基于组合规划与时间序列的投资选股问题求解

摘 要

随着国民经济的增长，有越来越多的人参与到投资选股中。但股市的投资需要遵循一定原则，在保证仓位要求的情况下首先需要选择收益高风险低的股票，然后选择购买仓位，是一个组合的优化问题。对于这一类问题，通常采用多种模型组合的方式进行求解。

进行优化之前需要根据当步股价预测未来股价，故我们首先根据历史收盘价数据应用 **ARIMA 模型**进行建模并预测。第一次规划是 100 支股票的选择，将其抽象为 **0-1 规划**求解，得到每次调仓时所选择的目标股票。第二次优化是收益最高的优化，但目标有中证指数的偏差限制和仓位限制，与前面的股票选择一同构成了**组合优化**。对于这一规划问题，除了使用传统的**动态规划**求解外，还使用了**遗传算法**进行优化求解，得到结果差异不大证明算法的可行性。最终我们还发现，按照月为周期进行调仓所得到的平均收益率是最高的。

模型形式简洁，不同模型在不同部分承担不同作用，协同配合，达成求解目标，展示出较为良好的效果。

关键词：ARIMA 模型，组合优化，0-1 规划模型，遗传算法，动态规划

目录

一、 问题背景

1.1 问题背景

目前，我国金融行业高度发达，许多公民与机构积极参与到股市投资中，为了反映沪深市场高股息股票的整体走势，中证指数有限公司发布了中证红利指数，从而为投资者提供新的业绩基准和投资标的。

中证红利指数是从沪深市场中选取现金股息率高、分红稳定、具有一定规模及流动性的 100 只上市公司证券作为指数样本，以反映沪深市场高红利上市公司证券的整体表现。研究相关问题，对于把控股市的整体走势有不容忽视的作用。

1.2 问题提出

我们需要从 100 只样本股票中选择不超过 10 只股票构建出一个资产组合配置，调仓时可重新选取股票，可以空仓，使该资产组合每月收益率不能偏离中证红利指数当月收益率上下 2%，并实现年度超额收益率最高。求其资产配置策略和求解算法，并对结果进行阐述和分析。

二、 问题分析

问题需要我们根据附件数据来对这些股票进行量化分析后选出最重要的 10 支，所以这里是一个组合投资问题。这一问题不仅需要得到选择哪 10 支，还需要确定这 10 支股票分别购买的比例，是一个连续和离散并存的组合优化问题。而比例的确定也与股票价格有关，但我们无法直接获取未来股票的价格，所以需要对股票价格进行预测才能对股市走向进行预判，所以还需要进行时间序列的预测。

附件表内由于很多数据是缺失的，所以需要对缺失值进行填充。这里我们使用了线性插值法进行填充以后再进行处理。模型应该分为序列预测和组合优化两个部分。

三、 模型假设

除了题目给出的模型假设以外我们还考虑如下假设

1、假设每两次调仓之间的时间间隔刚好是一周，这样有利于问题简化。

2、假设每次投资时都把上一次投资获取的收益计入下一次投资活动中，这会使得我们的问题得到简化。

3、可以使用时间序列的预测结果进行合理估计。

4、假设这里不考虑股票波动以及各种金融突发事件带来的股市风险，这将对我们的模型产生干扰。

四、 符号说明

符号	说明
y_t	一个时间序列
ε_t	白噪声序列
μ	在问题一表示均值，时间序列中表示常数
σ	标准差
λ	常数系数
α, β	时间序列项的系数
y	股票收盘价
r_f	无风险利率，2.4%
f	每次调仓带来的收益率
T	周期
w	仓位
r_i	收益率
D	选股策略的 0-1 矩阵

（注：若有变量未列入表内或与表内释义不一致请以原文解释为主）

五、 模型的建立与求解

5.1 序列预测模型的建立与求解

5.1.1 模型的建立

（1）利用 ARIMA 模型进行时间序列的分析与预测

为制定接下来的投资组合计划，需要对未来 100 支股票的变化趋势进行预测。

时间序列的预测方法有很多，例如广义线性模型(GLM)、自回归移动平均模型(ARMA)等[1]，近年来还诞生了基于机器学习算法（如支持向量机[2]）的时间序列预测方法和基于深度学习的循环神经网络时间序列模型(如 LSTM[3]、GRU[4]等)。由于数据量只有 240 周，数量并不大，使用神经网络和机器学习容易可能陷入参数训练不充分或者过拟合等问题，故这里我们选用差分自回归移动平均模型(ARIMA)预测。

