2023 年第四届“大湾区杯”粤港澳金融数学建模竞赛

题目：

摘要

一、问题重述

1.1 引言

证券投资的核心问题始终是如何获取收益与规避风险。在这一复杂且多变的领域中，准确评估证券价值是形成有效投资策略的关键所在。对于中国这个庞大的股票市场来说，其独特性和复杂性要求我们不能简单地应用传统的估值模型。据中国证监会主席易会满提出，要建立具有中国特色的现代资本市场，探索与之相符的估值体系，这无疑为我们提供了新的思考方向。这种具有中国特色的估值体系不仅考虑了公司的财务数据和市场前景，还充分融入了政策导向、价值投资、资本配置和风险管理等核心要素。

通过结合题目中所给出的相关热点问题、国家的重大战略方向等材料，结合所找寻到的相关证券信息进行分析，寻找出对具有中国特色的估指体系具有明显强相关的数据指标，建立一个基于2000年到2023年相关证券数据信息的中特估体系，完成对中特估股票画像的塑造，进而完成对中特估相对完整的定义并构建对所得到的中特估股票的分类模型。在已建立相关模型的基础上，分别构建中特估的短期、长期股票投资组合，进而预测所构建出的投资组合的风险收益特征。

1.2 问题提出

通过结合题目所提供的相关材料以及所寻找到的证券相关信息数据，建立具有中国特色的估指体系模型，并研究以下问题：

1.具有比较明确的政策背景和清晰的资本市场定位的中国特色估值体系存在一个问题，那就是其相对应的模型指标特征急需构建。可以对应于中特估概念的市场定位和专家解析，以及相关的政策背景，构建出对应于中特估股票的特征指标，进而给出中特估股票的画像，从而回答什么是中特估股票，给出中特估股票的定义。

2.根据所建立的模型特征指标数据，可以将限定区域内的证券市场(如沪深A股)的中特估股票进行筛选出来，并将其进行分类。在分类的基础上，分析分类所得到的股票的投资特点。

3. 证券市场的行为并不是孤立存在的，其很大程度上依赖于市场周围的环境，随经济环境的变化而变化。其中，经济环境的热点往往是影响股票走势的最敏感因素。针对所构建的中特估股票模型的特征指标，并结合典型的市场热点（如国际环境，资产重组，价值投资，以及舆论影响等热点事件），设计一个基于中特估评价体系的短期股票投资组合，并对其进行实测，得到其相对应的风险收益特征。

4. 基于所构建的沪深 A 股中特估的股票特征指标，设计一个长期股票投资组合模型，并分析该投资组合的风险收益特征。

二、问题分析

三、符号说明

四、模型假设

五、模型建立与求解

5.1 针对问题一的求解

5.1.1 数据预处理

本题中，题目并没有提供相应的数据，只是介绍了中特估的特点以及相对应的一些政策信息。因此，为了所构建模型的准确性，我们通过获取对应的沪深A股数据信息来进行模型的构建。

由于所得到的数据指标并不全部适用，因此需要对已有数据进行预处理，对数据缺失和年份不足2000到2023年的数据进行“剔除”，对“剔除”后的数据进行“筛选”，进而得到有效数据，并对得出的有效数据运用三次样条插值法来定义一个统一的标准。

5.1.1.1 “剔除”数据

根据所获取得到的股票的相应财务报告信息，股价的波动信息等数据，基于利用高频有效、特征明显的数据使用原则，需剔除所得到的数据中数据缺失和年份缺失的数据。具体数据缺失和年份缺失的数据所在的二级目录如下表所示：

|  |  |
| --- | --- |
| 二级目录 | |
| 数据缺失 | 年份缺失 |
| 各证券的资金流动比率、速动比率、现金比率、总负债同比增长率 | 细分行业数据 |
| 企业经济效益指标(月) | 证券财报公开季度数据 |
| 央行货币工具(日) | 公司现金流量 |
| 央行货币政策(日) |  |
| 企业日利息收入 |  |
| 拆借回购利率 |  |
| 每股收益 |  |
| 营业收入 |  |

通过剔除上表的二级目录中的数据后可知，可以进行操作的二级目录数据有政府资金支持比例(%),（持股比例）控制人所占股份比例(%),股票市盈率(%)，股票市净率(%)，股票盈利收益率(%),股票资产负债率(%)，现金流量覆盖率(%),国有企业性质。

5.1.1.2 “筛选”处理数据

由于秉着所选择的数据指标要高频有效，则需对剔除上述二级目录数据后的数据进行筛选。具体的筛选过程如下：

(1)政府资金支持比例（持股比例）：政府资金占公司总资本的比例，政府及相关机构持有的公司股份占总股本的比例，反映政府直接资金支持程度。相应的计算公式为：

(2)政策风险：评估政策变动对公司的潜在风险，包括政策不确定性、政策调整频率等。事实上，政策风险可应用泊松分布：

其中P(X=k)是在一个给定时间段内发生 k 次事件的概率。λ是往期政策调整的平均发生率。但由于不同行业政策调整不同次数的产生影响力不同，因此可将政策调整次数大于等于的概率作为风险指数

