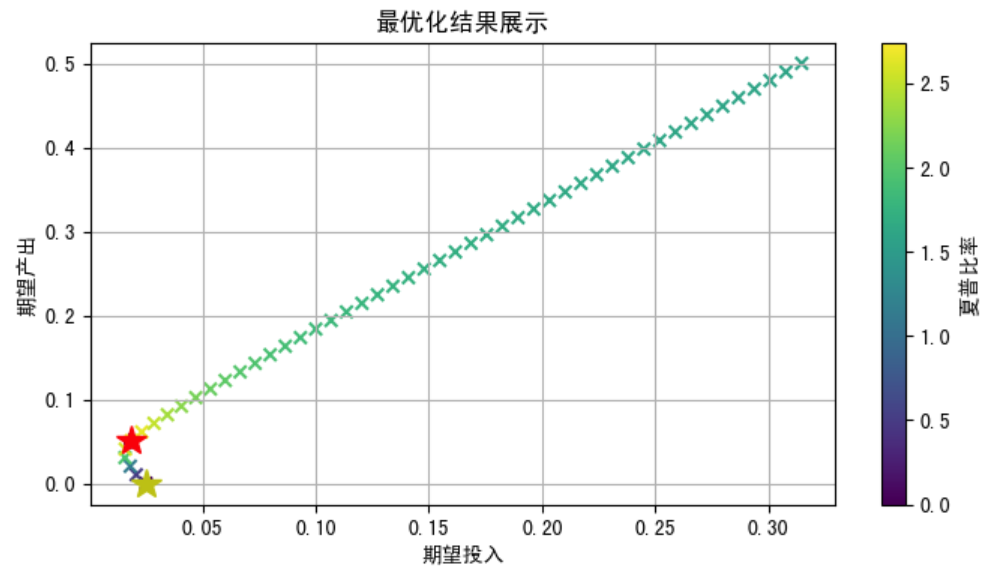
**4.4.2 LSTM算法预测未来五年各指数的投资组合**

对于问题四，我们已从5.2.3预测出了未来五年的经济情况（未来五年中国经济状态为从萧条到复苏时期），目前我们还需要预测未来五年各指数的夏普比率，与问题二一样，本文采用LSTM算法预测得到未来五年各指数夏普率，如下图。

IMG_256

**4.4.3 SLSQP算法得到最优投资组合**

对于问题四，我们需要先预测以及我们需要挑选出合适的大类资产指数构建投资组合，相当于将各类资产排列组合后，得到夏普率最高的投资组合。 本文采用scipy.optimize包来找到我的成本函数即夏普率的最大值。且SLSQP算法是一种基于梯度的优化算法，它期望目标和约束的导数是连续的，其特性是良好的 根据夏普比率用slsqp算法进行优化，得到的未来五年最优投资组合，如下图。



黄色五角星至红色五角星这一区间内是投资效果最好的组合，这一系列的点所组成的边界称为投资有效边界。这条边界的特点是指在同样的收益水平情况下风险是最小的，或者说，在同样风险的情况下获得的收益最大，也即夏普比率最大的地方。此外落在投资有效边界上，其在该点投资性价比最高。