ARIMA 模型是统计模型中最常见的一种用来进行时间序列预测的模型，只需要考虑内生变量而无需考虑其他外生变量，但要求序列是平稳序列或者差分后是平稳序列[5]。对 402 家企业的供货量和订货量均进行 Dickey-Fuller 平稳性检验[6]，结果显示序列均为平稳序列通过检验，可以使用 ARIMA 模型。

ARIMA 模型包含 3 个部分，即自回归（AR）、差分（I）和移动平均（MA）三个部分。对其每一个部分，都有其递推公式定义。

①自回归模型：

对于 p 阶自回归模型（AR），其递推公式形如：

$$y_t = \mu + \sum_{i=1}^{p} \alpha_i y_{t-i} + \varepsilon_t \tag{1}$$

②移动平均模型：

对于 q 阶移动平均模型（MA），其递推公式形如：

$$u_t = \varepsilon_t + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i} + \nu \tag{2}$$

③差分模型：

$$\begin{gathered} \nabla^{(1)} y_t = y_t - y_{t-1} \\ \nabla^{(d)} y_t = \nabla^{(d-1)} y_t - \nabla^{(d-1)} y_{t-1} \end{gathered} \tag{3}$$

即由自回归模型阶数 p、差分阶数 d 和移动平均阶数 q 就可以确定 ARIMA 模型的基本形式，我们将其简记为 ARIMA(p,d,q)

$$y_t = \sum_{i=1}^{p} \alpha_i \nabla^{(d)} y_{t-i} + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i} + \varepsilon_t \tag{4}$$

对于模型最优参数选择，我们使用 AIC 准则法分析。AIC 准则由 H. Akaike 提出，主要用于时间序列模型的定阶，而 AIC 统计量的定义如下所示[7]，其中 L 表示模型的最大似然函数：

$$AIC = -2\ln L + 2(p+q+d) \tag{5}$$

当两个模型之间存在较大差异时，差异主要体现在似然函数项，当似然函数差异不显著时，模型复杂度则起主要作用，从而参数个数少的模型是较好的选择。一般而言，当模型复杂度提高时，似然函数 L 也会增大，从而使 AIC 变小，但是参数过多时，根据奥卡姆剃刀原则，模型过于复杂从而 AIC 增大容易造成过拟合现象。AIC 不仅要提高似然项，而且引入了惩罚项，使模型参数尽可能少，有助于降低过拟合的可能性。

5.1.2 模型的求解

我们将几支股票的股价和收益率随时间变化的曲线图绘制在图 1 中：

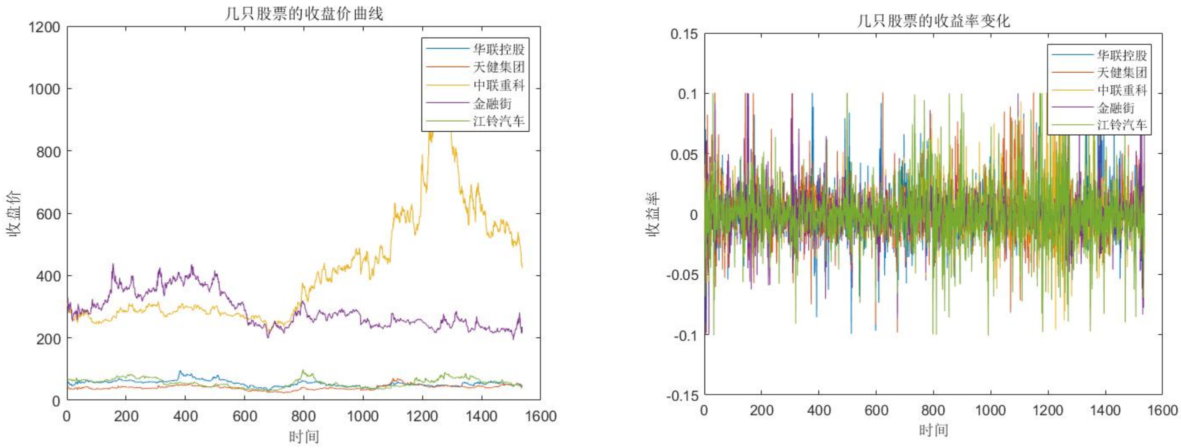

图 1. 几支股票的走向和收益率

可以看到几只股票大概走向并没有呈现一直上升的趋势而是有涨有跌，其中，中联重科的平均价格比其他几支都更高。我们也绘制了几支股票的收益率变化曲线，发现这是一个典型的平稳时间序列，通过了 ACF-PACF 检验。

