****京 江 学 院



**JINGJIANG COLLEGE OF J I A N G S U N I V E R S I T Y**

**本 科 毕 业 论 文**

基于Python和JoinQuant平台的选股模型构建与设计

**Construction and Design of Stock Selection Model Based on Python and JoinQuant Platform**

专 业： 计算机科学与技术

班 级： 15计算机2班

学 号： 20150301202

姓 名： 陈 裕

指导教师姓名： 王 海 燕

指导教师职称： 副 教 授

2019年 5 月

基于Python和JoinQuant平台的选股模型构建与设计

专业班级：15计算机2班 学生姓名：陈裕

指导教师：王海燕 职称：副教授

摘要 股票市场每天都会产生大量数据，如何从这些数据中分析出有价值的信息，是股票分析机构和股票投资者最为关注的问题。选股模型是将公司基本面分析与金融技术分析结合起来，以Python为开发语言，JoinQuant为测试平台，分析公司基本面（财务）数据及股票历史交易数据，构建出投资组合模型为投资者作出优质决策。具体操作：（1）通过获取的Wind数据对公司基本面进行分析，从净资产收益率（ROE）、净利润增长率、主营业务收入增长率三个指标入手，筛选出一批具有良好发展前景的公司；（2）使用金融技术分析：使用资本资产定价模型（CAPM）计算股票的Beta系数来探寻个股相对于大盘的波动性，将低于市场安全线的股票筛除；计算股票收益率的偏度和峰度来检验收益率数据的正态性，将数据特征偏离正态性较大的股票筛除。（3）使用Python语言构建分别以夏普比率和方差为指标的投资组合模型，从两个模型中分别选出排名前三的组合；将六种投资组合分别放入JoinQuant平台中回测，分析各组合的回测结果。从回测结果中得出使用夏普比率作为指标构建的模型效果更好。通过模型选出的投资组合质量较高，能有效规避风险，并获得较高收益，为选股开拓了新的视野。

关键词：Python JoinQuant 选股 CAPM MPT 夏普比率

Construction and Design of Stock Selection Model Based on Python and JoinQuant Platform

**Abstract** The stock market generates a large amount of data every day. How to analyze valuable information from these data is the most concerned issue for stock analysts and stock investors. The stock selection model combines the company's fundamental analysis with financial technology analysis, using Python as the development language, JoinQuant as the test platform, analyzing the company's fundamental (financial) data and stock historical trading data, and building a portfolio model for investors. Quality decision making. Specific operations: (1) Analysis of the company's fundamentals through the acquired Wind data, starting from the three indicators of return on net assets (ROE), net profit growth rate, and main business income growth rate, screening out a batch of good development Prospective company; (2) Using financial technology analysis: Calculate the stock's Beta coefficient using the Capital Asset Pricing Model (CAPM) to explore the volatility of individual stocks relative to the market, and to screen stocks below the market safety line; calculate stock returns The skewness and kurtosis are used to test the normality of the yield data, and the data features are deviated from the stocks with higher normality. (3) Using the Python language to construct a portfolio model with Sharpe ratio and variance as indicators, select the top three combinations from the two models separately; put the six portfolios into the JoinQuant platform for backtesting, and analyze each Combined backtest results. From the backtest results, the model constructed using the Sharpe ratio as an indicator is better. The quality of the investment portfolio selected by the model is high, which can effectively avoid risks and obtain higher returns, which opens up new horizons for stock selection.

**Key words:** Python JoinQuant Stock selection CAPM MPT Sharpe ratio

目 录

[第1章 绪论 1](#_Toc8852803)

[1.1 选题背景 1](#_Toc8852804)

[1.2 国内外研究现状 1](#_Toc8852805)

[1.3 研究目的 2](#_Toc8852806)

[1.4 研究意义 2](#_Toc8852807)

[1.5 研究方法 3](#_Toc8852808)

[1.6 本章小结 3](#_Toc8852809)

[第2章 相关理论与技术 4](#_Toc8852810)

[2.1 相关理论 4](#_Toc8852811)

[2.1.1对数收益率 4](#_Toc8852812)

[2.1.2资本资产定价模型（CAPM） 4](#_Toc8852813)

[2.1.3偏度和峰度 5](#_Toc8852814)

[2.1.4夏普比率（Sharpe Ratio） 5](#_Toc8852815)

[2.1.5现代投资组合理论模型（MPT） 6](#_Toc8852816)

[2.2 相关技术 6](#_Toc8852817)

[2.2.1基本面分析 6](#_Toc8852818)

[2.2.2金融技术分析 7](#_Toc8852819)

[2.2.3Python技术 7](#_Toc8852820)

[2.2.4JoinQuant平台介绍 8](#_Toc8852821)

[2.3 本章小结 8](#_Toc8852822)

[第3章 股票数据分析 9](#_Toc8852823)

[3.1 股票数据分析总体框架图 9](#_Toc8852824)

[3.2 基本面分析 9](#_Toc8852825)

[3.2.1基本面数据获取 10](#_Toc8852826)

[3.2.2基本面数据预处理 10](#_Toc8852827)

[3.2.3基本面数据分析步骤 11](#_Toc8852828)

[3.2.4程序设计 11](#_Toc8852829)

[3.2.5筛选结果 11](#_Toc8852830)

[3.3 金融技术分析 12](#_Toc8852831)

[3.3.1历史交易数据获取及预处理 12](#_Toc8852832)

[3.3.2资本资产定价模型（CAPM）应用 13](#_Toc8852833)

[3.3.3数据正态性检验 16](#_Toc8852834)

[3.3.4筛选股票并统计收益率 18](#_Toc8852835)

[3.4 投资组合模型构建 20](#_Toc8852836)

[3.4.1使用夏普比率构建模型 21](#_Toc8852837)

[3.4.2使用方差构建模型 22](#_Toc8852838)

[3.5 本章小结 23](#_Toc8852839)

[第4章 实证分析 24](#_Toc8852840)

[4.1 JoinQuant平台策略构建 24](#_Toc8852841)

[4.2 实证结果与分析 24](#_Toc8852842)

[4.2.1以夏普比率为指标的模型测试 24](#_Toc8852843)

[4.2.2以方差为指标的模型测试 26](#_Toc8852844)

[4.3 本章小结 29](#_Toc8852845)

[第5章 结论 30](#_Toc8852846)

[5.1 总结 30](#_Toc8852847)

[5.2 不足与展望 31](#_Toc8852848)

[致谢 32](#_Toc8852849)

[参考文献 33](#_Toc8852850)

# 第1章 绪论

## 1.1 选题背景

中国资本市场标准化，市场化和国际化的发展趋势不断提高了证券公司的竞争力和风险管理水平，使得证券市场日益规范，越来越多的人开始接触股票并且投资股票。纵观我国股票市场，伴随着我国经济的高速发展，股市也蓬勃发展起来，上证指数在2015年也突破了5000点的高位。但是随后而来的却是千股跌停，这对于投资者来说无疑是一个巨大的考验。股票分析主要分为基本面分析与金融技术分析，基本面分析需要投资者从企业的财务报表、资产重组与关联交易等信息中去探究企业的内在价值，基本面分析更能把握股票的长期趋势。金融技术分析是从股票的供需表现，市场价格和交易数量等市场因素上分析股票价格的波动规律，金融技术分析更能把握股票的短期趋势。只有将两种技术结合到一起，才能分析出值得投资的股票，再从这些值得投资的股票中构建投资组合模型。

随着信息技术的快速发展，股票市场每天都有很多数据产生，而这些数据就是帮助我们分析的最好资源，也是最珍贵的资源。并且随着数据分析技术的发展，特别是Python语言的发展，使得数据分析更加可行与简便，门槛也相对更低。通过数据分析技术，分析出股票市场潜在的有用信息与规律，指导投资者进行有效的投资，变得十分重要。同时也可以通过数据分析技术探究股票市场[1]。

## 1.2 国内外研究现状

近年来，计算机科学与高等数学、统计学的结合已经使金融领域产生了革命性的变革，数据挖掘也开始逐渐应用于股票建模中，以便做出更快更准确的决策。欧美国家的金融市场发展更早，他们对于数据挖掘技术的应用也更早。很多知名金融机构拥有一大部分从事数据挖掘和数据分析的专业计算机技术人员来为他们构建更有效的模型。2008年N S Thomaidis等人采用了统计假设检验的模型选择方法研究了基于神经网络模型的选股策略，对多个神经网络模型进行比较，实验表明神经网络模型和股权定价理论结合的效果更好。Zhang Yong等人在2015年将非监督学习特效结合了金融理论知识应用于选股策略中，构建出单一聚类策略，该策略可以挑选出最优的股票[2]。

本文将采用Python实现股票分析与选择。Python诞生于20世纪90年代初，如今Python已经发展为一种强大的编程语言，其可提供的开源库方便用户零成本入门；Python语言也很容易学习，代码简单，可读性高；像IPython这样有力的第三方工具和时间序列分析库Pandas等都能协助用户更有效的应用于金融分析中；当然Python处理金融数据时在性能上并不输于C，并且还能通过Numba库来提高性能；其强大的生态系统，已经确立了他在金融产业中的地位。[3]

Python的市场占有率也比较高，2012年第三季度伦敦地区（主要是金融服务）Python开发合同量为485，而2013年第三季度，数量已经达到了864，大型金融机构比如美国银行，美林和摩根有数百万行Python代码在运作，这些代码主要用于风险管理方面，Python在对冲基金行业也非常受欢迎。

## 1.3 研究目的

金融领域包含很多主题，例如资金，储蓄，投资和保险。本文将专注于股票投资，因为关于股票价格的数据非常丰富。根据学术理论，普通投资者不应该投资于个股，而应该是整个市场，比如一个国家里比较出名的大公司股票所组成的一篮子股票。因为金融市场具有随机性，通过挑选股票想击败平均篮子非常困难。并且个股随着价格波动而起伏，这些个股可以在篮子中得到平均，从而降低投资风险。

本文将使用Python作为开发语言，JoinQuant作为测试平台，结合金融理论知识，以及数据分析方面的知识，分析公司基本面数据、历史交易数据。构建和设计出有实用价值，测试效果良好的模型，从而挑选出投资风险低，收益率高的组合。并应用于实际市场，帮助投资者更有效的投资。

## 1.4 研究意义

如果使用传统的股票分析方法。对于基本面分析，需要对一定的指标，如净资产收益率（ROE），净利润增长率，股利收益率等，并且结合国家的财政政策，货币政策等进行分析，往往含有个人主观判断。对于金融技术分析，投资者往往需要一定的投资经验，难以在短时间内分析出有用的结果，并且投资者对于基本面的分析只能针对少数几支股票，难以对大量股票一起分析。

