题目 基于多目标优化的投资组合分析

摘 要

资产管理业务是我国金融体制改革过程中重要内容，平衡投资收益和系统性风险之间的关系，研究市场波动有利于有效监管和稳定市场。本文通过建立多目标优化，模糊-层次结构以及 *Prophet* 时间序列等模型来确定最优投资方案，研究未来的指数波动，并且给出一些投资建议和策略。

针对问题一，本文首先针对附件 10 只股票数据进行定性的可视化分析。波动趋势分析验证了部分参数的走势基本一致，并且利用 *Pearson* 相关性分析验证了参数之间的相关关系。通过正态分布 *Q-Q* 图得到了股价和成交量都是服从正态分布的，因此我们可以利用箱线图分析法提出异常值。而对于缺损数据的填补，我们利用基于最大似然理论的 *EM* 算法来填充缺损值。本文利用预处理之后的股票数据来确定最佳的投资组合方案。根据马科维茨资产组合理论，本文建立了基于固定风险水平，最大化收益的优化模型。考虑到，固定风险水平阈值一般较难设定，本文通过增加最小化投资风险价值目标将投资问题转化为多目标优化模型求解。对于多目标优化模型的求解，本文利用遗传算法进行寻优。最终模型求解结果为：最大的投资目标收益率为 4.3%，综合风险价值为 89.68。具体的投资选股方案见正文表 3。

针对问题二，本文选取操作性，求解难度，结果合理性，模型合理性，敏感度，建模难度等指标来评估问题一各种优化模型。通过给定相关判断矩阵结合层次排序算法给出了问题一各种模型的价值量化评估。然后本文将层次结构模型的量化评分作为模糊算子，基于最大隶属度原则，建立了模糊-层次结构模型。利用 *Matlab* 对上述模型进行求解，得到了问题一建立的多目标投资模型的定性评价为优，量化权重评分为 0.30075，优于另外四种选股投资模型，具体的模型评价结果见正文表 8。

针对问题三，本文首先针对成分股综合指数进行计算。利用股价矩阵和成交量占比矩阵，我们最终得到了 10 只成分股的综合指数。定性的波动趋势分析得到指数是一种波动上升的趋势。考虑到综合指数的周期性变化，本文建立了包含周期参数的 *Prophet* 时间序列模型。该模型具有自动寻参，开源有效的优点，目前广泛应用于金融预测领域。利用 *Prophet* 实现 *Prophet* 时间序列的求解，最终的指数趋势预测结果见正文图 25。结果表明未来一年的股价波动整体来说会相对平稳。2020-5 到 2020-9 这五个月时间，指数将会出现阶段性的反弹，这也符合新冠疫情基本控制住之后的经济复苏。本文进一步从经济复苏的机会把握和风险控制两个方面提出一些建议和策略。

关键词 极大似然 多目标优化 遗传算法 模糊层次结构 *Prophet* 时间序列

一、问题背景与重述

1.1 问题背景

中国的债券市场自 20 世纪 90 年代初开始发展，经过 30 年的快速发展，目前已经发展成为国民经济的重要构成部分。日益完善的市场交易制度、证券法规、信息披露制度，使得投资者更加理性，机构投资者发展规范、政府监管力度有效。而在证券市场中，波动是最为本质的属性和特点，它具有调节和监管作用，特别是对于股东权益最大化，人们风险收益、监督管理层的监管等都有重要影响，这主要源于证券市场的主要功能是积极引导资源的优化配置。因此研究证券市场波动原因和规律是理论研究和实证分析的重要内容，并且为上市公司、监管者、投资者提供决策依据[1]。

首先，上市公司作为我国证券市场的基础发挥着重要的作用，股票的价格围绕上市公司的内在价值上下波动。其次，在经济全球化的情况下，世界性的金融市场与证券市场的发展息息相关，例如，针对 2008 年出现的金融危机，政府部门出台了大量的宽松政策。再次，证券市场波动和我国经济发展相互影响，证券市场波动影响上市公司获得的经济效益，并且调控经济政策和规划未来发展策略，政府出台的政策与规划对于资金市场、人们的预期等都会发生作用，进而影响证券市场的波动。最后，国民经济的发展也会影响证券市场的波动。

1.2 问题重述

针对证券市场波动性，根据附件数据解决以下问题：

1. 补全附件数据中的缺损数据。对于附件的成分股数据，通过建立模型，给出合理选股方案和投资组合方案；

2. 给出合理的评价指标来评估问题一中的模型，并给出分析结果；

3. 通过附件股指数据和补充的数据，对当前的指数波动和未来一年的指数波动进行合理建模，并给出合理的投资建议和策略。

二、问题分析

2.1 问题一的分析

本题要求针对附件中的十只指数成分股相关数据进行预处理分析，主要包括异常值分析，缺失数据补齐等操作。并且利用预处理之后的数据进行建模，要求建立适当的数学模型来给出具体的投资组合方案，从尽可能的追踪指数，并且满足收益最大，风险最小的目标。

针对数据的预处理分析，本文首先选取一只股票进行相关参数的定性分析，主要分析该股票的波动趋势以及数据的分布状况。对于股票波动趋势的分析，可以通过添加趋势线来获取可视化的趋势分析。对于股票历史数据的预处理分析，本文主要通过箱线图分析进行异常值数据处理，对于缺损数据的补齐，常用的缺失值填充方法包括中位数填补，均值填补以及拟合插值填补。为了更加准确的进行缺失值数据的填补，本文考虑利

用基于最大似然估计的EM算法来进行缺失值的填充。通过上述数据预处理步骤，我们利用预处理之后的股票相关数据建立投资组合模型。

针对投资组合方案的选取，我们主要是根据附件十只股票数据来选取最佳的投资方案。对于投资组合方案的确定，本文可以通过建立组合优化模型来获取最佳的投资方案。一般来说，风险资产的投资首先需要解决的是两个核心问题：即预期收益与风险。如何测定组合投资的风险与收益和如何平衡这两项指标进行资产分配也是投资者迫切需要解决的问题。本文参考马科维茨理论，考虑证券组合预期收益、风险方差计算等建立了考虑一定收益水平下的均值－方差模型。由于马科维茨组合模型的主要目标是限制投资效用下的风险最小化。但是该模型需要设定一个投资收益率限制条件，因此本文考虑在马科维茨模型的基础上添加目标投资收益最大化，从而构建一个多目标优化模型。其中主要目标为投资风险方差最小化，投资收益最大化，比较重要的约束条件有每只股票投资比例大于等于0，股票投资不可以卖空等。主要的决策变量为每只股票的投资占比。对于多目标优化模型的求解，可以利用线性加权，主要目标法等转化为单目标求解。但是考虑到转化为单目标求解，无法得到一个较好寻优结果，本文考虑利用遗传算法等启发式算法来寻优，从而尽可能的搜索到一个全局最优解。

2.2 问题二的分析

本题要求给出合理的评价指标来评估问题一的模型，主要是通过建立综合评价模型来评价问题一投资组合模型。对于合理的评价指标选取，我们通过选取一些针对优化模型的检验指标来进行综合评估。对于指标的选取，我们主要从三个方面来选取，分别是是模型的合理性，模型求解结果，模型可操作性等三个角度选取不同的评价指标。

对于问题一模型的综合评估，我们可以根据相关指标数据来进行量化评分。但是考虑到很多评价指标无法获取量化数据，因此本文考虑构建一个层次结构模型，根据指标专家打分或者人为设定权重来评价问题一模型优劣。进一步根据利用模糊算子，基于最大隶属度原则，给出评价结果。通过层次结构模型确定目标权重，利用模糊评价法来获取针对问题一模型的定性评价。最终通过建立的模糊层次结构模型来获取最终的评价结果。对于评价目标，考虑到问题一针对投资组合方案建立的数学模型，我们可以抽出多个数学模型和本文建立的多目标优化模型进行评价对比，从而说明本文针对问题一建立投资组合模型的优劣程度。

2.3 问题三的分析

本题要求利用附件一的相关数据，给出当前指数的波动情况。并且利用已有数据来预测未来一年的指数波动状况，从而给出一些合理的投资建议和策略。

对于当前指数波动状况的分析，本文首先利用附件十只股票数据来量化一个股票综合指数，对于综合指数的量化，我们主要是利用十只成分股来集权计算出一个成分股指数。进一步针对该指数的变化趋势进行分析。而对于未来一年指数波动情况的预测，我们综合考虑指数波动的周期性，可以建立时间序列模型进行预测。对于时间序列预测模型，常用的包括自回归移动平均混合模型*ARIMA*，指数平滑法等。我们考虑建立基于周

期性变化的 *Prophet* 时间序列预测模型来预测未来的指数变化趋势，该模型参数设置简单，更加符合目前的时间序列建模趋势。为了重复说明时间序列预测指数波动的优越性，我们还建立了基于不同核函数的支持向量机回归（*SVR*）来验证 *Prophet* 时间序列预测的有效性[2]。

三、模型假设

结合本题的实际，为了确保模型求解的准确性和合理性，我们排除一些位置因素的干扰，提出以下几点假设：

1、假设附件所给的股票相关参数数据都是真实有效的；

2、假设投资者每一次投资选择的主要依据是某一时间内的证券收益的概率分布；

3、假设投资者的决定仅仅是依据证券的风险和收益，不考虑各种外界因素的影响；

4、假设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；

5、假设股票指数可以通过附近 10 只股票相关数据的加权计算来获取；

6、假设模型评价指标的选取都是合理有效的。

四、定义与符号说明

为了便于问题的求解，给出以下符号说明：

符号	说明
B_i	表示第 i 只股票的平均收益率
μ_i	各个月份每股的平均收益率
$B_{10\times1}$	十只股票的成交量占比矩阵
w_i	投资的第 i 种股票的权重
F	有效边缘的相切点 F
σ_i^2	股票 i 的方差
I_t	每股每月的股价
B_{it}	每股每月的收益率
r	投资组合的期望收益
μ_0	固定投资整体收益水平
$A_{i\times10}$	十只股票的价格指数矩阵
σ_0^2	满足投资者方差的一个常数
r_j	第 j 种风险资产的收益率的均值
a_{ij}	指标 i 和指标 j 的相对重要度之比

五、模型的建立与求解

5.1 问题一模型的建立与求解

本题要求针对附件中的十只指数成分股相关数据进行预处理分析，主要包括异常值分析，缺失数据补齐等操作。并且利用预处理之后的数据进行建模，要求建立适当的数学模型来给出具体的投资组合方案，从尽可能的追踪指数，并且满足收益最大，风险最小的目标。本题可以将问题确定为两个小问题，分别是针对附件股票数据的预处理分析以及具体投资组合方案的选取。

5.1.1 问题一的解题思路流程

针对数据的预处理分析，本文首先选取一只股票进行相关参数的定性分析，主要分析该股票的波动趋势以及数据的分布状况。对于股票波动趋势的分析，可以通过添加趋势线来获取可视化的趋势分析。对于股票历史数据的预处理分析，本文主要通过箱线图分析进行异常值数据处理，对于缺损数据的补齐，常用的缺失值填充方法包括中位数填补，均值填补以及拟合插值填补。为了更加准确的进行缺失值数据的填补，本文考虑利用基于最大似然估计的 EM 算法来进行缺失值的填充。通过上述数据预处理步骤，我们利用预处理之后的股票相关数据建立投资组合模型。

针对投资组合方案的选取，我们主要是根据附件十只股票数据来选取最佳的投资方案。对于投资组合方案的确定，本文可以通过建立组合优化模型来获取最佳的投资方案。一般来说，风险资产的投资首先需要解决的是两个核心问题：即预期收益与风险。如何测定组合投资的风险与收益和如何平衡这两项指标进行资产分配也是投资者迫切需要解决的问题。本文参考马科维茨理论，考虑证券组合预期收益、风险方差计算等建立了考虑一定收益水平下的均值－方差模型。由于马科维茨组合模型的主要目标是限制投资效用下的风险最小化。但是该模型需要设定一个投资收益率限制条件，因此本文考虑在马科维茨模型的基础上添加目标投资收益最大化，从而构建一个多目标优化模型。其中主要目标为投资风险方差最小化，投资收益最大化，比较重要的约束条件有每只股票投资比例大于等于 0，股票投资不可以卖空等。主要的决策变量为每只股票的投资占比。对于多目标优化模型的求解，可以利用线性加权，主要目标法等转化为单目标求解。但是考虑到转化为单目标求解，无法得到一个较好寻优结果，本文考虑利用遗传算法等启发式算法来寻优，从而尽可能的搜索到一个全局最优解。