在经过对数化处理和 ADF 单位根检验检验后我们认为我们的数据适合使用 ARIMA 模型进行处理。而最佳 ARIMA 模型的参数配置我们则采取网格搜索法，使用 AIC 准则作为评判标准。

对这 100 支股票数据分别利用 AIC 准则选择合适的参数组合进行预测可以得到其预测结果如下。图 2 为 ARIMA 模型的拟合效果与残差 Q-Q 图

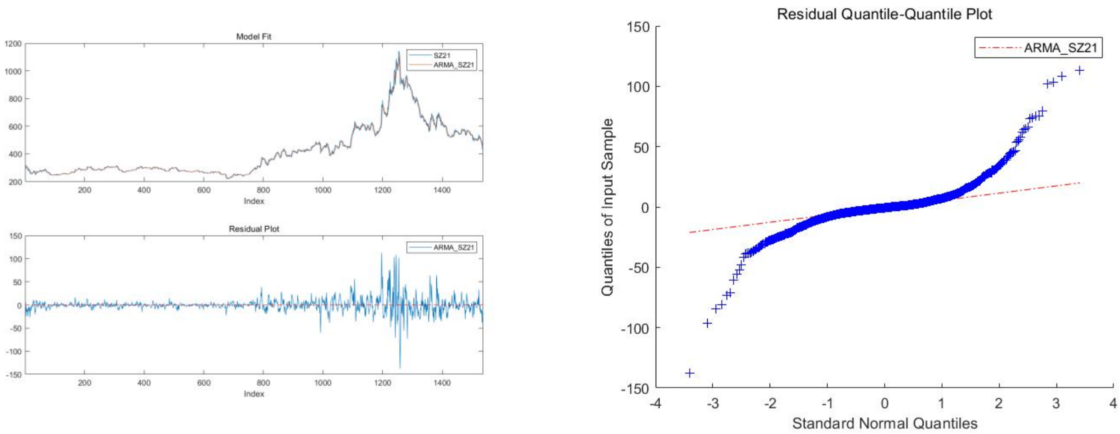

图 2　收盘价预测

可以看到，ARIMA 模型能够对模型进行很好的拟合。但在高峰区域拟合残差分布较大。这也与股市变化的不确定性有关。

5.2 投资组合模型的建立与求解

5.2.1 模型的建立

这一问题本质上是一个多目标规划的问题。决策变量是 10 支股票的投资金额和 100 支股票是否选中，约束条件一方面是月收益率与中证指数的关系，另一方面是决策变量要在我们预测值附近（由于历史数据通过了 K-S 检验具备正态性，我们认为数据应该满足正态性原则），还有就是每只股票的仓位不可以超过 20%。决策目标实际上就是年度超额收益率。

设每周用于投资的投资矩阵 D 是一个 0-1 矩阵，这一矩阵维度为(100,10)，每一行表示这 100 支股票中选中哪一只，所以每一行和为 1；由于一支股票不可能选两次所以每一列和也为 1，整个 0-1 矩阵求和为 1。预测对应股票序列一周以后的股价计算出周利率，也就是每一次调仓后的收益率向量。最后，仓位向量 w 需要满足每一维都不超过 20%。而对于约束条件，月利率与当月中证指数偏差不超过 2%。构造模型如式所示：

$$\max\ f_t = w^T D\left(\frac{\hat{Y}_{t+T} - Y_t}{Y_t}\right)$$

$$s.t.\ \begin{cases} 0 \leqslant w_i \leqslant 20\% \\ |r_t - r_{0t}| \leqslant 2\% \\ \sum\limits_i D_{ij} = 1 \\ \sum\limits_j D_{ij} = 1 \end{cases} \tag{6}$$

这里若选择以一周为周期进行调仓，那么月收益率可以表示为：

$$r_t = (1+f_t)(1+f_{t+T})(1+f_{t+2T})(1+f_{t+3T}) - 1 \tag{7}$$

因为月收益率可以视作四周周收益率之乘积。而我们为了保证年收益率最大实际上是基于一种贪心的策略，因为年收益率与周收益率有这样的一种关系：

$$R = \prod_{t=1}^{N}(1+f_t) - (1+r_f) \tag{8}$$

其中 rf 无风险利率可以取常用的 2.4%，是一个常量。我们的想法为如果保证了每一周收益都是最大的那么整体的利率必然是最大。

为了对问题进行简化，我们首先对 0-1 规划的部分进行简化抽象。我们认为，如果要使得当周收益率尽可能大，我们仍然保持局部贪心的策略。即：如果选择的十只股票利率都是最大的那么当周收益率必然最大。但这一问题并不能直接应用贪心策略，因为需要保证月利率与中证指数相差不超过 2%。我们将中证指数随着时间的变化曲线绘制如下：