红星即为夏普率最大的组合点，黄星即为方差最小投资组合点。

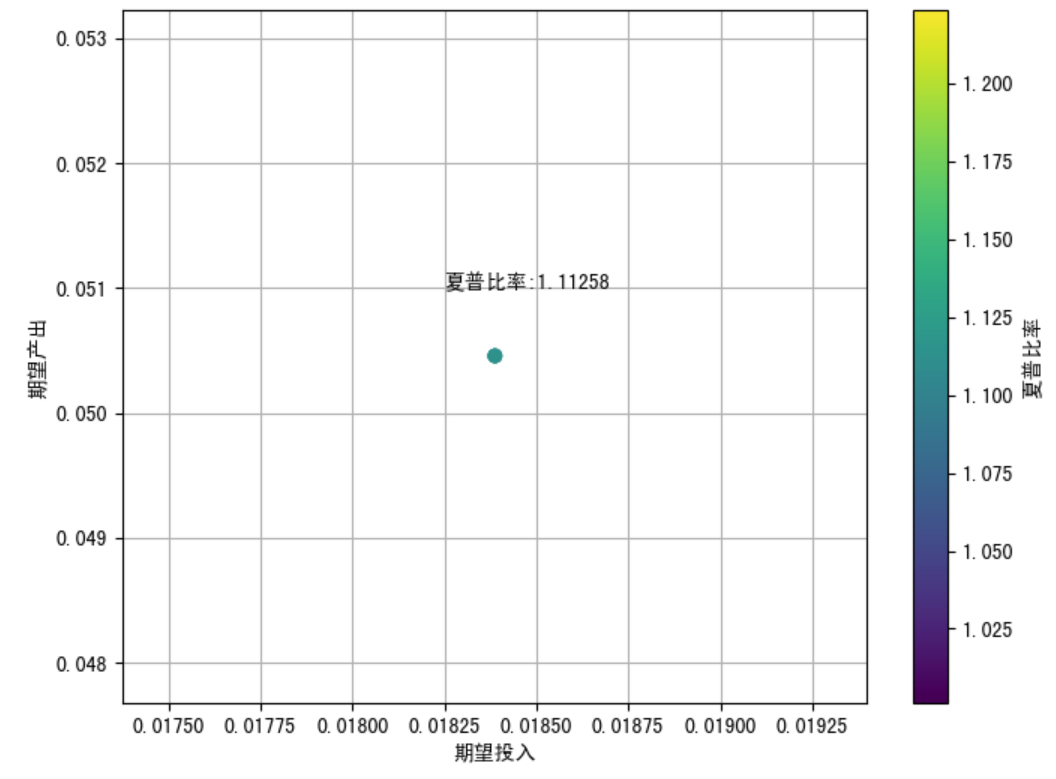
本文将投资组合性价比最高的组合定位最优投资组合。

因此，未来五年内最优投资组合为下图。

dbed339134f4cdea3ca9b2312cc1d43

**4.4.4 得到最优组合的夏普率**

由slsqp算法算出在最优投资组合下的夏普比率如下图所示



可见，在最优投资组合下，夏普比率达到1.1258。

# 五.模型的评价

**5.1 模型的优点**

1. 本文采用神经网络算法对未来经济情况和未来夏普指数走势进行预测，相比于传统的回归模型更具说服
2. 对于过多数据采用了PCA降维方法，使得数据后续好处理
3. 利用了熵权法挑选出合适的选项，具有客观性

**5.2模型的缺点**

1. 过多利用spsspro软件进行数据处理
2. 本文模型并没有考虑地面面俱到，部分情况考虑地不够细致
3. 宏观经济走向随机性太强，本文模型只能保证在理论上有效，并不能完全保证准确性

# 六.参考文献

[1] Scientific Platform Serving for Statistics Professional 2021. SPSSPRO. (Version 1.0.11)[Online Application Software]. Retrieved from <https://www.spsspro.com>.

[2] 甘浪雄,张怀志,卢天赋,等. 基于熵权法的水上交通安全因素[J]. 中国航海,2021,44(2):53-58.

[3] 郭京福，杨德礼.数据包络分析方法综述[J].大连理工大学学报，1998(02):116-121.

[4] 徐维超. 相关系数研究综述[J]. 广东工业大学学报,2012,29(3):12-17.

[5]周亮.Black-Litterman模型在大类资产配置中的应用：基于货币周期及风险平价策略的改进[J].运筹与管理,2021,30(08):198-204.

[6]陈其安,张媛,刘星.宏观经济环境、政府调控政策与股票市场波动性——来自中国股票市场的经验证据[J].经济学家,2010(02):90-98.DOI:10.16158/j.cnki.51-1312/f.2010.02.015.

[7]宋怡臻. 基于时间序列与LSTM模型的经济预测研究[D].西南交通大学,2019.DOI:10.27414/d.cnki.gxnju.2019.000222.

[8]张沐光,吴文倩.大类资产配置原理与相关指数的国内外实践分析[J].金融市场研究,2021(11):72-82.

[9]温彬,冯柏.2021年宏观经济形势展望与大类资产配置[J].金融市场研究,2021(03):62-73.

[10]杨爱斌.疫情冲击下的宏观经济及大类资产配置[J].债券,2020(04):70-74.

[11].中国宏观经济指标(2022年8月)[J].宏观经济研究,2022(10):176.

[12]王中兴,卢余刚.基于模糊熵和模糊夏普比率的多阶段投资组合模型及实证[J].广西大学学报(自然科学版),2019,44(06):1814-1821.DOI:10.13624/j.cnki.issn.1001-7445.2019.1814.

[13]赵影,周岳.CPI与实体经济状况、货币供应量的关系及对CPI未来走势的预测[J].金融经济,2010(10):130-131.

[14]蔡文捷. 基于风险平价策略的大类资产配置实证研究[D].浙江大学,2017.