我国股票市场虽没有欧美市场那样发达，但是我国股市发展很快，积累了较为齐备的金融数据，对于非金融行业的研究人员来说，数据就是最可靠的经验。使用Python来分析数据，选股视野可以变得更加开阔，还可以有效规避错误数据带来的不良影响。

如何利用好这些数据，如何从这些数据中找到隐含的有价值的信息，并且能找出风险较低，收益较高的投资组合成为本文最主要的研究意义。

## 1.5 研究方法

本文将使用文献与理论研究法以及工程实践法。

首先阅读国内外相关文献以及查阅相关资料，了解在现有投资模式下，国内外研究现状。并且学习一定的金融知识，了解一些经典、有名的金融理论，金融模型。阅读金融数据的分析相关文献，了解金融数据分析的原理及应用方法。

经过一些理论研究后，使用工程实践法。首先对公司的基本面数据进行分析，挑选出基本面数据符合要求的股票。然后获取股票历史数据，对数据进行处理及分析，筛除不合格的股票。再使用理论研究过的方法，构建模型。接着在JoinQuant平台上对模型选出的组合进行测试。分析测试结果。

## 1.6 本章小结

本章首先简单的介绍了选题背景，对国内外研究现状进行概述。接着概述了研究目的和研究意义，最后明确了研究方法。

# 第2章 相关理论与技术

## 2.1 相关理论

### 2.1.1对数收益率

在金融中使用较多的是简单收益率和对数收益率。对数收益率的公式如公式（1）：

 （1）

公式（1）中，表示各期股价。

对数收益率相较于简单收益率有很多好处。对数收益率是对称的，而简单收益率不对称。收益率是一个复利的概念，单纯使用简单收益率难以研究，而将收益率取对数后，降低了维度和复杂度，加快了运算速度和效率，使其具有可加性，采用对数收益率可以比较直观的感受到一段时间内的得利收益率。对数收益率之和的负数的绝对值越大，说明跌幅越大。

### 2.1.2资本资产定价模型（CAPM）

资本资产定价模型是现代金融学的理论基础。其定义如公式（2）：

 （2）

公式（2）中表示股票的预期回报率，表示无风险回报率，表示Beta系数，表示市场回报率,表示市场溢价。可以反映大众理论上可以接受的不承担任何风险的收益。[4]

Beta系数可以衡量股票相对于总体市场的波动性。其定义如公式（3）：

 （3）

其中表示股票与市场收益率之间的标准差，表示市场收益率的方差。

如果Beta系数为1，表示股票与总体市场的波动性一致，例如Beta系数为1.5，表示当市场波动10%时，股票会波动15%。可以指导投资者在市场价格下跌时，买入Beta系数较小的股票，在市场价格上涨是，买入Beta系数较大的股票。

资本资产定价模型告诉我们只有投资了高风险的股票，才能获得高回报。

资本资产定价模型也存在一些缺陷，首先是资本资产定价模型假设的条件太过美好，其次是对于缺乏历史数据的股票难以确定Beta系数。

### 2.1.3偏度和峰度

（1）偏度

偏度是一种衡量统计数据分布偏斜的程度[5]，其定义如公式（4）：

 （4）

公式（4）中表示三阶中心距，表示标准差。

偏度有三种情况，它能反应出数据是否对称。如果偏度为0，表示该数据分布对称，否则可能偏向左边或者右边。

（2）峰度

峰度可以衡量统计数据实际分布与理论正态分布相比的陡峭程度，其定义如公式（5）：

 （5）

公式（5）中表示四阶中心距，表示标准差。

峰度同样有三种情况，它能反应出数据是否符合正态分布。如果数据的峰度值为0，表示该数据符合正态分布，如果显著偏离于0，表示该数据不符合正态分布。

将偏度与峰度结合考虑，可以判断股票的对数收益率数据是否符合正态分布。

投资组合的建立可以从符合正态分布的股票中选取。

### 2.1.4夏普比率（Sharpe Ratio）

夏普比率是在资本资产定价模型上发展起来的，是一项基础性投资指标，其定义如公式（6）：

 （6）

公式（6）中表示投资预期报酬率，表示无风险利率，表示投资的标准差。

夏普比率可以同时考虑投资的收益与风险。它可以挑选出在固定风险水平下投资收益最大或者在固定投资收益下风险最小的股票。夏普比率为正时，表示股票投资收益高于风险，而当夏普比率为负时，表示股票投资收益低于风险。所以，夏普比率越高表示股票的投资性价比越高。夏普比率对于单只股票的衡量更准确。

### 2.1.5现代投资组合理论模型（MPT）

投资界常说风险分摊，或者是一篮子股票。其含义就是投资者选择两支及以上的股票形成的投资组合进行投资，投资组合中的每支股票不一定都会盈利，但是通过组合在一起，可以平摊风险，降低投资风险，高于市场平均盈亏水平。

期望效用假设认为投资者都是“风险厌恶者”，投资者都希望投资风险较小的资产。马科维茨提出的投资组合理论正是通过均值-方差（Mean-Variance）分析来构建出最小化收益率变化方差（风险），最大化投资收益的投资组合。这样的组合可以有效分摊风险。

双资产投资组合理论定义如公式（7）：



 （7）

公式（7）中和表示投资组合中的各股票权重，两者总和为1。表示投资组合中各股票的投资收益期望。

夏普比率就是综合考虑了风险和收益的指标，同样夏普比率也考虑了最大化投资收益且最小化收益变化方差。投资组合理论就可以通过使用夏普比率来优化模型。投资组合模型计算投资组合收益和收益方差的比值。该比值越高，说明该投资组合风险越小，收益越高。更适合组合在一起投资。

## 2.2 相关技术

### 2.2.1基本面分析

基本面分析的主要对象是宏观经济状况、利率水平、通化膨胀、公司的基本素质等[6]。分析的数据主要是公司的财务数据等。基本面分析适合长线投资的股民，不利于散户或长线投资的股民。

基本面主要包括宏观基本面和微观基本面。宏观基本面是研究国家政策，市场价值，行业价值等；微观基本面主要研究公司的发展情况。

微观基本面包含了一家公司的很多信息。通过微观基本面分析可以分析出企业的资产，负债和收益，它能反映出一家公司的健康状况和公司股票的内在价值。股票价格将围绕价值不断上下波动，投资者可以在股票价格低于内在价值的时候买入，在股票价格偏离内在价值的时候卖出[7]。股票的价值是真实的，理性的，抗跌性强的，而价格可能是虚化的，也可能是人为吹高的或者庄家做出来的。因此光分析价格风险较大，只有先分析出股票的内在价值才能更有利于投资。

### 2.2.2金融技术分析

金融技术分析主要是将股票的历史交易数据作为研究对象，金融技术分析专注于价格变动，成交量，时间等指标和各种分析工具来评估股票的优势和劣势。

金融技术分析基于三个假设：一、所有信息都将会在市场上得到反映。二、市场价格会沿着趋势进行波动。三、历史会不断重演[8]。使用相应模型进行金融技术分析，可以判断股票的大致走势，从而判断该股票值得投资。金融技术分析更适合做短期投资。[9]

基本面分析和金融技术分析各有优势，可以将两种分析方式结合在一起，分析出值得投资的股票，再从中构建投资组合模型。

### 2.2.3Python技术

随着信息时代的发展，Python作为一门可以服务于人工智能，大数据，Web开发等领域的语言备受关注。Python之所以能服务这么多领域，得益于其强大的扩展库。例如：NumPy，Pandas，SciPy，StatsModels等是数据科学的核心库。Numpy，Pandas都有其特有的数据类型。SciPy可以解决线代，概率，高数上的计算问题。StatsModels包含了常用的模型。同时在人工智能方面，Python拥有像Scikit-learn，TensorFlow这样优秀的库。这些库都可以很好的服务于用户构建自己想要的模型。为了将数据等实践效果展示出来，Python还提供了Matplotlib，Seaborn等绘图库。这些库提供的绘图样式多样，调用方法简单。同样，Python代码简单易学也是其一大特点。[10]

随着Python的发展，在数学界非常有名的MATLAB的光环越来越弱。Python完全免费，而MATLAB收费昂贵。Python的第三方库越来越完善，能做的事情更多。Python占用空间小，MATLAB占用空间大。这些特点，就足以让用户选择Python了。

使用TensorFlow可以轻松快速的构建自己的神经网络。TensorFlow擅长的任务就是训练深度神经网络。通过使用TensorFlow我们就可以快速的入门神经网络，大大降低了深度学习（也就是深度神经网络）的开发成本和开发难度。TensorFlow 的开源性，让所有人都能使用并且维护，巩固它。 使它能迅速更新，提升。

但是Python更应该规范其第三方库的生态环境。让用户有更好的体验。

### 2.2.4JoinQuant平台介绍

JoinQuant是一个量化交易平台。用户在平台上编写自己的交易策略，用平台的数据进行回测。通过回测，可以验证自己编写策略的有效性。JoinQuant提供对股票，期货，基金等品种的回测。[11]

交易策略使用Python编写。策略程序架构包括initialize，handle\_data，等函数，其中initialize为初始化函数，用来初始化一些全局变量；handle\_data在每个单位时间开始时会被调用一次；可以使用order函数来按股数下单；cancer\_order来撤单等。

使用JoinQuant平台可以快速，方便的构建自己的交易策略。回测结果以可视化界面显示出来，让用户更直观的查看策略有效性。

## 2.3 本章小结

本章主要介绍了研究所需的理论和技术。金融中相关理论有：对数收益率，资本资产定价模型（CAPM），夏普比率和现代投资组合理论（MPT）等。金融中的相关技术有：基本面分析和金融技术分析。还使用了热门的Python语言和JoinQuant平台。

# 第3章 股票数据分析

## 3.1 股票数据分析总体框架图

股票数据分析按基本面分析，金融技术分析，投资组合模型构建总体流程进行。金融技术分析中使用正态性检验和资本资产定价模型（CAPM）来筛选股票，再统计筛选后股票的收益率。在投资组合模型构建的时候分别使用夏普比率和方差作为构建指标。

股票数据分析总体框架图如图3.1所示。

图3.1 股票数据分析总体框架图

通过基本面分析筛选出一批具有较好前景，较高内在价值的股票供金融技术分析筛选。通过金融技术分析筛选出一批收益具有正态性特征且在市场中相对安全的股票供投资组合模型构建。最后构建投资组合模型来挑选最优投资组合。