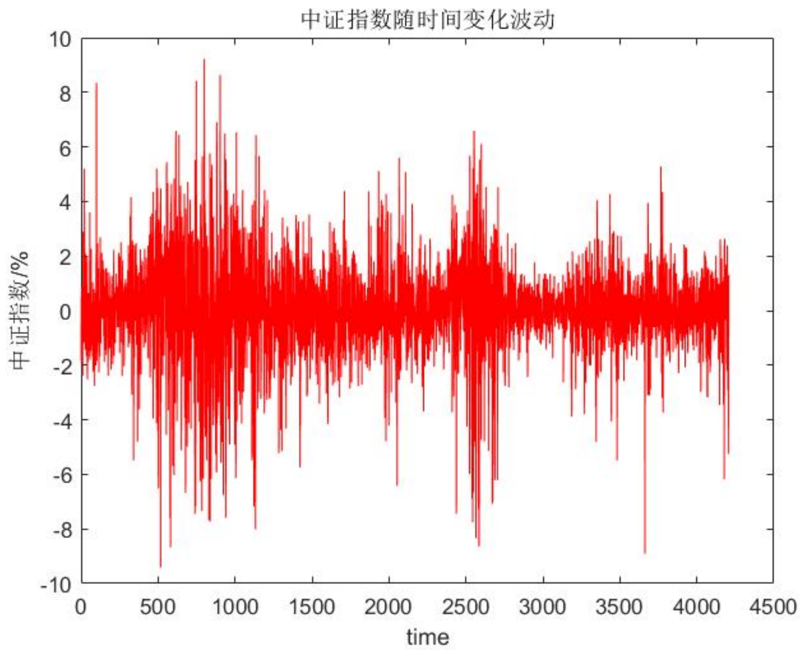

图 3. 中证指数变化

可以看到中证指数实际上是一个平稳时间序列。由于 2%的限制使得我们不可能做到贪心，但可以通过动态规划的方式进行选择。我们可以按照预测值和实际值的差额进行排序，获得每支股票收益率从大到小的序列。随后我们自大到小进行状态空间的划分可以对问题进行初步求解。即当步的收益函数可以表示为：

$$f(y_i|d_1,d_2,...,d_{i-1}) = f(y_{i+1}|d_1,...,d_{i-1},d_i=0) + f(y_{i+1}|d_1,...,d_{i-1},d_i=1) \tag{9}$$

即，当步收益为在前 i-1 支股票的选择条件下选择该股票的收益和不选择该股票的收益之和。这一操作可以划分状态空间树，从而在状态空间内分别进行更为容易的

求解。但状态空间过于庞大，我们可以利用 2%的浮动条件对状态空间进行剪枝，避免无效状态的生成。但绝对值条件并不易于我们处理，为了对绝对值进行简化，我们引入松弛变量，将条件变为：

$$\begin{cases} \xi_t = r_t - r_{0t} \\ -2\% \leqslant \xi_t \leqslant 2\% \end{cases} \tag{10}$$

状态空间树最大层数限制为 10 层，表示仅选择 10 支股票。于是我们可以设计下面的算法对问题进行求解：

```
Input: n series y, time step t
Variables: w,D,f
Output: max f
Algorithm: Dynamic Programming
for i in range(n):
    yhat[i]=ARIMA(p,d,q).predict(y[i],T=7)
    dy[i]=(yhat[i]-y[i])/y[i]
end
Sorted dy
save Index y_which
for i in range(1,length(dy)):
    choose[i]=choose[i+1,choose[i]=0]+choose[i+1,choose[i]=1]
    f=w@choose@dy
    Optimize f w
end

Algorithm: Optimize
f=w@choose@dy
max f
s.t.: 0<=w<=0.2
      xi=f_month-r0
      -0.02<=xi<=0.02
```

此外，我们还可以利用遗传算法进行全局式自动化求解，无需过多分析过程。事实上，这一问题可以抽象为一个 0-1 规划+线性规划问题，变量在一个 100 行 10 列的矩阵中排布，要求收益最大。

