

# 中药行业的量化投资策略

## ——基于XGBoost的多因子选股模型

曾凡一

zfygni@hotmail.com

2022年“国元证券杯”安徽省大学生金融投资创新大赛

省级一等奖作品



# 目录

01

项目背景

02

宏观经济

03

行业分析

04

量化选股

05

技术分析

06

财务分析

07

XGBoost

08

回测评价

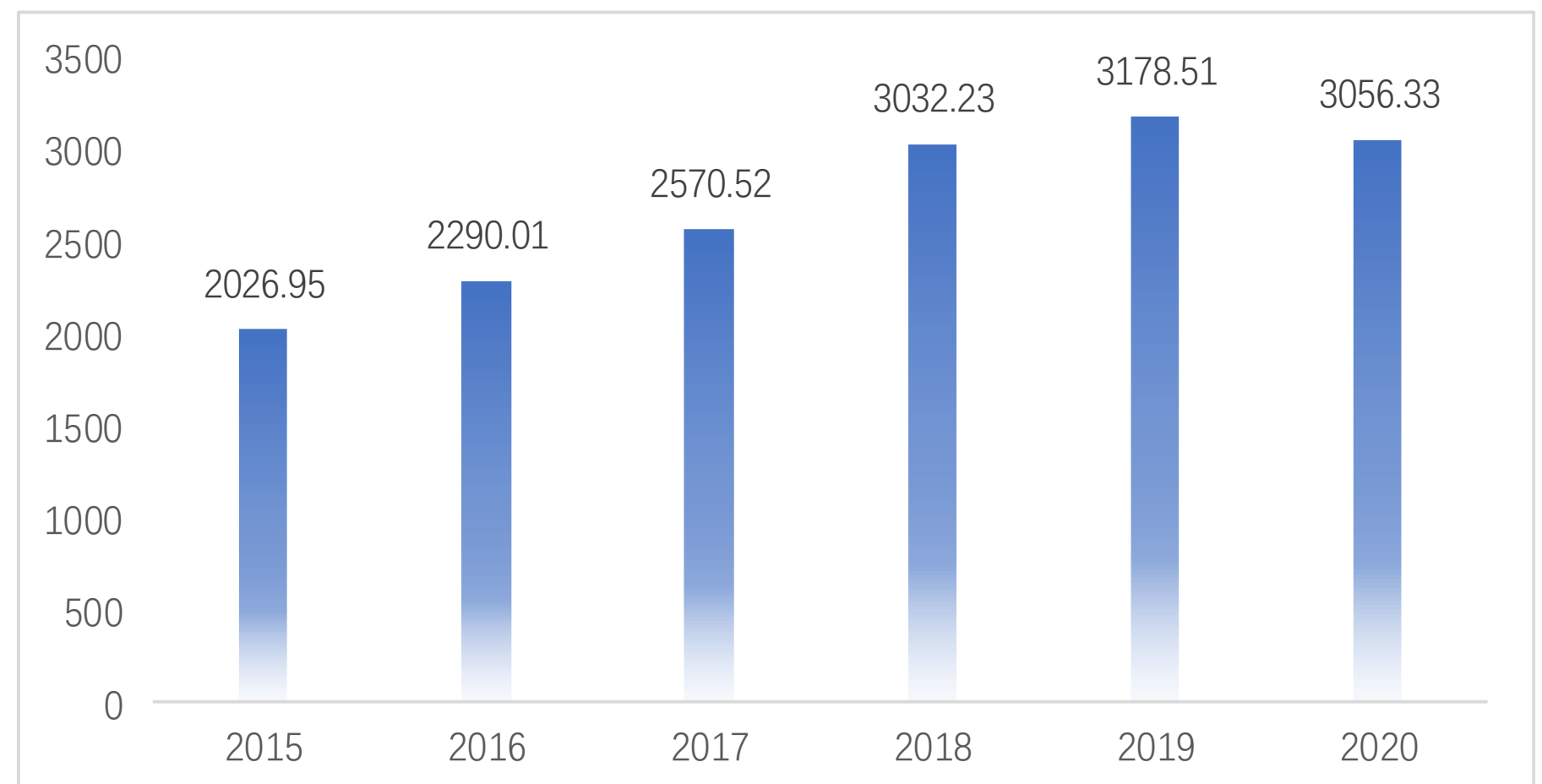
09

风险控制

01

项目背景

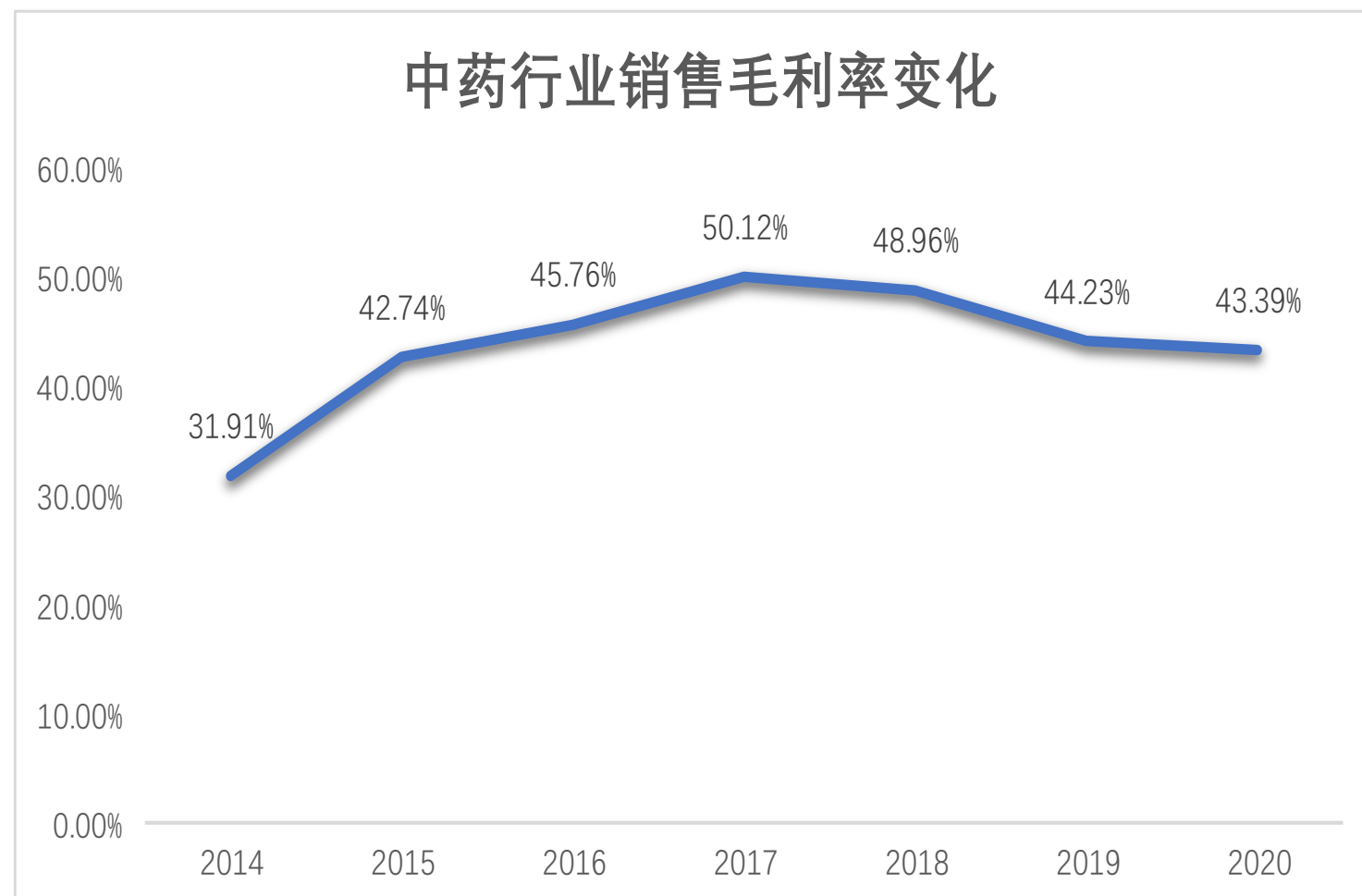
## 项目背景



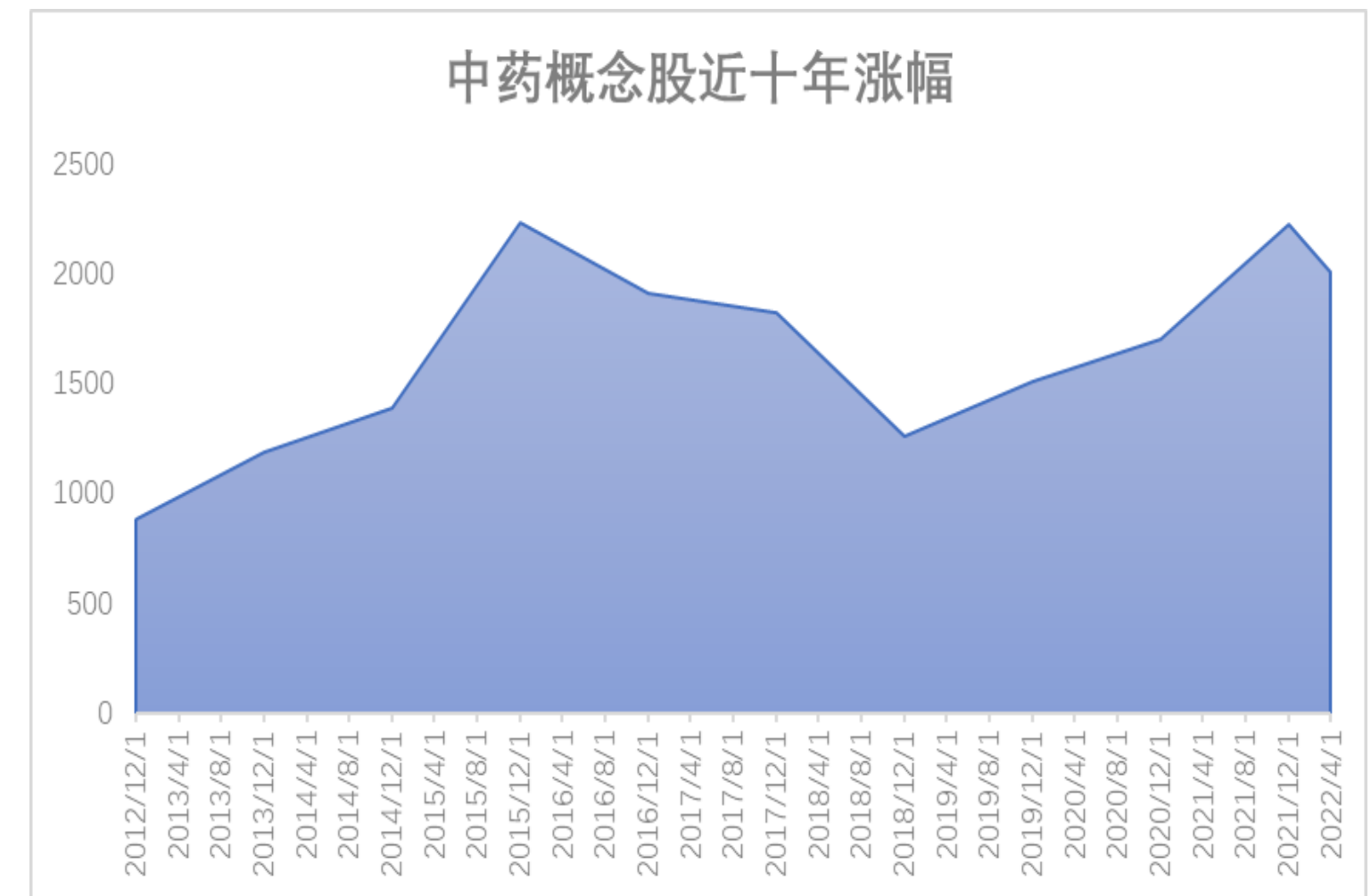
2015-2019年间，中药行业的营业收入呈连续增长趋势。2020年由于疫情的影响，限制医院治疗以及感冒药退烧药等处方药的购买，中药行业的营业收入下降了3.8%。



# 项目背景



2015-2021年间，中药行业的毛利率达到了40%以上，利润率高于我国大部分其他行业。



中药概念股整体趋势利好。在2020年大盘指数跌跌不止的时候，中药行业总体仍呈上涨趋势。

02

宏观经济

## 经济背景



国内疫情：  
对于消费的影响压力最大，同时工业生产、投资等活动也一定程度受限。

---



俄乌冲突：  
对整个全球经济体的发展是不理性的，对于中国经济的国际贸易部分会受到较大冲击。

---



国内通胀同时美联储加息：  
使得外资回流，北向资金净流出。



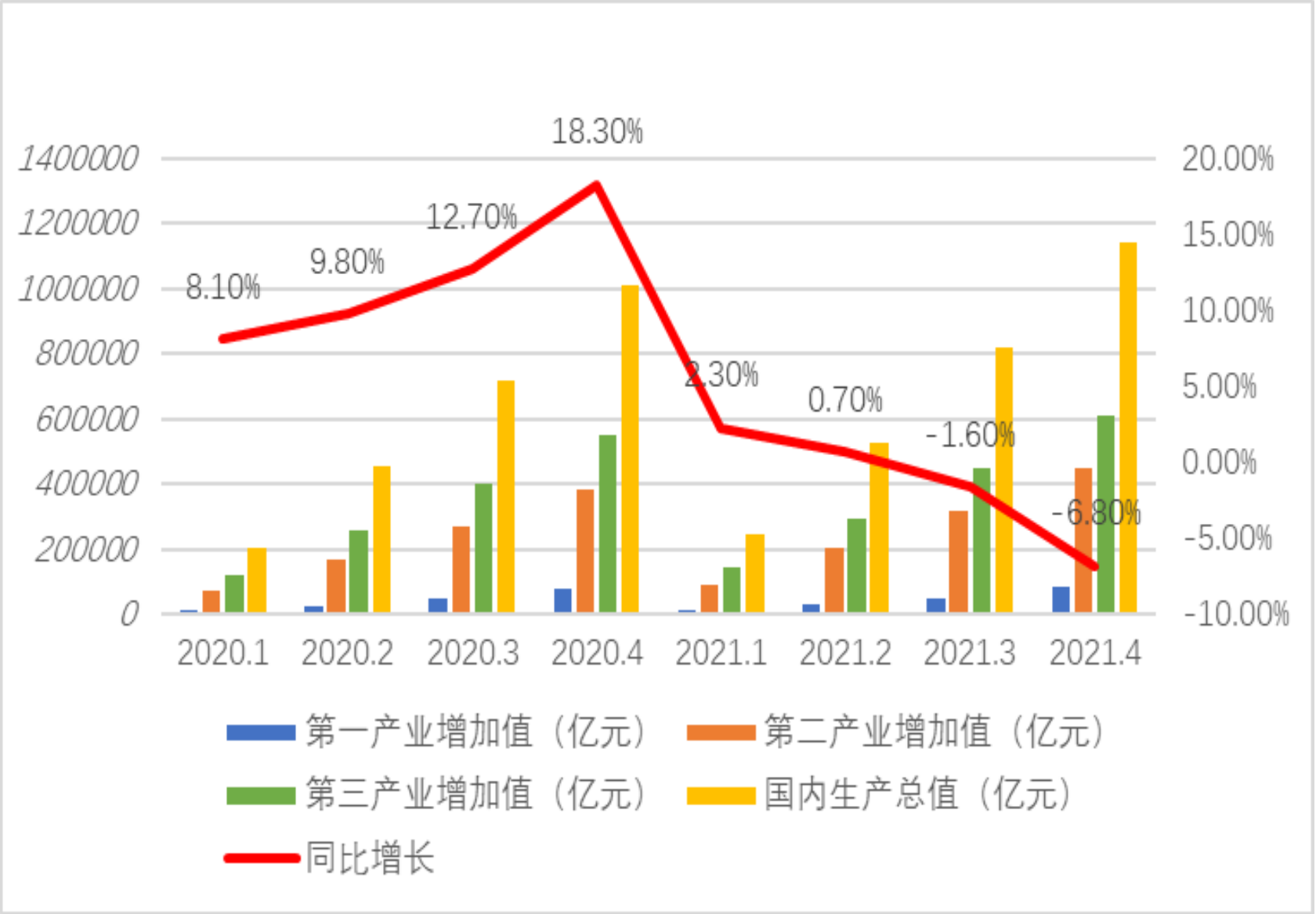
# 经济周期



宏观经济层面下，2021 年我国 GDP 增速为 8.1%，名义同比增速为 12.8%，GDP 平减指数为 4.38%



工业运行总体平稳，三大需求稳固增长，国内需求持续恢复，进出口仍维持较强的韧性