## 3.2 基本面分析

基本面有很多指标，如：每股收益（EPS），市盈率（PE），市净率（PB），每股净资产等。而比较经典且具有代表性的是净资产收益率（ROE），净利润增长率，主营业务收入增长率。

净资产收益率可以很好的衡量上市公司的盈利能力。是指净利润与平均净资产的比值。可以代表股东投入产出比，该指标越高，表示投资带来的收益越高，股东投入资本后创造的利润越多，指标越低，表示公司所有者权益的获益能力越弱。如果净资产收益率稳定，说明公司的盈利能力也比较稳定，投资带来的风险较小。

净利润增长率是公司当期利润较上期利润的增长幅度。净利润增长率越多，表示公司的盈利能力越强。净利润是公司的税后利润，是可供出资人分配的利润。它代表了一个企业的效益，净利润越多，公司效益越好。

主营业务收入增长率是公司当期主营业务收入较上期主营业务收入的增长幅度。它可以用来衡量一个公司的产品生命周期，可以判断公司所处阶段，如果主营业务收入增长率超过10%，表示该公司主营产品处于成长期；如果主营业务收入增长率在5%-10%之间，说明该公司主营产品正处于稳定期，不久将进入衰退期，需要着手研发新产品[12]。

通过基本面相关数据分析，可以筛选出一批具有良好发展前景、有较高内在价值的公司供之后做金融技术分析。

### 3.2.1基本面数据获取

基本面分析的数据来自于Wind数据库。该数据包含了中国深圳证券交易所和上海证券交易所内所有共计3608支A股数据。数据按照季度给出，同时将数据时间设定在了2015年至2018年共四年。数据共有52列，前4列为股票基本信息，分别为：证券代码，证券简称，上市时间和上市板。其中证券代码表示股票在证券市场中的代码，具有唯一性；证券简称表示股票的中文名称，也具有唯一性。上市时间列出了股票上市的年，月，日；上市板分为主板，创业板，中小板，和新三板，能反应公司的实力。5-20列为2015年至2018年各股票每季度的净资产收益率。21-36列为2015年至2018年各股票每季度的净利润增长率。37-52列为2015年至2018年各股票每季度的主营业务收入增长率。数据存放在Excel表中，且列名均为汉字。

### 3.2.2基本面数据预处理

列名为汉字不利于使用Python处理，因此将所有列名用英文表达。如：可将“证券代码”换为SecuritiesCode表达，将“净资产收益率ROE(平均)[报告期] 2015中报[单位] %”这样复杂的列名换为ROE2015\_2表达。同时，本数据为2015年-2018年的股票数据，其中存在一些公司是在2015年或2015年以后才上市的，这些公司的某些数据是空值，这也不利于后期处理。因此将所有空值的数据改为0值。同时为了之后的数据分析，添加了新列：年度净资产收益率变异系数，并将其命名为ROE\_C.V，变异系数为样本总体标准差除以样本均值。其可以反应年度净资产收益率的稳定性。以上处理操作均在Excel表中完成。

### 3.2.3基本面数据分析步骤

基本面分析的指标选择了上文提到的三种，即：净资产收益率（ROE），净利润增长率，主营业务收入增长率。现将规定如下筛选规则：一是连续四年年度净资产收益率不能低于10%，且每个季度的净资产收益率不能低于0%；二是年度净资产收益率变异系数小于0.5；三是连续四年年度净利润增长率不能低于10%，且每个季度的净利润增长率不能低于0%；四是连续四年年度主营业务收入增长率不能低于10%，且每个季度的主营业务收入增长率不能低于0%；五是只选择2015年之前上市的公司。

净资产收益率保持在15%说明这是一家绩优的公司，但是低于10%，公司就需要采取措施来提高业绩了。将净资产收益率设定在10%可以筛选出一批绩效不错的公司；将变异系数设定在0.5可以筛选出业绩稳定的公司；将净利润增长率设定在10%可以筛选出一批效益不错的公司；将主营业务收入增长率设定在10%可以筛选出一批成长型公司；同时将每季度增长率都设定在0%以上可以使筛选结果更稳定；将上市时间设定在2015年之前可以将数据信息较少的股票过滤掉。

通过筛选，可以找出一批业绩稳定良好，且处于成长阶段以及有较好内在价值的公司，在这样的公司中选股，可以有效规避风险，增加投资成功率。

### 3.2.4程序设计

定义名为fundamental\_analysis的函数用于处理数据。使用pandas库中的read\_excel函数读取Excel表格存于data变量中。将读取的数据转换为DataFrame数据类型存于df变量中。使用DataFrame数据类型可以方便筛选。通过DataFrame特有的数据筛选方式即使用pandas提供的loc函数来按以上提出的规则筛选数据。筛选的结果是DataFrame数据类型，通过tolist函数将SecuritiesCode和SecuritiesShortName两列数据，即证券代码和证券简称转换为list数据类型。转换为list数据类型是方便之后使用。最后返回筛选结果至infoClass类中的STOCKS，STOCKS\_ZH变量中。

### 3.2.5筛选结果

通过分析处理，共筛选出了19支股票有小天鹅A（000418），泸州老窖（000568），银轮股份（002126）等，如表1所示。

表1 基本面分析股票筛选结果

| 证券代码 | 证券简称 | 上市时间 | 上市板 |
| --- | --- | --- | --- |
| 000418.SZ | 小天鹅A | 1997年3月28日 | 主板 |
| 000568.SZ | 泸州老窖 | 1994年月9日 | 主板 |
| 002126.SZ | 银轮股份 | 2007年4月18日 | 中小企业板 |
| 002223.SZ | 鱼跃医疗 | 2008年4月18日 | 中小企业板 |
| 002372.SZ | 伟星新材 | 2010年3月18日 | 中小企业板 |
| 002677.SZ | 浙江美大 | 2012年5月25日 | 中小企业板 |
| 002717.SZ | 岭南股份 | 2014年2月19日 | 中小企业板 |
| 300003.SZ | 乐普医疗 | 2009年10月30日 | 创业板 |
| 300015.SZ | 爱尔眼科 | 2009年10月30日 | 创业板 |
| 300124.SZ | 汇川技术 | 2010年9月28日 | 创业板 |
| 300136.SZ | 信维通信 | 2010年11月5日 | 创业板 |
| 300207.SZ | 欣旺达 | 2011年4月21日 | 创业板 |
| 300244.SZ | 迪安诊断 | 2011年7月19日 | 创业板 |
| 600276.SS | 恒瑞医药 | 2000年10月18日 | 主板 |
| 600340.SS | 华夏幸福 | 2003年12月30日 | 主板 |
| 600419.SS | 天润乳业 | 2001年6月28日 | 主板 |
| 600566.SS | 济川药业 | 2001年8月22日 | 主板 |
| 603288.SS | 海天味业 | 2014年2月11日 | 主板 |
| 603368.SS | 柳药股份 | 2014年12月4日 | 主板 |

## 3.3 金融技术分析

### 3.3.1历史交易数据获取及预处理

金融技术分析数据来自于Yahoo金融，直接调用Yahoo金融接口即可获取相应股票的历史数据。可以指定获取数据的开始时间和结束时间，如果没有指定，则获取股票从上市至今的所有数据。

数据共有7列，其中Data表示时间；High表示最高价；Low表示最低价；Open表示开盘价；Close表示闭盘价；Volume表示成交量；Adj Close表示复权后的收盘价，可以更好的放映股票的表现。其中Adj Close和Volume是本研究中比较常用且有价值的数据。

数据来自Yahoo平台，拥有一定的完整性，无需过多处理。也可以通过dropna函数滤除缺失数据。

### 3.3.2资本资产定价模型（CAPM）应用

（1）数据分析步骤

首先获取股票历史数据，数据参考时间为2018年1月1日至2019年5月11日。根据资本资产定价模型（CAPM），获取到股票的数据后，分别计算每支股票的平均年化收益率和Beta系数。然后根据各股票的平均年化收益和Beta系数拟合出市场安全线（Security Market Line），市场安全线可以反应股票的定价是否合理。这里选择上证指数（000001）作为市场参考基准，因为上证指数能较好的反映出大盘的总体走势。最后将每只股票的Beta系数与市场安全线比较，低于市场安全线的列为不安全的股票，存入文件中供再次筛选。

（2）程序设计

经过基本面分析筛选后的股票证券代码已经存放在infoClass类中的STOCKS变量中。调用pandas\_datareader的data接口，使用get\_data\_yahoo函数即可获取相应股票在Yahoo金融中的历史数据。

定义calc\_beta函数来计算每只股票的Beta系数和平均年化收益率。函数实现方法如下：首先获取市场参考基准的历史数据，即上证指数的数据，然后获取股票的历史数据。然后将市场参考基准历史数据和股票历史数据整合成有三列的DataFrame数据类型。接着分别计算市场参考基准和股票的对数收益率，计算对数收益率的方法为先调用log函数求出对数，再使用diff函数，即差分法，即可求出数据的对数收益率。然后使用最小二乘法对市场参考基准和股票的对数收益率进行一元线性回归拟合，就可以得到Beta函数，最后返回Beta系数以及平均年化收益率。计算平均年化收益率的方法为：求出对数收益率的均值，乘以一年中大致股票交易天数，即250天，再乘以100，得到平均年化收益率百分百。

使用最小二乘法对所有股票的Beta系数以及平均年化收益率进行一元线性回归拟合，即可得到市场安全线。并将分析结果用散点图绘制[13]。

使用find\_not\_safe函数筛选出不安全的股票，实现方法如下：使用poly1d函数获取市场安全线对象，将Beta系数作为x变量值代入市场安全线方程中得到结果，再讲结果与市场安全线的理论值比对，如果低于市场安全线的理论值，则归为不安全股票。使用count\_save\_not\_safe计算所有模型中不安全股票出现的次数并保存在txt文件中。