对于这一规划问题，传统的算法例如匈牙利法、蒙特卡洛方法已经难以求解因为问题的变量维数太高。随着进化计算与群体智能算法的成熟，有越来越多的智能算法被用于函数优化问题的求解，其中，遗传算法是一类借鉴生物界自然选择和自然遗传机制的随机搜索算法，通过模拟生物的遗传、变异等自然现象搜索函数极值[8]。其主要特点是直接对结构对象进行操作，不存在求导和函数连续性的限定；具有更好的全局寻优能力；不需要确定的规则就能自动获取和指导优化的搜索空间，自适应地调整搜索方向[9,10]。

遗传算法借鉴了生物学的概念，首先需要对问题进行编码，通常是将函数编码为二进制代码以后，随机产生初始种群作为初始解。随后是遗传算法的核心操作之一——选择，通常选择首先要计算出个体的适应度，根据适应度不同来采取不同选择方法进行选择，常用方法有适应度比例法、期望值法、排位次法、轮盘赌法等[11]。

在自然界中，基因的突变与染色体的交叉组合是常见现象，这里也需要在选择以后按照一定的概率发生突变和组合。不断重复上述操作直到收敛，得到的解即最优。遗传算法基本思想如图 4 所示：

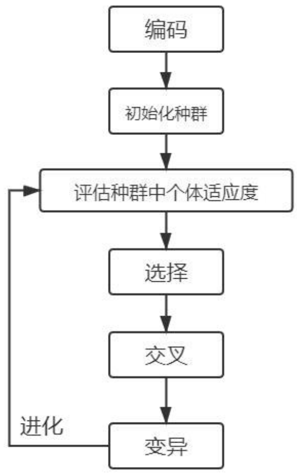

图 4 遗传算法的基本思想

最终通过遗传算法，设置迭代次数为 500 轮，最终结果如图 5 所示。比对动态规划的分配方案，两种方法求解差异不大，说明遗传算法的使用是可行的。

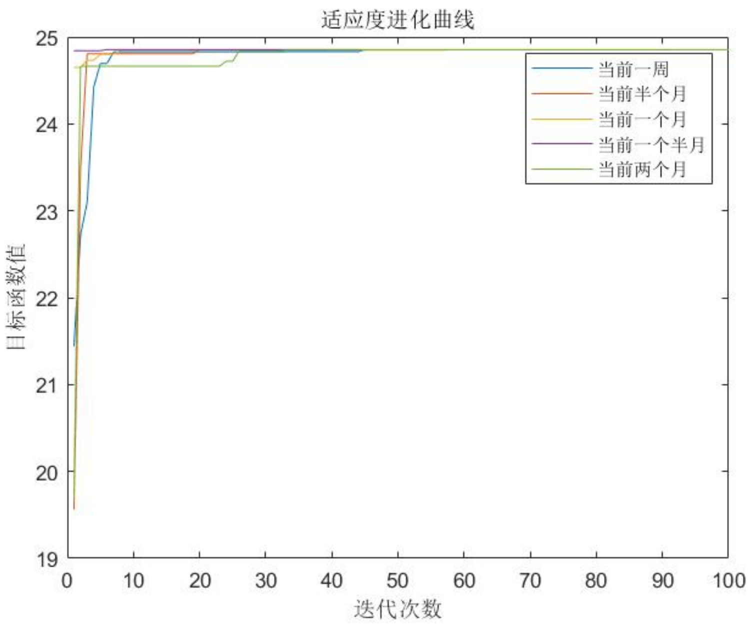

图 5　遗传算法的迭代曲线

5.2.2 模型的求解

由于模型求解相对较为复杂，这里我们分为三部分进行展示。

首先是股票的选择。我们以 2016 年 4 月 18 日为例，若在这一天进行调仓，我们采用 ARIMA 模型对齐进行了拟合预测，获得了一周以后（2016 年 4 月 25 日）的预测值，计算得到 100 支股票的预测收益率。