# 经济政策



- 宏观经济发展政策
- 中药行业相关政策

《“十四五”中医药发展规划》	国家中医药管理局 2022年3月29日	提出十项主要任务：1）建设优质高效中医药服务体系；2）提升中医药健康服务能力；3）建设高素质中医药人才队伍；4）建设高水平中医药传承保护与科技创新体系；5）推动中药产业高质量发展；6）发展中医药健康服务业；7）推动中医药文化繁荣发展；8）加快中医药开放发展；9）深化中医药领域改革；10）强化中医药发展支撑保障。
《推进中医药高质量融入共建“一带一路”发展规划（2021—2025 年）》	国家中医药管理局 2022 年 1 月	目标为“十四五”时期，与共建“一带一路”国家合作建设高质量中医药海外中心，颁布中医药国际标准，打造中医药文化海外传播品牌项目，建设中医药国际合作基地等
《国家医疗保障局国家中医药管理局关于医保支持中医药传承创新发展的指导意见》	国家医疗保障局； 国家中医药管理局 2021 年 12 月	《意见》指出：（1）充分认识医保支持中医药传承创新发展的重要意义；（2）加强中医药服务价格管理；（3）完善适合中医药特点的支付政策；（4）强化医保基金监管
《中医药创新团队及人才支持计划项目实施方案》	国家中医药管理局 2021 年 10 月	促进中医药多学科交叉创新团队建设，培养一批中青年多学科交叉创新人才。
《国家药监局关于促进中药传承创新发展的实施意见》	国家药监局 2020 年 12 月	《实施意见》：1.“指导思想”和“促进中药守正创新”2.“健全符合中药特点的审评审批体系”3.“强化中药质量安全监管”4.“推进中药监管体系和监管能力现代化”
《促进中医药传承创新发展的意见》	国务院 2019 年 10 月	健全中医药服务体系、发挥中医药在维护和促进人民健康中的独特作用、大力推动中药质量提升和产业高质量发展、加强中医药人才队伍建设、改革完善中医药管理体制机制等，明确了中医药传承创新发展的目标方向和具体举措。



