

参赛编号：202406900582

2024 年第五届“大湾区杯”粤港澳

金融教学建模竞赛

题目：基于 Markov 结合多指标的风险预警控制体系模型

摘要：

投资风险预警在实际金融交易中起到不可或缺的作用，为了更科学有效地实施风险管理，本文将围绕测度和检测系统性风险展开研究。

针对任务一，将风险量化分析为平均收益率，市场流动性(MFD)，市场情绪指标(投资者信心指数)三个指标，根据其经济意义，使用附件给出的沪深300的实际数据，分析各风险计量指标的优缺点，得出三个指标的可视化结果，且均能让投资者更清晰的看到该股（或市场）近期的大致情况即风险，具体指标计算结果及可视化见后文。

针对任务二，根据实际的市场情况增加了三个风险计量指标分别为高频交易影响比率，特殊事件风险指数，长短期波动率比例，将新增的三个指标结合任务一给出的三个指标，针对该多指标模型，用熵权法对各指标进行赋权，建立风险控制模型，用 Markov Chain 算法对风险进行预测，基于风险控制模型，建立风险状态转移概率矩阵和风险状态分布矩阵，得到风险预测模型，并以沪深300为例展示该模型风险预测结果，从模型实测结果可以看出，在抽样调查的结果下其在11年中8年基本完全吻合真实的风险水平，对于题目当中提到的需要构建考虑系统性风险的模型，模型当中也引用了六个因子，模型综合考虑较全面。

针对任务三，基于任务二建立的风险预测模型针对混合型权益类，设置止盈止损线，即通过预测未来风险状态，以投资者买入后盈利50%或亏损50%作为止盈止损线，若超过止盈线或止损线则直接进行空仓处理，考虑到使用回测使得回撤不超过0.7，故使用仓位调整及向上对冲的方式，将收益最大化或保住本金。在基于沪深300在2024年6月起始至10月的实测结果当中显示，其最大回撤为0.15，同时在波动的市场当中还有部分利润。

针对任务四，预测合理收益，使用蒙特卡洛进行随机模拟，基于任务一二设置的量化因子，增加溢价等多因子，结合十年国债年平均收益率3.281%使用蒙特卡洛模拟因子值的选取，用熵权法计算因子的权重，将它们加权求和得到预测的合理收益，通过10000次蒙特卡洛试验最后得到其收益预期在3%上下为较为合理的收益预期。

关键词：马尔科夫预测，熵权法，风控模型，蒙特卡洛模拟

一、问题重述

1.1 引言

随着中国金融市场的快速发展，金融体制经历了从大型国有金融机构主导到市场化和多元化发展的转变。这一过程中，金融市场的逐步发展和完善极大地促进了资本在不同产业间的流转和配置，增强了社会主义市场经济的活力。然而，金融市场也面临着系统性风险的挑战。从 2008 年金融危机到 2015 年股市大幅震荡，再到 2024 年初的股票市场剧烈波动，系统性风险成为研究和应对中国特色社会主义市场经济中越来越重要的问题。本文将探讨如何运用量化方法测度和监测系统性风险，以更科学有效地实施风险管理。

1.2 问题提出

现根据题目要求，参考市场因子设计，研究以下问题：

问题一：计算与分析风险计量指标。根据附件 1 提供的 2014 年至 2024 年数据，计算沪深 300 个股平均收益率、市场流动性、市场情绪指标这三个风险计量指标，分析各指标的经济意义、模型算法，指出它们各自的优缺点。

问题二：构建系统性风险预测的模型。设计三个或以上风险计量指标，基于所设计的风险计量指标，综合多个指标，构建一个对未来系统性风险具有预测能力的模型。

问题三：构建事前风控体系。国际惯例中，权益类共同基金的回撤风控线一般为 0.7，而我国权益类公募私募基金的回撤常跌破此标准。以 2014 年至今的 A 股历史数据为基础，构建出回测可以控制在 0.7 以内的事前风控体系（针对混合型权益类）。

问题四：合理收益预期设定。散户亏损的一个重要原因是对收益预期的非理性预期。以 2014 年至今的 A 股沪深 300 个股（历史数据）作为投资标的，参考 10 年期长期国债收益率为标准，设定一般投资者投资股权市场的合理收益预期。

二、问题分析

本题主要是分析在投资中如何在变动的投资市场中预测投资的风险，量化为不同指标帮助投资者预测风险。针对任务一，需要将题目给出的三个指标进行量化分析，建立数学模型，比较各指标之间对风险的预测情况和各自的优缺点；针对任务二，除任务一给出的三个衡量指标外根据实际的投资市场再增加三个衡量指标，以这六个指标为衡量依据，计算各指标所占的权重，建立多目标风险模型，在该模型的基础上，结合 *Markov Chain* 算法建立风险预测模型；第三问采用设置止盈止损线结合预测的风险进行调整和防止回撤；第四问采用通过蒙特卡洛模拟对参数进行随机抽取，最后得到一个较为合理的预期收益分布。

结合以上问题，分别作出如下分析：

2.1 任务 1 分析：

针对任务一，题目当中需要分别计算沪深 300 个股的平均收益率，市场流动性以及市场情绪指标这三个风险指标，为后文当中的风险量化做出铺垫，因此在该段当中选取的指标分别为沪深 300 的个股的平均收益率并通过附件当中给出的数据进行加权求和得到沪深 300 每日的收益率，同时对于市场流动性本文当中选取了 *MFI* 即资金流量指数作为标准，对于 *MFI* 大于 80 则市场存在超卖的情况出现，可能会在一段时间内由于回调而反弹，而小于 20 则说明存在超买的现象，同样也会出现反弹的情况，而对于市场情绪指标在该题当中选用投资者信心指数作为标准，投资者信心指数由成交量进行计算，可以很好的衡量投资者在沪深 300 当中投资的积极性。

2.2 任务 2 分析：

针对任务二，在任务二当中需要建立系统性风险预测模型，需要设计三个新风险计量指标，在这里考虑为高频交易影响比率，特殊事件风险指数以及长短期波动率比例，具体通过衡量三个角度的分析情况，事件风险如（“黑天鹅”，政策等问题），短期波动风险(多种原因构成，可能与企业有关)，交易风险（如投资者积极性，股票热门程度），同时在该问由于设计到多个指标的问题，为了对风险进行预测，在这里考虑使用 *Markov* 链对未来进行预测，通过一阶 *Markov* 推导，两种矩阵得到对实际风险的量化结果，并选用沪深 300 为标的，预测风险与真实风险差距。

2.3 任务 3 分析：

针对任务三，基于问题二当中的模型，同时设置止损止盈线，通过预测未来风险状态，超过止损线或者跌过止盈线则直接进行空仓，同时在该问当中仅需要在回测当中使得回撤不超过 0.7，因此还可以考虑使用仓位调整以及加上对冲等方式，通过对未来风险预测，对仓位以及对冲进行调整最后将收益最大化或者进行保本。

2.4 任务 4 分析：

针对任务四，在该问当中考虑使用蒙特卡洛模拟，对预期收益率进行建模，考虑到第一题以及第二题当中涉及的多个因子，同时以十年期国债的平均收益率为标的，以沪深 300 作为研究对象，使用模拟算法模拟因子值的选取，通过由熵权法计算得到的因子权重，加权求和可以得到其真实收益率大小。

基于以上总结为如下模型概念图：

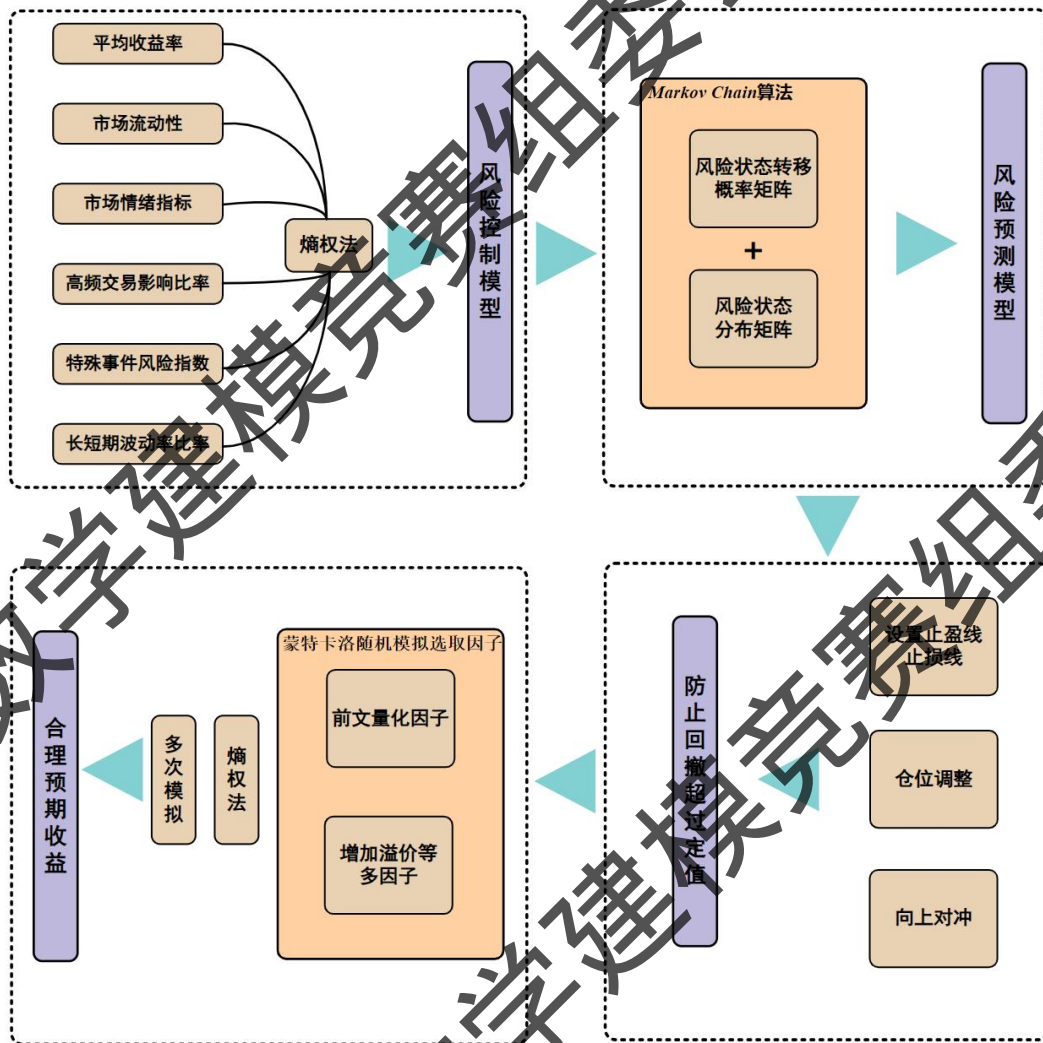


图 1 风险量化分析模型概念图

三、符号说明

符号	描述说明
$r_{i,t}$	第 i 只股票第 t 天的日收益率
$p_{i,t}$	第 i 只股票第 t 天的收盘价
Q_t	第 t 天的市场平均收益率
$K_{i,t}$	第 i 只股票第 t 天的典型价格
$\sigma_{i,t}$	正资金流量
$\tau_{i,t}$	负资金流量