通过动态规划剪枝，我们将选择的 10 支股票写入表 1 中：

表 1. 2016 年 4 月 18 日调仓选择的股票名称

编号	1	2	3	4	5
名字	金融街	旷达科技	上汽集团	大东方	佳士科技
编号	6	7	8	9	10
名字	中国石化	金地集团	海澜之家	绿地控股	工商银行

观察股票平均收益和风险的散点图如图 6 所示：

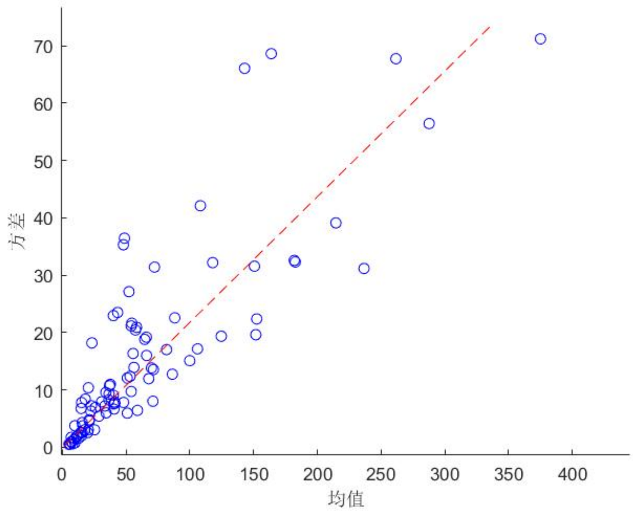

图 6. 收益-风险散点图

可以看到风险收益大致在两条直线之间分布，均值小的通常方差也小一些，这也与我们的直观认知相符合，即收益大的股票通常风险也更高。那么风险和收益近似呈现线性相关关系，R^2 为 0.94，说明拟合效果是比较好的。而我们选择的这十支股票，方差与斜率的比值也较低，说明这一选股策略保证了更高收益的同时也减低了风险。

我们认为我们选择的股票是较为合理的，基本上保证了均值高的情况下方差也较低，也就是既保证了高收益的同时还能够做到低风险。

其次是投资的仓位，以 2016 年 4 月 18 日的价格为例，若在这一天进行调仓，10 支股票的仓位如表 2 所示：

表 2. 投资的仓位分布

编号	1	2	3	4	5
名字	金融街	旷达科技	上汽集团	大东方	佳士科技
仓位	0.05	0.20	0.03	0.02	0.09

编号	6	7	8	9	10
名字	中国石化	金地集团	海澜之家	绿地控股	工商银行
仓位	0.04	0.11	0.06	0.20	0.20

可以发现，绿地控股、工商银行和旷达科技的股票是最值得买入的，仓位直接拉到了上限 20%。而例如上汽集团、大东方等则比较弱，若没有仓位限制很可能买入工商银行、旷达科技和绿地控股的仓位更高。