[15]张清晨,陆云峰.基于熵权-模糊综合评价的过程质量评价方法[J].项目管理技术,2021,19(12):40-45.

[16]周琪琪,邵振国,林韩.基于秩和比综合评价法的用户谐波危害分级评估[J].电力电容器与无功补偿,2018,39(05):116-122.DOI:10.14044/j.1674-1757.pcrpc.2018.05.021.

[17]张沐光,吴文倩.大类资产配置原理与相关指数的国内外实践分析[J].金融市场研究,2021(11):72-82.

[18]刘文宇. 动态风险预算下的大类资产配置研究[D].上海交通大学,2019.DOI:10.27307/d.cnki.gsjtu.2019.000960.

# 七.附录及支撑材料

因本文全部代码太多，且部分数据处理分析在spsspro平台上完成，因此附录仅包含涉及到的核心源码，其余材料详见支撑材料。

7.1支撑材料清单：





7.2部分核心代码

7.2.1引用程序及实现平台

Python3.8

#PCA(主成分分析)参考 <https://zhuanlan.zhihu.com/p/401953327?ivk_sa=1024320u>

#多变量LSTM模型 参考 <https://www.freesion.com/article/6924142815/>

#熵权法 参考 <https://blog.csdn.net/LuckyMon/article/details/121509072>

#夏普比率计算参考https://blog.csdn.net/stay\_foolish12/article/details/97371586

Spsspro平台 <https://www.spsspro.com>.

#熵权法

#数据包络分析(DEA)

#相关性分析

7.2.2 PCA(主成分分析)代码 question1.ipynb

import numpy as np

from sklearn.decomposition import PCA

#PCA方法使用sklearn处理，参考自 <https://zhuanlan.zhihu.com/p/401953327?ivk_sa=1024320u>

zhibiao=['GDP:现价', 'GDP:支出法']

col\_data = temp[zhibiao]

pca = PCA(n\_components=1) # 将维度降到一维

pca.fit(col\_data)

pcacol\_data=pca.fit\_transform(col\_data)

pca\_temp['国内生产总值GDP(年)']=pcacol\_data

for i in zhibiao:

del temp[i]

zhibiao=['实体经济部门杠杆率', '居民部门杠杆率', '非金融企业部门杠杆率', '政府部门杠杆率', '地方政府杠杆率',

'金融部门杠杆率(资产方)', '金融部门杠杆率(负债方)']

col\_data = temp[zhibiao]

pca = PCA(n\_components=1) # 将维度降到一维

pca.fit(col\_data)

pcacol\_data=pca.fit\_transform(col\_data)

pca\_temp['中国宏观杠杆率(季)']=pcacol\_data

for i in zhibiao:

del temp[i]

##......其余文件类似上述处理##

#保存处理后的数据

pca\_temp.to\_excel('pca\_data.xlsx')

7.2.3 多变量LSTM模型 question2.ipynb question4.ipynb

#划分训练用的数据集

def create\_dateset(sequences, n\_steps):

train\_X, train\_Y = list(), list()

for i in range(len(sequences)):

last = i + n\_steps

if last > len(sequences)-1:

break

train\_x, train\_y = sequences[i:last, :], sequences[last, :]

train\_X.append(train\_x)

train\_Y.append(train\_y)

return np.array(train\_X), np.array(train\_Y)

#训练LSTM

def train\_LSTM(dataset,n\_steps=1):

#dataset：数据标准化后的dataset，共有5个变量

# n\_steps：历史回溯大小，默认为1

train\_seq1= dataset[:,0].reshape((dataset.shape[0], 1))

train\_seq2= dataset[:,1].reshape((dataset.shape[0], 1))

train\_seq3= dataset[:,2].reshape((dataset.shape[0], 1))

train\_seq4= dataset[:,3].reshape((dataset.shape[0], 1))

train\_seq5= dataset[:,4].reshape((dataset.shape[0], 1))

dataset = np.hstack((train\_seq1, train\_seq2, train\_seq3, train\_seq4,train\_seq5))

train\_X, train\_Y = create\_dateset(dataset, n\_steps)

num\_zhibiao = train\_X.shape[2]

model = Sequential()

model.add(LSTM(500, activation='relu', return\_sequences=True, input\_shape=(n\_steps, num\_zhibiao)))

model.add(LSTM(500, activation='relu'))

model.add(Dense(num\_zhibiao))

model.compile(optimizer='adam', loss='mse')

model.fit(train\_X, train\_Y, epochs=200, verbose=1,shuffle=False)

model.save('lstm\_model.h5')

last\_input=np.array(dataset[-1:,:])

return train\_X,train\_Y,last\_input,num\_zhibiao,n\_steps

#对数据进行标准化处理

dataset=data[data.columns[1:]].to\_numpy()

dataset = dataset.astype('float32')

scaler = MinMaxScaler(feature\_range=(0, 1))

dataset=scaler.fit\_transform(dataset)