$H_{i,t}$	第 i 只股票第 t 天的高频交易影响比率
$L_{i,t}$	第 i 只股票第 t 天的长短期波动率
$E_{i,t}$	事件发生的七天之后其价格波动风险指数

四、模型假设

- 1、假设模型在考虑事件当中不需要考虑某些如战争等不可预测因素；
- 2、假设市场下一天的结果全部由上一天决定；
- 3、假设金融市场是有效的，即所有可用信息都已经反映在资产价格中；
- 4、假设不存在无风险的情况，天下没有免费的午餐；

五、模型建立与求解

5.1 问题 1 第 1 问：三种风险指标意义及算法

基于附件给出的数据，本问将对沪深 300 个股的平均收益率，市场流动性，市场情绪指标这三个风险指标进行量化分析，给出这三个指标的经济意义，及对应的模型算法和各自对应的优缺点。

5.1.1 收益率指标

在投资者实际采买股票的过程中，收益率是评估股票投资表现的关键指标，为了让投资者可以获取更深入的市场洞察，在本问中，将收益率指标分为单股和市场两个方面进行量化分析，并将市场方面拆分为市场平均收益率和市场年化收益率进行分析，如下文所示。

5.1.1.1 单股-日收益率指标

日收益率是指单股在特定交易日内的价格变动率，它反映了某支股票短期的表现。其计算公式为：

$$u_{i,t} = \frac{p_{i,t} - p_{i,(t-1)}}{p_{i,(t-1)}}$$

其中， $p_{i,t}$ 表示第 i 只股票第 t 天的收盘价，上式表明某股的第 t 天的收益率由当天的收盘价和前一天收盘价的差值除以当天的收盘价，得到该股当天的日收益率。

5.1.1.2 市场-日平均收益率

在这里将题目当中给出的平均收益率指标拆分为市场平均收益率即沪深 300 的日平均收益率和每年的年化收益率。为了更好的察觉股市的变化，对每日

的股票加权求和以得到该日市场的平均收益率。

$$Q_t = \sum_{i=1}^n \omega_i u_{i,t}$$

其中， ω_i 表示第*i*只股在该日在股市中所占的权重， $u_{i,t}$ 表示该股票在当天的日收益率,由此可以得到沪深300在该日的日收益率。以上即为对于每日的沪深300平均收益率，在下文再给出对于每年的平均收益率。

5.1.1.3 市场-年化收益率

对于市场整体的年化收益率，本文采取将该支个股在年初的开盘价以及年末的开盘价求差，再除以年初开盘价得到针对于该股的年化收益率；之后针对沪深300的全部股票进行加权求和求平均，得到对于沪深300而言的年化收益率，具体公式如下所示：

$$y_{i,t} = \frac{P_{i,last} - P_{i,1}}{P_{i,1}}$$

然后将该个股的年化收益率加权求和得到如下所示沪深300的年化收益率：

$$R_t = \sum_{i=1}^n \omega_i y_{i,t}$$

需要注意的是，这里的*t*与上文不同，这里的*t*指从2014年至2024年每年的数据，并非上文当中的每日。上式当中，*R*代表沪深300的年化收益率，而*y*则代表个股的年化收益率。

综上，根据市场平均收益率和市场年化收益率可以看出平均收益率的优缺点如下所示：

·优点：

通过计算各个股票的平均收益率，能够更直观易懂的给出资产的平均回报情况，能够帮助投资者更快速的了解大致的市场变化。

·缺点：

它忽略了在实际投资中的复利情况，可能会在波动性较大的市场中失真，使预测的准确性降低。收益的波动性，未能反映投资收益的波动情况，可能掩盖了潜在的高风险。

5.1.2 市场流动性指标-MFI

针对于市场流动性指标，在这里选取市场上通用的MFI作为衡量市场流动性的标准，MFI即资金流量指数，根据通用理论MFI在大于80的时候，市场呈现超买现象，当MFI小于20的时候市场呈现超卖的现象；当超买发生的时候可以在一定程度上说明资金流入要大于流出，以此类推。

5.1.2.1 典型价格

需要计算市场流动性指标需要先计算典型价格，典型价格如下式所示：

$$\kappa_{i,t} = \frac{high_{i,t} + low_{i,t} + close_{i,t}}{3}$$

在其中 $\kappa_{i,t}$ 代表第 i 只股票第 t 天的典型价格，而 $high$, low 以及 $close$ 分别为个股的最高价格，最低价格以及收盘价格，通过求得三者的平均得到该个股的典型价格。

5.1.2.2 两种资金流量

由上得到典型价格，根据理论说明若今日的典型价格高于昨日的典型价格，则今日的资金流量为正资金流量；若今日的典型价格小于昨日的典型价格，则说明当日的资金流量为负资金流量；需要注意的是，这里的正负并不代表数值的正负而是代表资金流入的方向的正负，可以理解为若投资者将资金转化为股份，则该资金流入方向为正。因此由上在这里可以得到两个集合，为正负的资金流量，如下所示：

$$flow_{i,t} = \begin{cases} \sigma_{i,t} & \kappa_{i,t} > \kappa_{i,(t-1)} \\ \tau_{i,t} & \kappa_{i,t} < \kappa_{i,(t-1)} \end{cases}$$

而资金流量的计算方式则如下所示：

$$flow_{i,t} = \kappa_{i,t} \times volume_{i,t}$$

二者计算方式皆相同，因此在这里不做过多赘述。

5.1.2.3 资金流量比率及资金流量指数 MFI

由上段当中得到资金流量，在这里计算资金流量比率，资金流量比率即为正向资金流量以及负向资金流量的比值，在这里称作为 MFR ，如式子所示：

$$MFR_T = \frac{\sum_{t=1}^{14} \sigma_{i,t}}{\sum_{t=1}^{14} \tau_{i,t}}$$

其中 $\sigma_{i,t}$ 代表正资金流量，而 $\tau_{i,t}$ 则代表负资金流量，注意的是在这里 T 为 14 天为一个周期，由于市场普遍选取 14 天作为标准，因此在这里服从市场通用计算方式。

得到 MFR 后即可通过如下式算出资金流量指数 MFI 。

$$MFI_T = 100 - \frac{100}{1 + MFR_T}$$

由此为市场流动性指标的模型，在前文当中也有简短说明，当 MFI 大于 80

则市场呈现超买状态，当 MFI 小于 20 则说明市场呈现超卖的状态，可以很好的权衡市场资金的流入以及流出，由于该 T 为滚动窗口，也可以写作如下：

$$MFR_{i,t} = \frac{\sum_{t=n}^{14+n} \sigma_{i,t}}{\sum_{t=n}^{14+n} \tau_{i,t}}$$

$$MFI_{i,t} = 100 - \frac{100}{1 + MFR_{i,t}}$$

该指标是为结合价格和成交量的技术分析指标，用来识别市场的超买和超卖状态。由于同时结合两种因素，因此其能比较全面的反映市场的买卖力量，同时在该值超过 80 则认为存在超买状态，小于 20 存在超卖状态，可以在一定程度上判断市场的潜在反转点，如下图所示：

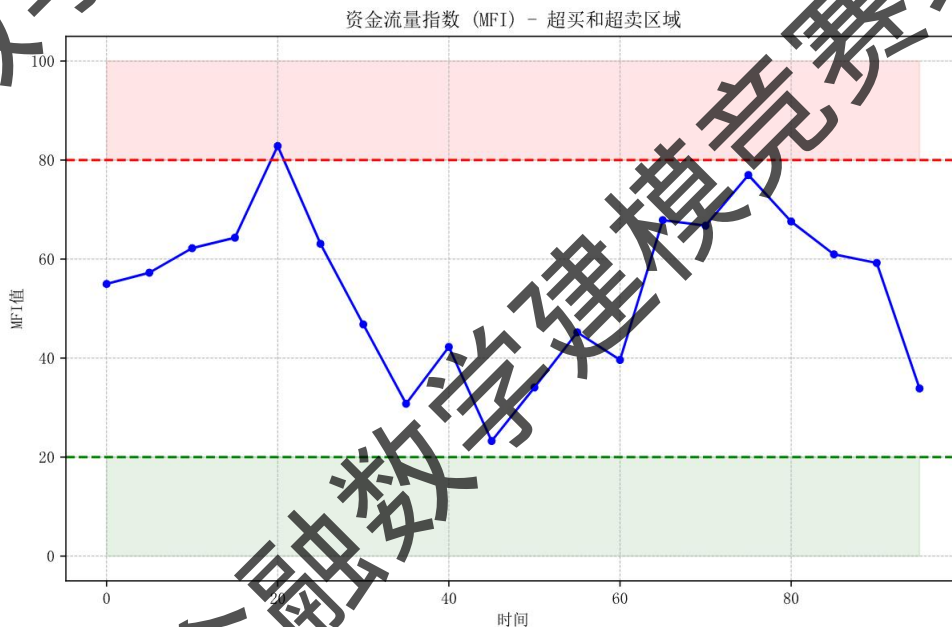


图 2：超卖超买状态划分

在上图中，绿色部分即为超卖部分 ($MFI < 20$)，红色部分即为超买部分 ($MFI > 80$)，需要注意的是，上图数据并非真实数据，为演示数据，展示 MFI 可以在一定程度上衡量超买或者超卖发生的情况，即能够让投资者更清晰的知道该股近期的大致情况，但是对于部分波动较大的市场条件下， MFI 可能会较长停留在其中区间，存在误导的可能性，且 MFI 的计算是基于过去的数据，存在一定的滞后性，以上即为 MFI 的优缺点分析。

5.1.3 市场情绪指标-投资者信心指数

投资者信心指数是一个重要的市场情绪指标，旨在反映投资者对市场未来走势的信心和预期。它不仅能帮助投资者了解当前的市场情绪，还能作为预测市场走向的有用工具。

5.1.3.1“0-1”变量

在这里设置“0-1”变量来确定该只个股是否上涨以及下降，由于计算方式相同，在这里仅解释针对是否该只个股是否上涨的“0-1”变量，如下所示：

$$\lambda_{i,t} = \begin{cases} 1 & \text{close}_{i,t} > \text{close}_{i,(t-1)} \\ 0 & \text{close}_{i,t} < \text{close}_{i,(t-1)} \end{cases}$$