最后整体的收益曲线，我们对比了不同调仓周期下的收益率曲线，发现按月进行调仓的效果是最好的。将收益曲线绘制在图 7 所示：

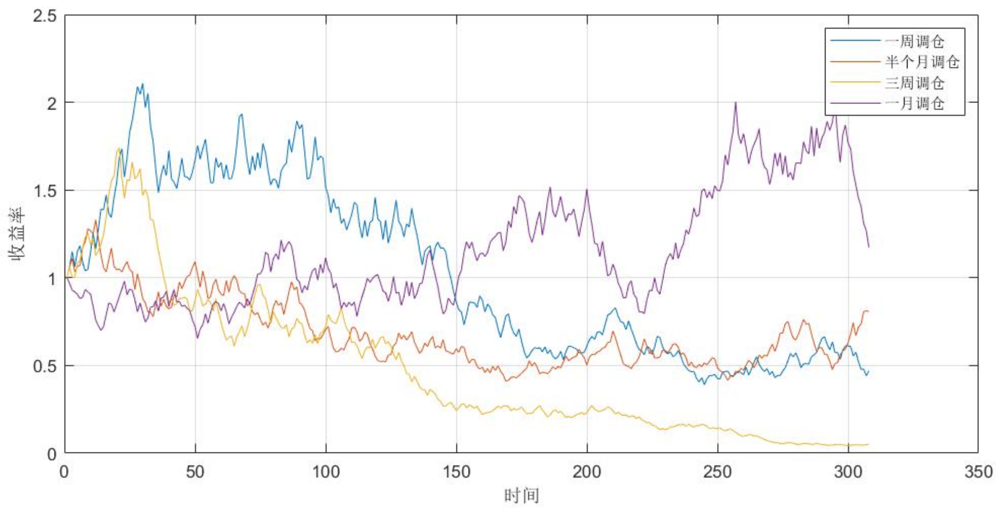

图 7. 总体的收益曲线

可以看到，按月为周期进行调仓交易则总体收益是最高的，不会造成像频繁调仓一样造成的大幅度亏损现象。

另外我们对比了每次调仓时投入利息和不投入利息的情况，认为投入利息会大大提高总体收益率。

六、 模型的检验与分析

6.1 模型的检验

为衡量模型的准确性，我们针对模型的一些步骤及其结果进行了相关检验：

1. ARIMA 模型平稳性的 Dicky-Fuller 检验：Dicky-Fuller 检验常被用来检验时间序列的平稳性，需要事先判断序列是否平稳。经检验，所有序列均为平稳序列。

2. ARIMA 模型残差的正态性检验：正态性检验的方法有很多，这里采取了 K-S 检验的方式，最终所有序列的残差均通过置信度为 0.05 的正态性检验，接受备择假设认为残差服从正态分布。

3. 遗传算法求解规划问题的结果检验：求解结果经检验均满足约束条件。

6.2 模型的定性分析

我们对每一步骤的模型选择与其它常用模型进行过对比，这里再定性分析一下这些模型的合理性：

1. 时间序列模型的选取：这里由于数据量是中规模体量，预测方法没有选用小规模数据常用的拟合与灰色预测，也没有使用大规模数据的机器学习和海量数据的深度学习方法。采用 ARIMA 模型，利用 AIC 准则确定最优参数，模型不仅满足平稳性，残差也是正态的，并且形式简洁防止过拟合。

2. 规划求解算法的选取：使用了传统算法和遗传算法两者进行对比，两种方法的结果差异不大，使得到的结果更具有可信度。

6.3 模型的灵敏度分析

为了进一步衡量模型的鲁棒性与泛化性能，我们做了一些参数相关的灵敏性分析。由于 ARIMA 模型的参数是利用 AIC 准则选取的最优参数组合，这里就主要针对规划问题的权重系数进行了讨论。

我们在问题中针对多目标规划的两个目标随机选取了 100 组权重参数，但惊奇发现最终的投资方案惊人地一致，说明多目标的参数选择对结果是不敏感的。

七、 模型的评价、改进与推广

7.1 模型的优点

综合上述的模型建立与求解，我们认为，我们的模型有如下优点：

1. 利用 ARIMA 模型的时间序列预测，从算法原理上属于时间序列预测中较优的方法。而从参数选择的意义上来看，使用 AIC 准则选择了合适的模型参数，既保障了模型拟合的效果，又降低了参数个数防止过拟合并且模型形式较简洁，训练开销得到了降低。

2. 两种规划组合，层层递进环环相扣，每一层都可以根据上一层的解来设计相应程序自动化求解。

3. 投资方案的求解使用遗传算法，相比传统算法而言提升了速度。并且这一想法可以扩展到任意高维自变量的函数极值求解。同时也使用传统贪心策略进行了对比，优先选择损失率最小的，这样求解的结果与遗传差异不大，进一步验证了结果可靠性。

7.2 模型的缺点

尽管模型已经表现出比较良好的特征，但仍存在一些问题：

1. 利用 ARIMA 模型预测的结果只能考虑内生变量的影响，预测结果是有一定偏差的，并且难以预测发生突发情况导致的订货量、供货量和损失率的变化。

2. 遗传算法在求解过程中可能并不能严格满足约束条件，由于初始化种群的随机性以及变异、重组操作的随机性，得到的结果可能不完全一致。这时需要人工进行一些微调，耗费精力。

7.3 模型的改进

若对模型进行改进，可从以下方面进行改进：

1. 尝试使用不同的群体智能算法对这一约束问题进行求解例如蚁群算法、粒子群算法、萤火虫群算法等。

2. 考虑在时间序列模型预测中加入一些隐藏外生变量或者考虑季节性因素，将其扩展为 SARIMAX 模型进行预测。

7.4 模型的推广

通过这一问题的求解，我们认为这一模型可以在如下方面进行扩展：

1. 未来一段时间的计划制定：指定未来一段时间的计划需要事先估计未来每个时间节点的相关数值，所以可以使用时间序列对历史数据进行建模再迁移到未来时间的预测效应。

2. 遗传算法用于高维自变量对应函数的极值求解：将矩阵每个 0-1 变量都看作自变量的一个维度，然后求解极值。这一做法主要是将二维矩阵展平为一维，这一想法可以迁移到任意高维自变量对应函数的极值求解。

八、 参考文献

[1]. 陈顺财. 基于支持向量机的时间序列预测研究[D].兰州理工大学,2008.

[2]. Hochreiter S , Schmidhuber J . Long Short-Term Memory[J]. Neural Computation, 1997, 9(8):1735-1780.

[3]. Cho K , Merrienboer B V , Gulcehre C , et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation[J]. Computer Science, 2014.

[4]. Asteriou D , Hall S G . ARIMA Models and the Box-Jenkins Methodology[M]. 2016.

[5]. Mcleod A I , Hao Y . Dickey-Fuller test[J].

[6]. 马永杰，云文霞．遗传算法研究进展[J]. 计算机应用研究,2012,29(04):1201-1206+1210.