根据上述解题思路阐述，我们给出本题得解题思路流程图如下：

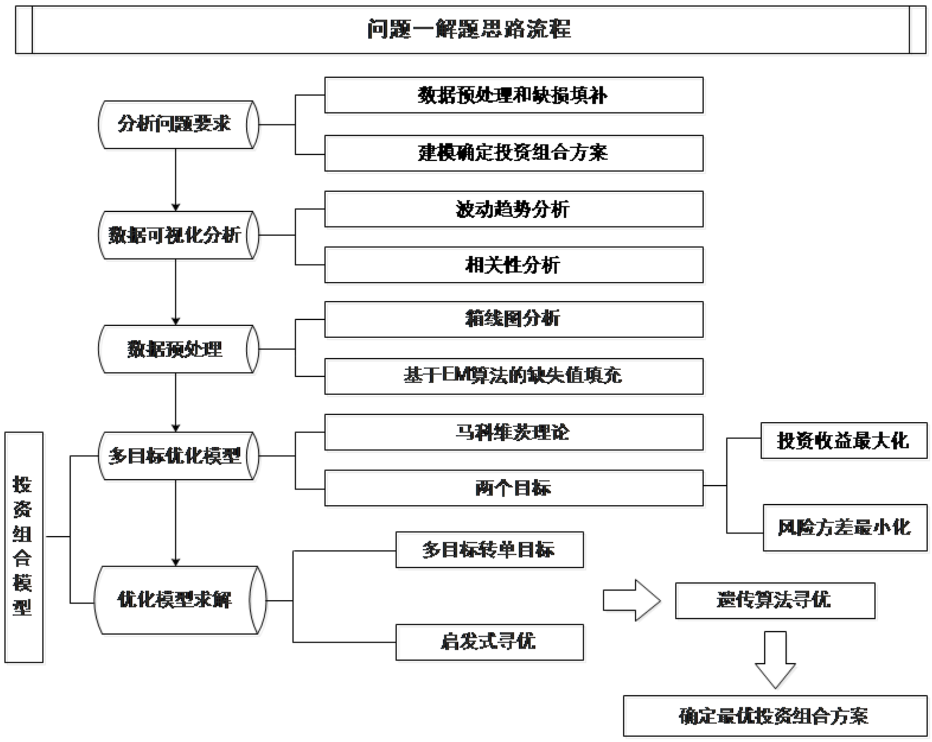


图 1 问题一解题思路流程图

根据上述思路流程，我们首先针对附件数据进行一些预处理分析。

5.1.2 附件数据的可视化分析

附件一的成分股数据中包含了一共 10 只股票，我们首先针对单一股票的相关走势进行分析。我们首先给出了其中一只股票 *abc*001 所有参数随着时间的变化趋势，具体的走势如下：

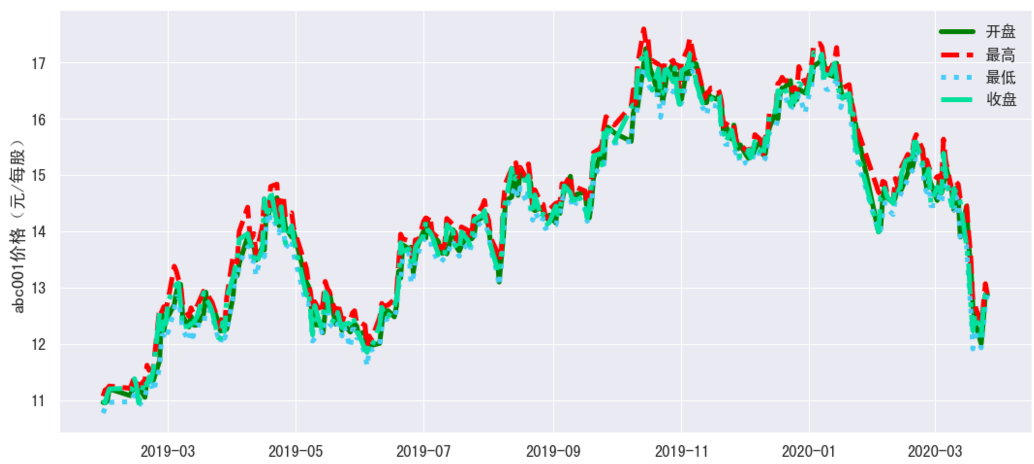


图 2 股票 *abc*001 所有价格参数走势分析

根据上图我们可以得到一只股票的开盘，最高，最低，收盘等参数的走势基本一致。因此本文后续主要针对股票的收盘价，成交量等参数进行分析。为了进一步验证相关参数走势基本一致的结论，本文利用皮尔逊相关性分析来研究各种参数之间的相关关系。

为了进一步的分析变量之间的线性关系，我们利用清洗好的数据集进行线性相关性分析。根据数据得到相关系数矩阵，计算它们的相关系数矩阵[4]：

$$R=\begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{pmatrix}$$

其中 $r_{ij}\left(i,j=1,2,\ldots,p\right)$ 表示是原变量与 x_j 的相关系数，并且 $X_{ij}=X_{ji}$。相关系数的计算公式为：

$$r_{ij}=\frac{\sum_{k=2}^{n}\left(x_{ik}-\bar{x}_i\right)\left(x_{kj}-\bar{x}_j\right)}{\sqrt{\sum_{k=1}^{n}\left(x_{ik}-\bar{x}_i\right)^2\sum_{k=1}^{n}\left(x_{kj}-\bar{x}_j\right)^2}}$$

根据上述公式，利用 *Python* 绘制相关关系矩阵如下：

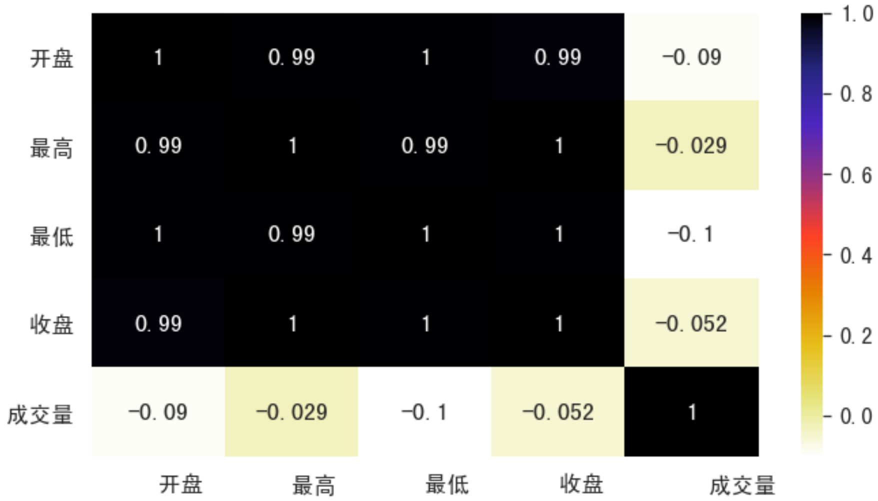
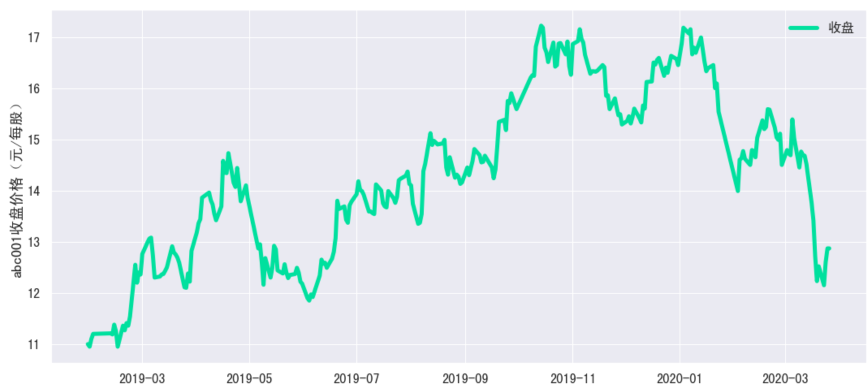


图 3 不同变量之间的相关关系矩阵

根据相关关系矩阵，我们可以看出开盘，最高，最低，收盘四个参数之间是一种显著正相关关系，相关系数在 0.99 以上，这也说明了四个参数趋势基本一致。进一步分析可以得到股票成交量和股票各种价格参数之间没有显著的相关关系，相关系数普遍低于 0.1。为了进一步分析股票价格走势情况，我们主要针对收盘价和成交量进行分析。给出了



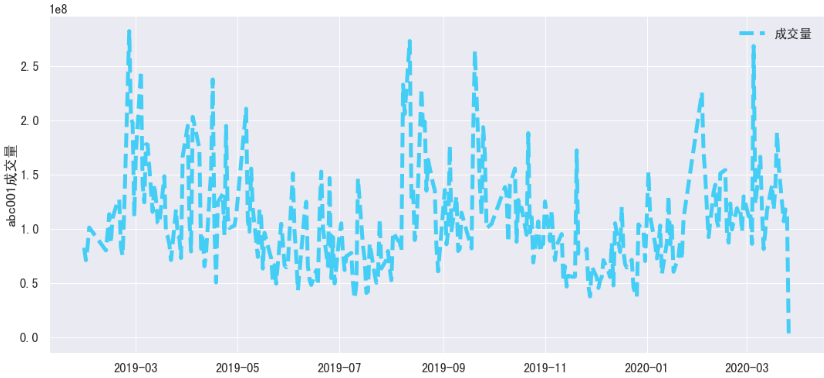


图 4 股票 *abc*001 收盘价和成交量随着时间的变化趋势

根据上图我们可以发现 *abc*001 股价走势在整体上是一种波动上扬趋势，股价波动较大，要想能够有所收益，最重要的就是抓住买入时机。而对于该股票得成交量指标，可以得到股票的成交量走势规律很难找到，但是根据成交量曲线的波动情况可以发现，成交量随着时间的走势呈现一种周期性的波动。

为了进一步分析股票收盘价与成交量指标，我们又给出了 *abc*008 相关参数指标随着时间的变化趋势。

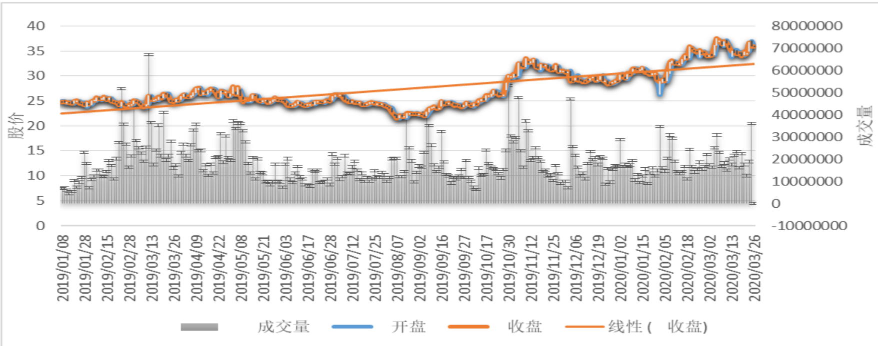


图 5 股票 *abc*008 参数趋势图

根据上图我们可以得到股票 *abc*008 最近一年的股价和成交量可视化趋势。为了更好的说明股票收盘价趋势，我们给出了线性拟合函数，从而比较直观的给出了股价的未来走势。可以看出股票 *abc*008 过去一年的走势也是一种比较明显的线性增长趋势。

上面的数据可视化分析只是从单一股票入手分析，本文利用附件 1 的十只股票数据进行综合分析。我们首先给出了十只股票在过去一年的收盘价曲线：

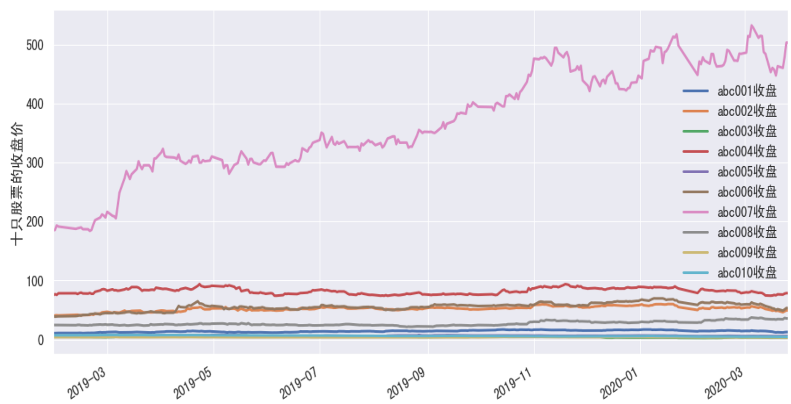


图 6 十只股票的收盘价曲线

根据上图，我们可以得到成分股十只股票的收盘价曲线中除了 *abc*007，其他的股票收盘价走势都较为平缓。其中股票 *abc*007 的股价相对于去年的股价整整翻了一倍多。这也说明了 *abc*007 股价波动较大，风险水平较高，而其他九只股票虽然收益较低，但是波动较小，股票风险也比较小。因此我们后续的选股模型需要以次分析作为依据。充分考虑股票投资的收益与风险，从两个角度出发，做到两者兼顾的投资选股。