当今日的收盘价大于昨日的收盘价可以在一定程度说明在今日该只个股存在上涨趋势，记为 1，反之为 0。由此在这里也可以得到两个集合，一个为在该日上涨的股票集合，以及在该日下降的股票的集合，如下所示：

$$V_{up,t} = \sum_i (\lambda_{i,t} \times \text{volume}_{i,t})$$

在这里下降的股票集合计算方式同理，但是需要重新设计下降的“0-1”变量，由于变量名称不易过于冗余，因此在这里不做过多赘述。

5.1.3.2 投资者信心指数

在上文当中得到在该日下降以及上升的股票集合，因此可以在这里计算出投资者的信心指数，投资者的信心与股票的上升下降直接相关，而投资者信心指数直接通过股票上升以及下降计算，因此可以在一定程度上衡量投资者的信心程度，如下式所示：

$$I_t = \frac{V_{up,t}}{V_{down,t}}$$

其中， $V_{up,t}$ 表示第 t 天所有上涨股票成交量总和， $V_{down,t}$ 表示第 t 天所有下跌股票成交量总和，得到的比值即为在该日的投资者信心，若该值大于 1 则说明投资者信心较强，若该值小于 1 则说明在该日投资者整体的信心较弱。

综上，可以得出投资者信心指数这一指标在预测中的优缺点。

·优点：

在短期和中长期中参考意义较大，能够反应短期市场波动的情绪变化，为观察长期市场趋势作参考。

·缺点：

投资者信心指数受外部干扰较大，若突发较大变化会导致投资者信心指数的准确性发生偏差。

5.1.4 数据预处理

对于题目当中给出的数据，首先发现在 2014 的成分股权重表当中并未给出各个的权重，因此将所有的成分股的权重全部均分，在这里设置如下：

$$\omega_i = \frac{1}{300}$$

同时题目当中给出的数据过于多样，在 MATLAB 处理数据的过程当中，分别设置两个 cell 数组，进行存储，通过并用 `strsplit` 的方法分割成分股数据。由于在前文当中详细介绍了数据的标准化过程，在这里不再过多赘述。

5.1.5 三种指标部分数据可视化结果

为了更直观的看出三个指标量化后的结果，在这里给出三个指标的可视化结果。

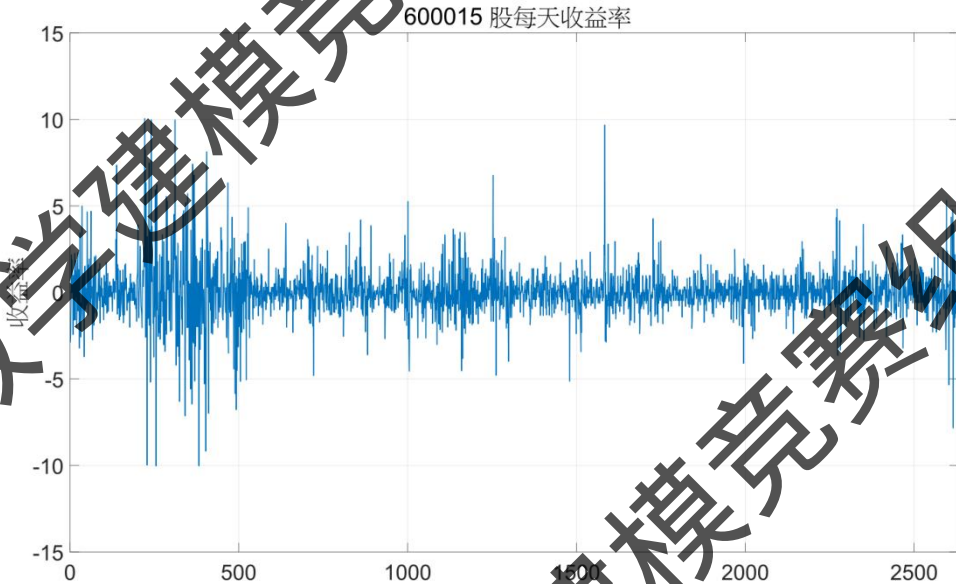


图 3: 某股日收益率示意图

由某股每日的收益率可视化可以看出该股在每日的收益率变化情况，从上图可以看出该股票收益率的正值和负值的个数相差不大，从该收益率可以看出该股票整体来说是呈现稳定趋势的，在 14-16 年左右的时间段该股波动较大。

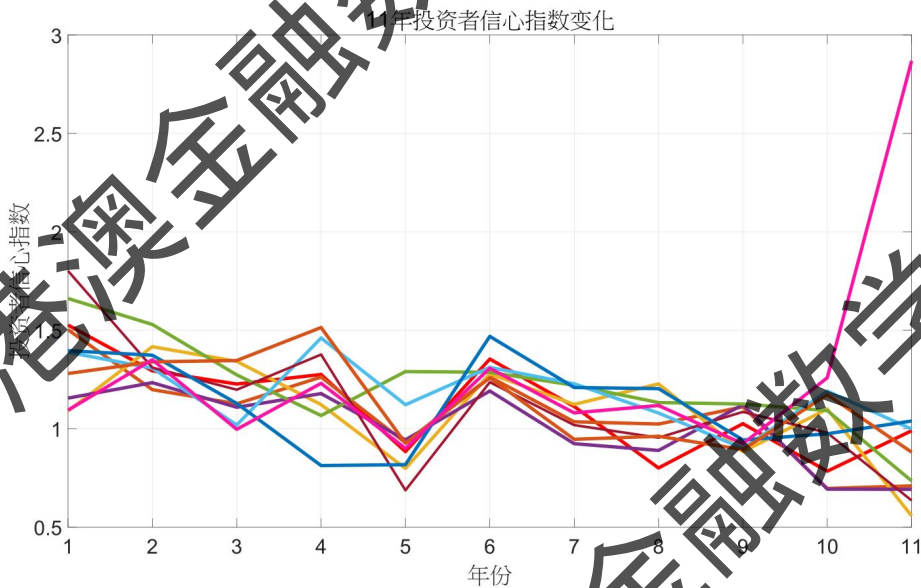


图 4: 投资者信心变化示意图

上图是随机选取了 10 个股票的投资者在 11 年间的信心指数变化的示意图，

该图可以看出大部分投资者的信心指数是呈现平稳趋势的，没有太大变化，部分存在投资者信心指数持续上升的情况。

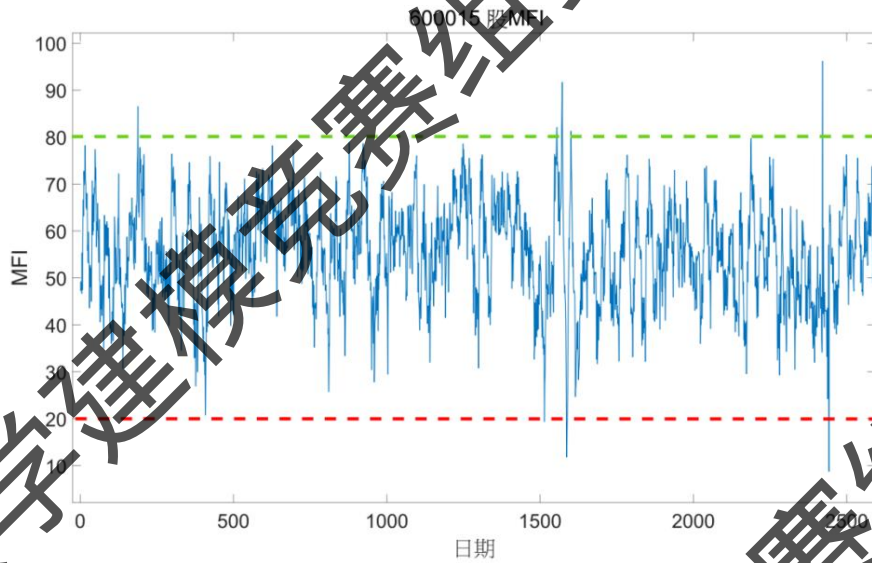


图 5: 个股 MFI 变化示意图

上图给出的是某个股票每日的 MFI 值，其中，若该股当日的值 >80 即超买， <20 即超卖，由该图可以看出个股的 MFI 是呈现平稳趋势的，但是存在部分超卖以及超买的状态，根据上文的模型可以说明其市场存在一定反弹的概率，从图中也可看出。

5.2 马尔可夫链结合多风险指标的风险预测模型

在该问当中本文将结合上文当中计算的三个风险指标，以及原题当中给予的部分指标，结合在后续的文章当中设立的两个风险控制指标，再结合多种指标对风险进行量化分析，将市场分为三种状态分别为高风险状态，中风险状态以及低风险状态，结合马尔可夫链理论，定义状态转移矩阵，即三种风险状态转换概率，通过该概率可以在一定程度上预测未来市场风险的波动，以对风险进行控制。

5.2.1 长短期波动率比例

在这里首先设置一个新的指标，在这里称为长短期波动率比例，目的是为了反映市场中短期和长期波动率的比例，若短期波动率显著高于长期，则说明市场正在经历短期波动或者不确定性，反之则说明市场稳定。具体公式如下所示：

$$L_{i,t} = \frac{\zeta_{i,t}}{\mu_{i,t}}$$

在上式当中 $L_{i,t}$ 为第 t 天的该只股票的长短期波动率，分子为短期波动率，分母为长期波动率，在这里在第一问当中给出了收益率的计算公式，因此在这里结合收益率计算其短期波动率以及长期波动率，如下所示：

$$s_{i,t} = \sqrt{\frac{1}{N-1} \sum_{t=1}^N (y_{i,t} - \bar{y}_i)^2}$$

在该式当中， N 为窗口内交易日数量，本文在这里将其设置为5天为一个窗口，通过计算五日内窗口收益率的标准差得到该只成分股的短期波动率；同理，计算长期波动率仅需要更改窗口即可，如下式所示：

$$M_{i,t} = \sqrt{\frac{1}{M-1} \sum_{t=1}^M (y_{i,t} - \bar{y}_i')^2}$$

在本文当中将长期的时间窗口设置为60天，即计算60天内收益率的标准差得到其该成分股长期的波动率，通过滚动窗口的方式得到每日股票的长短期波动率。在这里本文简单取2018年全年大盘长短期波动比率进行可视化，得到如下所示