[7]. MATLAB 中文论坛. MATLAB 智能算法 30 个案例分析[M]. 北京航空航天大学出 版社, 2010. P17-P22

[8]. 吴渝, 唐红, 刘洪涛. 网络群体智能与突现计算[M]. 科学出版社, 2012.

[9]. 李士勇，李研，林永茂. 智能优化算法与涌现计算[M]. 清华大学出版社，2020，P115-118.

附录

环境：Windows 10+ intel i7+ NVIDIA GTX1650

软件：jupyter notebook(python 3.8.2)+MATLAB 2017a

```
load data
%% 进行使用 ARIMA 进行预测的函数
function [forData,lower,upper] = Fun_ARIMA_Forecast(data,step,max_ar,max_ma,figflag)
    %  使用 ARIMA 进行预测的函数（使用 n 阶差分、不使用对数），可以直接调用，差分阶数自动确定。
    % 输入：
    % data 为待预测数据，一维数据，最小 11 个数据。但是数据长度处于 11~15 时依旧可能出现报错的情况。
    % step 为拟预测步数
    % max_ar 为最大 p 值
    % max_ma 为最大 q 值
    % figflag 为画图标志位，'on'为画图，'off'为不画
    % 输出：
    % forData 为预测结果，其长度等于 step
    % lower 为预测结果的 95%置信下限值
    % upper 为预测结果的 95%置信上限值
y_h_adf = adftest(Y)
y_h_kpss = kpsstest(Y)
% 一阶差分，结果平稳。如果依旧不平稳的话，再次求差分，直至通过检验
Yd1 = diff(Y);
yd1_h_adf = adftest(Yd1)
yd1_h_kpss = kpsstest(Yd1)
% ACF 和 PACF 法，确定阶数
figure
autocorr(aimY)
figure
parcorr(aimY)
% 通过 AIC，BIC 等准则暴力选定阶数
max_ar = 3;
max_ma = 3;
[AR_Order,MA_Order] = ARMA_Order_Select(Y,max_ar,max_ma,1);
Mdl = arima(AR_Order, 1, MA_Order);  %第二个变量值为 1，即一阶差分
EstMdl = estimate(Mdl,data);
[res,~,logL] = infer(EstMdl,data);   %res 即残差
stdr = res/sqrt(EstMdl.Variance);
figure('Name','残差检验')
subplot(2,3,1)
plot(stdr)
title('Standardized Residuals')
subplot(2,3,2)
histogram(stdr,10)
title('Standardized Residuals')
```

```
    subplot(2,3,3)
    autocorr(stdr)
    subplot(2,3,4)
    parcorr(stdr)
    subplot(2,3,5)
    qqplot(stdr)
    % Durbin-Watson 统计是计量经济学分析中最常用的自相关度量
    diffRes0 = diff(res);
    SSE0 = res'*res;
    DW0 = (diffRes0'*diffRes0)/SSE0 % Durbin-Watson statistic，该值接近 2，则可以认为序列不存在一阶相关性。
    %% 5.预测
    step = 300; %预测步数为 300
    [forData,YMSE] = forecast(EstMdl,step,'Y0',data);   %matlab2018 及以下版本写为 Predict_Y(i+1) = 
forecast(EstMdl,1,'Y0',Y(1:i));
    lower = forData - 1.96*sqrt(YMSE); %95 置信区间下限
    upper = forData + 1.96*sqrt(YMSE); %95 置信区间上限
    figure()
    plot(data,'Color',[.7,.7,.7]);
    hold on
    h1 = plot(length(data):length(data)+step,[data(end);lower],'r:','LineWidth',2);
    plot(length(data):length(data)+step,[data(end);upper],'r:','LineWidth',2)
    h2 = plot(length(data):length(data)+step,[data(end);forData],'k','LineWidth',2);
    legend([h1 h2],'95% 置信区间','预测值',...
            'Location','NorthWest')
    title('Forecast')
    hold off
    end
    %% 使用 Fun_ARIMA_Forecast 函数实现预测
    % 单步预测
    close all
    clear all
    addpath ../funs %将 funs 文件夹添加进路径
    load Data_EquityIdx   %纳斯达克综合指数
    len = 100;
    data = DataTable.NASDAQ(1:len); %如果要替换数据，将此处 data 替换即可。
    forData1 = zeros(1,len); %全部初始化为 0
    for i = 30:len
        forData1(i+1) = Fun_ARIMA_Forecast(data(1:i),1,2,2,'off');
    end
    figure()
    plot(31:len,data(31:len))
    hold on
    plot(31:len,forData1(31:len))
    legend('原始数据','单步预测数据')
```