## 大健康行业运行

全球及中国大健康行业运行情况

## 中药行业现状

产业组织 + 新冠的中药救治

## 中药产业结构

产业上下游 + 产业细分



全球：

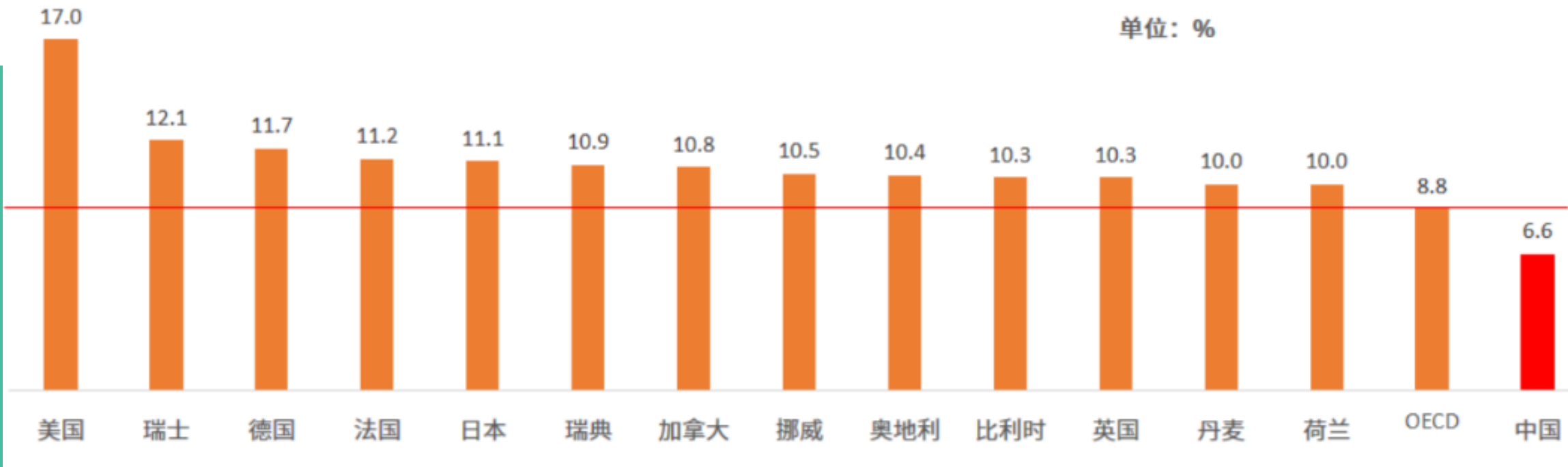
- 美国、瑞士、德国、法国等欧美发达国家医疗卫生支出占GDP的比重普遍超过10%
- 随着人们对健康标准的不断提高以及医疗研发技术的发展，全球大健康产业会在相当长的时期内持续保持增长态势



国内：

- 近年来政府高度重视人民群众的健康问题，中药等卫生健康事业获得了长足发展。
- 习近平同志在党的十九大报告中提出实施“健康中国战略”，指出“人民健康是民族昌盛和国家富强的重要标志”

2019年典型国家医疗费用支出及占GDP的比重  
Global medical expenditure and its proportion of typical countries in 2019



# 中国大健康行业运行情况



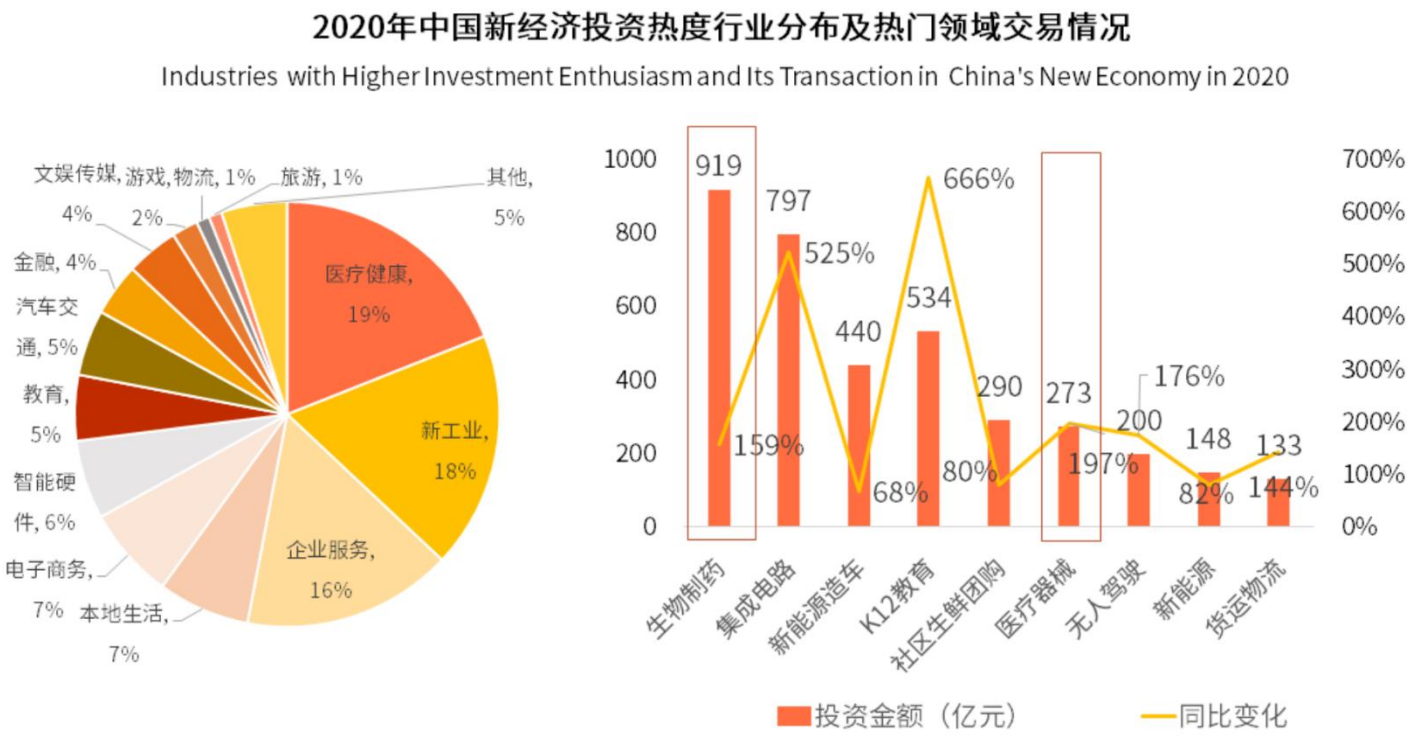
## 消费端

2014-2019年中国居民人均消费支出和医疗保健消费支出呈现同步增长的态势，2020年受疫情影响，人均居民消费支出略有下滑，人们的医疗保健消费支出也同步下降。



## 投资端

近两年中国投资热度最高的新经济行业是医疗健康，在热门投资领域中，投资金额最高的是生物制药，超过900亿元。



# 产业组织

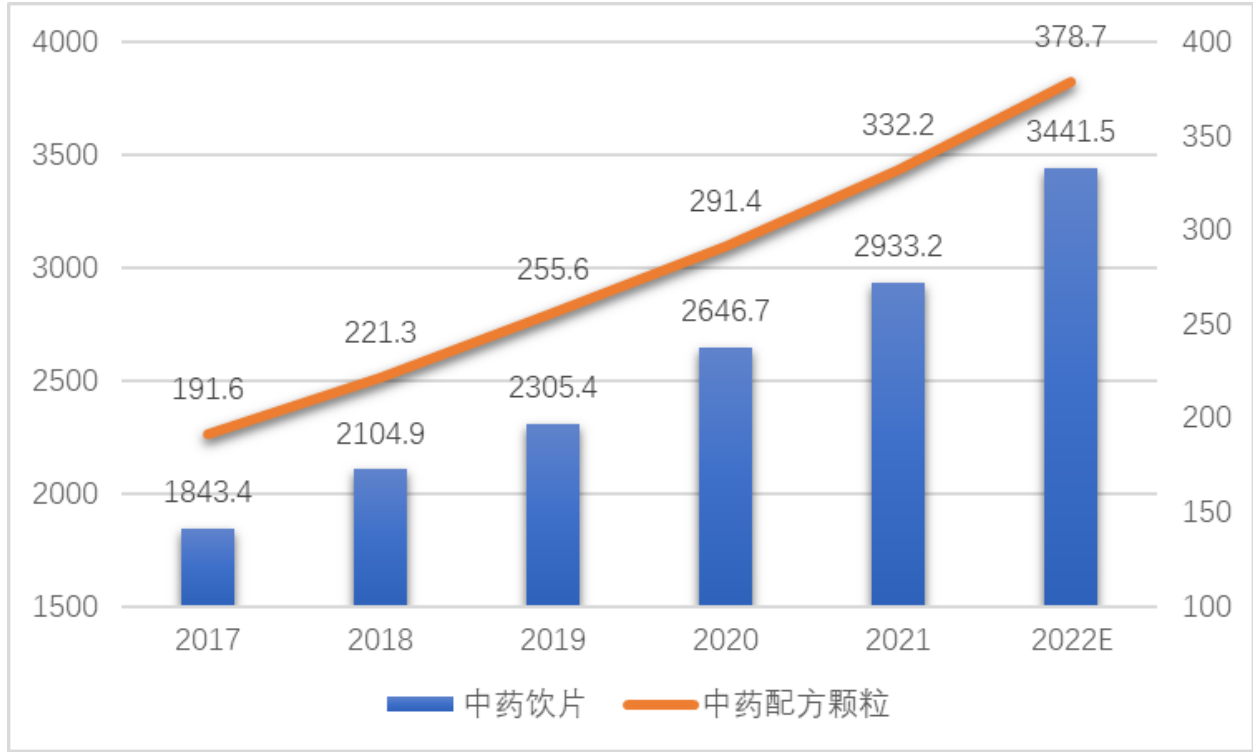


图 2017-2022年中国中药行业市场规模

- 中药企业呈现 “6 + N” 格局
- 2017 年-2020 年我国中药饮片市场规模由 1843.4 亿元增至 2646.7亿元，年均复合增长率为 8.9%
- 药配方颗粒市场规模由 2017 年的 191.6 亿元增长至 2020的 年的 291.4 亿元，年均复合增长率约 15%



# 新冠的中医治疗

表 新冠中医治疗机理	
功能	作用机理
免疫调节	中药在新冠肺炎治疗中发挥双向调节作用，能够系统性地调节新冠肺炎所引发的免疫紊乱，抑制促炎细胞因子的表达水平，调节细胞因子之间的平衡，从而改善炎症对组织和器官的损害。
炎症抑制	研究表明，中药对于病毒感染引发的炎症反应有抑制作用。如：连花清瘟胶囊可抑制不同流感病毒的增殖，同时能够减少病毒诱导的炎性细胞因子基因表达；热毒宁注射液可以有效减少促炎细胞因子的表达、炎症介质的释放和生成等。
抗病毒	中药在抗病毒和症状减缓等方面具备良好的效果，包括：治疗病毒性肺炎、降低巨细胞病毒复制水平以及调节病毒感染的免疫应答等。