我们进一步给出了十只股票成交量在夺取一年的走势分析：

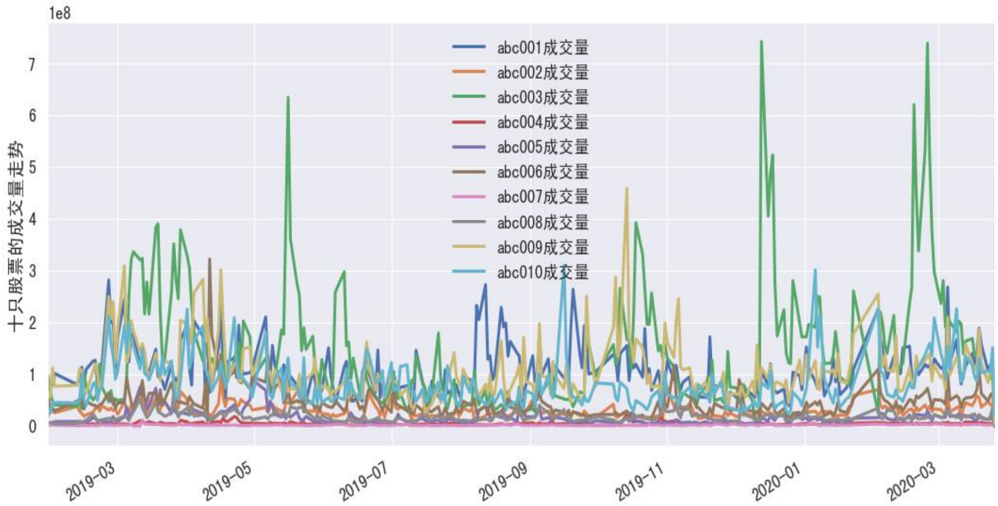


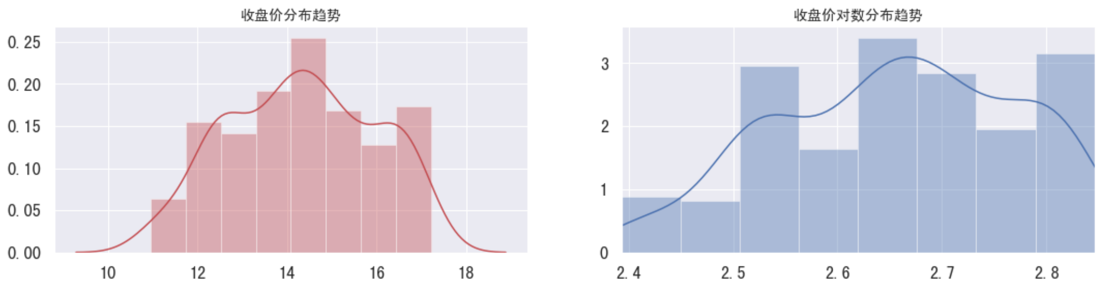
图 7 十只股票过去一年的成交量分析

根据上图我们可以发现各只股票成交量随着时间的变化趋势。其中 *abc*003，*abc*009 的成交量相对来说较大，但是股价走势较为平缓，这可以从侧面推出这两只股票的盘子比较大，相对较大的成交量很难影响到股价。也就是这两只股票相对于其他股票更加稳，波动风险水平较小。而股票 *abc*007 的股价在过去一年波动虽然很大，但是成交量水平确很低，也就是该股票可能属于创业板，盘子比较小，较小的资金成交量就可以导致股价大幅度的波动。

上述分析对于我们后续的投资选股都有十分重要的参考价值，为了方便后续模型的建立，我们还需要针对股票价格分布情况以及数据进行一些预处理分析。

5.1.3 异常值分析与缺失值填补

根据上述的可视化分析，我们接下来主要是针对股票数据得分布状况进行分析检验。我们首先给出 *abc*001 收盘价与成交量的分布直方图：



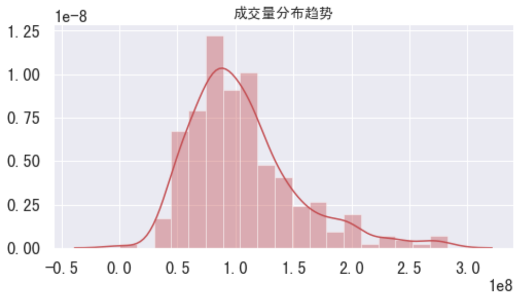


图 8 *abc*001 收盘价与成交量的分布直方图

根据上图我们可以得出，*abc*001 股票过去一年的收盘价格主要集中在 12-16 块一股之间。这也说明该股票的波动较小，对应的股票风险水平也不高[3]。为了方便进一步针对各只股票分布状态的研究，我们参考上述分布直方图可以发现，股价和成交量在一定程度上是呈现一种正态分布的，正态分布主要表现为数据主要分布于中央，属于中间大、两头小。为了进一步检验数据的分布状态，我们利用正态分布 *Q-Q* 图来进行检验。

分布检验 *Q-Q* 如下：

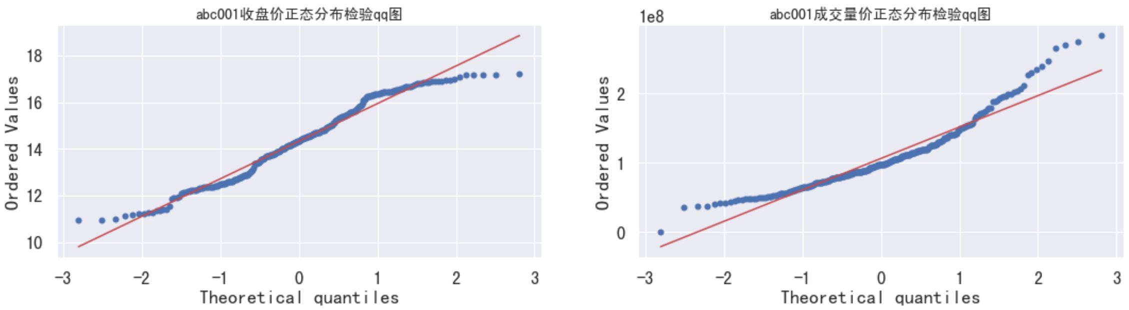


图 9 *abc*001 股价和成交量正态分布检验 *Q-Q* 图

可以看出 *abc*001 股价和成交量都是接近直线状态的，接近于直线 $y=x$ 的状态。根据上述正态分布检验，我们可以得到 10 只股票的成交量与收盘价历史数据都基本维持一种正态分布，这也为我们后续预测成分股指数趋势以及数据填充和处理奠定基础。

根据上述针对数据分布状态的分析，本文接下来主要是利用箱线图分析法和相关数据补全算法来对附件的 10 只股票数据进行预处理分析。最终利用上述预处理之后的数据来进行选股方案和投资组合方案的研究。

5.1.4 基于马科维茨理论的多目标优化模型建立

针对投资组合方案的选取，我们主要是根据附件十只股票数据来选取最佳的投资方案。对于投资组合方案的确定，本文可以通过建立组合优化模型来获取最佳的投资方案。一般来说，风险资产的投资首先需要解决的是两个核心问题：即预期收益与风险。如何测定组合投资的风险与收益和如何平衡这两项指标进行资产分配也是投资者迫切需要解决的问题。

为了确定最优的股票投资组合策略，此处以投资效用最大化为目标，建立投资效用最大化模型来研究基金公司的股票投资组合，它依据每类证券的预期收益率、方差和证券之间的协方差矩阵，通过计算求出投资组合的有效边界，再依据投资者的效用无差异曲线，最终得出最佳投资组合。

根据马科维茨资产组合理论，根据投资组合概念，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[6]。根据马科维茨对于投资组合的概念，得知该模型也在追求投资组合风险最小化，符合本文对于最优投资目标的实现。通过查找资料可以发现投资效用由两个因素所决定，投资者在投资股票时会考虑到风险以及收益的因素，所有的投资者都会希望收益越大越好，但是收益越大时，所面临的风险也会越大，会面临亏损的风险。因此投资者也希望风险越小越好。

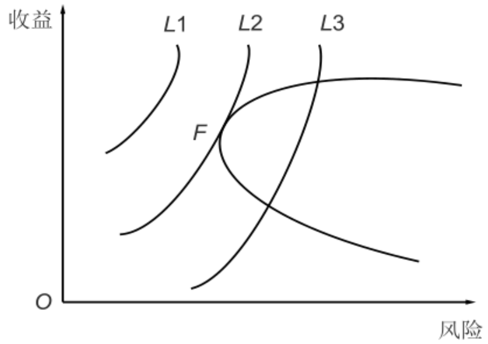


图 10 无差异曲线与有效边缘的组合

根据经济学原理，无差异曲线上反映着风险和收益呈正向变动，$L1$ 的投资偏好是相对于 $L2$ 偏高的，它与有效边缘的相切点 F 点作为他的最佳投资组合。

本文首先做出以下假设：

（1）投资策略要求基金公司在 2019 年初买入股票后一直持有；

（2）以每支股票的收益率的均值来衡量未来实际收益率的水平，以每支股票的收益率的方差来衡量每股的风险程度，当标准方差越小时，说明风险程度也比较低。

马科维茨投资模型把投资组合的价格变化看作随机变量，用它的平均值来衡量收益，并用它的方差来衡量风险。在此基础上，目标为效用最大化，且结合题目所给定的要求，不考虑股票之间的相关性以及股票不允许卖空，建立效用最大化模型：

$$max\, r = \sum_{i}^{n} \sum_{j=1}^{n} w_i r_j$$

其中，r 表示投资组合的期望收益，$r_j\left(j=1,2,\cdots n\right)$表示第 j 种风险资产的收益率的均值,w_i 表示投资的第 i 种股票的权重。

单纯考虑收益在证券资产配置策略中是行不通的，往往收益越大投资其风险也就越高。本题从方差的角度对风险进行量化。及考虑在某一风险下的最高收益。故需要满足的约束条件为：

$$\sum_{i=1}^{n} w_i \sigma_i^2 \le \sigma_0^2$$

其中 σ_i^2 表示股票 i 的方差，σ_0^2 表示满足投资者方差的一个常数。

并且各股票投资比例之和为 1，故得到约束条件为：

$$\sum_{i=1}^{n} w_i = 1$$

结合题目要求与我国证券行业的实际情况，在我国股票是不允许进行卖空操作的，故此时可以得知分配系数的约束条件为：

$$w_i \geq 0$$

故此投资收益最大化模型为：

$$max\ r = \sum_{i}^{n} \sum_{j=1}^{n} w_i r_j$$

$$s.t \begin{cases} \sum_{i=1}^{n} w_i \sigma_i^2 \leq \sigma_0^2 \\ w_i \geq 0 \\ \sum_{i=1}^{n} w_i = 1 \end{cases}$$

在马科维茨投资理论指导下建立的数学模型的主要目标是投资收益最大化。而风险控制主要是依靠投资股票价格方差参数维持在某一特定水平下来确定的。而我们的目标是寻求一个最优的投资组合策略，因此本文考虑增加一个优化目标，也就是最小化投资风险价值[7]。

其中风险价值可以通过投资组合方案中每只股票风险价值量化总和来衡量。每只股票具体的风险价值可以通过股票收盘价的方差来确定，利用风险价值 VAR 模型来量化。

最终的最小化风险价值目标函数为：

$$min\ \phi = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i \sigma_j$$

对于每只股票的风险价值计算，我们可以利用风险价值模型进行量化。也可以通过每只股票收益率的方差来衡量每只股票的离散波动情况，也就是每只股票的风险程度。本文在这里选用了收益率方差来进行求解。最小化风险价值目标函数可以简化为：

$$min\ \phi = \sum_{i=1}^{j} \sum_{j=1}^{j} w_i w_j Cov\left(B_i, B_j\right)$$

其中 B_i 表示第 i 只股票的平均收益率，最终的多目标优化模型如下：

$$min\ \phi = \sum_{i=1}^{j} \sum_{j=1}^{j} w_i w_j Cov\left(B_i, B_j\right)$$

$$max\ r = \sum_{i}^{n} \sum_{j=1}^{n} w_i r_j$$

$$s.t \begin{cases} \sum_{i=1}^{n} w_i \sigma_i^2 \leq \sigma_0^2 \\ \sum_{i=1}^{10} w_i \mu_i \geq \mu_0 \\ w_i \geq 0 \\ \sum_{i=1}^{n} w_i = 1 \end{cases}$$

对于上述模型的求解，本文可以采用主要目标法和分层序列法将多目标优化模型转化为单目标模型。多目标转化为单目标求解可以将模型转化为以下两种情况：

（1）固定风险水平 σ_0^2，最大化投资收益目标。

模型可以转化为经典的马科维茨投资组合模型：

$$max\ r = \sum_{i}^{n}\sum_{j=1}^{n} w_i r_j$$

$$s.t\begin{cases}\sum_{i=1}^{n} w_i \sigma_i^2 \le \sigma_0^2 \\ w_i \ge 0 \\ \sum_{i=1}^{n} w_i = 1\end{cases}$$

（2）固定投资整体收益水平 μ_0，最小化投资风险价值。

模型可以转化为最小化风险价值目标：

$$min\ \phi=\sum_{i=1}^{j}\sum_{j=1}^{j} w_i w_j Cov\left(B_i, B_j\right)$$

$$s.t\begin{cases}\sum_{i=1}^{n} w_i \mu_i \ge \mu_0 \\ w_i \ge 0 \\ \sum_{i=1}^{n} w_i = 1\end{cases}$$

但是对于考虑到转化为单目标求解，需要固定一个目标参数。从而导致无法得到一个较好寻优结果，本文考虑利用遗传算法等启发式算法来寻优，从而尽可能的搜索到一个全局最优解。

5.1.5 基于遗传算法的优化模型求解

对于上述建立的多目标优化模型，我们需要利用合适的算法进行求解。未来求解出最优投资方案。我们首先要对一些参数进行求解，比如每只股票的平均收益率以及每只股票的风险价值计算。