（3）结果分析

经过数据分析及处理，得到所有经基本面筛选后股票的Beta系数，平均年化收益率值，如表2所示。

表2 股票Beta系数及平均年化收益率

| 证券代码 | 证券简称 | Beta系数 | 平均年化收益率（%） |
| --- | --- | --- | --- |
| 000418.SZ | 小天鹅A | 0.983913 | 35.06878 |
| 000568.SZ | 泸州老窖 | 1.076294 | 29.05923 |
| 002126.SZ | 银轮股份 | 1.135663 | 6.406247 |
| 002223.SZ | 鱼跃医疗 | 1.005194 | 9.932455 |
| 002372.SZ | 伟星新材 | 1.122485 | 31.15374 |
| 002677.SZ | 浙江美大 | 0.97677 | 16.45511 |
| 002717.SZ | 岭南股份 | 1.032216 | 9.123661 |
| 300003.SZ | 乐普医疗 | 1.019776 | 18.3092 |
| 300015.SZ | 爱尔眼科 | 0.99018 | 34.42607 |
| 300124.SZ | 汇川技术 | 1.199477 | 11.76268 |
| 300136.SZ | 信维通信 | 1.20618 | 33.09365 |
| 300207.SZ | 欣旺达 | 1.13969 | 21.12557 |
| 300244.SZ | 迪安诊断 | 1.09298 | 0.200092 |
| 600276.SS | 恒瑞医药 | 0.727581 | 37.08674 |
| 600340.SS | 华夏幸福 | 0.980691 | 8.085196 |
| 600419.SS | 天润乳业 | 1.028023 | 8.003297 |
| 600566.SS | 济川药业 | 0.940907 | 14.46449 |
| 603288.SS | 海天味业 | 0.84769 | 34.0195 |
| 603368.SS | 柳药股份 | 1.052306 | 6.568389 |

同时给出数据分析结果的散点图，如图3.2所示。

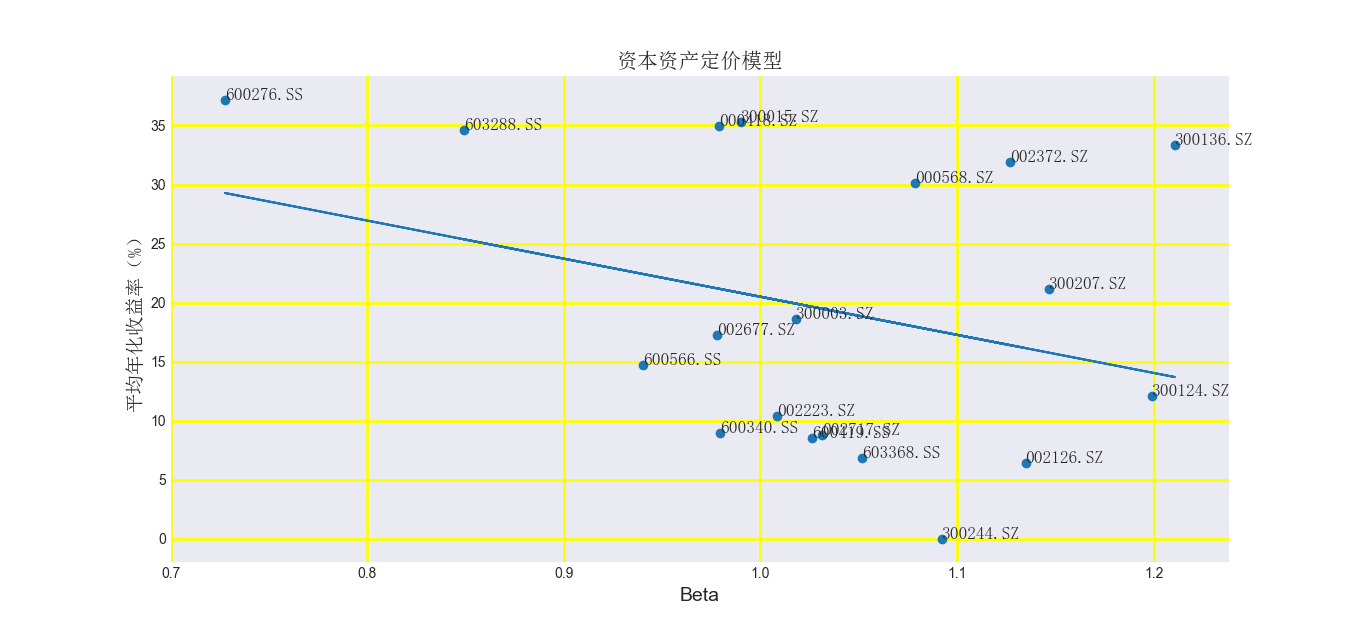


图3.2 CAPM模型应用

市场安全线方程如公式（8）：

 （8）

图3.2横坐标表示股票的Beta系数，纵坐标表示平均年化收益率。由于图像大小限制，并没有把证券简称一一打出，如600276.SS表示恒瑞医药。如果Beta系数为1，表示股票与总体市场的波动性一致，Beta系数绝对值小于1，表示股票波动性小于市场波动性。平均年化收益率可以反映股票的收益情况。

从图3.2和表2中可以看出，在市场安全线以上的共有8支股票，分别为恒瑞医药（600276）、海天味业（603288）、小天鹅A（000418）、爱尔眼科（300015）、信维通信（300136）、伟星新材（002372）、泸州老窖（000568）、欣旺达（300207）。

在市场安全线以下的不安全股票共有11支，分别为银轮股份（002126）、鱼跃医疗（002223）、浙江美大（002677）、岭南股份（002717）、乐普医疗（300003）、汇川技术（300124）、迪安诊断（300244）、华夏幸福（600340）、天润乳业（600419）、济川药业（600566）、柳药股份（603368）。

平均年化收益率较高（大于30%）的共有6支，分别为恒瑞医药（600276）、海天味业（603288）、小天鹅A（000418）、爱尔眼科（300015）、信维通信（300136）、伟星新材（002372）。股票价格波动小于大盘指数（上证指数），即Beta系数绝对值小于1的共有7支，分别为小天鹅A（000418）、浙江美大（002677）、爱尔眼科（300015）、恒瑞医药（600276）、华夏幸福（600340）、济川药业（600566）以及海天味业（603288）。

经过资本资产定价模型应用并综合分析后，认为恒瑞医药（600276）、海天味业（603288）、小天鹅A（000418）、爱尔眼科（300015）为较优秀的股票。

### 3.3.3数据正态性检验

（1）数据分析步骤

偏度和峰度可以用来检验数据是否具有正态性特征。具有正态性特征的金融数据非常利于研究。

首先获取各股票的历史数据，因为样本的增加会减少偏度值和峰度值的标准差，最终会拒绝条件假设，所以数据参考时间为2018年12月1日至2019年5月11日各计100个样本。计算各股票的对数收益率。使用字典类型变量将所有股票的对数收益率数据保存起来。

接着计算各股票的偏度和峰度，将数据结果展现在散点图上。通过计算各股票的偏度值与总偏度值的标准差的比值得到Z评分，用以评价偏离合理性。对峰度值同样操作，在在α=0.05的检验水平下，若Z评分在±1.96之间，则可以认为数据服从正态分布。[14]

根据正态性检验的规则，将不符合一定要求的股票认为是不安全的股票，筛选出这些股票放入文件中。

（2）程序设计

使用get\_data\_yahoo函数获取股票的历史数据，使用同样方法计算股票的对数收益率。接着使用skew和kurtosis函数分别计算股票对数收益率的偏度和峰度，然后将数据结果展现在散点图上。

使用np.std函数计算偏度与峰度的标准差值，将各股票偏度与偏度标准差值比值的绝对值（偏度值Z评分）超过1.96，或者各股票峰度与峰度标准差值比值的绝对值（峰度值Z评分）超过1.96的认为是不安全的股票，将其筛选出来，保存在txt文件中。

（3）结果分析

经过分析处理，将证券代码，证券简称，偏度与偏度标准差的比值（偏度值Z评分），峰度与峰度标准差的比值（峰度值Z评分）的结果展现在表3中。

表3 股票对数收益率的偏度和峰度与标准差的比值

| 证券代码 | 证券简称 | 偏度/偏度标准差  （偏度值Z评分） | 峰度/峰度标准差  （峰度值Z评分） |
| --- | --- | --- | --- |
| 000418.SZ | 小天鹅A | 0.786127 | 1.440091 |
| 000568.SZ | 泸州老窖 | 0.771076 | 0.552267 |
| 002126.SZ | 银轮股份 | -1.52059 | 2.239661 |
| 002223.SZ | 鱼跃医疗 | 0.683055 | 0.129124 |
| 002372.SZ | 伟星新材 | -0.137297 | 0.437838 |
| 002677.SZ | 浙江美大 | 0.562264 | 0.013354 |
| 002717.SZ | 岭南股份 | -2.282202 | 3.546691 |
| 300003.SZ | 乐普医疗 | -1.177584 | 0.881929 |
| 300015.SZ | 爱尔眼科 | 0.015038 | 0.210253 |
| 300124.SZ | 汇川技术 | 0.724419 | 1.12725 |
| 300136.SZ | 信维通信 | -1.358709 | 2.497759 |
| 300207.SZ | 欣旺达 | -0.231955 | 1.810048 |
| 300244.SZ | 迪安诊断 | 0.753857 | 0.904525 |
| 600276.SS | 恒瑞医药 | -0.619842 | 0.272449 |
| 600340.SS | 华夏幸福 | 0.084004 | 0.654772 |
| 600419.SS | 天润乳业 | -1.816624 | 3.04946 |
| 600566.SS | 济川药业 | -1.061546 | 0.869607 |
| 603288.SS | 海天味业 | 1.196109 | 0.508304 |
| 603368.SS | 柳药股份 | -0.086713 | 0.512724 |