中药对于新冠病毒的治疗主要在于免疫调节、炎症抑制和抗病毒三个层面

# 中药行业现状



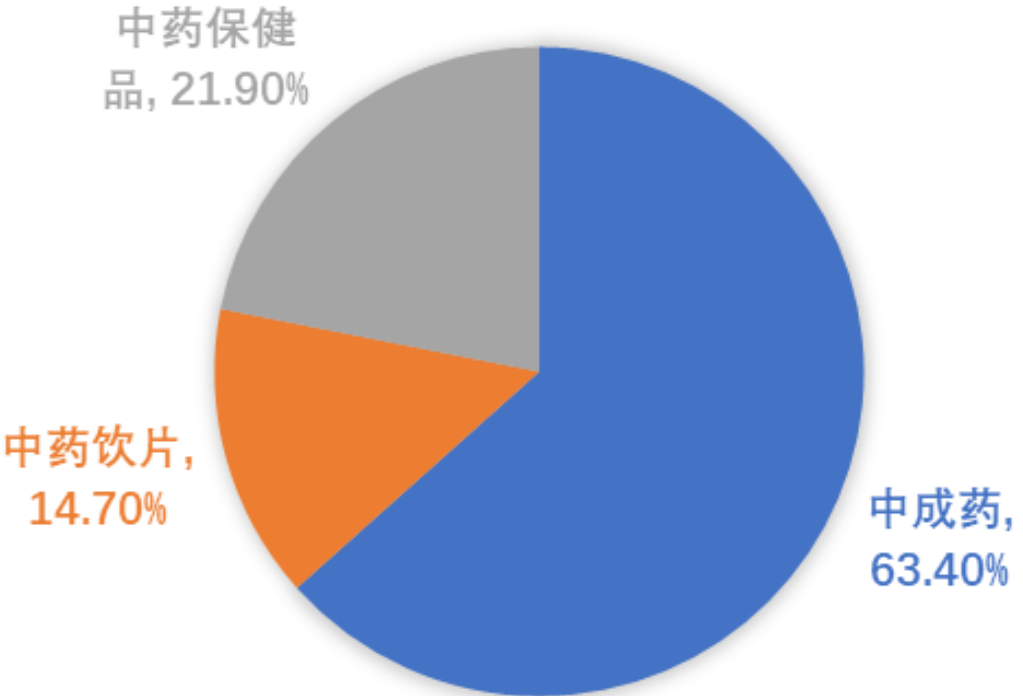
## 产业上下游

产业链	子行业
上游	中药材
中游	中药饮片
	中成药
下游	销售终端

- 中医药产业链包含上游中药原材料端、中游中药材加工制造端和下游中医类机构服务端
- 由此引申出的中医药市场可大致细分为中成药、传统中药饮片和中药保健品。



## 产业细分



2021年中成药占中国传统中医药市场的比例为 63.4%，传统中药饮片占比为14.7%，中药保健品占比21.9%

# 中药产业结构

# 04

## 多因子选股模型介绍

利用历史数据，选取某些可能影响股价波动的因素去选择股票买入持有一段时间，然后分析收益率、回撤率等相关数据，最终确定哪些因素能够显著的影响股价

## XGBoost算法介绍

XGBoost 算法是一种提升算法，在筛选股票因子过程中，可以利用XGBoost算法的随机选取特征，对股票因子的好坏进行判断，得出因子分类结果。

# 一、多因子选股模型介绍

(1) 定义：多因子选股模型是量化选股版图中的一个非常重要的组成部分，所谓多因子选股，就是利用历史数据，选取某些可能影响股价波动的因素去选择股票买入持有一段时期，然后分析收益率、回撤率等相关数据，最终确定哪些因素能够显著的影响股价，此后我们就可以利用这些因素进行投资。

## (2) 因子分类

公司层面因子

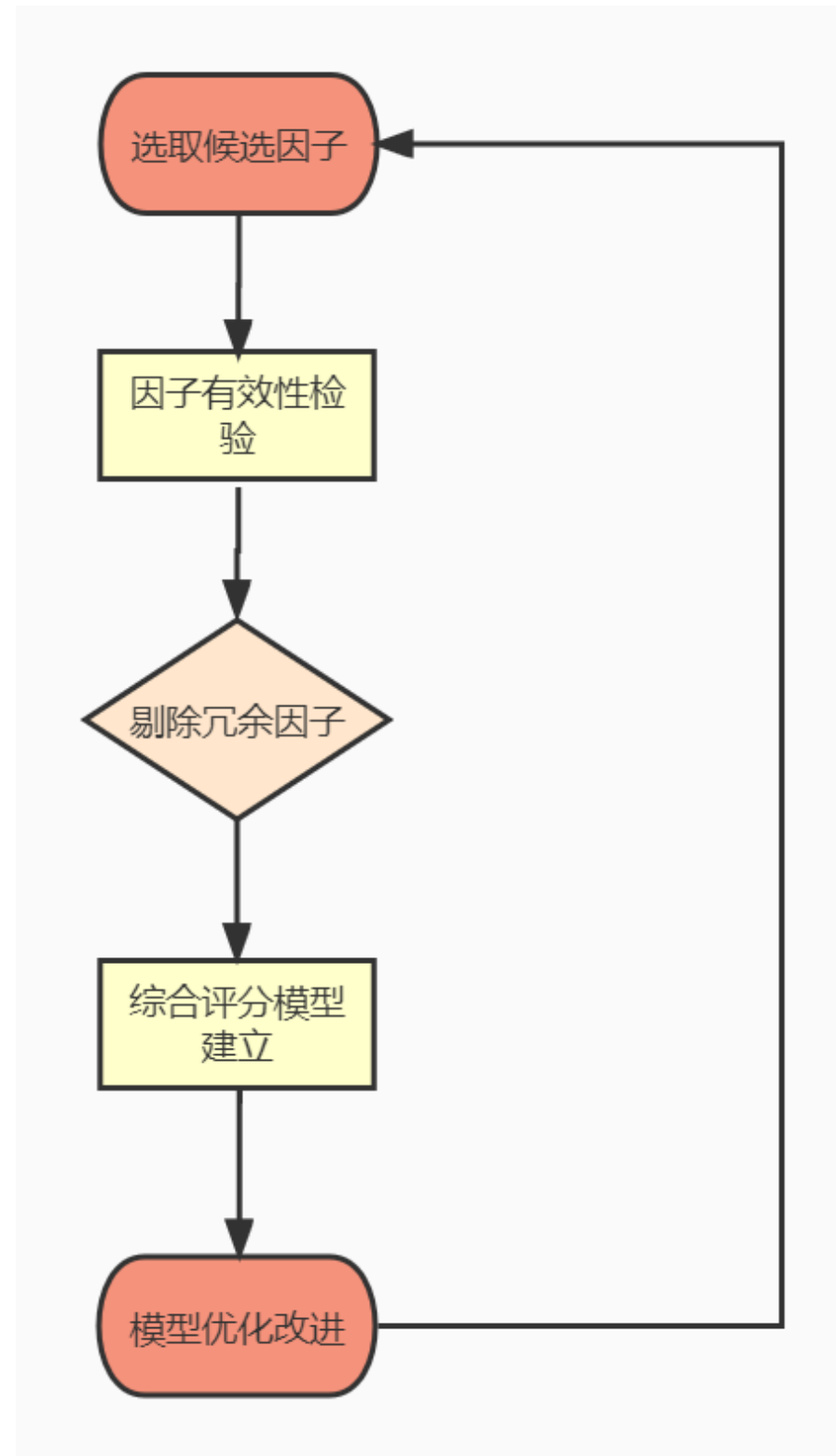
外部环境因子

市场表现因子



# 一、多因子选股模型介绍

## (3) 流程图



## (4) 优缺点

优点:

1. 简化数据
2. 解释性更好
3. 减少了人为定义权重的随意性
4. 操作性极强

缺点

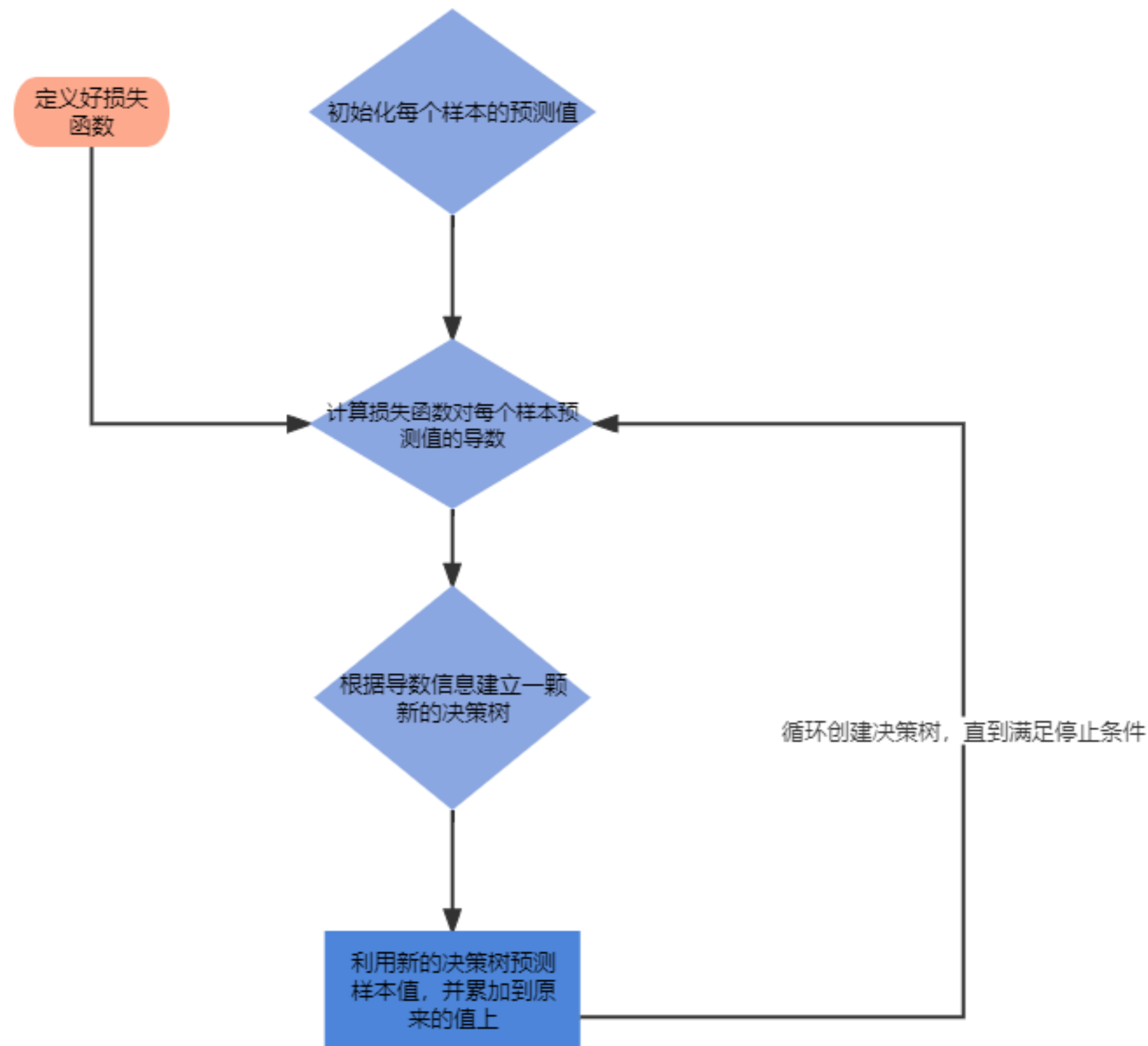
1. 方法有时会无效
2. 错误不易被发现

# 一、XGBoost算法介绍

(1) 定义：XGBoost是一个优化的分布式梯度增强库，旨在实现高效，灵活和便携。

它在 Gradient Boosting 框架下实现机器学习算法。XGBoost提供并行树提升，可以快速准确地解决许多数据科学问题。相同的代码在主要的分布式环境（上运行，并且可以解决数十亿个示例之外的问题。

流程图



# 一、多因子选股模型介绍

## (4) 优缺点

优点:

- 1.既可用于非线性分类，也可用于线性分类
- 2.提升泛化能力
- 3.提高计算效率
- 4.较好的容忍异常值

缺点

- 1.由于其最开始的代码实现是由C++写出，所以在如Python、R等机器学习语言上安装异常麻烦
- 2.可供研究的材料较少

# 05

## 方案设计

量化策略整体思路

## 因子池构建

因子池构建 + 数据预处理

## 算法优化

XGBoost 算法参数优化详解及因子筛选

# 方案设计框架

本文的方案设计框架主要分为：

- 因子池构建与数据预处理
- 分类算法选择
- 模型及回测构建
- 模型评估及实施建议

利用XGBoost算法，提高因子筛选的精度有效性

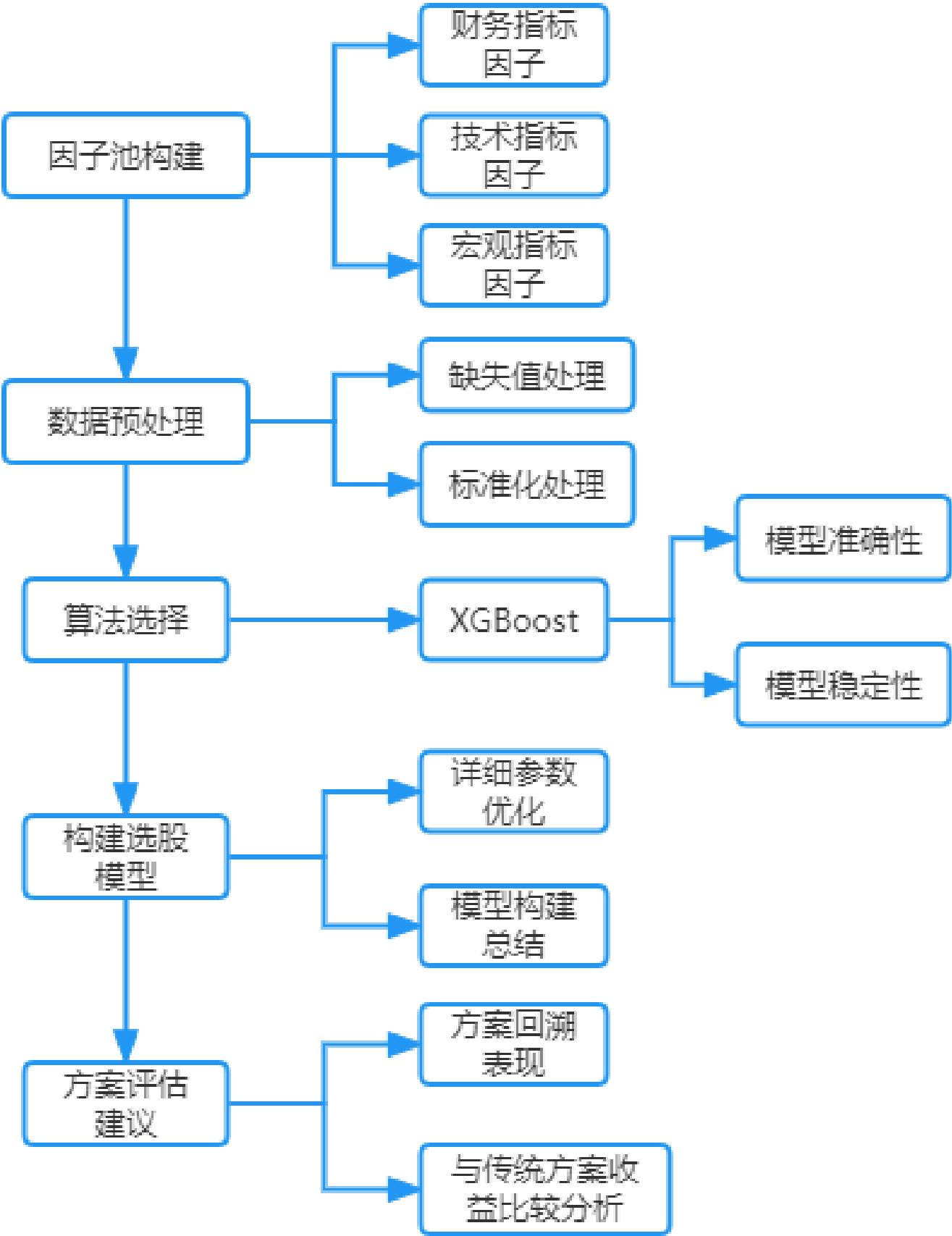


图 方案设计技术框架

# 数据处理

## 缺失值处理

本文的因子数据主要来源于的 1200 只股票的财务数据和行情数据，以及同期的宏观、房地产和债券等方面的数据，其包含的行业比较广泛，因此，数据中可能会存在某些因子确实的情况，本文采取的方式是缺失值过多就放弃该只股票。缺失值较少的话就用插值的方法进行补全。

$$L_n(x) = \sum_{i=0}^n l_i(x) y_i$$
$$l_i(x) = \prod_{\substack{j=0 \\ j \neq i}}^n \frac{x - x_j}{x_i - x_j}$$

## 异常值处理

在处理异常值是不能一味的删除异常值，因为在数据量较小的时候，删除异常值可能会影响整个数据的结构，进而影响后续的建模结果。在多数情况下，一般要结合具体问题具体分析。

## 数据标准化

为了消除不同指标间的量纲的不同，需要将数据进行标准化处理，常用的办法是将每个指标按照一定比例的缩放到 0 到 1 之间。

1、最小-最大规范化

$$x^* = \frac{x - \min}{\max - \min}$$

2、零-均值标准化

$$x^* = \frac{x - \mu}{\sigma}$$

3、小数定标规范化

$$x^* = \frac{x}{10^k}$$

## 数据转换

将连续的数值转换为定类数据，有两种方式，其一是按照极差定距分为一定数量的类别，其二是按照数值的数量，将特征分类每个类别有相同个数一定数量的类别。

# 模型优化

1. 选择较高的学习速率。为了更为精细化的优化参数，需要选出参数组合能够使得模型的泛化；
2. 对于给定的学习速率和决策树数量，进行决策树特定参数调优；
3. 降低学习速率，确定理想参数。