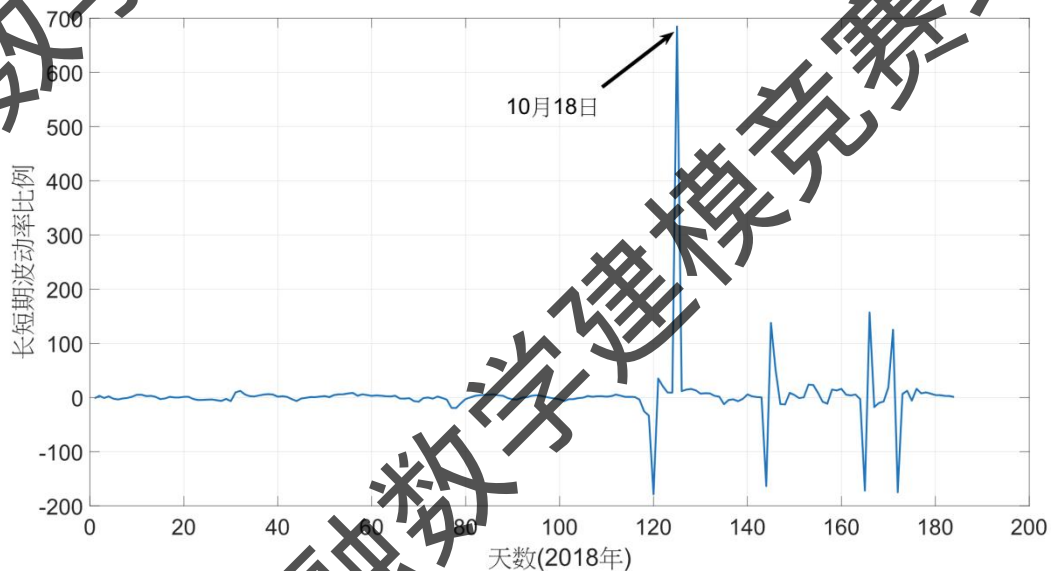


图6. 2018年沪深300长短期波动率比例

上图取为2018年全年大盘长短期波动比率示意图可以看到，在2018年10月18日股市的长短期波动比率较大，在计算当中长短期波动比率为短期波动比长期波动，因此说明短期内该段时间有剧烈变动，根据查询资料显示，18日市场在当天经历了极端波动，午后市场以及央行的3000亿操作导致指数狂飙一度上涨超5%，尾盘再次冲高回落闪崩收盘，十五分钟指数下跌超过3%，也可以在一定程度上看出该指标能很好的衡量短期内股票的波动效应。

5.2.2 高频交易影响比率

在前文当中定义了长短期波动比率，在这里定义高频交易影响比率，设置这个指标的主要原因是需要监控高频交易对市场波动性以及流动性的影响，在这里同样需要设置一个动态滑窗来计算每日的成交量是否高于或者低于一个基准期的平均水平，如下式所示：

$$H_{i,t} = \frac{V_{i,t}}{\sum_{t=n}^{n+29} V_{i,t} / 30} \quad n=1, \dots, N; i=1, \dots, 300$$

在这里通过将当日的成交量和基准期的平均交易量的比值作为高频交易影响比率，在通过上式计算之后，若该值大于一定程度，则说明当日的交易量高于平均水平，可能说明高频交易的活动较为频繁，可以在一定程度上暗示市场风险的增大，因此该值为负向指标。

5.2.3 特殊事件风险指数

在真实的交易市场当中，往往会出现公司并购，财报发布，政策变化以及自然灾害以及战争发生多种情况，会让市场呈现不稳定的一个状态，为此本文还设定一个特殊事件风险指数，旨在衡量特殊事件发生时对市场波动，价格变化的影响。由于根据先验经验可以得知，如 2024 年美国联邦财政主席鲍威尔在九月份发布的降息准则时，美股市场在事件发生的七天之后以及七天之前有剧烈变化，因此可以一定程度上认为在事件发生的时候，市场的变动往往存在一定的延迟，市场对事件的反映会存在一定的偏差，因此在这里选用价格来计算事件风险指数，具体如下式所示：

首先需要计算事件前价格的平均值，在这里取事件发生时前七天作为一个滑动窗口，计算价格平均值，如下所示：

$$p_{i,pre} = \frac{1}{7} \sum_{t=n}^{n+7} p_{i,t}$$

上式当中的 $p_{i,pre}$ 代表事件发生前七天价格的均值，在这里本文仍然取收盘价作为价格；以上，再计算事件发生后的价格标准差，同样取得七天作为滑动窗口，如下式所示：

$$p_{i,post} = \sqrt{\frac{1}{7-1} \sum_{t=n}^{n+7} (p_{i,t} - \bar{p}_{i,t})^2}$$

然后将二者求比值，得到如下式：

$$E_{i,t} = \frac{p_{i,pre}}{p_{i,post}}$$

需要注意的是，上式子当中的 $E_{i,t}$ 代表事件发生的七天之后其价格波动风险指数，该 t 从事件发生后的七天开始计算。

5.2.4 现存风控因素综述

在上一问当中建立了三大指标，在这里结合上一问当中建立的三大风险计量指标，分别为平均收益率，市场流动性-MFI，市场情绪指标-投资者信心指数，共六大指标对风险进行量化，建立如下式所示模型：

$$Risk_t = \gamma_1 Q_t + \gamma_2 MFI_t + \gamma_3 I_t + \gamma_4 H_t + \gamma_5 E_t + \gamma_6 L_t$$

需要注意的是，在这里 H_t, E_t, L_t 都为单下标，此为将上面求得成分股对应的指标，加权求和为沪深300的对应指标，因此在这里 $Risk$ 风险仅针对大盘进行分析，由于计算方法一样，在这里可通过查阅 5.1.1.1~5.1.1.2 中从成分股的平均收益率 $u_{i,t}$ 向沪深300平均收益率 Q_t 的转变过程，在这里不做过多赘述。

5.2.5 基于 Markov Chain 预测结合熵权法多指标模型构建

在前文当中建立了量化风险的模型，其中设计到六个指标对应的权重，为了得到客观的赋权，在这里考虑使用常用的客观赋权方法-熵权法，对六大指标进行赋权，在这里简单介绍熵权法步骤，如下所示：

(1) 去量纲化处理

在这里构建指标矩阵，如下所示：

$$X_{j,t} = [Q_t, MFI_t, I_t, H_t, E_t, L_t]$$

对其进行标准化得到如下标准化指标矩阵：

$$X_{j,t} = [x_{1,t}, x_{2,t}, x_{3,t}, x_{4,t}, x_{5,t}, x_{6,t}]$$

如上所有值，其上需要对其进行区分正向负向指标，比如当平均收益率越大，则说明投资者从中的获利较大，因此市场可能为牛市或者存在一段上升期，风险相较于收益率较少的部分较低，在这里总结为下表所示：

表 1：六大指标方向示意表

指标名称	指标符号	指标方向
平均收益率	Q_t	负向型指标
资金流量指数	MFI_t	负向型指标
投资者信心指数	I_t	负向型指标
高频交易影响比率	H_t	正向型指标
特殊事件风险指数	E_t	正向型指标
长短期波动率比例	L_t	正向型指标

对极小型指标和极大型指标本文用如下两种变化方式：

$$X'_{j,t} = \begin{cases} \frac{X_{j,t} - \min(X_j)}{\max(X_j) - \min(X_j)} & \text{正向} \\ \frac{\min(X_j) - X_{j,t}}{\max(X_j) - \min(X_j)} & \text{负向} \end{cases}$$

(2) 求指标在其他指标上的比重

在这里通过如下公式进行计算各指标其对应的比重，如下所示：

$$p_{j,t} = \frac{X'_{j,t}}{\sum_{j=1}^6 X'_{j,t}} \quad (j = 1, 2, \dots, 6)$$

(3) 计算各指标信息熵

$$U_j = -\ln(n)^{-1} \sum_{i=1}^n p_{j,t} \ln(p_{j,t}) \quad (\text{若 } p_{j,t} = 0, \text{ 则 } U_j = 0)$$

(4) 计算各指标权重

由上信息熵可以通过如下方式计算各个指标的权重，如下所示：

$$\gamma_j = \frac{1 - U_j}{6 - \sum U_j}$$

得到各个指标权重如下表所示：

表 2：六大指标权重表

指标名称	指标符号	指标对应权重
平均收益率	Q_t	0.2007
资金流量指数	MFI_t	0.2008
投资者信心指数	I_t	0.0098
高频交易影响比率	H_t	0.1977
特殊事件风险指数	E_t	0.1941
长短期波动率比例	L_t	0.1969

因此根据上文所述，在这里可以将风险控制模型为：

$$Risk_t = 0.2007Q_t + 0.2008MFI_t + 0.0098I_t + 0.1977H_t + 0.1941E_t + 0.1969L_t$$

由此本文将在下文使用 *Markov Chain* 方法对其进行预测。

5.2.5.1 Markov 状态转移概率矩阵及状态分布矩阵

在使用 *Markov* 时，涉及到两大矩阵，分别为状态转移概率矩阵以及状态分布矩阵，在这里根据题意，股票市场存在三种状态，分别为高风险状态，中风险状态以及低风险状态，如下示意图所示：

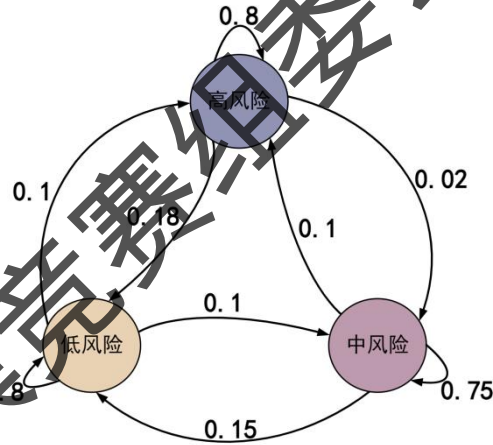


图 7: 风险状态转化示意图

在上文当中状态转移的概率为说明使用,并非为后文当中实际用到的状态转移概率矩阵,而状态转移概率矩阵即为从当前状态转化至下一个状态的变化概率,在这里引入一个思想即频次接近概率思想,当频次高的时候可以在一定程度上认为该状态转换的概率较大,因此本文在这里首先构建转移概率矩阵,如下所示:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

而如何求得该状态转移概率矩阵,本文用如下算法求得状态转移概率矩阵,在这里以沪深 300 为例:

Step1: 通过上述量化风险模型,计算全年沪深 300 每日风险大小,通过 $sort(risk(t))$ 函数对其进行排序,将 33% 分位数下的风险设置为低风险阈值,将 33%~67% 分位数区间的设置为中风险阈值,超过 67% 的设置高风险阈值。

Step2: 得到风险阈值后将风险状态值低中高分别设置为 1, 2, 3, 分别根据区间对其每日赋值风险状态即 $risk_state(t) = (low\ or\ mid\ or\ high)$ 。

Step3: 设置滑动窗口大小为 14 个开盘日 $window_size = 14$, 从开始计数的第一日算,在这里开启第一日为年初开盘后 30 日,由于预测模型需要吸收一定量数据才能进行预测,因此将年初开盘后的 30 日作为数据读取日,在后面计算得到状态转移分布矩阵也需要用到。

Step4: 通过统计该滑窗内状态变化的频数,如前一日状态为低风险,今日状态为高风险,则在频数矩阵的对应位置加 1,然后对每行进行归一化,得到 t 日向 $t+1$ 日状态转换的状态转移概率。

而对于初始状态转移分布矩阵,即用到年初开盘后 30 日的的数据,对该 30 日的的数据进行统计频次,得到在第 30 日其初始状态分布概率矩阵如下所示:

$$init_state = [a_1, a_2, a_3]$$

上矩阵当中的 a_1, a_2, a_3 分别代表当日是低风险的概率，当日是中风险的概率以及当日是高风险的概率，该值通过将前 30 日的出现低中高风险频次进行归一化得到。