其中偏度的标准差值为0.6403459121979265，峰度的标准差值为1.637677882193852。

数据分析结果展示在散点图上，如图3.3所示。

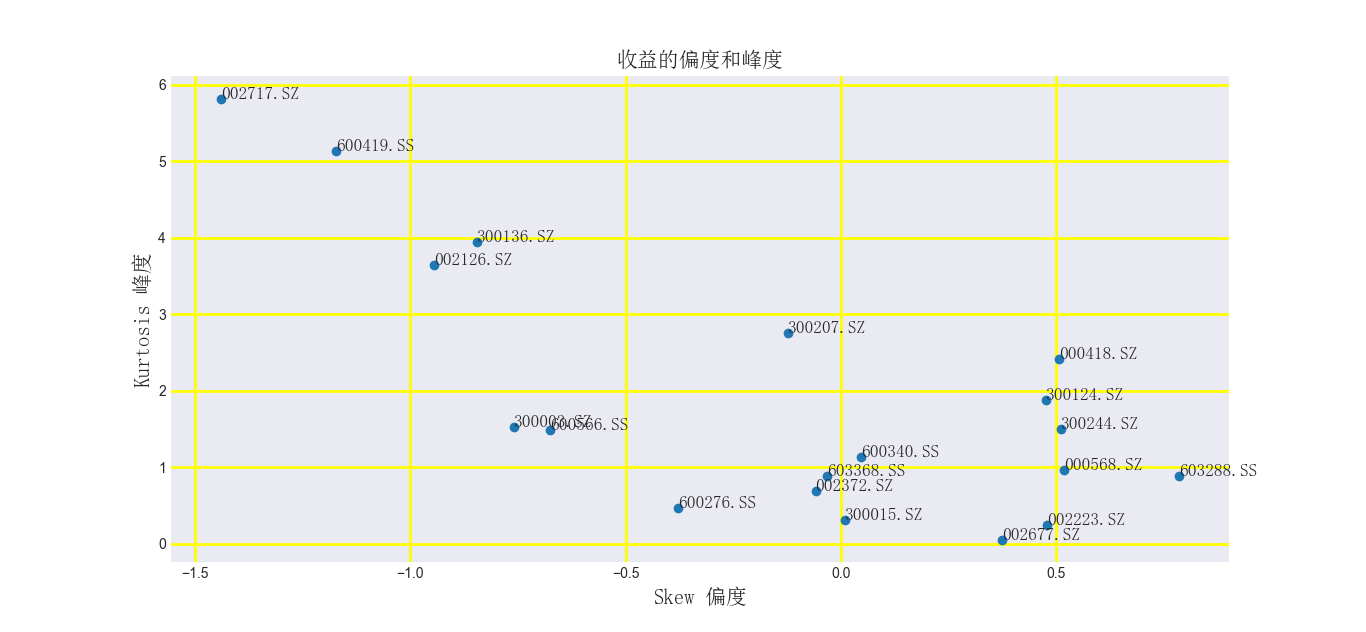


图3.3 正态性检验

图3.3横坐标表示偏度，纵坐标表示峰度。由于图像大小的限制，并没有将证券简称一一列出，如002717.SZ表示岭南股份。经过分析，找出偏离正态性太多的股票为：银轮股份（002126），岭南股份（002717），信维通信（300136），天润乳业（600419）。

### 3.3.4筛选股票并统计收益率

经过金融技术分析，已经将一批认为是不安全的股票集中在了一起，现将低于市场安全线或者偏离正态性太大的股票筛除掉。

然后统计一下所有筛选后股票相较于市场参考标准（上证指数）对数收益率的情况及各筛选后股票对数收益率总和的比对。

（1）程序设计

读取存放不安全股票的文件，将文件中所有出现过的股票在STOCKS变量中筛除掉。

数据参考时间设定在2018年1月1日至2019年5月11日。统计收益率时，首先使用get\_data\_yahoo函数获取各股票的历史数据，求出各股票的对数收益率，求所有对数收益率的总和并使用matplotlib提供的barh函数绘制条形图，以展现更好的比对效果。

将所有股票的对数收益率以及市场参考基准的对数收益率结合在一起绘制成折线图，以展现股票与市场参考基准比对的情况。

（2）结果分析

经过金融技术分析再次筛选后得到的股票有小天鹅A（000418），泸州老窖（000568），伟星新材（002372）等，如表4所示。

表4 金融技术分析筛选后结果

| 证券代码 | 证券简称 | 上市时间 | 上市板 |
| --- | --- | --- | --- |
| 000418.SZ | 小天鹅A | 1997年3月28日 | 主板 |
| 000568.SZ | 泸州老窖 | 1994年月9日 | 主板 |
| 002372.SZ | 伟星新材 | 2010年3月18日 | 中小企业板 |
| 300015.SZ | 爱尔眼科 | 2009年10月30日 | 创业板 |
| 300207.SZ | 欣旺达 | 2011年4月21日 | 创业板 |
| 600276.SS | 恒瑞医药 | 2000年10月18日 | 主板 |
| 603288.SS | 海天味业 | 2014年2月11日 | 主板 |

各股票收益率与上证指数收益率对比结果如图3.4所示。

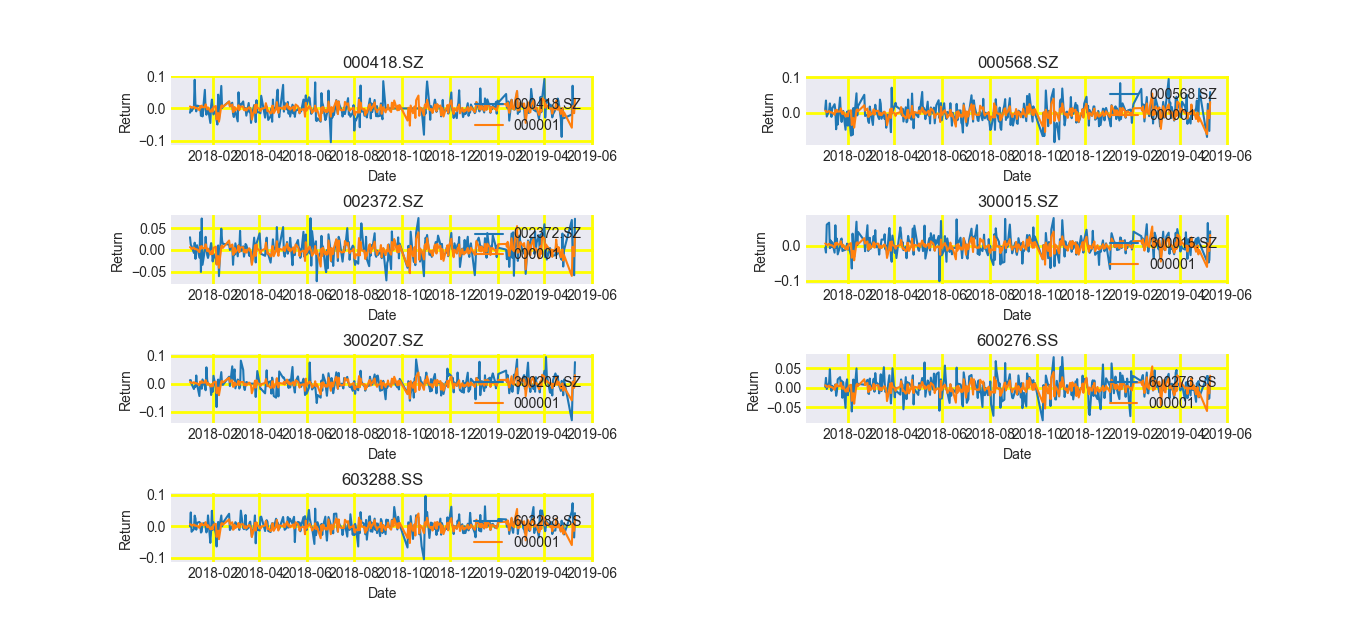


图3.4 各股票对数收益率与上证指数对数收益率对比

图3.4共有7个子图，每个子图代表一只股票与上证指数对数收益率的对比图。每个子图的横坐标代表时间，纵坐标代表对数收益率。可以从图中看出，每只股票的收益率都能超过市场参考基准（上证指数）的收益率。

将筛选后每只股票的证券代码，证券简称，股票收益率总和展现在表5中。

表5 各股票收益率总和

| 证券代码 | 证券简称 | 股票收益率总和 |
| --- | --- | --- |
| 000418.SZ | 小天鹅A | -0.13960884221777856 |
| 000568.SZ | 泸州老窖 | 0.1083104290312816 |
| 002372.SZ | 伟星新材 | 0.2600414920231694 |
| 300015.SZ | 爱尔眼科 | 0.5551806247034758 |
| 300207.SZ | 欣旺达 | 0.23953516622834398 |
| 600276.SS | 恒瑞医药 | 0.38072968172230515 |
| 603288.SS | 海天味业 | 0.5685424142252188 |
| 000001.SS | 上证综指 | -0.13031934114694468 |

将各股票收益率总和展现在图3.5中。

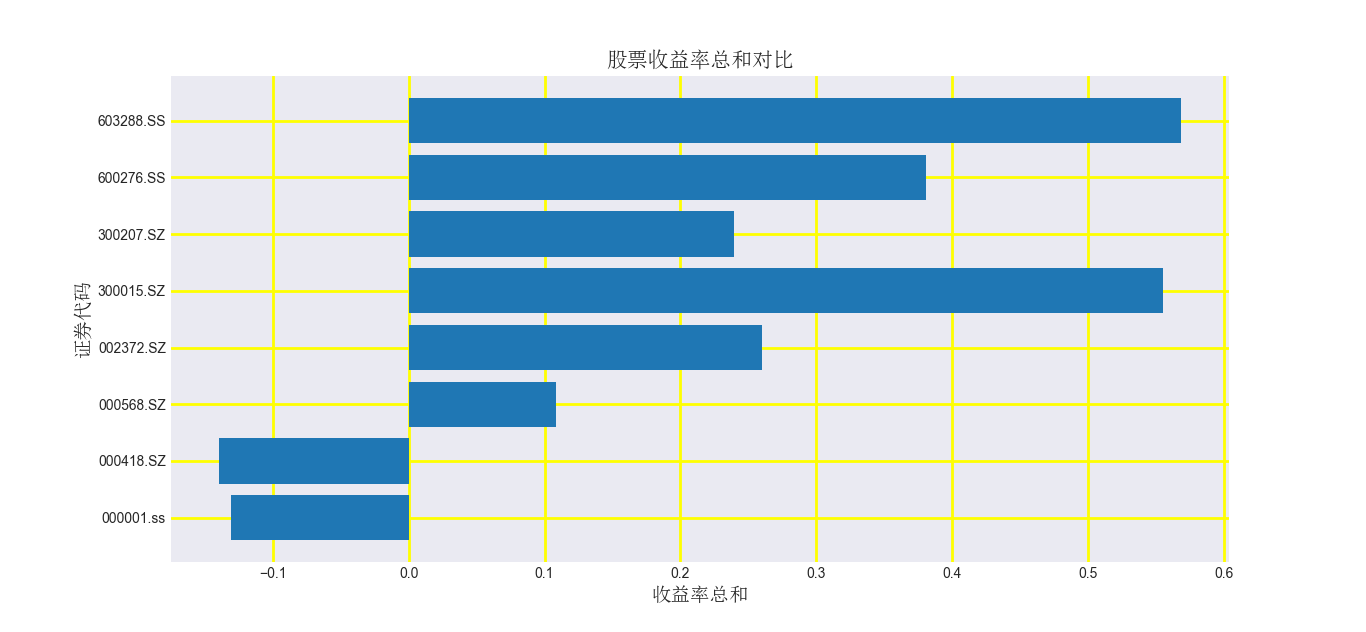


图3.5 股票收益率总和对比

图3.5横坐标表示收益率总和，纵坐标表示各证券代码。可以从图中看出经过筛选后的股票近一年多的收益率总和都为正的。且远远高于市场参考基准。

## 3.4 投资组合模型构建

根据现代投资组合理论（MPT），构建投资组合可以有效规避风险，提高投资收益。将经过基本面分析和金融技术分析后筛选出的股票两两组合，构成双资产投资组合，并且设置每只股票的投资权重相等。最后构建出等权重双资产投资组合。[15]