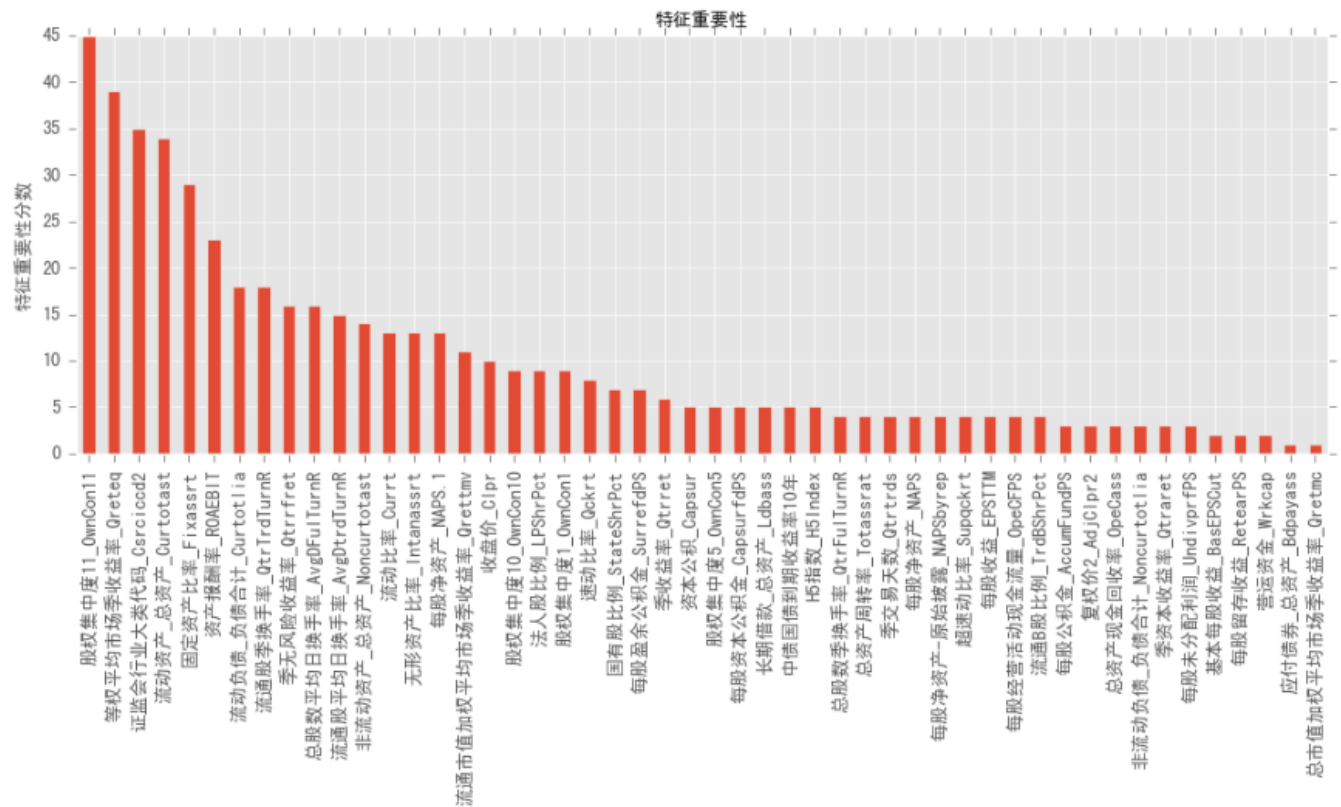


图 因子重要性输出1

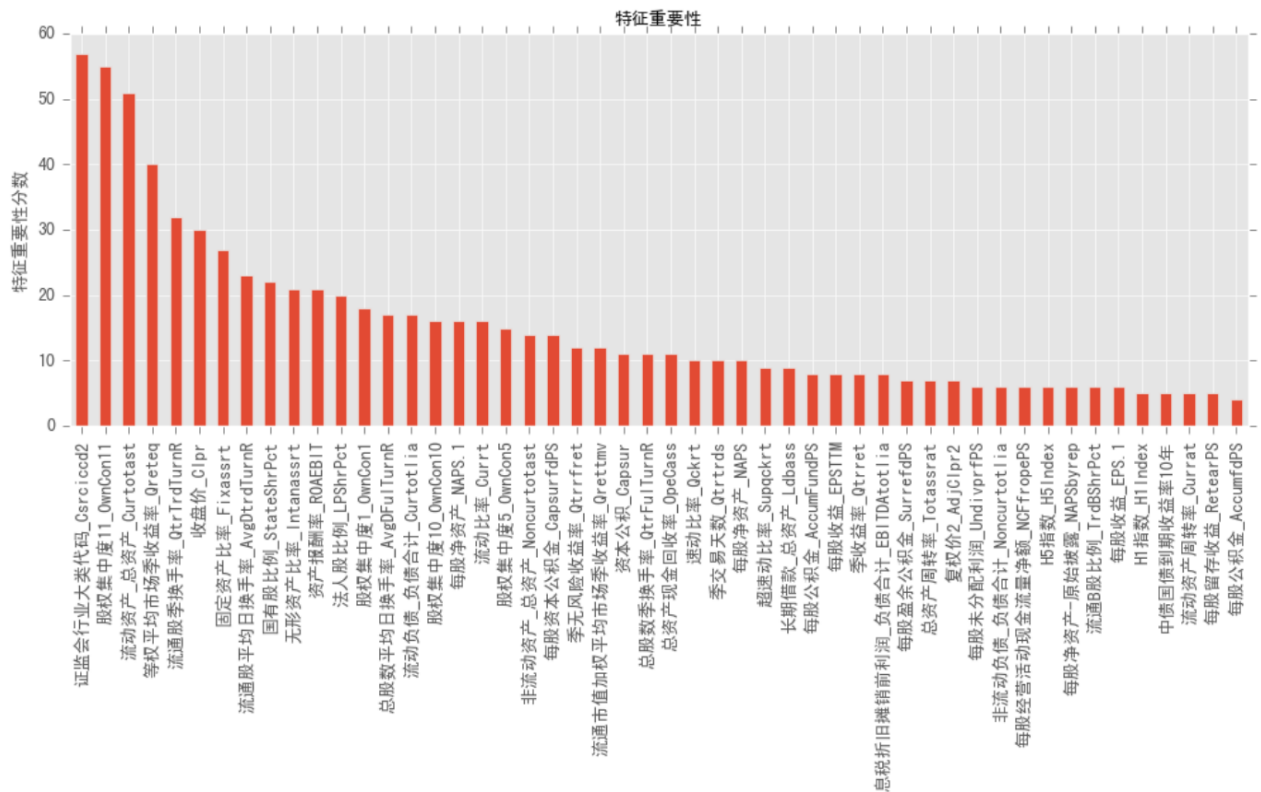


图 因子重要性输出2

定义学习速率 → max\_depth 和 min\_weight 参数调优 → gamma 参数调优 → 降低学习速率



# 股票筛选结果

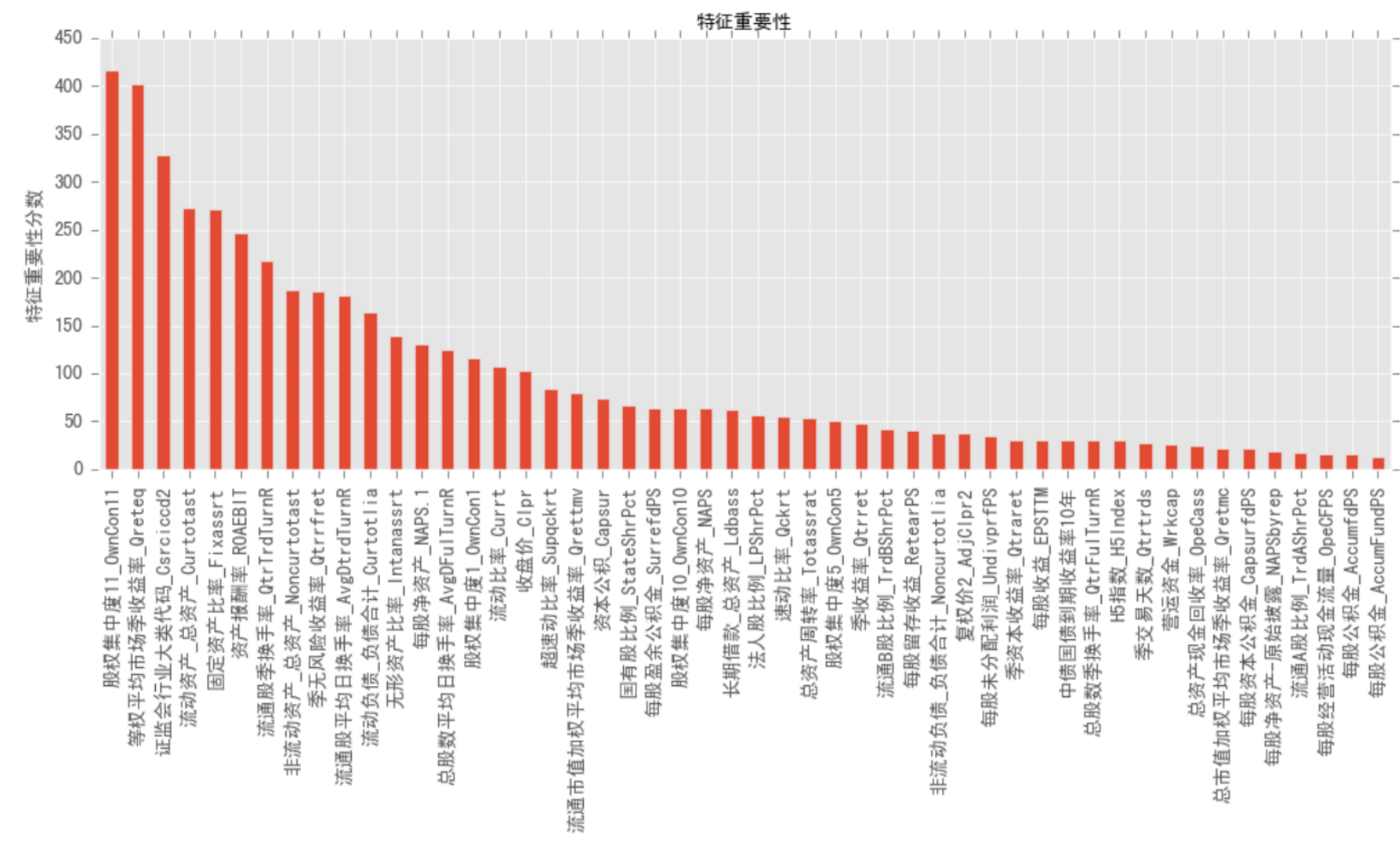


图 因子重要性输出最终结果

根据XGBoost模型的输出结果，选取最终得分排名前3的股票，  
即为我们筛选出的股票投资组合。

表 最终得分表			
排名	股票代码	股票名称	综合得分
1	600436	片仔癀	86.2342
2	000538	云南白药	83.5420
3	600332	白云山	75.3800
4	600479	千金药业	69.3513
5	300039	上海凯宝	69.1347
6	002166	莱茵生物	62.8025
7	600085	同仁堂	60.0200
8	002566	益盛药业	58.6025
9	603858	步长制药	55.8165
10	603998	方盛制药	53.8646