5.2.5.2 Markov 预测原理

对于 Markov 的预测原理，在这里由 Markov Chain 组成，即后一日的状态肯定与当日有关，与前一日也有些许关系，但是强关系肯定为离其最近一日的价格，在 5.2.5.1 当中通过模型计算可以得到转移概率矩阵 A 以及状态转移分布矩阵 $init_state$ ，因此在这里 Markov Chain 的递进关系如下示意图所示：

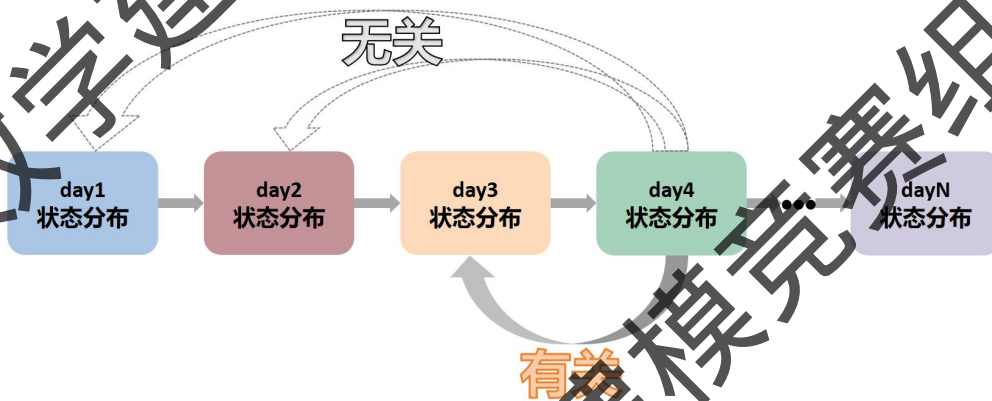


图 8：马尔可夫链示意图

因此由上图可以推理出如下公式，由于明日的状态分布矩阵与今日有关，因此可通过 $init_state$ 求出第二日的状态转移分布矩阵，如下所示：

$$\begin{aligned} state(1) &= init_state \\ state(2) &= init_state \times A(1) \\ state(3) &= state(2) \times A(2) \end{aligned}$$

其为一个链式结构，在这里状态转移概率矩阵是由前文提到的滑窗统计 14 日内的频数后归一化得到的，其经过每天将会变化，因此根据一阶 Markov Chain 理论，得到模型如下：

$$state(a_1, a_2, \dots, a_N) = state(a_1) \sum_{n=2}^N state(a_n | a_{n-1})$$

这里由于高阶 Markov Chain 过于复杂，股票市场可能涉及到非常高阶的马尔科夫链，由于过于复杂在这里仅以一阶作为研究对象，即前一日对其的影响最大，服从上文流程图所示。

5.2.5.6 以沪深 300 为例的 Markov 多指标模型预测结果展示

在上文当中，首先通过建立三大指标分别为高频交易影响比率，特殊事件风险指数，长短期波动率频率三者再结合问题一当中涉及到的三大衡量市场波动变化的指标分别为平均收益率，资金流量指数以及投资者信心指数；通过熵权法求得在客观情况下各个指标对应的权重；之后结合 *Markov* 预测的方法设立三大状态高低中风险，再通过收集预测前一个月的所有数据得到状态转移概率矩阵以及状态分布矩阵，通过以上两矩阵推得后面 14 天的风险状态，并与真实的风险状态进行对比，使用抽样调查得到本问的预测结果，如下所示：

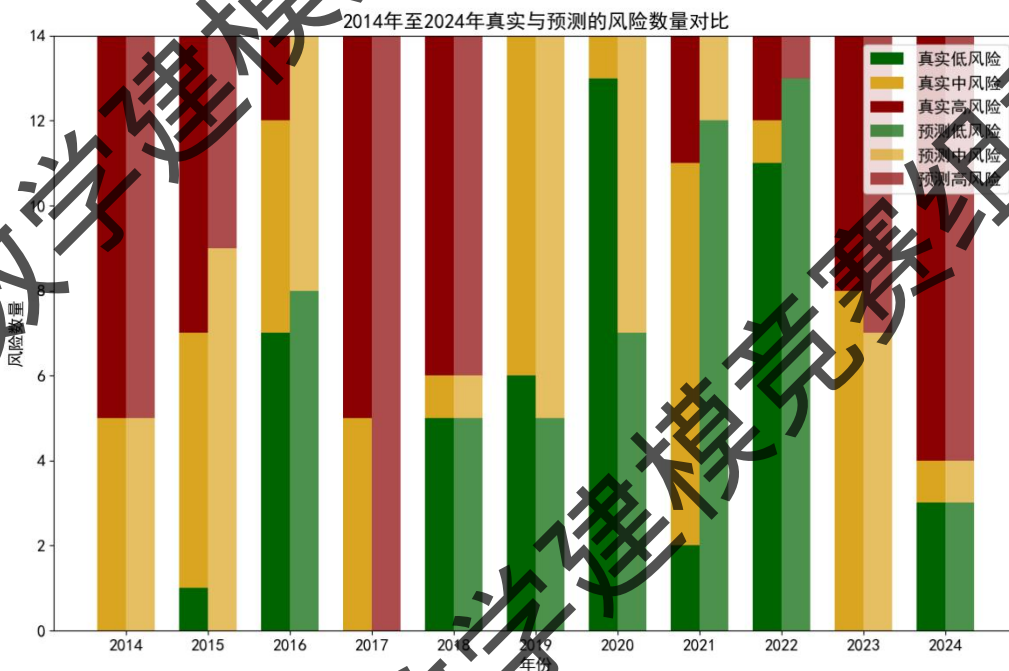


图 9 *Markov* 抽样调查预测结果

• 注：在上图当中收集的数据为每年开始交易的前 30 个交易日数据，预测的为 30 个交易日之后的 14 个交易日的风险；

从上图当中也可以看出，模型整体预测效果较好，在 2014、2015、2016、2018、2019、2022、2023、2024 年的效果都表现的较好，只有极少数年份会出现偏差，因此模型在一定程度上能准确的预测市场当中未来的风险，模型整体效能较好。具体抽样调查结果如下表所示：

表 3：风险预测模型抽样调查结果

年份	真实			预期		
	低风险	中风险	高风险	低风险	中风险	高风险
2014	0	5	9	0	5	9
...
2023	0	8	6	0	7	7
2024	3	1	10	3	1	10

5.3 结合 Markov 的风险控制动态交易调整体系(针对混合型权益类)

针对第三问，由于在前文当中已经通过六大指标分别为平均收益率，MFI，投资者信心指数，高频交易影响比率，特殊事件风险指数以及长短期波动比率构建了风险预测模型，通过滑动窗口以及标签等方法对每日的风险进行排序且选出高风险中风险以及低风险的风险，通过阈值分别设置标签 1, 2, 3 为低中高风险，通过 Markov 原理设置状态分布矩阵以及状态转移概率矩阵，通过该原理推导过程可向后进行预测，推得短期时间内市场的风险情况。而对于该问，在这里设计一个基于风险发生前的风控动态交易体系。

5.3.1 回撤

根据题意，在该问当中需要计算最大回撤，需要让最大回撤可以控制在 0.7 的事前风控体系，设置 0.7 为回撤风控线，而回撤的计算如下式所示：

$$Drawdown_t = \frac{\max(high_t) - cur_t}{\max(high_t)} \leq 0.7$$

在上式当中，当前回撤的计算为历史最高值减去当前值再比上历史最高值，需要始终保持该风控线，让其中所有的交易的回撤不会超过 70%。

5.3.2 交易信号-止盈止损线设置

在这里由于需要建立动态交易，因此在这里考虑设置止盈止损线，在内部有仓位的时候，若持续上涨超出止盈线，不考虑后续的盈利直接触发止盈，防止风险发生，同时结合第二问当中的 Markov 预测模型，预测下一个交易日的风险选择是否持仓加仓或者减仓，止损同理，如下图所示：

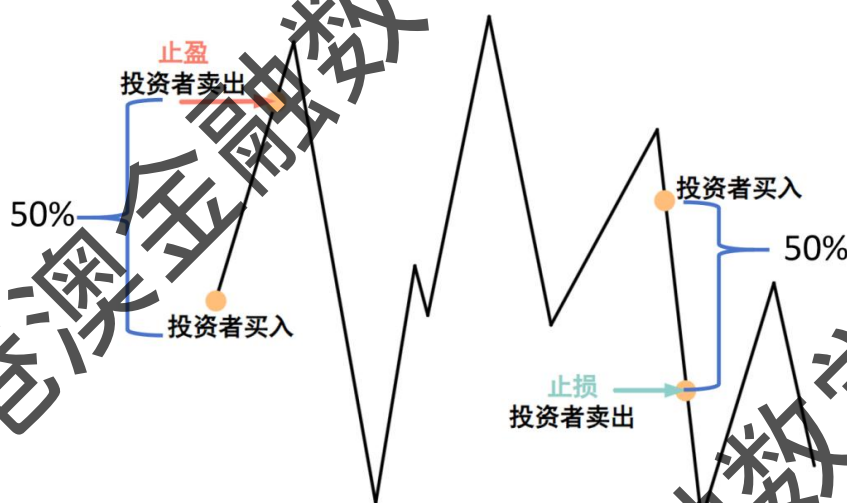


图 10：止损止盈示意图

在这里初始化止盈线和止损线都为 0.5，由于 0.5 小于回撤风控线 0.7 且该值也能保证一定的收益，在真正的实盘预测当中通过最后风险预测，及时动态变化止盈线以及止损线的变化，在这里建立如下所示：

$$SL = TP = 0.5$$

在上式当中的 SL 代表止损(*Stop Loss*), TP 则代表止盈(*Take Profit*),若该止损止盈线过于小则会导致可能无法承担股市风险中带来的收益,过于大则可能导致无法承受股票大风险波动带来的影响,比如诸多黑天鹅事件。

5.3.3 风险预测的仓位调控算法

Step1: 首先基于任务二当中建立的 *Markov* 预测模型,通过短期预测(在这里为 14 天),得到短期内每日的市场风险概率情况,选择该日概率最大的为当日的预测风险状态 *risk state*。

Step2(开仓选择): 在这里决定是否开仓有两个因素,其一为当日状态是否为中低风险,且下一日的风险同样需要为中低风险才允许开仓,如下式所示:

$$Open_t = \begin{cases} 1 & \text{if } state_{t+1}, state_t \in \{1,2\} \\ 0 & \text{else} \end{cases}$$

将开仓价格记为历史最高价格,后续若新价格高于开仓价格则即新价格为历史最高价格。同样将该价格记为历史最低价格,后续若新价格低于当前价格则替换为新价格为历史最低价格。