下面是针对模型求解的一些参数计算：

*Step*1：求出每股每月的收益率 B_{it}

$$B_{it} = \frac{I_t - I_{t-1}}{I_{t-1}}\ t = 2,3,\cdots,12$$

其中 I_t 表示的是每股每月的股价。

*Step*2：求出各个月份每股的平均收益率，即数学期望 μ_i

$$\mu_i = \frac{1}{11}\sum_{t=2}^{12} B_{it}\ i = 1,2,\ldots,10$$

求得 μ_1=0.039, μ_1=0.0159, $\cdots$, μ_{10}=0.0386 。

Step 3：计算出协方差矩阵 $Cov\left(B_i, B_j\right)$

$$Cov\left(B_i, B_j\right)=\frac{1}{n}\sum_{t=1}^{n}\left(B_{it}-\mu_i\right)\left(B_{jt}-\mu_j\right)$$

其中 Cov 表示股票 i 和股票 j 之间的协方差，得到：

$$Cov=\begin{Bmatrix} 2.47 & \cdots & 12.3 \\ \vdots & \ddots & \vdots \\ 12.3 & \cdots & 2.47 \end{Bmatrix}$$

表示 10 个股票相互之间组合的相关性，即得到 $Cov(B1, B1)=2.47$，$Cov(B1, B10)$ =12.3 等数据。当协方差的绝对值越大时，每支股票的相互影响的概率就越大，风险也就越大。为了使风险能进一步降低，我们需要在制定投资策略中使得协方差进一步降低。

其中 σ^2 表示的是 10 只股票，两两组合的方差即组合投资的方差，w_i、w_j 表示第 i、j 股票投资的比例，μ_i 表示每支股票的平均收益率。

根据上述算法步骤，我们首先计算出来每只股票每个月的平均收益率，进而获取每只股票的综合平均收益率如下：

表 1 十只股票的综合平均收益率与方差

股票	综合平均收益率	方差
*abc*001	0.018865	2.63322
*abc*002	0.018851	17.54344
*abc*003	-0.01702	0.982076
*abc*004	0.003246	25.98487
*abc*005	0.007582	0.254715
*abc*006	0.028848	45.48436
*abc*007	0.075775	506.4598
*abc*008	0.028712	14.63643
*abc*009	-0.00408	0.061625
*abc*0010	-0.02502	0.74779

进行排序并且可视化，最终得到十只股票的综合平均收益率排序可视化结果如下：

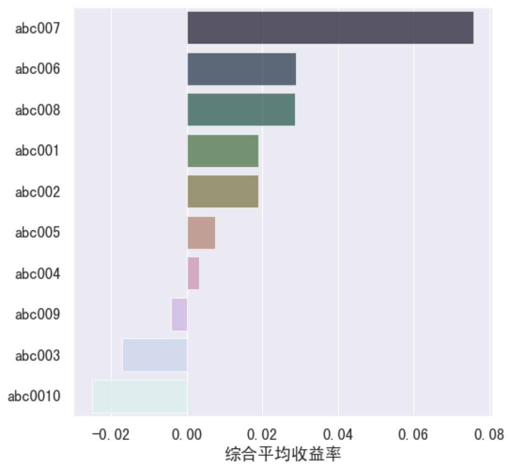


图 11 十只股票的综合平均收益率排序可视化

根据上图我们可以发现 *abc*007 的综合收益率较高，年化平均收益可以达到 8%。这十只股票的收益率差异较大，整体收益率几乎分布在-0.02 至 0.1 之间。其中综合平均收益率最低的是 *abc*010，最低收益率-2.5%。为了进一步分析十只股票股价波动风险水平，我们针对十只股票的方差进行计算分析。最终十只股票的波动方差计算结果如下图：

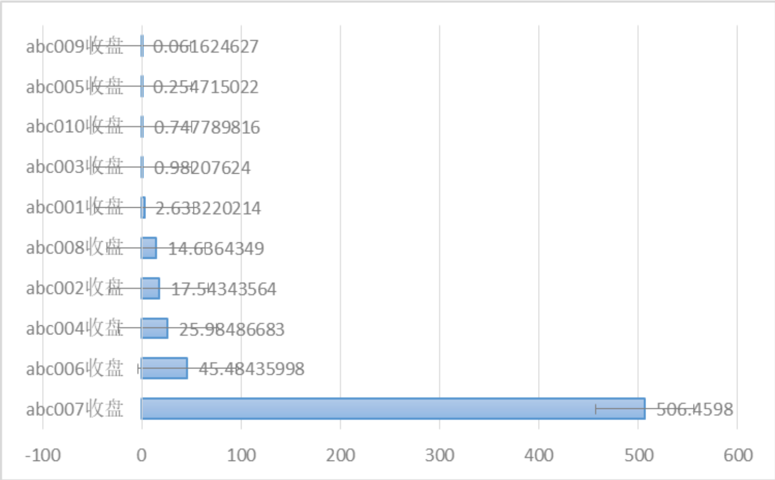


图 12 十只股票的波动方差可视化

根据上图我们可以发现除了 *abc*007 的方差较为异常之外，其他股票的波动方差基本维持在 50 以内，这基本是一个可控范围。*abc*007 的方差加大，这也说明该股票的风险较大。考虑到该股票综合收益率也是最高的，这也说明了投资中一个非常经典的理念：风险越大，收益越大。根据上述针对平均收益率以及波动方差可视化分析的结果我们可以找出波动风险较小而且收益较为客观的几只股票为：*abc*008，*abc*006，*abc*001。

根据上述定性分析虽然也可以选出较好的股票组合方案，但是无法确定出具体股票选择占比。因此本文从基金理财投资的角度[5]建立了多目标优化模型来确定最优选股方案。

1、对于多目标优化模型的求解，我们主要利用上述求解参数记性求解。考虑到多目标优化模型求解的难度，本文主要利用启发式算法进行寻优。解就是遗传基因的组合。为减少组合数量，将图像中央进行分块。把每一块看成一个基因，进行组合优化计算；

2、初始群体的生成：随机产生 N 个初始串结构数据，每一串结构数据叫做一个个体。有 N 个体，组成一个群体。*GA* 以这 N 个结构数据作为初始点迭代。参数 N 要根据问题的规模来确定；

3、杂交：由交换概率挑选的，每两个父代之间通过将相异的部分基因进行交换，产生新的个体。这样得到新的一代个体，新的个体组合了父辈个体的特性。体现了信息交换的思想；

4、适应度值评估监测：计算交换得到的新个体的适应度，适应度可以用来度量种群中的个体优劣（符合条件的程度）的指标值；

5、选择：选择是为了能够从交换后的群体中选择出优良的个体，让它们有机会作为父代来为下一代繁衍子孙。遗传算法通过选择过程体现这一思想，进行选择的原则是

适应性强的个体为下一代贡献的概率大，选择实现了达尔文的适者生存的原则；

6、变异：变异首先在群体中随机选择一定数量个体，对于选中的个体以一定的概率随机地改变串结构数据中某个基因的值。

本文给出了遗传算法的一般流程图如下：

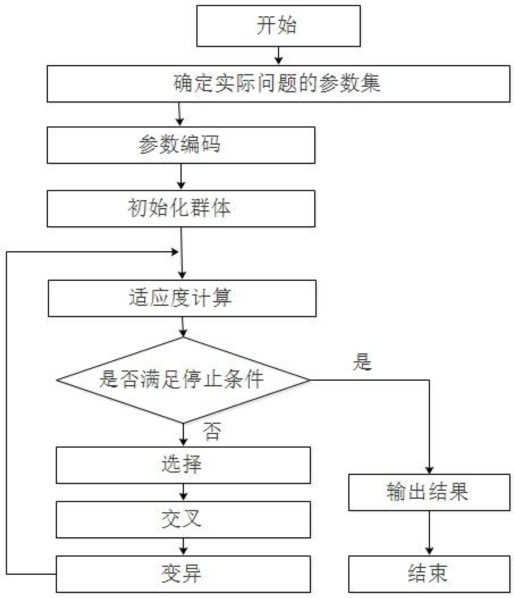


图 13　遗传算法运算流程图

遗传算法的具体实现步骤：

Step 1：将要求问题的参数范围来进行评估，并对其进行编码；

Step 2：随机产生有多个个体组成的种群；

Step 3：对种群中所有个体进行解码，再通过解码后的参数来对适应度函数进行求解。从而评估出个体的适应度；

Step 4：判断收敛条件，如果求出最优解则停止收敛选择，让适应度值比较大的个体越来越多，而适应度较小的个体会被淘汰；

Step 5：交叉，使两个个体以某一概率进行交叉操作；

Step 6：变异，让某个个体以一定的概率进行变异操作，使个体特性发生改变；

Step 7：重复 *Step* 3 到 *Step* 7，直到参数收敛为止。

根据以上所建遗传算法求解目标规划模型，利用 *Matlab* 遗传算法工具箱来求解目标规划模型。

遗传算法的具体参数值如下表所示：

表 2 遗传算法具体参数值

遗传算法参数	种群规模	迭代次数	交叉概率	变异概率
数值	200	200	0.9	0.01

同时为了更为直观的看出目标函数同迭代次数之间的关系，本文运用 *Matlab* 求解得到目标函数与迭代次数的图像如下图所示：

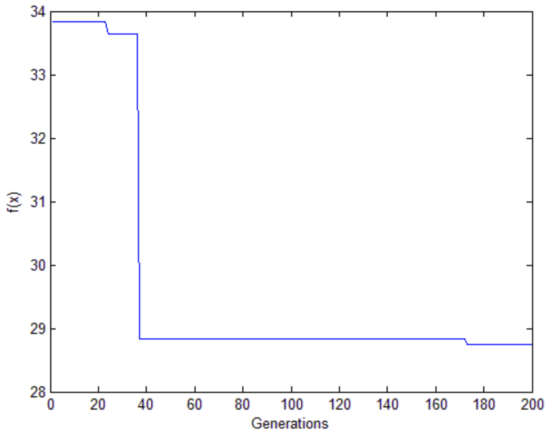


图 14 遗传算法适应度函数收敛图

根据目标函数与迭代次数关系图可以发现，遗传算法经过六次迭代就能得到问题的最优解，因此遗传算法的收敛性较好。算法的收敛性与算法的迭代过程有关，本文给出了遗传算法的迭代过程图，具体迭代过程图如下：

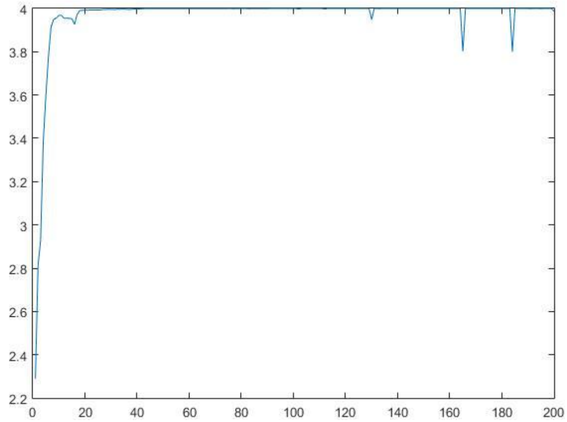


图 15 遗传算法迭代收敛过程

从图可以看出，随迭代次数的增加，算法逐渐收敛。说明遗传算法的收敛性较好。最终的求解结果为：最大的投资目标收益率为 4.3%，综合风险价值为 89.68.

最终的投资选股方案如下：

表 3 最终的投资选股方案（股票投资占比）

股票	投资占比	股票综合收益率
*abc*001	0.203	0.018865
*abc*002	0	0.018851
*abc*003	0	-0.01702
*abc*004	0	0.003246
*abc*005	0	0.007582
*abc*006	0.108	0.028848
*abc*007	0.346	0.075775
*abc*008	0.343	0.028712
*abc*009	0	-0.00408
*abc*0010	0	-0.02502

对最终的投资选股方案进行可视化可以得到结果如下图：

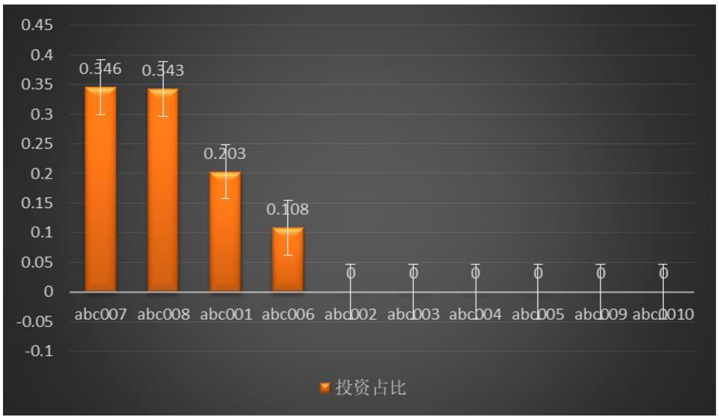


图 16 最终的投资方案

根据上述针对优化模型的求解过程，我们可以得到一个较好的投资方案，其中投资收益率将达到 4.3%，该收益率仅仅低于全部购买 *abc*007 获取的综合收益率 7.5%。而且进一步优化了投资风险价值，最终的投资风险价值为 89.68，该风险值基本在可控范围之内，可控风险值要远远小于只选择 *abc*007 时的风险价值。

5.2 问题二模型的建立与求解

本题要求给出合理的评价指标来评估问题一的模型，主要是通过建立综合评价模型来评价问题一投资组合模型。对于合理的评价指标选取，我们通过选取一些针对优化模型的检验指标来进行综合评估。对于指标的选取，我们主要从三个方面来选取，分别是是模型的合理性，模型求解结果，模型可操作性等三个角度选取不同的评价指标。