## BOLL指标

BOLL指标可以研究股票股价在市场波动变化情况，从而推测股价的未来走势的一种重要技术分析工具。

## 矩形整理

矩形整理体现为股价在两条平行而且横向的直线之间来回上下摆动，在价格上升到上半部分位置时，又回落回到下半部分，这样来回摆动。

## 双重顶

双重顶即人们所熟知的M头，双重顶顾名思义是共出现两个顶，也就是有两个相同高度的高点。

# BOLL指标



在3月3日价格经过一波下跌之后，上下轨分别向上向下运动，形成一个“双开”的喇叭口，预示此时股票行情下行趋势会停止，随后云南白药的股价变动趋于平缓。



在4月10日通过上图可以看到股票整体开始回落，随后价格不断出现一定程度跌幅



4月25日跌至最低点，跌破布林线的下轨。但在4月26日，BOLL的上下轨的开口增大，此时预示该股随后有上涨趋势，后面连续3日股价向BOLL线中线靠拢，这个时候就可以在低价时买入并当它升到一定高价时抛出。



资料来源：同花顺

# 矩形整理



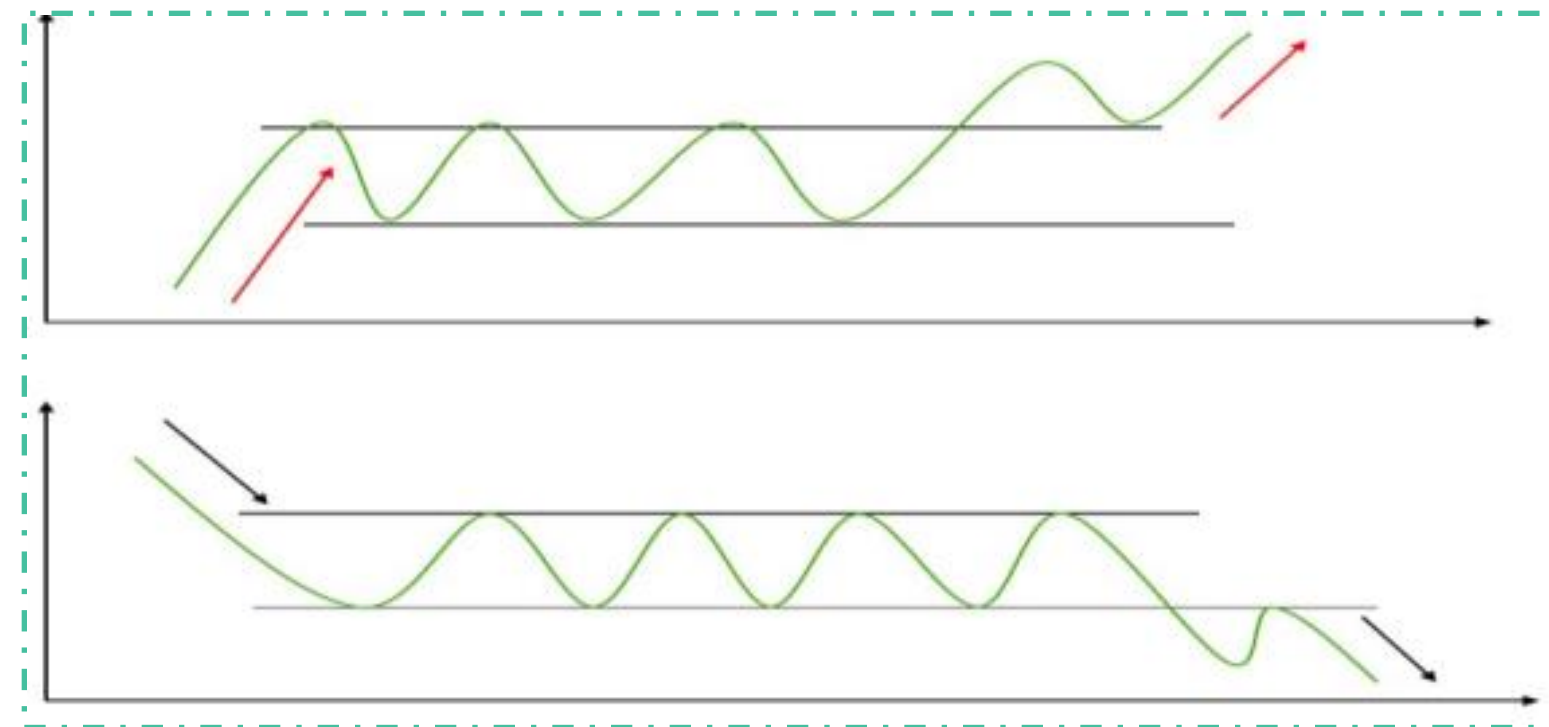
如果一个投资者是中长线投资者，那么在上升趋势没有改变并且出现矩形整理形态时，可以多持股;反之，需要等待更低的买入点。



白云山：2022年6月15日到2022年5月15日呈现出矩形整理的形态



在4月29日最终跌到最低点，后续根据矩形整理形态，会出现一定程度回升



## 突破形态之双重顶



片仔癀(600436) 2020年6月至2022年5月13日期间周K线图



该股在2022年3月中旬之前的这段时间内，出现了一个标准的双重顶形态即走出M型的形态



于3月初至今是呈平缓走态，拟形成第二个M型的形态。从此股随后的走势中可以看到，双重顶形态上升是非常可信的“空转多”信号，一旦出现此形态可考虑进场，以获取收益。



资料来源：同花顺

07

财务分析



# 财务分析——盈利能力



股票	股票代码	每股收益
片仔癀	600436	4.03
云南白药	000538	2.21
白云山	600332	2.288

这几支股票的**盈利能力较强**，同时因为疫情的原因，对中药的需求迅猛增长，有较大的发展潜能，**具有相对较高的投资价值**。



股票	股票代码	营业利润率（%）
片仔癀	600436	30.72
云南白药	000538	7.69
白云山	600332	5.75

中药行业的平均营业利润率为**15%**，说明片仔癀的业务能力很强的同时创造利润的能力也很强，经营风险相对较小。另外两支股票的营业利润率虽低于行业平均值，但也处于**5-10%之间，有不错的盈利能力**。



股票	股票代码	净资产收益率（%）
片仔癀	600436	27.68
云南白药	000538	7.58
白云山	600332	13.46

片仔癀对股东投入资本的利用效率较高，**即运用公司自有资本的效率较高**。而云南白药和白云山在运用自有资本效率方面相较于片仔癀较低。

# 财务分析——成长能力



股票	股票代码	营业收入增长率（%）
片仔癀	600436	23.2
云南白药	000538	11.09
白云山	600332	11.9

三支股票的营业增长率均超过10%，说明这些公司的产品在**市场中都有一定的份额**。



股票	股票代码	净利润增长率（%）
片仔癀	600436	45.46
云南白药	000538	-49.17
白云山	600332	27.6

片仔癀和白云山都有一个较高的净利润增长率，**说明企业发展情况良好**，产品销售增长快，市场竞争力强。而云南白药的净利润增长率为负，主要系**股票投资等公允价值变动损益影响**，随着疫情情况好转及公司减持股票，云南白药后续经营状况有望恢复增长。

# 财务分析——偿债能力



股票	股票代码	现金流量比率
片仔癀	600436	0.216
云南白药	000538	0.411
白云山	600332	0.176

三支股票的现金流量比率都小于1，应该需要采取一定的筹资措施从而满足企业日常的基本需要，**需增强其还债能力**，使市场上的投资者更加充满信心。



股票	股票代码	速动比率
片仔癀	600436	3.835
云南白药	000538	2.919
白云山	600332	1.314

白云山的速动比率是最接近于1的，尽管偿债能力一般，但是是**适合企业发展的**。相比之下，片仔癀和云南白药的速动比率远大于1，企业在速动储备资产占用太多的资金，大大减少了企业获取收益的能力，很大程度上会成为企业快速发展的一个阻碍。



# 财务分析——营运能力



股票	股票代码	总资产周转率
片仔癀	600436	0.707
云南白药	000538	0.677
白云山	600332	1.097



股票	股票代码	存货周转率
片仔癀	600436	1.708
云南白药	000538	2.736
白云山	600332	5.522



股票	股票代码	应收账款周转率
片仔癀	600436	13.21
云南白药	000538	5.162
白云山	600332	5.008

三支股票的总资产周转率都大于50%，说明三支股票的企业经营期间全部资产投入到产出的速度较快，且产品销售能力强，**资产的利用效率较好，企业的经营管理较好。**

2014-2021年中国中药行业的存货周转率一直处于震荡形态，区间在2.4-4.4。由上表可以看出，三支股票的存货周转率都在这区间左右，说明**企业整体的存货变现能力较为良好。**