本节将分别使用夏普比率与方差来作为构建模型的依据，通过实际验证选出使用哪种方案最佳。

### 3.4.1使用夏普比率构建模型

夏普比率综合考虑风险与收益，使用夏普比率构建模型可以找出风险最小化，收益最大化的组合。

（1）构建步骤

首先获取经过基本面分析和金融技术分析后筛选过的各股票的历史数据，接着计算每一只股票对数收益率的平均收益和标准偏差。然后分别计算投资组合收益期望与方差。接着计算投资组合收益期望与方差的比值。并使用热图将各组合的比值表示出来。

构建了投资组合后，使用TensorFlow求出组合中的最大值，即风险最小，且收益最大的组合。

（2）程序设计

首先用grid\_list函数将筛选后的股票构建出一个二维笛卡尔乘积，再使用sorted函数将笛卡尔积转换成一个三维列表，该列表将记录所有组合的夏普比率。

使用get\_data\_yahoo函数获取股票历史数据。提取Adj Close列数据，求出每只股票的对数收益率，用mean函数求对数收益率的期望，用std函数求标准差。

接着求投资组合的期望收益，使用各支股票的期望值相加再除以二即可得到每个组合的期望收益。使用如下方法求出投资组合的方差：分别求出投资组合中各支股票的对数收益率，然后使用numpy提供的corrcoef函数求出两只股票对数收益率的相关系数，最后根据上文提到的投资组合收益方差公式（权重各为0.5）即可算出投资组合的方差。

然后求出投资组合期望收益与方差的比值，即是投资组合最后的系数。

使用TensorFlow求最大比率时，先将存放比率数据的数组转换成Tensor张量。接着使用TensorFlow的reduce\_max函数直接求出最大值，可以使用eval函数将张量转换成普通数据类型，再使用argmax函数求出最大值的索引值，通过索引值可以指导股票的一切信息。

（3）结果分析

经过模型的构建，使用夏普比率作为依据的各投资组合比率如图3.6所示。

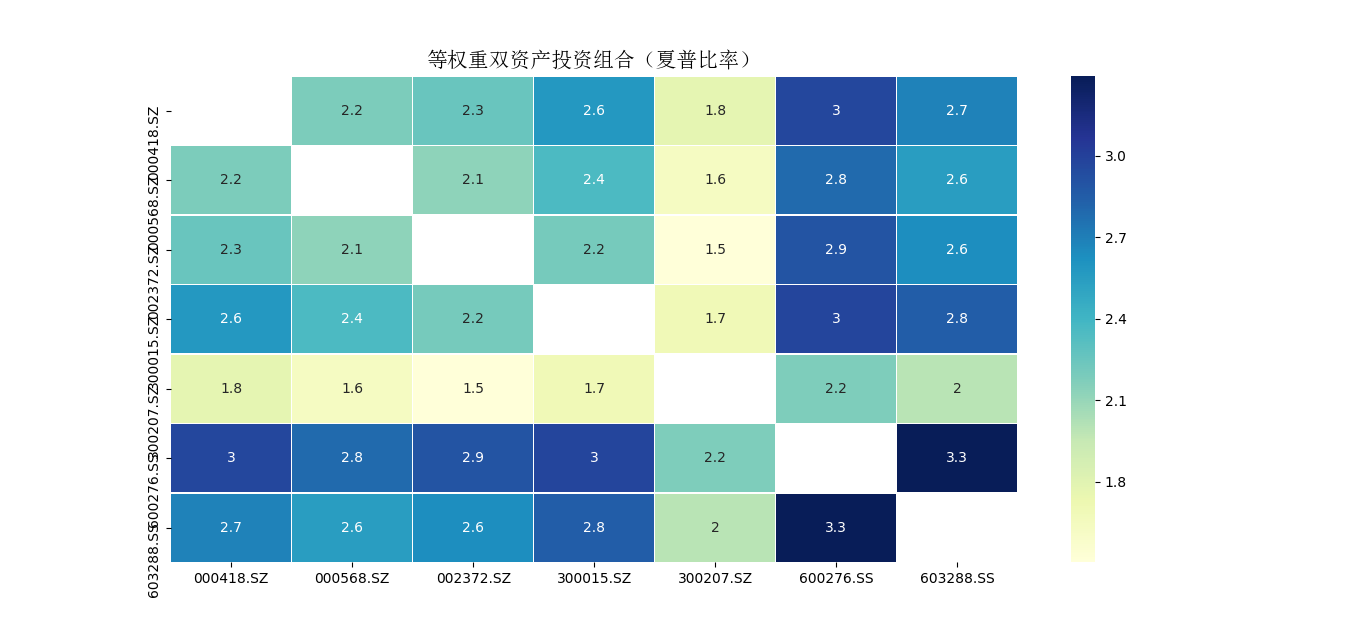


图3.6 使用夏普比率构建投资组合

图3.6横坐标及纵坐标表示各股票证券代码，每一个方块代表两个股票构成的投资组合，方块中的数字表示每个投资组合的夏普比率，对角线为空白是因为相同的股票构不成投资组合，所以也没有值。方块的数字越大（颜色越深），该投资组合的夏普比率越大，也就代表该投资组合的风险越小且收益越大。

经过TensorFlow运算，该图42个投资组合中恒瑞医药（600276）与海天味业（603288）构成的投资组合的夏普比率最大，为3.2933273。同样，还能看出其他两个夏普比率较大的投资组合为恒瑞医药（600276）与爱尔眼科（300015），该投资组合的夏普比率为3，还有一个为恒瑞医药（600276）与小天鹅A（000418），该投资组合的夏普比率也为3。

### 3.4.2使用方差构建模型

方差能代表投资风险。使用最小方差构建模型可以找出风险较小的组合。

（1）构建步骤

该模型构建步骤与使用夏普比率构建模型步骤一致，只不过使用方差作为依据构建模型不需要计算各股票的收益期望，只需要计算各股票对数收益率的标准偏差与投资组合的方差。也同样将方差值在热图上表示出来。

构建了投资组合后，使用TensorFlow求出组合中的最小值，即风险最小的组合。

（2）程序设计

程序设计与使用夏普比率构建模型的程序基本一致，只需要计算投资组合的方差。

（3）结果分析

经过模型的构建，各项投资组合的方差如图3.7所示。

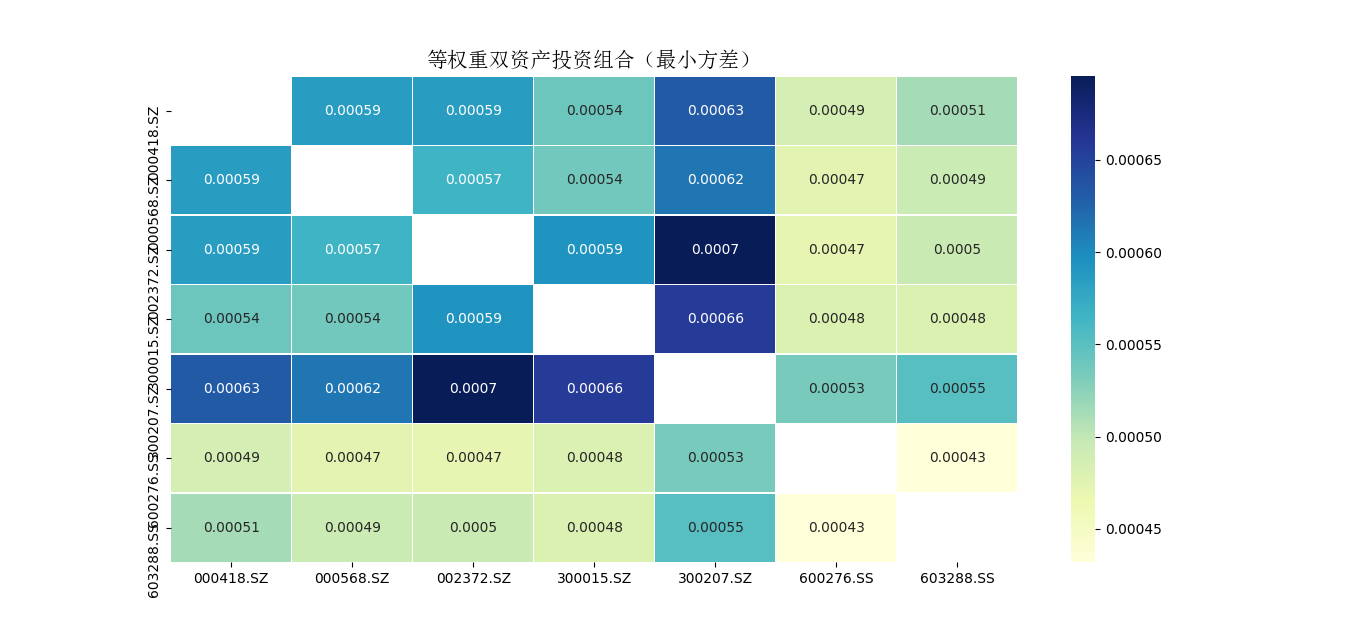


图3.7 使用方差构建投资组合

图3.7横坐标及纵坐标表示各股票的证券代码。每一个方块代表一个投资组合，方块上的数字表示该投资组合的方差。对角线空白因为相同的股票构不成投资组合。方块的数字越小（颜色越浅），表示投资组合的方差越小，表示投资组合的风险越小。

经过TensorFlow的运算，使用方差作为依据的投资组合中方差最小的投资组合为恒瑞医药（600276）与海天味业（603288），方差为0.00043182008。同样从图中还能看出其他两个方差较小的组合为恒瑞医药（600276）与伟星新材（002372），其方差为0.00047。还有一个组合为恒瑞医药（600276）与泸州老窖（000568），其方差同样为0.00047.