5.2.1 问题二解题思路流程分析

对于问题一模型的综合评估，我们可以根据相关指标数据来进行量化评分。但是考虑到很多评价指标无法获取量化数据，因此本文考虑构建一个层次结构模型，根据指标专家打分或者人为设定权重来评价问题一模型优劣。进一步根据利用模糊算子，基于最大隶属度原则，给出评价结果。通过层次结构模型确定目标权重，利用模糊评价法来获取针对问题一模型的定性评价。最终通过建立的模糊层次结构模型来获取最终的评价结果。对于评价目标，考虑到问题一针对投资组合方案建立的数学模型，我们可以抽出多个数学模型和本文建立的多目标优化模型进行评价对比，从而说明本文针对问题一建立投资组合模型的优劣程度。

本题可以看看作是问题一的延伸，我们通过建立评价模型来评估对比不同优化模型实现选股投资。根据上述分析，我们首先根据相关指标以及问题一的模型方案建立一个层次结构评估模型。

5.2.2 层次结构模型的建立与求解

层次结构模型(*AHP* 法) 是一种解决多目标的复杂问题的定性与定量相结合的决策分析方法。该方法将定量分析与定性分析结合起来，用决策者的经验判断各衡量目标能否实现的标准之间的相对重要程度。最终可以得到各种方案的综合评价得分，因此该方

法是有效的将定性定量结合的一种有效方法。

本文给出了层次结构模型建模求解的基本流程如下图：

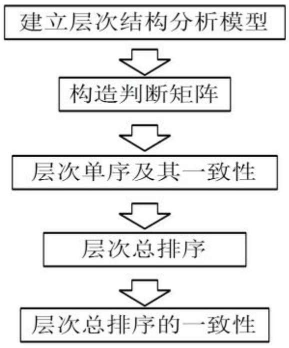


图 17 层次结构模型建模流程

层次结构模型使用的第一步就是构建一个层次结构模型，主要就是确定层次结构模型的目标层，准则层，方案层等。

对于层次结构模型的构建，主要参考前面构建的评价指标体系，从而可以得到层次结构模型中的目标层是模型评估结果。

主要的方案层就是问题一投资选股所考虑的一系列投资组合模型。最终可以得到层次结构模型如下：

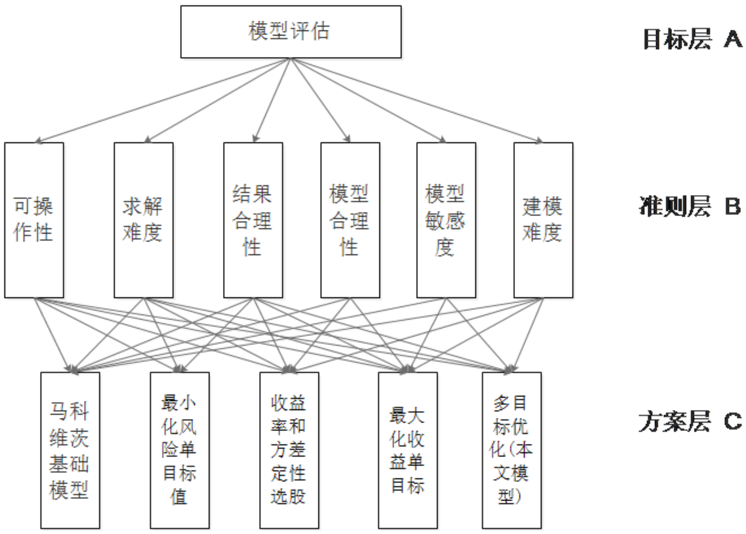


图 18 层次结构图

判断矩阵的构造是根据元素间两两比较得到。一般需要邀请专家进行分析，给出各指标间对于总目标的重要性的比例关系[9]。本文通过参考相关文献，加入对于判断矩阵的确定主要是主观意志的确定来确定判断矩阵的大小。

本文假设判断矩阵为 A，那么假设在判断矩阵中是具有元素 a。表示指标 i 和指标 j 的相对重要度之比，且有：

$$a_{ij} = \frac{1}{a_{ji}}$$

通过查找相关的文献资料可以得到，一般常用的确定指标间相对重要度的方法是 1—9 比率标度法(以下简称比率标度法)，比率标度法的取值标准具体见下表所示：

表 4 比率标度表

比率标度	同样重要	稍微重要	重要	比较重要	极其重要	稍微不重要	不重要	较不重要	极不重要
a_{ij}	1	3	5	7	9	1/3	1/5	1/7	1/9

根据上表我们可以给出各指标间对于总目标的重要性的比例关系。然后就可以得到如下判断矩阵：

$$A=\begin{bmatrix} \frac{w_1}{w_1} & \frac{w_1}{w_2} & \cdots & \frac{w_1}{w_n} \\ \frac{w_2}{w_1} & \frac{w_2}{w_2} & \cdots & \frac{w_2}{w_n} \\ \cdots\cdots & & & \\ \frac{w_n}{w_1} & \frac{w_n}{w_2} & \cdots & \frac{w_n}{w_n} \end{bmatrix}$$

根据以上建立的判断矩阵列出准则层，指标层的判断矩阵。然后进行层次单排序，即本层各因素对上层某因素的重要性排序。可以用判断矩阵的特征向量 W 来表示。将该题解向量 W 经过归一化处理如下：

$$W_i=\frac{W_i}{\sum_{i=1}^{n}W_i}$$

本文首先对 6 个准则层的权重进行赋权计算，根据自己的偏好进行打分最终就可以得到准则层的判断矩阵：

表 5 准则层判断矩阵表

A	*B1*	*B2*	*B3*	*B4*	*B5*	*B6*
B1	1	1	1	4	1	1/2
B2	1	1	2	4	1	1/2
B3	1	1/2	1	5	3	1/2
B4	1/4	1/4	1/5	1	1/3	1/3
B5	1	1	1/3	3	1	1
B6	2	2	2	3	3	1

根据以上方法，首先根据自己的偏好确定判断矩阵，对各指标作成对比较,比较出每对指标之间的相对重要性,构造判断矩阵,然后再确定权值的大小。

按照上述方法将判断矩阵规范化之后，就可得到单一准则下所求指标的权重向量：

$$w_i=\frac{\left(\prod_{j=1}^{n}a_{ij}\right)^{\frac{1}{n}}}{\sum_{i=1}^{n}\left(\prod_{j=1}^{n}a_{ij}\right)^{\frac{1}{n}}}$$

求得的向量 W 即为各指标的权重系数。需要对判断矩阵的一致性进行检验，即要计算随机一致性比率。一致性检验是要检验在准则间两两比较的过程，对重要性的判断标准是否前后一致，这是检验人为因素带来误差的手段。一致性检验比率 CR 作为衡量标准数学原理已经证明，判断矩阵具有一致性的条件是矩阵的最大特征根:

$$\lambda\text{max}$$

与阶数 n 相同。一致性指标如下：

$$CI=\frac{\lambda_{\max}-n}{n-1}$$

给定随机一致性指标 RI ，学者 *Saaty* 给定了一致性指标数据如下表[18]：

根据以下一致性比率计算公式进行求解：

$$CR=\frac{CI}{RI}=\frac{\lambda_{\max}-n}{RI\cdot(n-1)}$$

只有求解得到 $CR<0.1$ 时，该层次排序的结果才认为是满意的，否则需要重新判断矩阵元素的取值。根据以上所建立层次分析法的权重求解模型可得该指标的权重系数。

根据上述分析，可以得到方案层相对于各个准则层指标的层次单排序结果如下表：

表 6 方案层对于准则层的的权重计算结果

	层次单排序权重 (*B1*)	层次单排序权重 (*B2*)	层次单排序权重 (*B3*)	层次单排序权重 (*B4*)	层次单排序权重 (*B5*)	层次单排序权重 (*B6*)
C1	0.1925	0.263	0.2636	0.4390	0.4585	0.2215
C2	0.0812	0.475	0.4758	0.2636	0.2558	0.2734
C3	0.1738	0.055	0.0538	0.0894	0.0809	0.2731
C4	0.2105	0.090	0.0981	0.1465	0.1500	0.0556
C5	0.3360	0.110	0.1087	0.0615	0.0547	0.1764
CR（一致性比率）	0.04359<0.1	0.018<0.1	0.0161<0.1	0.0284<0.1	0.0472<0.1	0.0596<0.1

最终通过不断的改动判断矩阵使最终的一致性比率都通过了一致性检验，并且都得到了层次单排序权重结果，下面我们在进行最后的层次总排序，从而可以得到 5 个具体选股模型的权重评估。

根据前文的层次单排序结果，本文接下来进行层次总排序，从而确定各个职业岗位的具体定量价值评估。

其中层次总排序的具体计算方法如下：

$$W_i=\sum_{j=1}^{6}B_{ij}A_j$$

其中 B 是方案层相对于准则层的权重，A 是准则层相对于目标层的层次单排序结果，这样两种结果相乘就可以得到层次总排序的结果。进一步的对层次总排序的一致性检验进行验证就可以得到基于层次结构模型的模型价值评估结果。

最终对前文的层次单排序权重进行汇总可以得到以下的层次总排序汇总表如下：

表 7 层次总排序

准则		可操作性	求解难度	结果合理性	模型合理性	模型敏感度	建模难度	层次总排序权重
准则层权重		0.1507	0.1792	0.1886	0.0472	0.1464	0.2879	
方案层层次单排序权重	最小化风险单目标	0.1925	0.263	0.2636	0.439	0.4585	0.2215	0.277469
	多目标优化模型	0.0812	0.475	0.4758	0.2636	0.2558	0.2734	0.300754
	收益率和方差定性选股	0.1738	0.055	0.0538	0.0894	0.0809	0.2731	0.126028
	最大化收益单目标	0.2105	0.09	0.0981	0.1465	0.15	0.0556	0.158997
	马科维茨基础模型	0.336	0.11	0.1087	0.0615	0.0547	0.1764	0.165529

根据上表的结果可以得到基于层次结构模型的问题一模型价值评估，为了清晰地分析层次总排序的权重结果，将上述结果用直方图（见附录 1.3）表示如下：

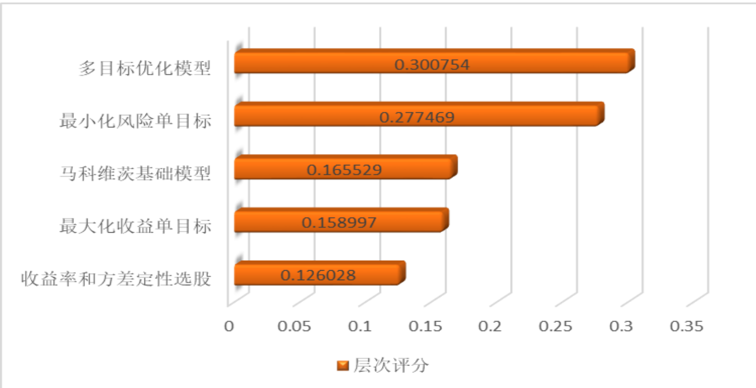


图 19 基于层次结构模型的问题一模型价值评估量化

根据上图我们可以得到基于层次结构模型的问题一模型价值评估。其中，本文建立的多目标优化模型获得了最高的层次评分，模型评分为 0.30075。为了进一步问题一模型的定性评价结果，本文利用模糊评价法来评价模型优劣。

5.2.3 模糊层次评价法定性评价模型

设有 n 个评价等级，m 个一级评价指标，每一个一级指标又包含了多个二级指标，并用 U、V、V_i 等符号来表示，即：

等级论域为：$U=(u_1, u_2,\cdots, u_n)$；因素论域为：$V=(V_1, V_2,\cdots, V_m)$；因子论域为：$V_i=(V_1, V_2,\cdots, V_k)$；

由于 U 与 V 之间存在模糊关系 R，可表示为模糊矩阵形式：

$$R=\begin{bmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1n} \\ r_{21} & r_{22} & r_{23} & \cdots & r_{2n} \\ r & r & r & \cdots & r_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & r_{m2} & \cdots & r_{ij} \end{bmatrix}\begin{array}{l} V_1 \\ V_2 \\ V_3 \\ \\ V_4 \end{array}$$

其中r_{ij}可表示第i个评价因素V_i对第j个等级的隶属度，它依赖于V_i所包含的各个因子对各等级的隶属度及各因子对因素的权重，设V_i所包含的第P个因子对第g个等级的隶属度为$S_{pq}^{i}(p=1,2,\cdots,k;\ q=1,2,\cdots,n)$，第$P$个因子对该因素的权重$W_p^i$，则：

$$(r_{i1},\ r_{i2},\cdots,\ r_{in})=(W_1^i,\ W_1^i,\cdots,\ W_1^i)\begin{bmatrix} S_{11}^i & S_{12}^i & \cdots & S_{1n}^i \\ S_{21}^i & S_{22}^i & \cdots & S_{2n}^i \\ \vdots & \vdots & \ddots & \vdots \\ S_{k2}^i & S_{k2}^i & \cdots & S_{kn}^i \end{bmatrix}$$

这样就确定了模糊关系矩阵。

记一级评价因素的权重为，$A=(A_1,\ A_2,\cdots,\ A_m)$则综合评价结果为：$B=A$。$R=(b_1,\ b_2,\cdots,\ b_n)$，若$b_k=Max(b_1,\ b_2,\cdots,\ b_n)$，则评价对象属于$k$类。

