

基于粤港澳大湾区指数的策略分析

一、摘要

针对第一问，依据题意，利用题目给出的指数增强策略，根据收益率、斜率、 R^2 逐个进行单项排名，将排名累计后取综合排名前十支股票，此为强势股。确定选股后建立模型，将粤港澳指数大湾区指数 30 支成分股的每天的收益率，对应得出 2011-2020 年的收益率曲线，进而对该方案做出了在收益预测方面，表现良好，累计收益达到 36.1918%，但亏损次数偏多，胜率偏低的评价。将我们的收益率曲线与粤港澳指数大湾区指数曲线进行对比，发现从 2011 年开始连续两年该策略基本跑赢粤港澳指数大湾区指数，在 2013 年后，策略的收益率无法跟上粤港澳指数大湾区指数走势，说明该策略在小熊市时期表现良好；但在单边上涨大行情中表现不佳，仍有很大改进空间。

针对第二问，题目给定策略交易的时机，因此只需要通过仓位分配来提高收益并且尽量降低风险。因此我们采用经典的仓位分配理论哈里马尔维茨的均值-方差投资组合理论来进行仓位的分配。我们利用最小方差模型和最大夏普率模型分别进行交易，比较两种交易策略在 2011 年-2020 年的收益率、波动率、方差和夏普率。采用 SLSQP 优化算法，在一定误差范围内和限制条件下，从大量随机数据组合中得出最小方差模型和最高夏普率模型的权重分配，采取该权重分配方案并购入股票。最小方差模型以年收益率 12% 的微弱优势胜于最大夏普率 10.2%。

针对第三问，题目已经给定了调仓时间和仓位的分配策略了，仓位分配策略为单只股票的投资额为本金 10%。因此我们只需要选择自己的选股策略。在选股上我们使用了多因子分析法选股。首先我们从因子池中根据因子分类选出了总市值因子、净市率、权益回报率、资产回报率、换手率相对波动率、6 日乖离率这六个因子，之后分别分析六个因子的单独收益，然后加权每支股票六个因子的 IC 值，演算后对其进行排序，选出前十支股票，此即我们的选股策略，最后从 2011 年回测到 2020 年最终获得 220.21% 的收益率，0.61 的夏普比率。

针对第四问，第四问给了一个非常自由的交易环境，可以让我们自由搭配交易策略。因此我们采用了和第三题一样的多因子模型来选股和海龟交易策略来进行交易的择时和仓位的分配。通过海龟策略设置一个亏损区间和上涨加仓的策略，得到收益率 56.72% 与较低回撤率 13.56%。这样的一个交易策略可以让我们在照顾收益率的同时降低风险。

关键词：Python、仓位管理、最优风险资产组合、多因子模型、海龟交易策略

目录

一、摘要.....	1
二、问题重述.....	3
三、问题分析.....	5
四、名词解释.....	6
五、模型的建立与求解.....	8
六、参考文献.....	19
七、附录.....	19

二、问题重述

指数增强策略采用量化增强模型，追求高于标的指数回报水平的投资，同时力求进行有效的风险控制、降低交易成本、优化投资组合。指数增强策略不会对跟踪标的成份股进行完全复制，而是会对部分看好的股票增加权重，不看好的股票则减少权重，甚至完全去掉。通过对交易成本模型的不断监测，尽可能让交易成本降到最小。综合来看，就是既做到超额收益，又控制主动风险。某证券公司选取三十支大湾区金股(表 1)构建了大湾区指数 (399999)，指数行情如图 1 所示，附录一为成份股的行情数据，附录二为大湾区指数行情数据。

表 1 大湾区指数成份股

公司名称	股票代码	公司名称	股票代码	公司名称	股票代码
分众传媒	002027	华侨城 A	000069	塔牌集团	002233
亿纬锂能	300014	金地集团	600383	粤水电	002060
立讯精密	002475	保利地产	600048	顺丰控股	002352
风华高科	000636	招商积余	001914	中顺洁柔	002511
国星光电	002449	中国平安	601318	美盈森	002303
生益科技	600183	瀚蓝环境	600323	珠江啤酒	002461
德赛电池	000049	广电运通	002152	中炬高新	600872
顺络电子	002138	海信家电	000921	白云山	600332
长盈精密	300115	华帝股份	002035	丽珠集团	000513
华发股份	600325	格力电器	000651	国药一致	000028