对于泊松分布，事件发生次数大于或等于 的概率可以通过以下公式计算：

（3）市盈率（PE ratio）：市场价格与公司过去年度平均盈利之比，反映公司的估值水平。

市净率（PB ratio）：市场价格与该股票公司的净资产之比，反映公司资产估值水平。

这里我们使用一个综合估值比率来比较客观的反映综合衡量公司的估值情况：

负债率：公司的负债占该公司现有总资产的比例，一般来说，较低负债率表示该公司的财务情况较为稳健。

现金流量覆盖率：同构该数据可以了解公司的自由现金流量是否足以覆盖债务和分红。

平均年复合增长率：过去几年的盈利年复合增长率，反映公司的长期盈利增长潜力。

（3）国家战略关联度：公司业务与国家战略方向的相关性程度，该指标可以用产业关联度指标来衡量。事实上，使用产业关联度指标可以考虑公司业务与国家战略方向相关性的程度。可以通过皮尔逊相关系数来计算

其中， 代表公司业务和国家战略方向相关的变量值, 代表对应变量的均值。另外，与需由构建的公司业务向量与国家战略向量所构成，并且需要采用L2范数正则化处理。

国有企业性质：公司是否属于国有企业，以及国有股比例。公司是否属于国有企业，以及国有股比例，反映国有资本的影响程度。该指标可由：公式计算可得。

波动率（Volatility）：公司股价的历史波动率，可通过计算日收益率的标准差来衡量。

其中：是收益率观测次数；是第天的收益率 是平均收益率。

贝塔系数（Beta）：公司股价相对于市场的变动，可通过回归分析计算。

其中：是股票收益率与市场收益率的协方差，是市场收益率的方差。

市场泡沫风险：分析公司估值相对于行业和市场估值的比例，可使用市盈率相对行业平均值等指标进行计算：。

资金风险：分析公司的财务风险水平，包括债务水平和偿债能力，可以使用负债率、利息保障倍数等指标进行表示，。

5.1.1.3 三次样条插值法进行数据采集

根据“筛选”处理数据后，得到高频有效的宏观经济指标。但由于数据多样、复杂、时间不统一，故需定义一个标准。同时考虑到所提供的数据中，以“月”来提供的数据的占比较多,所以我们团队选择运用三次样条插值法采集数据,将“日”“周”、“季”与“年”所提供的数据都转化为“月”来进行分析。

三次样条插值函数S（x）是一个分段三次多项式，需要求出三次样条插值函数S（x）。可在每个小区间[,]上确定4个待定参数，用表示它在第t个子区间[,]上的表达式，则其表达式如下:

其中，、、、为每个小区间的4个待定参数。

具体数据转化计算的推导过程如下：

（1）由于所有点必须满足插值条件，可得：

同时，所有个内部点的每个点都满足：

(2)根据n-1个内部点的一阶导数需连续，即在第区间的末点和第区间的起点是同一个点，则它们的一阶导数也需相等，即：

同时，内部点的二阶导数也需连续，即：

(3)根据推导过程（1）可得：

(4)以表示步长，结合tuidao过程(1),可得：

(5)根据推导过程(2)可得：

(6)设=2可得：

(7)根据推导过程(4)可得：

(8)将所推导得的、、、代入到推导过程(5)中，可得：

其中，表示第t个数据指标所对应的时间，表示第t个数据指标的时间下所对应的数值。

综上所述，通过上述的推导过程，我们构造出一个以M为未知数的线性方程组，导入数据，从而将“日”与“年”所提供的数据都转化为“月”来进行分析。相应的数据结果如下表所示：（部分，具体详情可见附录支撑材料）

|  |  |  |  |
| --- | --- | --- | --- |
| **时间** | **流动比率** | **速动比率** | **现金比率** |
| 2007-03 | 0.127655 | 0.073317 | 0.012559 |
| 2007-06 | 0.121009 | 0.067158 | 0.015398 |
| 2007-09 | 0.117993 | 0.064408 | 0.014266 |

5.1.1.3 对指标数据进行标准化处理

由于数据预处理后确定出的证券数据指标属于不同类型的数据，包括百分率，具体金额数据等量纲不统一的问题，因此在正式进行分析前需要对数据进行无量纲化处理，消除量纲对结果可能带来的影响。此处理方法是将原始数据标准化，即先对数据做如下均值和标准差处理:

再对处理完的数据进行标准化，即：

其中，表示由三次样条插值法采集得到的t时刻第j个证券指标数据;表示第项证券经济指标的均值；表示第项证券经济指标的方差。

通过上述的处理，可以得到t时刻第j个证券指标的标准化数据。