本文主要是利用最大隶属度的原则进行模糊综合评价，主要是通过矩阵运算计算出一个评价数值。根据各个评价等级赋值，通过计算出，最终得到一个数值，用以综合评价目标的整体状况。

根据上述针对模糊评价法的理论分析，利用层次结构评分作为模糊算子，结合最大隶属度原则。最终可以得出问题一五个模型方案的综合评价结果如下：

表 8 问题模型的模糊综合评价结果

问题一模型	模糊定性评价
最小化风险单目标	良
多目标优化模型	优
收益率和方差定性选股	中
最大化收益单目标	差
马科维茨基础模型	良

通过上述建立的模糊-层次结构模型对问题一的五个具体选股方案的评价分析，最终的模型评价结果表明，本文建立的多目标优化模型可以较好的满足风险和收益目标，模型的最终评价结果为优，量化权重评分为 0.30075，好于另外四种选股投资模型。

5.3 问题三模型的建立与求解

本题要求利用附件一的相关数据，给出当前指数的波动情况。并且利用已有数据来预测未来一年的指数波动状况，从而给出一些合理的投资建议和策略。

对于当前指数波动状况的分析，本文首先利用附件十只股票数据来量化一个股票综合指数，对于综合指数的量化，我们主要是利用十只成分股来集权计算出一个成分股指

数。进一步针对该指数的变化趋势进行分析。而对于未来一年指数波动情况的预测，我们综合考虑指数波动的周期性，可以建立时间序列模型进行预测。对于时间序列预测模型，常用的包括自回归移动平均混合模型 *ARIMA*，指数平滑法等。

我们考虑建立基于周期性变化的 *Prophet* 时间序列预测模型来预测未来的指数变化趋势，该模型参数设置简单，更加符合目前的时间序列建模趋势。为了重复说明时间序列预测指数波动的优越性，我们还建立了基于不同核函数的支持向量机回归（*SVR*）来验证 *Prophet* 时间序列预测的有效性。

5.3.1 成分股指数量化分析

对于当前指数波动状况的分析，本文首先利用附件十只股票数据来量化一个股票综合指数，对于综合指数的量化，我们主要是利用十只成分股来集权计算出一个成分股指数。本文首先针对部分股票在过去一年的价格走势进行分析[10]。

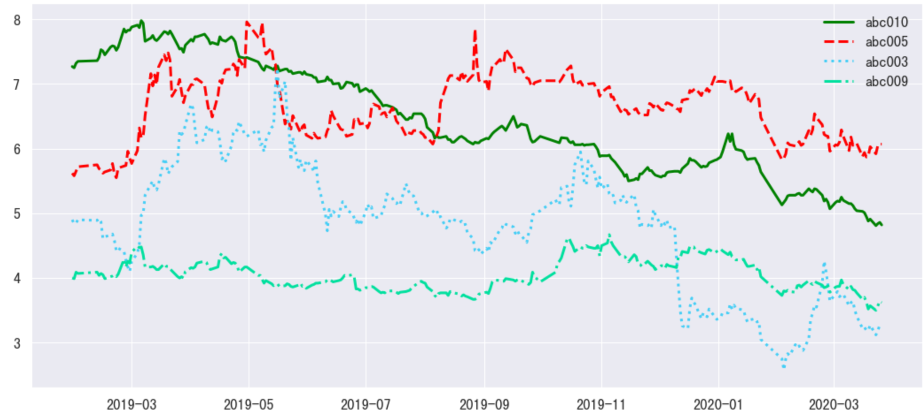


图 20 部分股票过去一年的价格走势

根据上图我们可以发现这些成分股单一的走势都非常没有规律，因此考虑针对成分股综合指数进行量化。对于成分股综合指数的量化，我们主要利用成交量占比和十只股票的股价来进行计算。最终的成分股综合指数计算公式如下：

$$\lambda_{i\times 1} = A_{i\times 10} \times B_{10\times 1}$$

其中 $A_{i\times 10}$ 表示十只股票的价格指数矩阵， $B_{10\times 1}$ 表示 10 只股票的成交量占比矩阵。最终通过矩阵相乘运算可以得到成分股指数随着时间的变化趋势。

利用上述成分股指数计算方法得到附件 10 只股票的成分股综合指数变化趋势如下：

图 21 过去一年的成分股综合指数变化趋势

根据上图我们可以得到附件十只股票综合计算得到的成分股综合指数再过去一年的波动情况。指数波动幅度较大，但是总体上可以看出该指数是一种波动上升的趋势。相信该指数在未来也会有较好的表现。为了更加具体的研究该指数的波动情况，我们利用过去一年的指数波动情况来预测未来一年的指数波动。

5.3.2 基于 prophet 时间序列的指数波动预测

而对于未来一年指数波动情况的预测，我们综合考虑指数波动的周期性，可以建立时间序列模型进行预测。对于时间序列预测模型，常用的包括自回归移动平均混合模型 *ARIMA*，指数平滑法等。

我们考虑建立基于周期性变化的 *Prophet* 时间序列预测模型来预测未来的指数变化趋势，该模型参数设置简单，更加符合目前的时间序列建模趋势。建立时间序列模型之前，我们首先针对成分股指数历史数据进行定性的自相关分析。在统计学中，通常需要衡量两个变量之间是如何相关变化的（协方差函数）以及不同事件之间的相关程度（相关系数）。而在时间序列中，需要研究序列相关性，此时需要度量同一事件在不同时期的相关程度以及相关性变化，类比协方差函数和相关系数的概念，自协方差函数以及自相关系数的定义如下;

假定一个时间序列 $\{X_t, t \in T\}$，对于任意的 $t, s \in T$,序列 $\{X_t\}$ 的自协方差函数 $\gamma(t,s)$ 定义如下：

$$\gamma(t,s) = E(X_t - \mu_t)(X_s - \mu_s)$$

自相关系数 $\rho(t,s)$ 定义如下：

$$\rho(t,s) = \frac{\gamma(t,s)}{\sqrt{DX_t \bullet DX_s}}$$

在时间序列分析中，自协方差函数以及自相关系数是常用的分析方法，根据上述自相关系数的求解方法，我们得到不同时间间隔的自相关性分析可视化如下图：

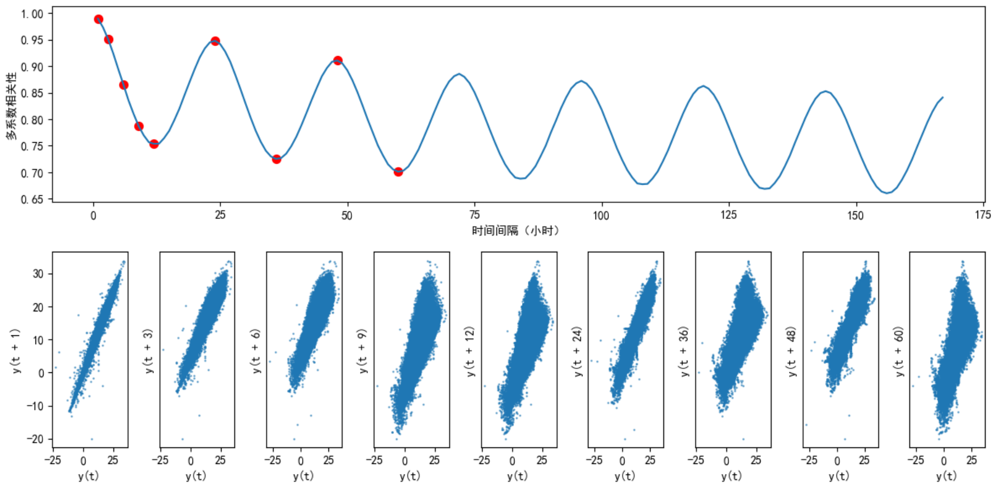


图 22 过去一年指数波动的自相关分析

顶行显示温哥华自相关图。下方的散点图对应于自相关图上标记的红点。最左边的图显示了 t 时刻的指数相对于 $t+1$ 时刻散点图的指数。指数的自相关分析可以很好的解

释短时刻的指数变化情况，该种现象伴随着一定的周期性。为了进一步明确分析它的周期性质，我们通过尺度缩放得到指数信号的自相关系数：

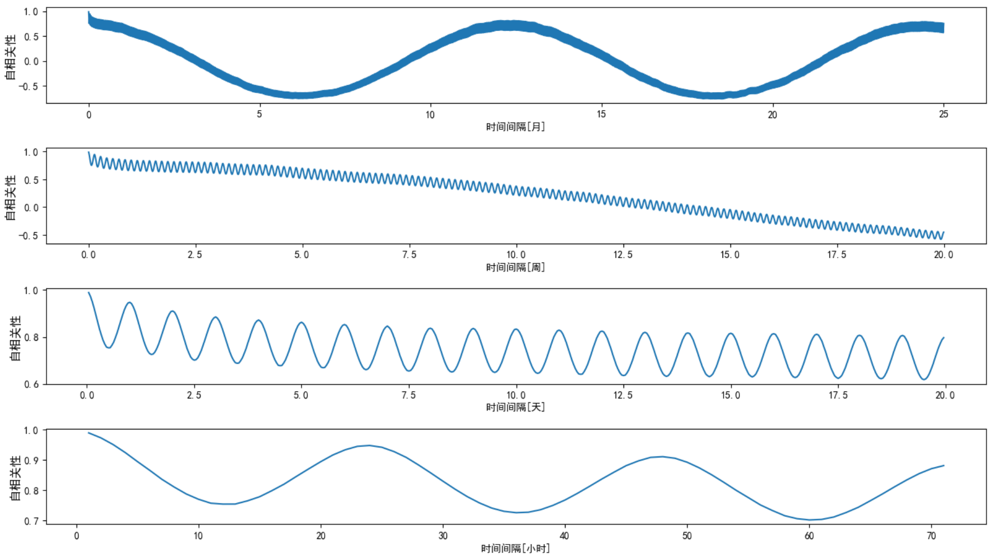


图 23 时间尺度缩放的自相关分析

根据上图我们可以清楚地看到这两个时期：

1、顶部图像的年度期间（12 个月期间）周期性质；

2、两个底部图像的每日时段（24 小时）周期性质；

当我们查看数据时，我们也看到了这两个周期，但这些自动图更平滑，因为它们代表了信号所有时间点的汇总数据。

根据上述分析我们可以很清楚的得到指数波动的周期性，因此本文建立基于周期变化的 *Prophet* 时间序列模型来预测未来一年的指数波动情况。

Prophet 模型的一般结构为：

$$y(t)=g(t)+s(t)+h(t)+\varepsilon$$

$g(t)$：是趋势函数，用来分析时间序列中非周期性的变化；

$s(t)$：代表周期性的变化，例如一周或一年周期性；

$h(t)$：代表节假日等偶然一天或几天造成的影响；

ε：代表随机误差，随机误差不可控；

该模型具有操作灵活，拟合速度快，预测效果好等优点。因此正在被广泛应用于商业分析的领域。

该模型的增长类似于 *Logistic* 回归增长模型，该模型的基本形式如下：

$$g(t)=\frac{C}{1+\exp(-k(t-m))}$$

C：承载能力（饱和值）

k：增长率

m：偏置参数

随着时间 C，k 会发生变化，所以在时间序列中设置一系列转变点 $s_j, j=1,\dots,S$，其

增长率会发生变化，δ_j 是在时间 t_j 处的变化量，从而构建向量 $a(t) \in \{0,1\}^S$

$$a_j(t) = \begin{cases} 1, t \geq s_j \\ 0, otherwise \end{cases}$$

则增长率 k 在时间 t 处的表达式为：

$$k = k + a(t^T)\delta$$

偏置函数 m 根据增长率的变化做出相应改变，以此与时间片段的尾部连接，在转折点 j 处，偏置量适当调整，其公式如下：

$$\gamma_j = (s_j - m - \sum_{l<j} \gamma l)(1 - \frac{k + \sum_{l<j} \delta l}{k + \sum_{l \leq j} \delta l})$$

于是可以得到分段的 *logistic* 趋势模型，公式如下：

$$g(t) = \frac{C(t)}{1 + \exp(-(k + a(t)^T)\delta)(t - (m + a(t)^T)\gamma)}$$

当进行预测的问题可再增长时（即没有饱和），此时分段的连续增长模型效果在某些时候很好，公式如下：

$$g(t) = (k + a(t)^T \delta)t + (m + a(t)^T \gamma)$$

与之前的符号定义类似 k：增长率；δ：增长率的调整值；m 是偏置参数。

当依据从前的方式来对未来的增长率进行预测，但是需要将 τ 换成一个根据旧数据推断出的方差。运用贝叶斯架构，能够对 τ 设置多层先验分布得到后验分布，也可以使用增长率尺度参数的极大似然估计：

$$\lambda = \frac{1}{s} \sum_{j=1}^{S} |\delta_j|$$

为了保证转折点的平均频率符合历史分布，未来的转折点是用以下方式随机选取的：

$$\forall j > T, \begin{cases} \delta_j = 0, w.p., \frac{T-S}{T} \\ \delta_j, Laplace(0,\lambda), w.p., \frac{S}{T} \end{cases}$$

通过这种方式可以设定为历史和未来的变化趋势设置相同的频率和幅度，进一步测量不确定性。我们的研究中依靠傅里叶级数创建灵活性高的周期性模型，设置 P 作为时间序列的规则周期长度（如：时间序列设置为以天为单位，年度数据 P 设为 365.25，周数据 P 设为 7）。得出任意平滑周期效应的估算值为：

$$s(t) = \sum_{n=1}^{N} (a_n \cos(\frac{2\pi nt}{P}) + b_n \sin(\frac{2\pi nt}{P}))$$