#last\_input为最后一行的数据，用于预测未来的输入,num\_zhibiao是选择的指标数：5，n\_steps是历史回溯大小

#输出：testPredict为预测训练数据，testY为真实数据

train\_X,train\_Y,last\_input,num\_zhibiao,n\_steps=train\_LSTM(dataset,n\_steps=1)

#拟合效果评估

def train\_Predict(train\_X,train\_Y,scaler):

model=load\_model('lstm\_model.h5')

trainPredict = model.predict(train\_X)

testPredict = scaler.inverse\_transform(trainPredict)

testY = scaler.inverse\_transform(train\_Y)

count=0

for i in ['央行货币政策(日)', '超额存款准备金率(超储率)-金融机构', '中国大宗商品价格指数-总指数', '人民币存贷款利率','公开市场操作(周)']:

y\_true,y\_pre = testY[:,count], testPredict[:,count]

#评估拟合程度

print("MAPE :",np.mean(np.abs((y\_true - y\_pre) / y\_true)) \* 100)

print("RMSE :",np.sqrt(metrics.mean\_squared\_error(y\_true, y\_pre)))

print("MAE :",metrics.mean\_absolute\_error(y\_true, y\_pre))

print("R2 :",np.abs(r2\_score(y\_true,y\_pre)))

#展示拟合效果图

plt.plot(y\_true,label='True')

plt.plot(y\_pre, color='orange', label='Predict')

plt.xlabel( '年份')

plt.ylabel( '趋势')

plt.title( 'LSTM拟合效果图 2001-2021 %s'%i)

plt.legend() # 显示图例

plt.savefig('./问题2图片/LSTM拟合效果图 2001-2021 %s趋势.jpg'%i)

plt.show()

count+=1

return testPredict,testY

testPredict,testY=train\_Predict(train\_X,train\_Y,scaler)

def future\_predict(n\_step\_forward,x,num\_zhibiao,n\_steps,scaler,testPredict,testY):

model=load\_model('lstm\_model.h5')

predict\_list=[]

predict\_list.append(x)

# n\_step\_forward预测未来n-1年的情况

while len(predict\_list) < n\_step\_forward:

x = predict\_list[-1].reshape((-1, n\_steps, num\_zhibiao))

ypre = model.predict(x, verbose=0)

#预测未来的值

predict\_list.append(ypre)

#返回逆归一化后的预测值

Predict\_forword = scaler.inverse\_transform(np.array([ i.reshape(-1,1)[:,0].tolist() for i in predict\_list]))

return Predict\_forword[1:,:].tolist()

y\_pre=future\_predict(6,last\_input,num\_zhibiao,n\_steps,scaler,testPredict,testY)

#数据处理

predictdata=pd.DataFrame(range(2023,2028),columns=['年份'])

predictdata=pd.concat([predictdata,pd.DataFrame(y\_pre,columns=data.columns[1:])],axis=1)

data=pd.concat([data,predictdata])

data = data.reset\_index(drop=True)

#展示预测的指标未来5年趋势

for i in data.columns[1:]:

plt.plot(data['年份'].values[0:22],data[i].values[0:22],label=i)

plt.scatter(data['年份'].values[22:27],data[i].values[22:27],label=i+'-预测',c='orange',marker='x')

plt.legend()

plt.xlabel( '年份')

plt.ylabel( '趋势')

plt.title( '2001-2027 宏观经济环境-%s趋势'%i.replace(':','-'))

plt.savefig('./predict\_fig/2001-2027 宏观经济环境-%s趋势.jpg'%i.replace(':','-'))

plt.show()

7.2.4 熵权法 question3.ipynb

# 熵权法计算指数权重 参考 https://blog.csdn.net/LuckyMon/article/details/121509072

def cal\_weight(x):

x = x.apply(lambda x: ((x - np.min(x)) / (np.max(x) - np.min(x))))#归一化

rows = x.index.size

cols = x.columns.size

k = 1.0 / math.log(rows)

lnf = [[None] \* cols for i in range(rows)]

x = array(x)

lnf = [[None] \* cols for i in range(rows)]

lnf = array(lnf)

for i in range(0, rows):

for j in range(0, cols):

if x[i][j] == 0:

lnfij = 0.0

else:

p = x[i][j] / x.sum(axis=0)[j] #占一列数总和的比例

lnfij = math.log(p) \* p \* (-k)

lnf[i][j] = lnfij

lnf = pd.DataFrame(lnf)