Step3(空仓信号): 在这里触发空仓将会有三个原因,分别为止损和止盈即前文当中涉及的初始止损止盈线,当上涨幅度超过 50%则即刻触发止盈,全部空仓;当下跌幅度超过 50%则即刻触发止损同样全部空仓;此外还有一种即回撤幅度超过 0.7, 即刻空仓,但是在止损和止盈线的前提下该种情况不会发生。

Step4(止损止盈调整): 在前文当中记录短期内,即 14 天内所有触发止损线止盈线次数,若触发止损线次数过多则在下一个 14 天当中将该止损线上移即 0.4 甚至 0.3; 反之亦然,若该周期触发止盈线过多则考虑上移,调整至 0.7 为最大止盈线。

5.3.4 以 2015 年沪深 300 为例的动态交易调整结果

在第三问当中涉及了第二问当中的风险预测模型,再通过算法描述当中分别对开仓选择,空仓信号以及止损止盈调整等步骤通过不断的进市和退市达到从中保本赚钱的目的,并且为了保本,通过不断的调整止损止盈线以及保证回撤不会超过题目当中的限制 0.7, 将 2015 年 6 月份起始至今的日期作为研究对象,使用模型进行风险交易寻找交易信号,得到如下结果可视化:

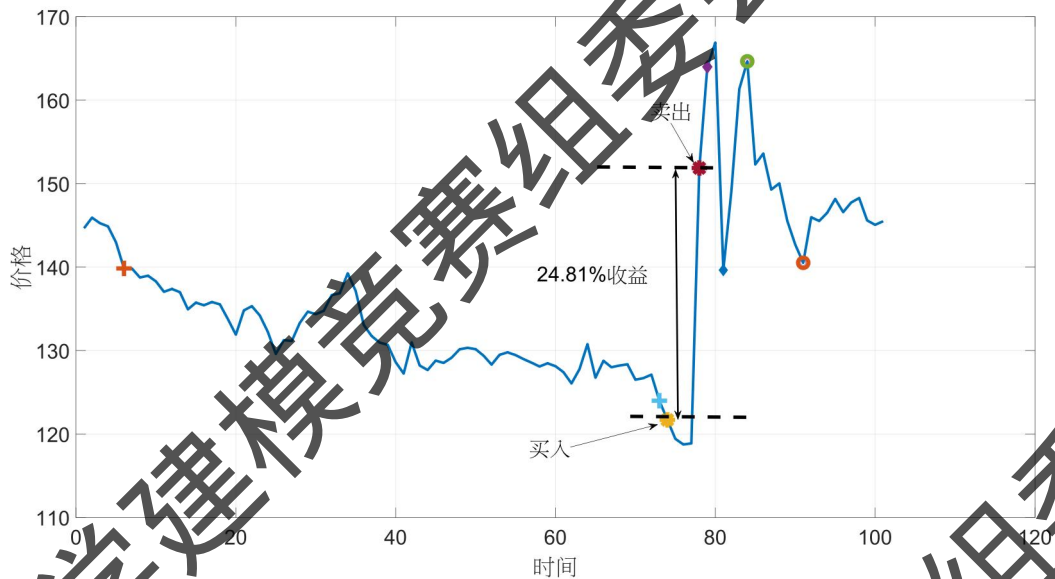


图 11: 以 24 年 6 月起始至今模型交易信号

注: 上图当中的两种相同符号随着日期的变化一前一后分别为全仓和空仓

可以从上图当中看出由于止损止盈线的设置, 其可能承担一定的风险即可能会触发止损如上图当中十字符号当中演示; 在“*”型的交易信号当中, 其通过止损线, 在中间获得 24.81% 的收益, 由此也可以看出由于预测模型的存在, 模型能捕捉到市场当中的波动, 即使会存在损失但是由于止损的原因不会导致亏损过多, 而在止盈的过程当中其可以防止一定程度下的市场反弹, 因此模型具有较高可信度, 在这里给出上图当中涉及到的具体开仓清仓和具体收益率及日期, 如下表所示:

表 4: 回测控制 0.7 的事前风控体系最终结果

以 2024/06/01 至 2024/10/17 模型具体交易过程					
进市次数	收益率	开仓时间	清仓时间	开始价格	结束价格
1	-11.34%	2024/06/11	2024/09/12	139.8380	123.9850
2	24.81%	2024/09/13	2024/09/23	121.6842	151.8681
3	-14.85%	2024/09/24	2024/09/26	163.9917	139.6327
4	-2.35%	2024/10/08	2024/10/17	148.9861	145.4873

以上回测当中也可看出, 其最大回撤为 0.15, 因此可以在一定程度上认为本段当中的事前风控体系具有一定可信度以及可行性, 同时市场波动较大的情况下也能进行部分套利交易。

5.3.5 模型优化方向

在前文当中设置了动态的止损止盈线, 本文仍然考虑到模型的优化方向, 通常防止回撤超过 0.7 可以使用对冲, 在市场预测为高风险时, 增加期货空头头寸

或者购入更多的看跌期权；在市场预测明日为低风险的时候不使用对冲方案，这可以降低组合的总体波动性以达到更稳定的收益。

此外对于仓位调整，购入沪深 300 的时候可以动态调整投资组合的持仓比例，比如如下策略：

①未来低风险：可以维持较高仓位，捕捉市场上行收益；

②未来中风险：适当降低仓位如 30%~50%，控制下行风险，同时保证盈利空间；

③未来高风险：进一步降低仓位至 10%~29%，甚至清仓等待进场时间，而非直接通过止损和止盈来全仓和清仓，这样可以比较好的避免市场下跌带来的较大回撤；同时可加大权重在低风险资产比如说债券(国债等...)的配置，从而提高整体组合的稳定性。

5.4 多维度合理收益预期模型

在第四问当中需要设定合理的收益预期，散户比较容易在波动市场当中由于对自身的收益预期不准确，导致亏损，因此在这里本文考虑结合多维度对合理收益预期进行设定，首先题目当中需要以 10 年期长期国债收益率为标准，查询可得知其平均收益率为 3.281%，以此为无风险收益，再结合其他维度进行模型的构建，具体见下文所述。

5.4.1 基准收益率

根据题目要求，需要以 2014 年至今的 A 股沪深 300 个股的历史数据作为投资标的，且参考 10 年期长期国债收益率为标准，国债通常为无风险情况收益，在这里可以认为保底收益，通过在行情中心查询到过去十年期长期国债平均收益率为 3.281%，因此有如下式：

$$R_f = 3.281$$

这里的 R_f 代表无风险收益率即 10 年期国债的年均收益率，通过查询得出。

5.4.2 溢价风险量化

在第二问当中对风险进行了量化，选取了六个指标，使用 Markov 预测模型对未来风险进行预测，在该问当中需要设定投资者投资股权市场的合理收益预期，因此需要考虑到溢价的因素，溢价在很大一部分的情况下影响了散户对收益的预期，而溢价风险又分为多种，下文一一对其进行建模，将第二问的模型优化为多维度收益预期模型。

5.4.2.1 市场风险溢价

市场风险溢价即市场整体的收益率与无风险收益率的差值，代表超出预期的部分，因此市场风险溢价的计算公式为：

$$R_m = R_t - R_f$$

在上式当中， R_t 为沪深 300 的年化平均收益率，通过计算将年化收益率减去基准收益率得到市场的溢价风险 R_m 。

5.4.2.2 流动性溢价

流动性在第一问当中选取了现金流量指数作为流动性指标，在本问计算流动性溢价的时候仍然选取 MFI 作为标准，当 MFI 大于 80 的时候认为市场为超卖状态，小于 20 则认为市场为超买状态，有一定的反弹风险，在这里不过多赘述，直接引用 MFI 最后计算的公式，如下所示：

$$\begin{cases} MFR_t = \frac{\sum_{t=n}^{14+n} \sigma_t}{\sum_{t=n}^{14+n} \tau_t} \\ MFI_t = 100 - \frac{100}{1 + MFR_t} \end{cases}$$

注意的是，这里所有的下标都为单下标，为沪深 300 的数据，通过求得每股的收益率加权求和得到沪深 300 大盘收益率再计算得到以上流动性溢价。

5.4.2.3 市场情绪溢价

市场情绪溢价则是沿用在第一问当中使用的投资者信心指数，该指数由上文的使用当中也可看出其能很好的权衡市场中投资者的信心，可以较好的复刻市场当中出现的如“黑天鹅”或者“牛市熊市”等大波动的出现，如投资者信心指数很好的刻画了前段时间由于美联储降息且外资回流导致中国 A 股大涨的现象，从图像当中的后部分也可说明，如下所示：

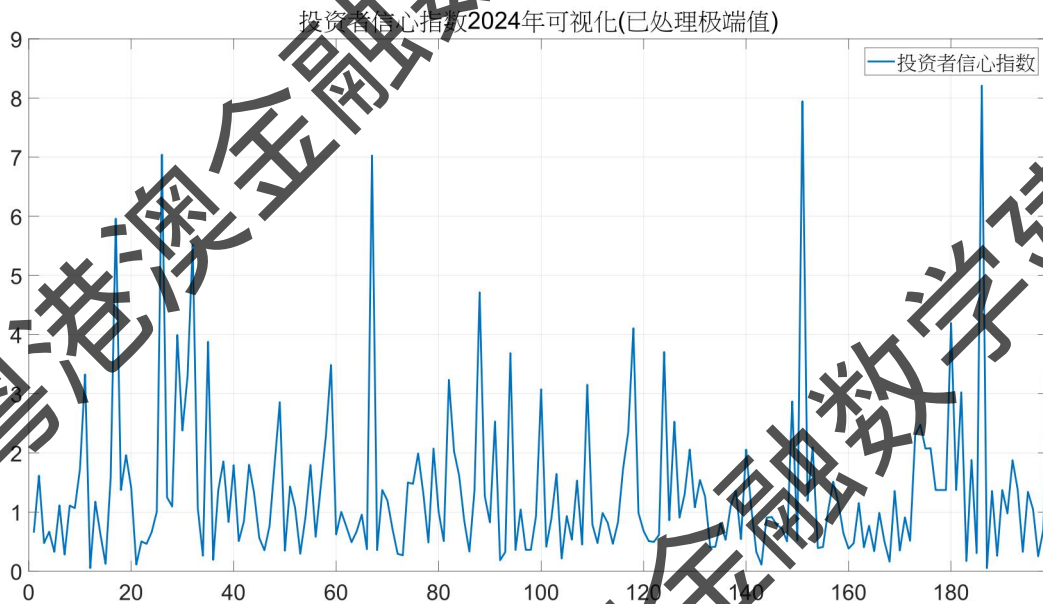


图 12: 投资者信心指数 2024 年可视化

因此在这里直接给出计算投资者信心指数的公式，如下所示：

$$I_t = \frac{V_{up,t}}{V_{down,t}}$$

在上式当中通过用所有上涨的股票集合除以下跌的股票集合得到沪深 300 第 t 日的投资者信心指数，若下跌的股票为 0，则设为 1。

5.4.2.4 事件风险溢价

这里的事件风险溢价使用第二问当中的事件风险指数进行计算，事件的影响可以很大程度对市场的波动起到一定推波助澜的作用，在这里不做过多赘述，直接引用前文对事件风险溢价的描述，如下式所示：

$$\begin{cases} p_{pre} = \frac{1}{7} \sum_{t=n}^{n-7} p_t \\ p_{post} = \sqrt{\frac{1}{7-1} \sum_{t=n}^{n-7} (p_t - \bar{p}_t)^2} \\ E_t = \frac{p_{pre}}{p_{post}} \end{cases}$$