为了拟合周期性，需要测评 $2N$ 个参数 $\beta = [a_1, b_1, ..., a_N, b_N]$。通过为历史和未来的每一个 t 值构造了一个季节性的向量矩阵来实现，例如设置每年的周期性 N=10，

$$X(t) = [\cos(\frac{2\pi(1)t}{365.25}), ..., \sin(\frac{2\pi(10)}{365.25})]$$

季节性组分为：

$$s(t)=X(t)\beta$$

在我们生成的模型中，采用 $\beta \sim Normal(0,\sigma^2)$ 对季节性添加了一个先验分布。

进一步给出 *Prophet* 建模预测的一般性流程图：

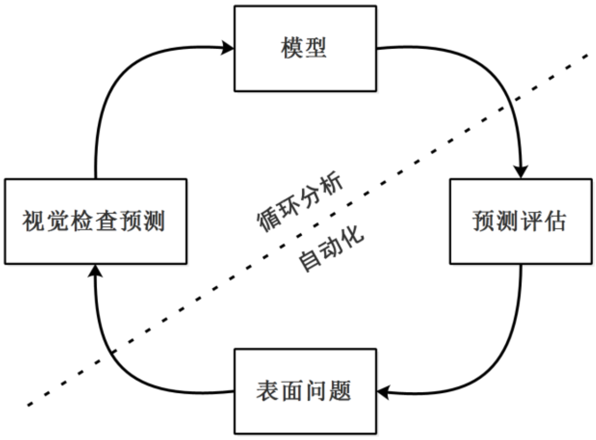


图 24 *Prophet* 建模预测流程

根据上述建模预测流程，我们给出了未来一年的成分股指数波动预测结果如下图：

图 25 成分股指数未来一年的预测结果可视化

具体的成分股指数按照月份的预测结果见附录，根据上图的未来一年成分股趋势我们也可以看出模型预测结果较为准确。未来一年的股价波动趋势符合过去一年的指数波动发展趋势。而且未来一年的股价波动整体来说会相对平稳，整体上也和过去一年一样会呈现一种波动上升的趋势。同时，阶段性的回调和政策，经济大环境影响下的指数走势也是不可避免的。比如从 2020-5 到 2020-9 这五个月时间，指数将会出现阶段性的反弹，这也符合新冠疫情基本控制住之后的经济复苏。因此，如果想回去较高的收益，本文建议应当把握住这一波经济复苏带来的股市反弹。但同时，也要警惕各种黑天鹅事件对于成分股指数的影响。本文接下来主要从疫情之后的经济复苏以及风险控制等方面提出一些投资建议和策略[8]。

5.3.3 投资建议与策略

本文通过建立 *Prophet* 时间序列模型对于未来一年的成分股指数做出了较好的预测。根据未来指数波动和预测结果，并且集合问题一针对最优投资组合方案的建模求解。本文从经济复苏和机会把握和风险控制提出以下建议。

1、抓住疫情之后经济复苏带动的股市反弹，根据预测结果附件所给成分股指数将

会在 2020-5 到 2020-9 这五个月时间内持续股价反弹；

2、抓住机遇的同时还要进行风险控制，切记不可一把梭或者玩杠杆炒股，不可预测的黑天鹅时间将会很大程度上影响指数波动；

3、如果不能很好的进行自我投资风控和选股，建议投资偏向保守型，或者购买一些指数基金，紧跟中国发展红利；

接着提出以下投资策略：

1、*CTA* 策略：波动行情下首选的投资策略

当前市场可能会面临两种情况：第一，疫情完全得到控制，经济强烈反弹，股市大幅上涨；第二，疫情形势更加严峻或疫情的冲击对经济下行产生了较大影响，导致股市下跌。无论哪种情况发生，未来市场都将会面临较多的不确定性因素，因此在这样的高波动行情下，相对于挑选资产，更应该考虑挑选策略，像 *CTA* 这样在高波动中表现较好的策略适合未来的行情。

2、*CTA* 投资实践思路：多频率、多品种、多策略

多频率：不仅要观察日频、周频、月频策略，也要进行分钟级、小时级的策略观察；

多品种：在投资股指期货之外也可以进行不同商品的分散投资；

多策略：中长期、日内策略的占比分配，趋势策略、反转策略的组合等。

3、投资者：转换传统投资思路

当前疫情情况下，投资者不能准确预测股市行情，建议不要轻易抄底。此外，投资者还需转换固有的投资思路，既不能因为股市暴涨暴跌而进行激动投资，也不能因为股市波动大而不敢投资，要充分利用各种衍生品工具，找到在各种行情下适合的投资策略。

针对投资策略分析，本文参考问题一多目标优化确定投资组合方案。给出了以下投资策略，对于投资而言，收益和风险一样重要。但是对于不同的人而言，需要确定出一个主目标。比如在一定风险水平下，最大化收益，这就要求投资者能够承担一定的本金损失，因此针对具体的策略，需要针对投资者的具体情况来实行。针对本文而言，可以通过建立多目标优化模型，同时选出一个风险可控，收益较好的投资方案。

六、模型检验

6.1 针对问题三 Prophet 模型预测结果分析

首先分析模型的拟合效果误差，得到两种模型在测试集上的预测误差对比表如下：

表 9 时间序列与随机森林预测对比表

建模预测方法	模型拟合优度	测试评估得分
Prophet 模型	0.974506813	0.8863644452
加权随机森林回归	0.906365663	0.7364795256

根据上表可以得到加权随机森林回归模型在已有数据集的测试上明显的比时间序列模型效果要差。这主要是因为时间序列建模虽然只能考虑单一时间维度要素，并且从中发现规律，但是因为考虑指数变化的周期性因此效果比较好。而且时间序列虽然能够发现指数变化的一般趋势，但是对于短时间预测的效果较差，进行长期趋势预测效果很好。

本文建立的 *Prophet* 模型不仅可以考虑周期变化，而且还可以定量的计算这些指标对于预测结果的重要程度。可以说该模型是预测未来指数的一种较好方法。

七、模型的优缺点与改进

7.1 模型的优缺点

7.1.1 模型的优点

1、马科维茨模型利用数理方法给出了组合选择最基础完备的框架，具有开创性；

2、本文建立多目标优化模型更好的考虑收益和风险，最终选出一个最优的投资方案；

3、评价模型对问题一的五种投资方案都进行了很好的分析，评价结果较为可信；

4、本文综合考虑成分股指数的波动周期性，建立 *Prophet* 时间序列模型对于指数的波动趋势给出了较好的预测结果和分析。

7.1.2 模型的缺点

1、没有时间尝试更过的启发式算法对多目标优化模型进行寻优；

2、问题三针对成分股指数趋势预测，没有考虑一些特殊因素影响，只是单纯从时间角度出发；

八、参考文献

[1]陆涛.我国证券市场国际化分析[J].合作经济与科技,2019(06):58-59.

[2]张峰瑞.论股指期货对我国证券市场的影响及建议[J].中国市场,2019(05):54-55.

[3]张晓宇.我国证券市场现状分析及监管的研究[J].中国市场,2018(17):40-42.

[4]吴宁宁. 中国证券市场投资者羊群行为及其市场影响研究[D].合肥工业大学,2018.

[5]孙君威.我国证券市场股价波动同步性研究[J].时代金融,2018(06):148+152.

[6]佘镜怀,生蕾.我国证券市场波动风险预警模型研究[J].金融理论与实践,2017(12):37-42.

[7] 张琳,张军,王擎.宏观经济信息发布对股票市场收益率及其波动的影响[J].系统工程理论与实践,2020,40(06):1439-1451.

[8]闵诗筠.经济政策不确定性对我国股市波动的影响——基于 2007—2019 年上证指数的实证分析[J].江苏商论,2020(03):88-92.

[9]韩晨宇,王一鸣.中国股票市场波动率的多重分形分析与实证[J].统计与决策,2020,36(01):136-140.

[10]丁硕.宏观经济因素对我国股价波动的影响分析[J].环渤海经济瞭望,2019(11):44-45.

九、附录

9.1 问题一数据预处理与可视化分析代码

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import matplotlib as mpl
import os
# sns.set(style="darkgrid") #这是 seaborn 默认的风格
mpl.style.use('seaborn')

plt.rcParams['font.sans-serif'] = ['SimHei']   # 中文字体设置-黑体
plt.rcParams['axes.unicode_minus'] = False   # 解决保存图像是负号'-'显示为方块的问题
sns.set(font='SimHei',font_scale=1.5)   # 解决 Seaborn 中文显示问题并调整字体大小

df = pd.read_excel('data/附件：十支股票参数.xlsx',encoding='utf_8',sheet_name='abc001')
df.head()
print(df.columns)

fig = plt.figure(figsize=(11, 6))
ax = fig.add_subplot(111)
# sns.heatmap(data.corr(), ax=ax, annot=True, linewidths=0.05, fmt= '.2f',cmap="magma")
print(df.iloc[:,1::].corr())
sns.set_context({"figure.figsize":(8,8)})
sns.heatmap(df.iloc[:,1::].corr(),annot=True,cmap="CMRmap_r")

df['date'] = pd.to_datetime(df['时间'])
df.set_index("date", inplace=True)
plt.figure(figsize=(18, 8))
plt.plot(df['开盘'], label='开盘', color='green', linestyle='-',linewidth = '5')
plt.plot(df['最高'], label='最高', color='red', linestyle='--',linewidth = '5')
plt.plot(df['最低'], label='最低', color='#44cef6', linestyle=':',linewidth = '5')
plt.plot(df['收盘'], label='收盘', color='#00e09e', linestyle='-.',linewidth = '5')
# plt.plot(df['时间'], df['现存疑似'], label='现存疑似', color='#00e09e', linestyle='-.',linewidth = '5')
#添加图例
plt.legend()
#添加网格
plt.grid(True)
```

```
print(type(df['时间']))
plt.ylabel("abc001 价格（元/每股）")
df.head()

df['date'] = pd.to_datetime(df['时间'])
df.set_index("date", inplace=True)
plt.figure(figsize=(18, 8))
plt.plot(df['收盘'], label='收盘', color='#00e09e', linestyle='-',linewidth = '5')
plt.ylabel("abc001 收盘价格（元/每股）")
#添加图例
plt.legend()

df['date'] = pd.to_datetime(df['时间'])
df.set_index("date", inplace=True)
plt.figure(figsize=(18, 8))
plt.plot(df['成交量'], label='成交量', color='#44cef6', linestyle='--',linewidth = '5')
plt.ylabel("abc001 成交量")
#添加图例
plt.legend()

#实验样本索引查看
df1 =df.iloc[:,1:-1]
#实验数据导出
# df1.to_csv('C:/Users/13109/desktop/df1.csv',encoding="gb2312")
#绘制原始的箱线图，并返回异常值字典
p = df1.plot.box(sym='r+',return_type = 'dict')

df0 = pd.read_excel('data/附件：十支股票参数.xlsx',encoding='utf_8',sheet_name='汇总')
df0.head()
print(df0.columns)
df0

df_shoupan = df0.loc[:,['时间','abc001 收盘','abc002 收盘','abc003 收盘','abc004 收盘','abc005 收盘','abc006 收盘','abc007 收盘','abc008 收盘','abc009 收盘','abc010 收盘']]
df_shoupan['date'] = pd.to_datetime(df_shoupan['时间'])
df_shoupan.set_index("date", inplace=True)
df_shoupan.head()

df_shoupan.plot(figsize=(18, 8),linewidth = '3')
```

```
plt.ylabel('十只股票的收盘价')
plt.xlabel('')
plt.show()

df_cheng.plot(figsize=(18, 8),linewidth = '3')
plt.ylabel('十只股票的成交量走势')
plt.xlabel('')
plt.show()


fig, ax = plt.subplots(1, 2, figsize=(18,4))

time_val = df['收盘'].values

sns.distplot(time_val, ax=ax[0], color='r')
ax[0].set_title('收盘价分布趋势', fontsize=14)

ax[1].set_xlim([min(time_val), max(time_val)])
sns.distplot(np.log(time_val), ax=ax[1], color='b')
ax[1].set_title('收盘价对数分布趋势', fontsize=14)
ax[1].set_xlim([min(np.log(time_val)), max(np.log(time_val))])
plt.show()

#导入依赖库
import pandas as pd
from scipy import stats

#绘制并打印 QQ 图
fig, ax = plt.subplots(1, 2, figsize=(18,4))
stats.probplot(df['收盘'], dist="norm",plot=ax[0])
ax[0].set_title('abc001 收盘价正态分布检验 qq 图', fontsize=14)

stats.probplot(df['成交量'], dist="norm",plot=ax[1])
ax[1].set_title('abc001 成交量价正态分布检验 qq 图', fontsize=14)

plt.show()

df_avg = df_shoupan.resample('m')['abc001 收盘','abc002 收盘','abc003 收盘','abc004 收盘','abc005 收盘','abc006 收盘','abc007 收盘','abc008 收盘','abc009 收盘','abc010 收盘'].mean()
# df_avg[['abc001 收盘','abc002 收盘']].apply(lambda x: (x - np.min(x)) / (np.max(x) -
```