## 3.5 本章小结

本章主要阐述了使用Python语言研究的过程。首先对公司基本面进行分析，筛选出一批股票。接着使用金融理论对剩下的股票进行金融技术分析，再筛选掉一批股票。最后对剩余的股票分别以夏普比率和方差为指标构建了投资组合模型。

# 第4章 实证分析

## 4.1 JoinQuant平台策略构建

上一章中通过基本面分析和金融技术分析筛选出有投资价值的股票，并从有投资价值的股票中分别使用夏普比率和方差作为指标构建出投资组合模型。从两个模型中分别选出了前三个最优组合。两种指标构建的模型找到的最优组合是一样的，但是其余前两个优秀组合都不一样。

下面将使用JoinQuant平台，将以夏普比率和方差作为指标构建的模型选出的前三个最优组合依次放入平台进行测试。分析每个模型的挑选情况，并分析出最优模型。

定义initialize初始化函数，初始化一些全局变量，首先设置沪深300（000300）作为基准。因为沪深300指数结合了沪深两地较为优秀的300家公司的股票，高于市场平均水平，具有一定指导意义。接着进行股票交易的相关设定，如买入的手续费以及卖出的手续费，以模拟更真实的交易。最后设置投资组合的股票。

定义handle\_data函数来模拟每个单位时间所执行的策略。定义weights变量记录每个单位时间开始时按当前总权益等权重分配后的权益。接着计算各股票所需调整的仓位。最后用order\_target\_value函数将股票持仓调整至要求的仓位。

## 4.2 实证结果与分析

### 4.2.1以夏普比率为指标的模型测试

对以夏普比率作为指标的模型得出的前三个最好的组合依次进行回测。回测时间都设定为2015年1月1日至2018年12月31日（四年），交易单位时间设定为每天，且初始金额都为10万元整。

恒瑞医药（600276）与海天味业（603288）构成的组合回测结果如图4.1所示。

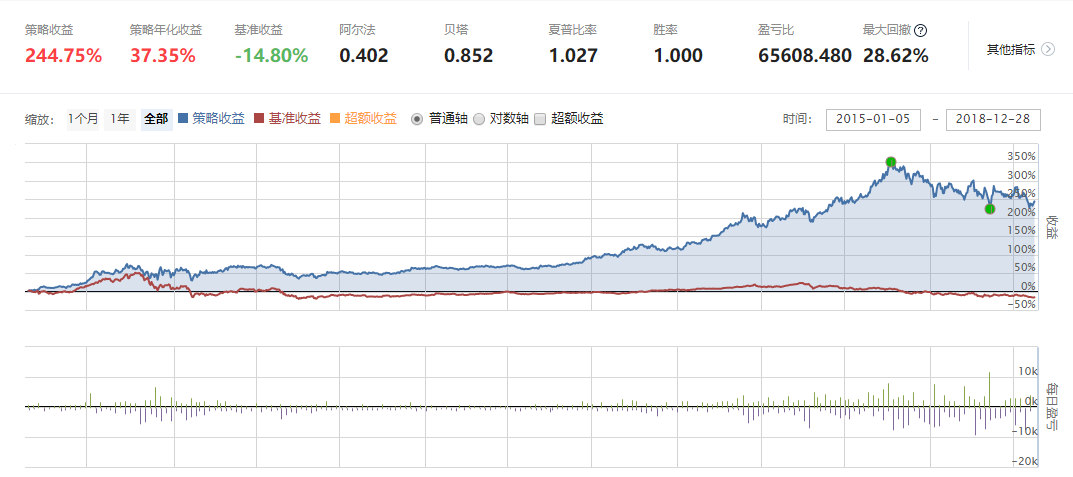


图4.1 恒瑞医药+海天味业投资组合回测效果

恒瑞医药（600276）与爱尔眼科（300015）构成的组合回测结果如图4.2所示。

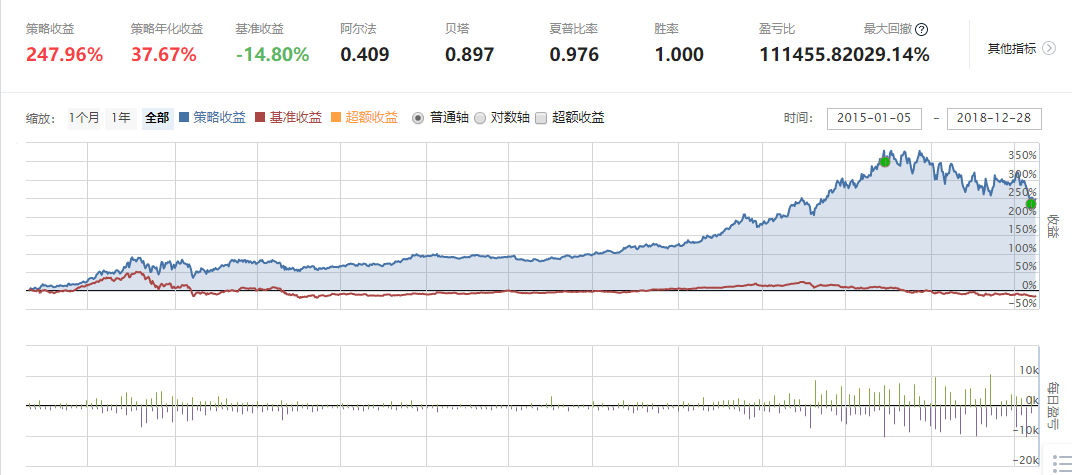


图4.2 恒瑞医药+爱尔眼科投资组合回测效果

恒瑞医药（600276）与小天鹅A（000418）构成的组合回测结果如图4.3所示。

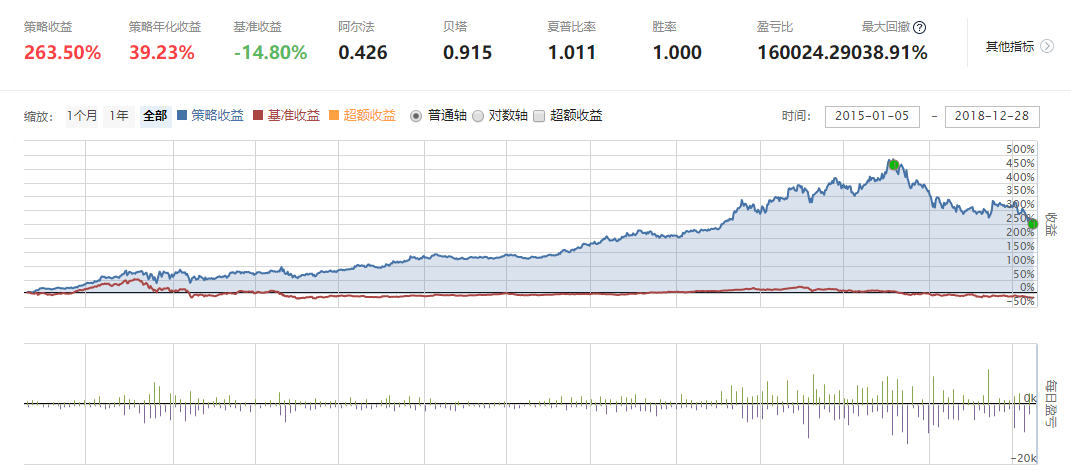


图4.3 恒瑞医药+小天鹅A投资组合回测效果

图4.1，图4.2，图4.3的横坐标表示时间，纵坐标表示收益情况。蓝色的线表示投资组合的收益情况，红色的线表示市场基准（沪深300）的收益情况。图中还给出了投资组合其他的评价指标。如：Beta系数，夏普比率，最大回撤等。

单从收益上看，收益最高的是恒瑞医药（600276）与小天鹅A（000418）构成的组合。

单从Beta系数上看，Beta系数最好的，即Beta系数的绝对值距离1最大，为恒瑞医药（600276）与海天味业（603288）构成的投资组合。Beta系数越好，表明投资组合相较于市场的波动性越小。

单从夏普比率来看，夏普比率最高的是恒瑞医药（600276）与海天味业（603288）构成的组合。夏普比率越高，表明投资组合的收益越高，风险越小。

单从最大回撤来看，最大回撤最小的，依然是恒瑞医药（600276）与海天味业（603288）构成的组合。最大回撤率表示买入产品后，可能出现的最糟糕情况，最大回撤率越低，表示组合的投资风险越小。且使用夏普比率作为指标的组合更优秀。

综上分析，使用夏普比率为指标构建的模型选择的理论最优投资组合为恒瑞医药（600276）与海天味业（603288）。通过实际验证，实际最优投资组合为恒瑞医药（600276）与海天味业（603288）。因此，使用夏普比率构建的模型是正确的。且选股质量较高。

### 4.2.2以方差为指标的模型测试

对以方差为指标构建的模型前三个最优组合进行回测，回测时间设定为2015年1月1日至2018年12月31日（四年），交易单位时间设定为每天，且初始金额都为10万元整。