上式子当中的 E 为特殊事件风险指数，通过将事件后七天的价格波动标准差与事件前七天的平均价格作比，能在一定程度上衡量风险的发生。

5.4.3 多维度收益预期模型的建立

结合上文当中的四大风险溢价，模型结合如下所示：

$$Profit_t = R_m + w_1 MFI_t + w_2 I_t + w_3 E_t$$

上式当中的 $Profit$ 代表预期收益率，在受到市场波动影响的情况下得到的预期收益，注意的是，其中一个参数为市场风险溢价，在里面包含了十年期国债平均收益率，因此仍然是以国债为标的求得预期收益率。

5.4.3.1 熵权法求解权重

在上文的多因子模型当中结合了六个指标，而在该问当中需要结合四个指标进行预期收益率的求解，分别为市场风险溢价，事件风险溢价，市场情绪溢价以及流动性溢价，总结如下所示：

表 5：风险溢价指标

指标名称	指标符号	指标方向
市场风险溢价	R_m	正向型指标
事件风险溢价	E_t	正向型指标
市场情绪溢价	I_t	正向型指标
流动性溢价	MFI	正向型指标

需要注意的是，这里几个指标与上文方向不同的原因在该问当中需要求得为收益率，而非风险，相较于风险而言，市场风险溢价越高说明风险越高，且收益可能会更高，当然可能需要承担较大亏损，对于收益而言以上皆为正向指标，但是存在市场反弹的可能性。

$$R_{j,t} = \frac{X_{j,t} - \min(X_j)}{\max(X_j) - \min(X_j)}$$

同时用该式对正向指标进行转换，将其转化为适应熵权法的指标。由于熵权法过程相同，在这里仅列出公式，不做过多赘述，如下所示：

$$\begin{cases} p_{j,t} = \frac{X'_{j,t}}{\sum_{j=1}^3 X'_{j,t}} & (j=1,2,3;) \\ U_j = -\ln(n)^{-1} \sum_{t=1}^n p_{j,t} \ln(p_{j,t}) & (\text{若 } p_{j,t} = 0, \text{ 则 } U_j = 0) \\ w_j = \frac{1 - U_j}{3 - \sum U_j} \end{cases}$$

最后得到四大风险溢价指标对应权重如下所示：

表 6：四大风险溢价指标权重

指标名称	指标符号	对应权重
市场风险溢价	R_m	无
事件风险溢价	E_t	0.1798
市场情绪溢价	I_t	0.6883
流动性溢价	MFI_t	0.1319

由此根据权重，预期收益率模型可以写成如下式：

$$Prof_{i,t} = R_m + 0.1798MFI_t + 0.6883I_t + 0.1319E_t$$

由此对预期收益率的模型已经完成，可以使用后文的算法对其进行预期收益率测试。

5.4.4 基于蒙特卡洛算法结合收益预期模型

在该问当中需要使用蒙特卡罗模拟算法，通过历年数据求得其四大指标的波动区间，后文当中会具体介绍；通过该区间在蒙特卡洛循环当中随机取指标大小，最后得到每次的收益率分布，即考虑风险情况下的预期收益率大小。

5.4.4.1 基础参数设置

在这里需要对蒙特卡洛模拟当中的基础参数进行一些设置，在下文当中本文将进行 10000 次蒙特卡洛模拟；对于上文当中的四个指标，取其区间为如下所示：

$$R_m = 3.281$$

$$E_t \in [\bar{E} - 3\sigma, \bar{E} + 3\sigma]$$

$$I_t \in [0, \bar{I} + 3\sigma]$$

$$MFI_t \in [\overline{MFI} - 3\sigma, \overline{MFI} + 3\sigma]$$

在这里使用 **3sigma** 原则剔除极端值，取除极端值之外的指标作为其的波动区间，通过模拟 10000 次试验得到最后预期收益率。需要注意的是在上式当中的皆为分析沪深 300 个股收益率加权求和得到大盘收益率，以个股为标的进行分析。

5.4.4.2 算法具体流程

Step1: 首先根据上文当中设定的参数，在每次的蒙特卡洛循环当中，在区间内随机取值，将值代入模型总体计算最后预期收益率的公式；

Step2: 记录该次所得预期收益率，对该模型循环 10000 次，取 10000 次不同的参数组合，得到每次的预期收益率；

Step3: 对 1000 次最后的预期收益率分布进行可视化，并分析与十年期国债的平均收益率差距，展示收益率分布情况。

5.4.5 结果展示及可视化

在该问当中通过蒙特卡洛循环模拟四大参数分别为市场风险溢价，事件风险溢价，市场情绪溢价，流动性溢价，结合指标使用熵权法求解得到每个指标其对于的权重，通过 **3sigma** 原则规范四大参数其范围，注意的是，对于市场情绪溢价，其在 $[0,1]$ 区间代表沪深 300 在下跌，在 $(1,+\infty]$ 区间的时候其为沪深 300 上涨，市场整体情绪较高，因此在处理该指数时，将其值全部映射至 $[-1,1]$ 区间内，最后通过蒙特卡洛随机模拟参数变化，每次模拟得到不同参数组合，最后对其 10000 次蒙特卡洛模拟下的预期收益率进行可视化得到如下直方图：

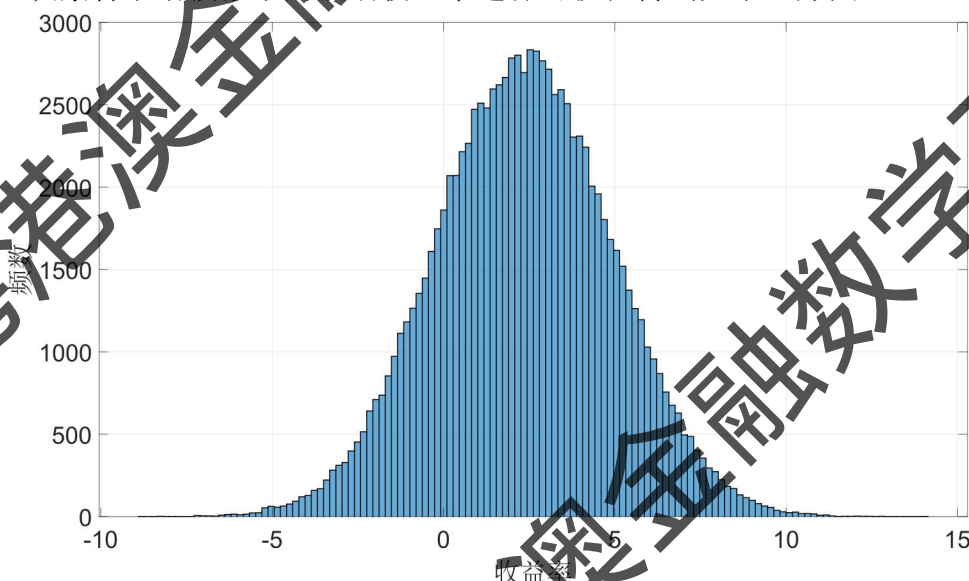


图 13: 合理收益预期可视化结果

从图中可以看出,其收益率基本维持在[0,5]区间内,在这里取均值,对于收益预期可以认为设定在 3%上下为较为合理的收益率预期,在承担一定风险的情况下能有较为合理的收益为本题的目标所在。

六、模型的评价与推广

6.1 模型优点

1. 该模型考虑了多种因素并且将其进行组合,鲁棒性较强,可以适应各种不同的交易体制。
2. 模型整体处理数据,运行效率较高。

6.2 模型缺点

1. 模型并未加入仓位调整等其他方面的交易策略,没有完全利用到预测风控的能力。

6.3 模型推广

模型适用性较强,可以用于不同市场,如美股或者虚拟货币市场,可以应对多种可能的情况,能在极端事件出现的情况下仍然具有一定的风险量化能力,可以使用该模型进行套利或对市场的分析以及量化。且该模型未来前景较好,可以从多种不同方面优化该模型,模型提升性能较大。

七、参考文献

[1] 张宗新,黄梓健.资本市场系统性风险监测及风险跨市场溢出研究——基于金融压力指数视角[J].证券市场导报,2024,(07):57-67+79.

[2] 巢文,钱晓涛.基于隐马尔可夫模型的台风风险评估与巨灾债券定价研究[J]. 绵阳师范学院学报,2024,43(09):7-14+37.DOI:10.16276/j.cnki.cn51-1670/g.2024.09.002.

[3] 姚宏亮,江永生,杨静,等.基于情绪向量的隐半马尔可夫模型股市拐点预测方法[J].合肥工业大学学报(自然科学版),2024,47(10):1335-1340.

[4] 李琳.基于高斯过程回归和蒙特卡洛模拟的期权定价算法研究[D].贵州财经大学,2023.DOI:10.27731/d.cnki.ggzcj.2023.000134.