```
np.min(x))) ##对每一列数据标准化

def calculate_rate(df_avg):
    rate_data = pd.DataFrame()
    for j in range(10):
        rate_list = []
        for i in range(len(df_avg)-1):
            rate = (df_avg.iloc[i+1,j]-df_avg.iloc[i,j])/df_avg.iloc[i,j]
            rate_list.append(rate)
        rate_data['abc00'+str(j+1)] = rate_list
    # rate_data['date'] = pd.date_range('2019-2', '2020-4',freq='M')
    rate_data.set_index(pd.date_range('2019-2', '2020-4',freq='M'), inplace=True)
    #            print(rate)
    return rate_data


rate_data = calculate_rate(df_avg)
rate_data

rate_mean = rate_data.mean()
# rate_data.to_csv('data/question1/十只股票月平均收益率.csv',encoding='utf_8_sig')
# rate_mean.to_csv('data/question1/十只股票的综合平均收益率.csv',encoding='utf_8_sig')
# rate_df = pd.DataFrame(rate_mean)
rate_mean
df_cumulate = pd.DataFrame()

df_cumulate['综合平均收益率']=rate_mean.sort_values(ascending=False)
df_cumulate
f,ax=plt.subplots(figsize=(8,8))
sns.barplot(y=df_cumulate.index,x =df_cumulate['综合平均收益率'],data=df_cumulate,palette='cubehelix',
            ax=ax,ci=85,errcolor='yellow', errwidth=5, capsize=0.1,alpha=0.7)

dd = df_shoupan[['abc001 收盘','abc002 收盘','abc003 收盘','abc004 收盘','abc005 收盘','abc006 收盘','abc007 收盘','abc008 收盘','abc009 收盘','abc010 收盘']].var()
# df_avg[['abc001 收盘','abc002 收盘']].apply(lambda x: (x - np.min(x)) / (np.max(x) - np.min(x))) ##对每一列数据标准化
df_var = pd.DataFrame()
df_var['十只股票的波动方差']=dd.sort_values(ascending=False)
# df_var.loc['abc007 收盘','十只股票的波动方差']=506.4598
```

```
df_var.to_csv('data/question1/十只股票的波动方差.csv',encoding='utf_8_sig')
df_var
f,ax=plt.subplots(figsize=(8,8))
sns.barplot(y=df_var.index,x =df_var['十只股票的波动方差'],data=df_var,palette='cubehelix',
            ax=ax,ci=85,errcolor='yellow', errwidth=5, capsize=0.1,alpha=0.7)

df_cm = pd.DataFrame()
df_cm['十只股票的平均成交量']=df_cheng[['abc001 成交量','abc002 成交量','abc003 成交量','abc004 成交量','abc005 成交量','abc006 成交量','abc007 成交量','abc008 成交量','abc009 成交量','abc010 成交量']].mean()
# df_var.loc['abc007 收盘','十只股票的波动方差']=506.4598
df_cm.to_csv('data/question2/十只股票的平均成交量.csv',encoding='utf_8_sig')
f,ax=plt.subplots(figsize=(8,8))
sns.barplot(y=df_cm.index,x =df_cm['十只股票的平均成交量'],data=df_cm,palette='cubehelix',
            ax=ax,ci=85,errcolor='yellow', errwidth=5, capsize=0.1,alpha=0.7)
df_cm

df_cmrate = df_cm.apply(lambda x:x/sum(x))
df_cmrate.to_csv('data/question2/十只股票的平均成交量占比.csv',encoding='utf_8_sig')
df_cmrate
np.dot(df_shoupan.iloc[:,1:].values,df_cmrate.values)
df_shoupan['成分股指数'] = np.dot(df_shoupan.iloc[:,1:11].values,df_cmrate.values)
df_shoupan.to_csv('data/question3/成分股综合指数计算结果.csv',encoding='utf_8_sig')
df_shoupan.head()
```

9.2 问题一 lingo 单目标优化代码（马科维茨均值方差模型）

```
model:
sets:
month/1..12/;
STOCKS/1,2..55/:X;
link(month,STOCKS):R;
stst(Stocks,stocks):cov;
endsets
data:
TARGET=0.66;
R= ;
enddata
[OBJ]MIN=@sum(stst(i,j): cov(i,j) *x(i) * x(j));
```

```
[ONE]@SUM(STOCKS: X)= 1;
[TWO]@SUM(stocks: mean* x)>=TARGET;
end
```

9.3 遗传算法求解多目标优化

```
close all
clear all
clc
x=[0 250 500 700];
y=[96.5465 148.44 129.214 115.6626];
cs=spline(x,[0 y 0]);
xx=linspace(0,700,100);
plot(x,y,'o',xx,ppval(cs,xx),'-');
Program 2 :Genetic algorithm main program
clc
clear all
close all
ts=AA(:,1);
Hs=sqrt(AA(:,2).^2+AA(:,3).^2);
xxx=AA(:,2);
yyy=AA(:,3);
[PM PN]=size(AA);
for times=1:1
    % for m=1:PN
    %       A(m,PN)=0;
    % end
    n=200;
ger=200;
pc=0.9;
pm=0.01;
    % [Ax,Ay]=size(A)
    % 生成初始种群
    v=init_population(n,2);
    [N,L]=size(v);
disp(sprintf('Number of generations:%d',ger));
disp(sprintf('Population size:%d',N));
disp(sprintf('Crossover probability:%.3f',pc));
disp(sprintf('Mutation probability:%.3f',pm));
    % 计算适应度，并画出图形
    [fit,xx,yy,QS]=Sun_Shadow_fun(v,ts,AA(:,2),AA(:,3));
    % 初始化
```

```
vmfit=[];
it=1;
vx=[];
        %C=[];
        % 开始进化
while it<=ger
                %Reproduction(Bi-classist Selection)
vtemp=Sun_Shadow_roulette(v,fit);
                %Crossover
                v=Sun_Shadow_crossover(vtemp,pc);
                v=Sun_Shadow_mutation(vtemp,pm);
                %Results
                [fit,xx,yy,QS]=Sun_Shadow_fun(v,ts,AA(:,2),AA(:,3));
                [sol,indb]=min(fit);
v(1,:)=v(indb,:);
media=mean(fit);
vx=[vx sol];
vmfit=[vmfit media];
it=it+1;
end
        %%%% 最后结果
disp(sprintf('\n'));        %空一行
        % 显示最优解及最优值
v(indb,:
        % 图形显示最优及平均函数值变化趋势
figure(1);
plot(vx);
        %title('最优,平均函数值变化趋势');
xlabel('Generations');
ylabel('f(x)');
hold on;
        %plot(vmfit,'r');
hold off;
fa(times)=180*v(1,1)-90;
H(times)=2*v(1,2)+1;
fit(1)/20;
Fit(times)=fit(1)/sum(Hs)/2;
end
runtime=toc
res=[fa
H
```

```
Fit];
[b,bint,r,rint,stats]=regress(Hs,QS(:,1))
rcoplot(r,rint)
```
Program 3 :Polynomial fitting solution code
```
clear;
x=[0 250 500 700];
y=[96.5465 148.44 129.214 115.6261];
p=polyfit(x,y,3);
px=poly2str(p,'x');
pv=polyval(p,x);
plot(x,y,'*',x,pv,'k-')
xlabel('x')
ylabel('y')
p
px
x_aver=sum (x)/length(x);
y_aver=sum (y)/length(y);
fori=1:length(x)
        Lxx_1(i)=(x(i)-x_aver)^2;
        Lyy_1(i)=(y(i)-y_aver)^2;
        Lxy_1(i)=x(i)*y(i)-x_aver*y_aver;
end
Lxx=sum(Lxx_1);
Lyy=sum(Lyy_1);
Lxy=sum(Lxy_1);
b_guji=Lxy/Lxx;
a_guji=y_aver-b_guji*x_aver;
y_guji_x_200=a_guji+b_guji*200;
sigamma_guji=1/(length(x)-2)*(Lyy-b_guji^2*Lxx);
rhoxy_guji=Lxy/(sqrt(Lxx)*sqrt(Lyy));
disp(['R=',num2str(rhoxy_guji)])
```
Program 4 :Residual test code
```
close all
clear all
clc
x=[143 145 146 147 149 150 153 154 155 156 157 158 159 160 162 164]';
X=[ones(16,1) x];
Y=[88 87 88 91 92 93 93 95 96 98 97 96 98 99 100 102]';
[b,bint,r,rint,stats]=regress(Y,X);
rcoplot(r,rint);
```

9.4 问题二模糊层次结构模型的 matlab 代码

```
%/准则层到目标层的层次单排序
clc;
clear;
A=[1 1 1 4 1 1/2;
    1 1 2 4 1 1/2;
    1 1/2 1 5 3 1/2;
    1/4 1/4 1/5 1 1/3 1/3;
    1 1 1/3 3 1 1;
  2 2 2 3 3 1];%因素对比矩阵 A，只需要改变矩阵 A
[m,n]=size(A);%获取指标个数
RI=[0 0 0.58 0.90 1.12 1.24 1.32 1.41 1.45 1.49 1.51];
R=rank(A);%求判断矩阵的秩
[V,D]=eig(A);%求判断矩阵的特征值和特征向量，V 特征值，D 特征向量；
tz=max(D);
B=max(tz);%最大特征值
[row, col]=find(D==B);%最大特征值所在位置
C=V(:,col);%对应特征向量
CI=(B-n)/(n-1);%计算一致性检验指标 CI
CR=CI/RI(1,n);
if CR<0.10
    disp('CI=');disp(CI);
    disp('CR=');disp(CR);
    disp('对比矩阵 A 通过一致性检验，各向量权重向量 Q 为：');
    Q=zeros(n,1);
for i=1:n
        Q(i,1)=C(i,1)/sum(C(:,1));%特征向量标准化
end
    Q%输出权重向量
else
    disp('对比矩阵 A 未通过一致性检验，需对对比矩阵 A 重新构造');
end
```

9.5 问题三时间序列预测指数波动的 python 程序

```
import statsmodels.api as sm
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import matplotlib as mpl
import os
```

```
# sns.set(style="darkgrid") #这是 seaborn 默认的风格
mpl.style.use('seaborn')

plt.rcParams['font.sans-serif'] = ['SimHei']   # 中文字体设置-黑体
plt.rcParams['axes.unicode_minus'] = False   # 解决保存图像是负号'-'显示为方块的问题
sns.set(font='SimHei',font_scale=1.5)   # 解决 Seaborn 中文显示问题并调整字体大小

def predict_plot(df,data,label):
    plt.figure(figsize=(18, 8))
    plt.plot(df.loc[:,label], label='实际'+label,linestyle='-',linewidth = '5')
    plt.axvline(x=df.index[-1], alpha=0.5, c="r", ls="--", lw=3)
    # plt.plot(test['count'], label='Test')
    plt.plot(data['Prophet'+label], label='预测'+label,linestyle='--',linewidth = '5')
    plt.legend(loc='best')
    plt.show()

fit1 = sm.tsa.statespace.SARIMAX(df_3.loc[:,'成分股指数'], order=(2, 1, 3), seasonal_order=(1, 1, 1, 7)).fit()
date = pd.date_range('2020-3', '2021-3',freq='M')
data = pd.DataFrame(index=date)
data['Prophet'+'成分股指数'] = fit1.predict(start="2020-3", end="2021-3", dynamic=True)
predict_plot(df,data,'成分股指数')
data

import statsmodels.api as sm
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import matplotlib as mpl
import os
```

```
# sns.set(style="darkgrid") #这是 seaborn 默认的风格
mpl.style.use('seaborn')

plt.rcParams['font.sans-serif'] = ['SimHei']   # 中文字体设置-黑体
plt.rcParams['axes.unicode_minus'] = False   # 解决保存图像是负号'-'显示为方块的问题
sns.set(font='SimHei',font_scale=1.5)   # 解决 Seaborn 中文显示问题并调整字体大小

def predict_plot(df,data,label):
    plt.figure(figsize=(18, 8))
```

```
    plt.plot(df.loc[:,label], label='实际'+label,linestyle='-',linewidth = '5')
    plt.axvline(x=df.index[-1], alpha=0.5, c="r", ls="--", lw=3)
    # plt.plot(test['count'], label='Test')
    plt.plot(data['Prophet'+label], label='预测'+label,linestyle='--',linewidth = '5')
    plt.legend(loc='best')
    plt.show()

df = pd.DataFrame()
df['成分股指数'] = df_shoupan['成分股指数']
df['date'] =pd.date_range('2019-4', '2020-1-5',freq='d')
df.set_index("date", inplace=True)

fit1 = sm.tsa.statespace.SARIMAX(df.loc[:,'成分股指数'], order=(4, 1, 4), seasonal_order=(1, 1, 1, 36)).fit()
date = pd.date_range('2020-1-5', '2021-1-5',freq='d')
data = pd.DataFrame(index=date)
data['Prophet'+'成分股指数'] = fit1.predict(start="2020-1-5", end="2021-1-5", dynamic=True)
predict_plot(df,data,'成分股指数')
data
```

时间	*Prophet* 预测成分股指数
2020/3/31	16.66742
2020/4/30	18.26752
2020/5/31	18.65654
2020/6/30	18.27996
2020/7/31	18.05164
2020/8/31	19.96202
2020/9/30	16.92872
2020/10/31	16.4768
2020/11/30	18.89673
2020/12/31	18.8735
2021/1/31	17.78728
2021/2/28	17.68791