三支股票的应收账款周转率都在行业平均值上下，表明企业在回收账款和资金管理上处于行业平均水平，**拥有较好的发展前景。**



## 模型精度

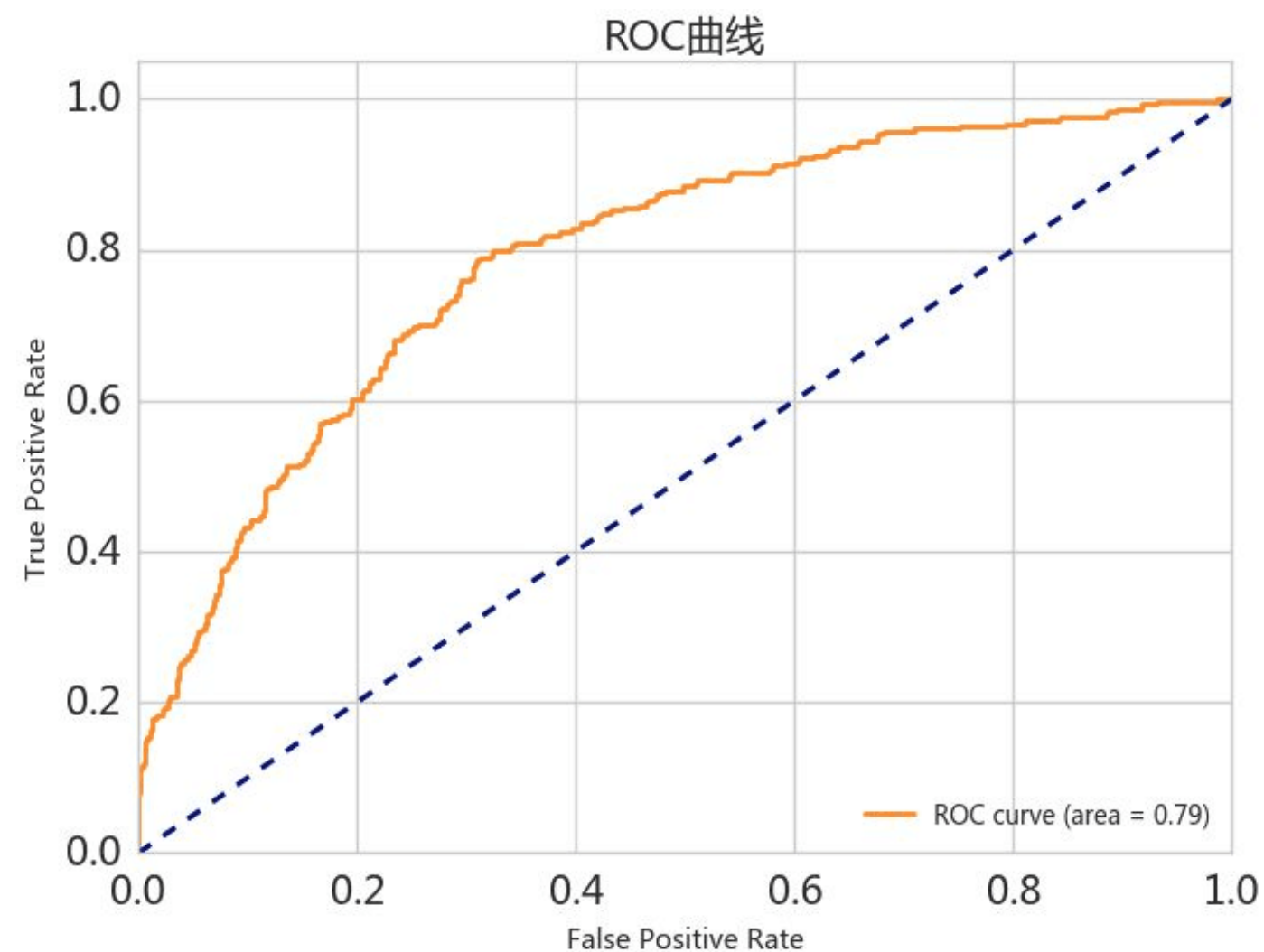
利用ROC 曲线和 AUC 值来判断模型的好坏

## 回测分析

历史回测绩效评价

## 模型准确度

经过上一节的数据处理和参数优化，最终建立了训练和测试模型，结果为模型输出结果为训练模型准确率 86.97%，AUC 分数为 0.79



## ROC 曲线

ROC 曲线要高于 45 度角，说明模型结果模型好于随机猜测，但是本方案并没有涉及到卖空，只涉及买入，因此在本方案中只关心正例在何种情况下被预测为正例的准确性最大。

# 回测绩效指标选择



## 收益率指标

### 年化复合收益率

年化收益率是一种将当前收益率转换成年限下所得的收益率，其本质是一种理论的收益率，非真正取得的收益率。年化收益率可较为直观地比较两种产品直接收益的高低。



## 风险度量指标

# 风险性指标



## 贝塔系数

用来衡量个别股票或股票基金相对于整个股市的价格波动情况。



## 夏普比率

夏普比率比较的是量化选股的投资组合的收益率与市场无风险收益率。



## 最大回撤

最大回撤率是代表了投资人所能忍受亏损的极限。



## 信息比率

以马克维茨的均异模型为基础，用来衡量超额风险带来的超额收益。



# 历史回测绩效评

## 回测结果

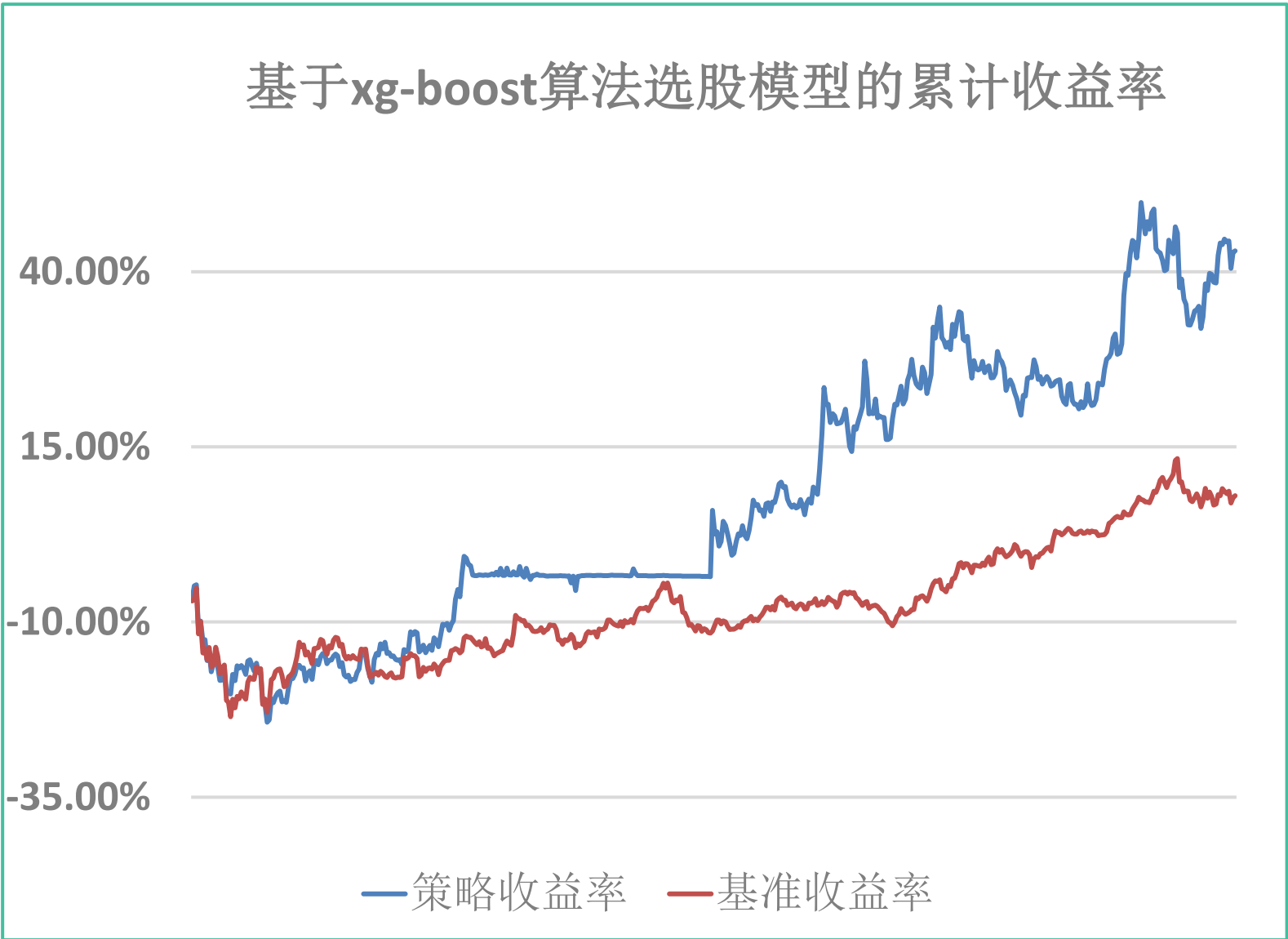


表 基于 xgboost 算法选股模型的回测结果			
年化收益率	20.09%	基准年化收益率	4.04%
阿尔法	0.1614	贝塔	0.83
夏普比率	0.68	波动率	0.2448
信息比率	0.79	最大回撤	0.2055
年化换手率	0.02		



## 风险类型

风险是市场运行的不确定性。期货市场的风险可以用市场上发生的所有影响市场运行的风险集合表示

## 中药行业风险

投资中药行业的风险既有中药行业特有的不能预期的职业投资风险，同时也有投资前对项目的评估、防控、分析不够充足带来的风险

## 风险控制

风险控制是指利用有效的方法和措施，来减少发生风险的概率以及降低风险所带来的损失。

# 一、风险类型

**系统风险：**是指受市场变化影响较大的风险，是不可分散的，也叫市场风险。

**内容：**利率风险、通货膨胀  
通货紧缩风险、流动性风险、  
政治、心理、人才、法律、  
道德和品种风险

**非系统风险：**指与市场无关的,由于某个因素变化引起市场中某种收益损失的可能性。它只对市场中的某一种产品有影响，非系统风险可以用投资分散化来减轻、消除。



二、中药行业风险类型



政策风险



土地成本风险



资本成本风险



违法违规带来的风险



中药保险风险

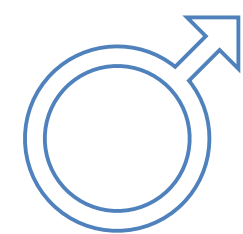


管理风险

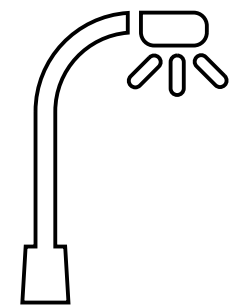


技术人才流失风险

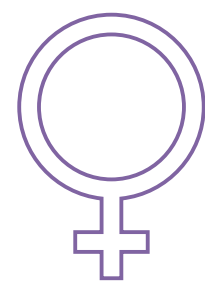
### 三、风险控制



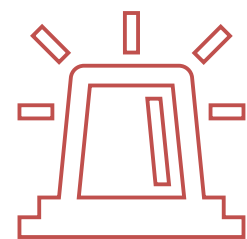
提前做好相关政策法规的了解



明确项目投资管理的目标



做好对中药项目团队的考察



投资决策规范化



制定有效的立足长远的投资管理制度



制定有效的人才管理机制



投资决策的可行性分析

# 谢谢观看