实测报告

一、测试目的

在该题当中需要对问题而当中的模型进行实测检验，问题二为基于多因素及马尔科夫链的风险预测模型，通过滑窗收集一部分数据，在这里收集 2024 年 10 月 1 日至 2024 年 10 月 31 日期间所有开盘日作为数据基础，通过统计低中高风险数据出现的频率分布，以频率接近概率的理论使用马尔科夫链对后续风险进行预测，通过正文当中可以清楚得知在这里文章分为三个风险分别为低中高风险，通过马尔科夫链预测未来风险等级，可以在一定程度上防止风险过高带来的损失过大，在这里测试的主要目的通过真实风险以及预测风险进行对比，测出模型是否能在一定程度上预测市场未来的风险等级，帮助投资者以及企业更好的调整投资方案。

二、测试环境

代码环境: *Matlab2023*

运行环境: *Windows11*

三、测试内容

在该段当中基于论文中提出的基于多因素结合马尔科夫链的预测模型，在这里给出如下汇总式：

首先是在第二问当中提炼的六大指标以及其对应权重，在这里如下所示：

$$\left\{ \begin{array}{l} Q_t = \sum_{i=1}^n \omega_i A_i \\ MFI_t = 100 \frac{100}{1 + MFR_T} \\ V_t = \frac{V_{up,t}}{V_{down,t}} \\ H_t = \frac{V_t}{\frac{\sum_{t=n}^{n+29} V_t}{30}} \quad n = 1, \dots, N; i = 1, \dots, 300 \\ E_t = \frac{P_{pre}}{P_{post}} \\ L_t = \frac{\zeta_t}{\mu_t} \end{array} \right.$$

通过 *MATLAB* 实现熵权法求解各指标权重，得到如下表所示：

表 1: 六大指标权重表

指标名称	指标符号	指标对应权重
平均收益率	Q_t	0.2007
资金流量指数	MFI_t	0.2008
投资者信心指数	I_t	0.0098
高频交易影响比率	H_t	0.1977
特殊事件风险指数	E_t	0.1941
长短期波动率比例	L_t	0.1969

并且将其组合为风险控制模型，如下式所示：

$$Risk_t = 0.2007Q_t + 0.2008MFI_t + 0.0098I_t + 0.1977H_t + 0.1941E_t + 0.1969L_t$$

该模型可以测试第 t 天的风险程度，对于本研究，风险程度一共划分为三类，分别为低中高风险，而对于实际风险划分的值在这里以分位数的方式反复调整取得预测结果较好的风险划分方式，在这里通过收集全年开盘日的所有数据，将实盘数据按风险从低到高的方式进行排序，首先通过 33%，67% 分位数划分大体低中高风险，然后通过对比真实风险，通过使用机器学习中梯度下降的思想将误差降到最低，微调划分最后得到具体低中高划分为：

表 2: 低中高风险划分

风险程度	风险阈值
低风险	$0 < Risk_t < 25.01$
中风险	$25.01 \leq Risk_t \leq 33.2$
高风险	$Risk_t > 33.2$

得到风险阈值划分后，便结合一阶马尔可夫预测模型，在这里将预测前 30 个开盘日作为数据收集阶段，在这里需要预测 2024 年 11 月 1 日至 2024 年 11 月 15 日的实盘数据，因此需要收集的数据即从 2024 年 10 月 1 日始至 10 月 31 日之间所有开盘日的数据，求得风险状态转移概率矩阵以及初始风险状态分布矩阵，以频次接近概率的思想，频次高的时候可以在一定程度上认为其状态转移的概率较大，因此通过收集过往 30 日数据，得到初始风险概率分布如下所示：

$$init_state = [0.4000, 0.2667, 0.3333]$$

可以看到在过去的 30 天内，按上述风险划分得到的低风险概率为 40%，中风险为 26.67%，高风险为 33.33%，同时在这里可以得到风险状态转移概率矩阵，即低风险向中风险或者高风险转化的概率，同理中风险和高风险也如此，因此得到行列皆为 3 的矩阵如下所示：

$$A = \begin{bmatrix} 0.6 & 0.4 & 0 \\ 0.25 & 0.25 & 0.5 \\ 0 & 0.4 & 0.6 \end{bmatrix}$$

在这里的第一行为低风险向低风险，中风险，高风险转化的概率，同理后两行，因此可以从风险状态转移矩阵简单进行分析看出，低风险向低风险转移的概率较高，高风险向高风险状态转移的概率较高，市场有一定的连续性，比较符合客观规律。得到以上两个矩阵后便可通过马尔可夫的链式结构，得到如下推导式：

$$\begin{aligned} state(1) &= init_state \\ state(2) &= init_state \times A(1) \\ &\dots \\ state(11) &= state(10) \times A(10) \end{aligned}$$

即第二天的市场风险一定与第一天的市场风险有关，以此类推，通过初始状态分布矩阵 $init_state$ 推导出第二日的状态分布矩阵，最后选取概率最大的为该日的风险等级。由此根据模型便可得到每日的风险等级，在测试结果分析段会详细展示结果以及可视化。

四、测试结果分析

特别需要注意的是，在实测数据当中给出的数据，其存在与 2024 沪深 300 个股权重表不对应的情况，在这里首先将其进行处理，处理得到如下表所示：

表 3：沪深 300 成分股对应表

沪深 300 剔除	沪深 300 剔除	沪深 300 新加入	沪深 300 新加入
000069	300763	000807	600027
000877	600606	001965	600161
002202	600754	300418	600415
601155	603486	300442	603296
601615	688065	300832	688009
603290	688561	600026	688082

上表所表示的为沪深 300 中 12 支年初存在的成分股替换为其它如今存在的成分股，考虑到成分股发生了替换，故在此将所有已经存在的成分股按原权重分配，新更新的成分股按其总权重剩余部分进行平均分，以处理计算各个指标当中需要用到各成分股权重的问题。

同时给出预测结果并进行分析，如下所示：

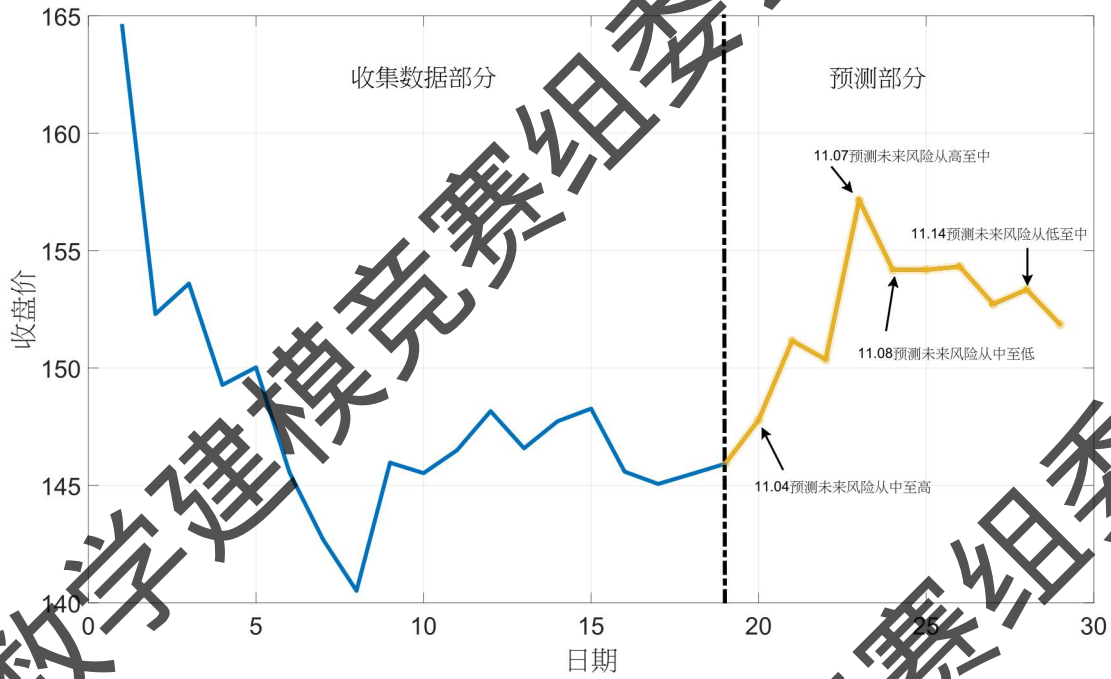


图 1: 11月1日至11.15 沪深 300 预测可视化

上图所展示的为根据中蓝色收集数据部分为10月1日至10月31日的实际数据，黄色预测部分为根据所建模型预测得到的预测数据，从该可视化可以清晰的看出预测风险的变化，在11月4日时预测未来市场风险从中等级转变成高等级，从图上也可清晰看到，该收盘价的斜率变大，说明市场变化的更快，同理还可以看到在11月8日的预测当中，预测市场未来的风险从中风险转化为低风险，确实在后续的市场变化当中斜率整体较为平稳，没有出现特别大的波动，因此上图可以在一定程度上说明模型的预测准确度较为有效，能较直观且综合的预测市场未来的风险走向。

表 4: 沪深 300 成分股对应表

日期(2024年)	低风险概率	中风险概率	高风险概率	预测风险	真实风险
11.01	0.2727	0.5500	0.1773	中	高
11.04	0.2700	0.3500	0.3800	高	高
11.05	0.2677	0.3000	0.4333	高	中
11.06	0.2625	0.3500	0.3875	高	中
11.07	0.2571	0.4167	0.3262	中	低
11.08	0.3500	0.3500	0.3000	低	低
11.11	0.3500	0.3500	0.3000	低	低
11.12	0.3500	0.3500	0.3000	低	低
11.13	0.3500	0.3500	0.3000	低	中
11.14	0.3000	0.4000	0.3000	中	中
11.15	0.2600	0.4500	0.2900	中	中

上表给出了通过模型预测的 11 月 1 日到 11 月 15 日低中高风险的不同概率，并给出了真实风险概率进行对比，在这里做详细分析：

在上表的预测当中，可以看出本模型的预测存在部分误差，无法做到完全的与真实情况相符，这是由于在使用 *Markov* 链进行预测的过程当中，存在市场的突发情况且在做预测的时候存在市场信息的延后性，比如说在 10 月 5 日与 10 月 6 日时，其预测准确率皆为高风险，可能是受到前几日连续高风险的原因，因此模型对上述提到的情况即短时间内的大风险波动比较难以预测，但是在大多数情况下可以进行准确的市场预测，在上文当中对图解释的时候即可发现，模型在预测未来市场的风险具有较高的准确性，其预测的价格波动基本符合实际情况，可能仅存在由于低中高风险划分导致的问题，但是其预测的基本方向正确。

五、模型未来改进方向

虽然本模型基本上符合了市场的变动，但仍存在部分问题，针对该问题做出以下分析：

1. 市场变动对模型的影响：市场变动对模型的影响较大。例如，美国降息导致的 A 股大幅上升可能会影响模型的预测准确性。这是因为模型可能没有充分考虑到外部经济政策变化对市场的影响，但由于模型无法预测一些在实际生活中突然发生的情况如“美国降息”等，在对市场量化的过程当中，考虑收集对“重要事件”发生的因子，由于在这里无法直接对该事件进行预测，只能综合过去十年的因素进行综合量化，比如每年在十一月份由于美国大选，定会导致市场的波动，这是根据时间进行量化，又比如说在发生战争的时候，由于在战前一定会存在两国的工业体制方向转变，从市场当中也可以看出来，由此便是根据事前进行量化，防止战争导致的对市场的波动。

2. 同时在使用该模型的过程当中，在本文的第三问采用的是止损止盈的设置，在实际的过程当中，可能融入动态补仓以及风险对冲的方式对模型进行优化，可以更加全面的将风险模型转化为套利模型。

3. 在模型当中，本文用的是一阶 *Markov* 模型，该模型具有一定局限性，在这里规定该天的风险与前一天有关以此类推，因此无法对市场进行更加全面的描述，因此在这里考虑将提升 *Markov* 的阶数以提高前几日风险波动对该日风险波动的影响，即使与该天风险相关的一定是前一天，但由于市场的持续性，考虑因素的过程当中也需要考虑前几日对市场风险的影响。

综上所述，模型在预测沪深 300 市场风险时具有一定的准确性，但仍有改进空间。模型需要不断地调整和优化，以更好地适应市场的变化和规律。