E = lnf

d = 1 - E.sum(axis=0)

w = [[None] \* 1 for i in range(cols)]

for j in range(0, cols):

wj = d[j] / sum(d)

w[j] = wj

w = pd.DataFrame(w)

return w

#股票资产指数权重

w = cal\_weight(data[gupiao])

w.index = gupiao

w.columns = ['weight']

Print(w)

#大宗商品资产指数权重

w = cal\_weight(data[dazongshangpin])

w.index = dazongshangpin

w.columns = ['weight']

Print(w)

7.2.5 夏普比率计算 question4.ipynb

#夏普比率计算 参考自https://blog.csdn.net/stay\_foolish12/article/details/97371586

# 预测未来5年指数回报率

risk\_return = data[data.columns[1:]].iloc[21:27].pct\_change()

risk\_return = risk\_return.iloc[1:6]

risk\_return

#给不同资产随机分配初始权重,所有的权重系数均在0-1之间

weights = np.random.random(4)

weights /= np.sum(weights)

weights

# 投资组合优化使得夏普率最大

#记录重要的投资组合统计数据（收益，方差和夏普比）

def stats(weights):

weights = np.array(weights)

weights /=np.sum(weights)

port\_returns = np.sum(risk\_return.mean()\*weights)

port\_variance = np.sqrt(np.dot(weights.T, np.dot(risk\_return.cov(),weights)))

return np.array([port\_returns, port\_variance, port\_returns/port\_variance])

#最小化夏普指数的负值

def min\_goal(weights):

return -stats(weights)[2]

#设定参数权重

x0 = 4\*[1./4]

#将权重限制在0和1之间。

bnds = tuple((0,1) for x in range(4))

#使权重总和为1。

cons = ({'type':'eq', 'fun':lambda x: np.sum(x)-1})

#优化函数，最小化夏普比率的负值

opts = sco.minimize(min\_goal,

x0,

method = 'SLSQP',

bounds = bnds,

constraints = cons)

opts

#最优投资组合权重向量，小数点保留5位

print('最优投资组合权重')

print(opts['x'].round(5))

stats(opts['x']).round(5)

# 最小化投资组合方差

def min\_variance(weights):

return stats(weights)[1]

optv = sco.minimize(min\_variance,

x0,

method = 'SLSQP',

bounds = bnds,

constraints = cons)

optv['x'].round(3)

stats(optv['x']).round(5)

#在不同目标收益率水平（target\_returns）循环时，最小化的一个约束条件会变化。

target\_returns = np.linspace(0.0,0.5,50)

target\_variance = []

for tar in target\_returns:

#给定限制条件：给定收益率、投资组合权重之和为1

cons = ({'type':'eq','fun':lambda x:stats(x)[0]-tar},{'type':'eq','fun':lambda x:np.sum(x)-1})

res = sco.minimize(min\_variance, x0, method ='SLSQP', bounds = bnds, constraints = cons)

target\_variance.append(res['fun'])

target\_variance = np.array(target\_variance)

#下面是最优化结果的展示。

#叉号：构成的曲线是有效前沿（目标收益率下最优的投资组合）

#红星：sharpe最大的投资组合

#黄星：方差最小的投资组合

plt.figure(figsize = (8,4))

#圆点：随机生成的投资组合散布的点

plt.scatter(port\_variance, port\_returns, c = port\_returns/port\_variance,marker ='o')

#叉号：投资组合有效边界

plt.scatter(target\_variance,target\_returns, c = target\_returns/target\_variance, marker ='x')

#红星：标记夏普率最大的组合点

plt.plot(stats(opts['x'])[1], stats(opts['x'])[0],'r\*', markersize =15.0)

#黄星：标记方差最小投资组合点

plt.plot(stats(optv['x'])[1], stats(optv['x'])[0],'y\*', markersize =15.0)

#plt.annotate('夏普比率:'+str(shape\_r[0].round(5)),(0.01825,0.051))

plt.grid(True)

plt.title('最优化结果展示')

plt.xlabel('期望投入')

plt.ylabel('期望产出')

plt.colorbar(label = '夏普比率')

plt.savefig('./问题4图片/最优化结果展示.jpg')