恒瑞医药（600276）与海天味业（603288）构成的组合回测结果如图4.4所示。

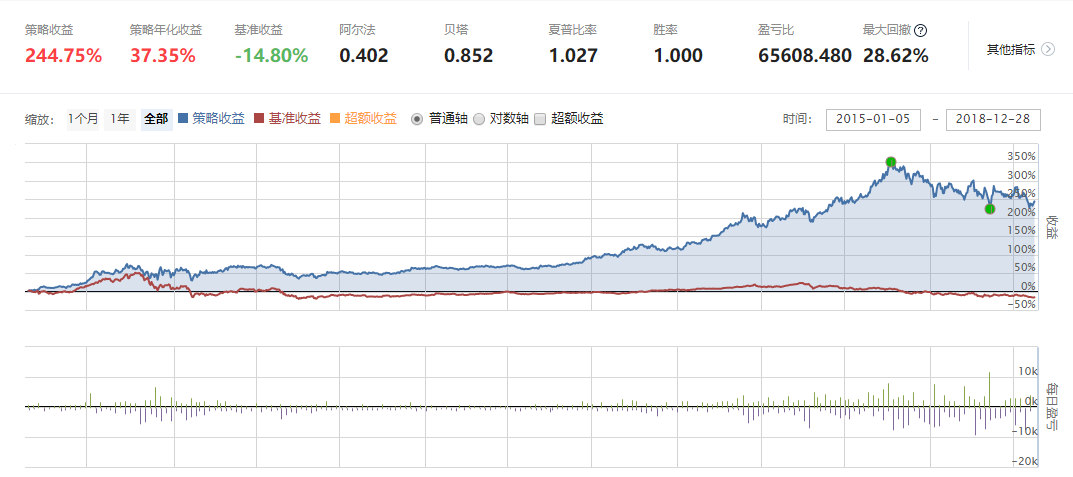


图4.4 恒瑞医药+海天味业投资组合回测效果

恒瑞医药（600276）与伟星新材（002372）构成的组合回测结果如图4.5所示。

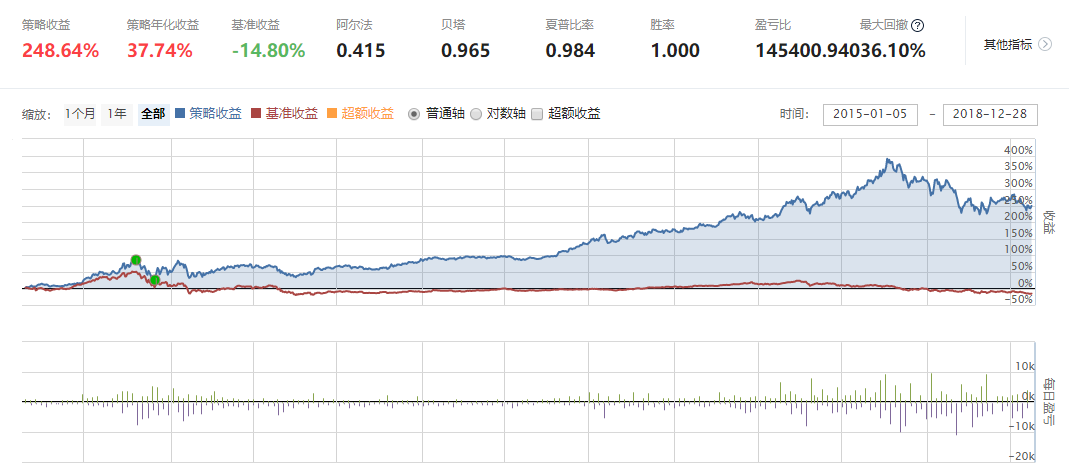


图4.5 恒瑞医药+伟星新材投资组合回测效果

恒瑞医药（600276）与泸州老窖（000568）构成的组合回测结果如图4.6所示。



图4.6 恒瑞医药+泸州老窖投资组合回测效果

图4.4，图4.5，图4.6的横坐标表示时间，纵坐标表示收益情况。蓝色的线表示投资组合的收益情况，红色的线表示市场基准（沪深300）的收益情况。图中还给出了投资组合其他的评价指标。如：Beta系数，夏普比率，最大回撤等。

单从收益上看，收益最高的是恒瑞医药（600276）与伟星新材（002372）构成的投资组合。

单从Beta系数上看，Beta系数最好的为恒瑞医药（600276）与海天味业（603288）构成的投资组合。

单从夏普比率来看，夏普比率最高的是恒瑞医药（600276）与海天味业（603288）构成的组合。

单从最大回撤来看，最大回撤最小的，依然是恒瑞医药（600276）与海天味业（603288）构成的组合。

综上分析，使用方差为指标构建的模型选择的理论最优投资组合为恒瑞医药（600276）与海天味业（603288）。通过实际验证，实际最优投资组合为恒瑞医药（600276）与海天味业（603288）。因此，使用方差构建的模型是正确的。且选股质量较高。

4.2.3模型比对分析

使用夏普比率为指标的模型选出的前三投资组合测试结果汇总如表6所示。

表6 以夏普比率为指标的模型测试结果汇总

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 投资组合名称 | 收益率 | Beta系数 | 夏普比率 | 最大回撤率 |
| 600276+603288 | 244.75% | 0.852 | 1.027 | 28.62% |
| 600276+300015 | 247.96% | 0.897 | 0.976 | 29.14% |
| 600276+000418 | 263.50% | 0.915 | 1.011 | 38.91% |

使用方差为指标的模型选出的前三投资组合测试结果汇总如表7所示。

表7 以方差为指标的模型测试结果汇总

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 投资组合名称 | 收益率 | Beta系数 | 夏普比率 | 最大回撤率 |
| 600276+603288 | 244.75% | 0.852 | 1.027 | 28.62% |
| 600276+002372 | 248.64% | 0.965 | 0.984 | 36.10% |
| 600276+000568 | 187.56% | 0.952 | 0.796 | 39.72% |

单从收益率来看，收益最高的是恒瑞医药（600276）与小天鹅A（000418）组合，使用夏普比率作为指标选出的其他组合总体优于使用方差作为指标选出的其他组合。

从Beta系数来看，Beta系数最优的是恒瑞医药（600276）与海天味业（603288）。且使用夏普比率作为指标选出的其他组合总体更优秀。

从夏普比率来看，夏普比率最优的组合为恒瑞医药（600276）与海天味业（603288）。且使用夏普比率作为指标选出的其他组合总体更优。

从最大回撤来看，最大回撤最小的依然是恒瑞医药（600276）与海天味业（603288）构成的组合。且使用夏普比率作为指标选出的其他组合总体更优秀。

综上，使用夏普比率作为指标构建的模型效果更好。且选出的投资组合质量更高。

## 4.3 本章小结

本章主要对投资组合模型选出的组合分别在JoinQuant平台上进行回测，接着分别分析每个模型的选股情况，最后比较两个模型的选股情况。

# 第5章 结论

## 5.1 总结

本文通过文献与理论研究法以及工程实践法。使用Python语言，研究分析了公司基本面数据和股票历史交易数据，从这些数据中发现一些有用的信息，并且使用这些数据探究市场，构建出投资组合选股模型，最后用JoinQuant平台测试验证模型。

本文的主要研究内容如下：（1）通过获取的公司基本面数据，对A股所有3608家公司进行基本面分析，选出一批具有良好发展前景的公司。（2）接着使用金融技术分析，分析股票的历史交易数据：首先从Yahoo金融上获取历史数据并进行预处理。使用资本资产定价模型（CAPM）确定股票的Beta系数以及市场安全线，将低于市场安全线的股票筛除掉；同时对股票的收益率进行正态性检验，将偏离正态性过大的股票筛除掉。统计筛选后股票的收益率总和情况并绘制于条形图上。（3）对筛选后的股票使用现代投资组合理论（MPT）构建等权重的双资产投资组合，分别使用夏普比率与方差作为指标构建模型。最后分别对模型选出的最优组合在JoinQuant平台上进行回测。对比分析测试结果。

通过研究比较得出，使用夏普比率作为指标构建的模型选出的投资组合质量较高，能有效规避风险，并获得较高收益。

通过本文的研究，了解到了以下知识：首先是基于Python的数据科学，从Python第三方库如NumPy，pandas等入手，了解了数据科学基础知识，并对Python及数据科学核心库有了较深入的了解。同时对量化投资，数据科学产生了浓厚的兴趣。其次，通过本文的研究，对于金融知识，证券市场也有了很多了解，知道如何获取数据，哪些数据源比较好等；了解到分析股票可以使用基本面分析和金融技术分析；对于有名的金融理论也有所耳闻，比如：现代现代投资组合理论（MPT），资本资产定价模型（CAPM）等；对于数学知识也有了新的认识，比如：差分，方差，相关系数等；对量化平台JoinQuant也有所了解。

## 5.2 不足与展望

通过一系列的研究，发现自己还存在很多问题。由于自己并不是金融专业出身，对于金融领域知识有所欠缺，导致很多有效的方法在之后才接触到，比如：在基本面分析中可以构建多因子模型，本研究中选取的仅仅只是三个指标，而构建了多因子模型之后，可以对基本面数据有更详细，准确的分析。其次是一些前沿技术，本研究仅仅只是使用了金融经典理论来构建模型，并没有使用如机器学习，神经网络技术。如果使用这些前沿技术搭建模型，准确度会更高，效果会更好。在JoinQuant构建策略进行验证回测的时候，用的只是最简单的策略，并没有涉及买入，卖出信号。

在之后的研究中，希望能获取到更精确，有丰富的数据来研究。同时要了解前沿知识与技术，将这些技术运用到研究中。不能做坐享其成者，对于新的知识，要勇于突破。希望能将这个研究一直做下去，一直完善，也能完成一些实盘交易。在大数据时代，不能只盯着计算机一门学科看，还有很多学科的有趣数据有待研究，要将多学科联系在一起，才能更好的发展。

# 致谢

大学四年在写完这篇论文时也就要宣告结束了，四年的生活过得很快。

在此，我要感谢我的指导老师——王海燕老师。王海燕老师亦师亦友，待人和蔼。大学四年给了我很多帮助。第一次上她的课是大二的时候学的数据结构，在之后的交流中慢慢熟悉了这个老师，觉得这个老师很好。王海燕老师在我参加蓝桥杯比赛时也耐心培训过我，虽然我没有拿到好的名次，但是老师还是经常鼓励我。之后，我就经常喊她海燕姐。在我考研期间因为学校临时换科目还和王海燕老师聊天，王海燕老师给我分析了形势，给我推荐了其他的学校，这让我用最短的时间适应了换科目这件事，重新投入到备考中。虽然第一次考研没有考上，我也和老师说过会二战，老师也非常支持我。在毕业季的时候，王海燕老师还给我推荐了一份实习的工作。这样的老师真的太好了，是我求学道路上的贵人。

同样，大学四年我也要感谢我的好朋友们给我的支持和鼓励，在我遇到人生转折困难的时候，始终能站在我身边。

感谢教我打球的同学，让我有了一个强身健体的爱好。

感谢学校和老师对我的培养，以及室友的理解与宽容。

感谢大学四年出现过的人。

# 参考文献

[1]陈金佑. 数据挖掘在股票分析中的应用研究[D].华南理工大学,2014.

[2]孟叶,于忠清,周强. 一种安全的量化选股策略——来自A股市场的实证[J].金融理论与实践,2018(08):102-107.

[3][德]伊夫·希尔皮斯科. Python金融衍生品大数据分析[M]. 北京:电子工业出版社, 2017. 1-6.

[4]蔡立耑. 量化投资以Python为工具[M]. 北京:电子工业出版社, 2017. 298-305.

[5]叶陈勇. 投资者情绪对金融风险度量的影响[D].浙江工商大学,2012.

[6]黄智翔. 数据挖掘技术在股价分析中的应用研究[D].云南大学,2016.

[7]冯强. 价值投资在我国的有效性研究[D].电子科技大学,2010.

[8]王朝龙. 股票交易决策模型研究[D].西南石油大学,2016.

[9]张侠. 数据挖掘技术在股票分析中的应用研究[D].重庆大学,2017.

[10][美]韦斯·麦金尼. 利用Python进行数据分析（原书第2版）[M]. 北京:机械工业出版社, 2018. 9-1

[11]王征,李晓波. Python量化交易实战入门与技巧[M]. 北京:中国铁道出版社, 2018. 19-44.

[12]张晓彤. 股票市场预期收益率研究[D]. 沈阳工业大学, 2015.

[13]严玉星. Python金融实战[M]. 北京:人民邮电出版社, 2017. 101-136.

[14][德]伊夫·希尔皮斯科. Python金融大数据分析[M]. 北京:人民邮电出版社, 2015. 275-288.

[15][印]伊凡·伊德里斯. Python数据分析实战[M]. 北京:机械工业出版社, 2017. 159-185.