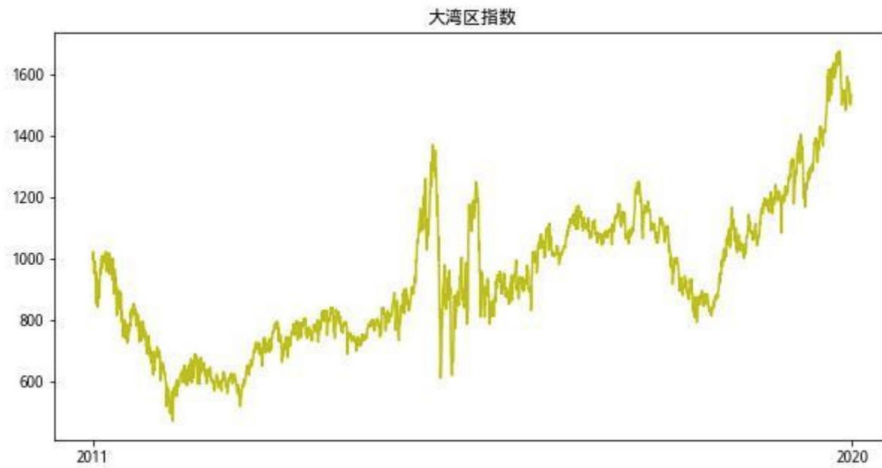


图1 大湾区指数行情实况(2011-2020)

目前该公司采用的大湾区指数增强策略为：在 30 支成份股中选取上一周最强势的 10 支股票，作为下一周投资标的，每周星期一换仓，以当天的收盘价为准考虑收益，每支股票投资额固定为本金的 10%，手续费万分之 2.5。

- 1) 请计算该投资策略从 2011-2020 年的收益率曲线，并且参照大湾区指数评价其投资成效。
- 2) 如果调仓时间不变，但单支股票的投资额可以灵活变化，请找出该投资策略最优的收益率曲线。
- 3) 调仓时间不变，单支股票的投资额为本金的 10%，根据市场交易数据建模，设计你自己的选股策略。
- 4) 如果调仓时间和单支股票的投资额都可以灵活变化，根据市场交易数据建模，设计你自己的选股策略。要求：前三个交易日完成建仓，仓位不低于 50%，单支股票不超过总资产 10%。

三、问题分析

1. 第一问：

根据指数增强模型里对强势股的定义，选择收益率，拟合直线斜率及 R^2 为判断依据选出强势股。从第二周开始以一周为单位，计算出上一周每支股票相应数据并择出 10 支强势股，作为下一周投资标的。通过 python 编写具体运作的代码，运行得出 2011 到 2020 年的收益率曲线并与大湾区指数对比，通过策略累计收益、年化收益、最大回撤率等数据及风险控制模型、交易模型等评价该策略的效果。

2. 第二问：

本题目中，我们掌握了一种风险资产，且股票组合、换仓时间等已确立，只通过仓位分配来提高收益和降低风险。可利用 Harry M. Markowitz 的均值-方差投资组合理论分析 10 支股票的收益率、波动率、方差和夏普率。通过大量随机权重分配组合得出可以得到最高夏普率的权重分配，采取该权重分配方案并购入股票。

3. 第三问：

据题目描述，只需进行选股而不需进行仓位分配，所以选择利用多因子策略进行选股。利用 IC 值计算排名，得出所选股票组合。

4. 第四问：

在本题中，我们有相当高的自由度设计策略。由前三问的结果可知，多因子策略在选股策略中有相当好的效果。但仅选用多因子策略且无仓位分配时回撤率也较高，说明此方法不够稳定，若大盘行情不佳则很可能亏损。所以拟加入海龟策略，在对收益率影响不大的情况下尽可能控制风险。

四、名词解释

名词	含义
指数增强策略	指数增强策略是指基金经理在构建投资组合时，运用“指数跟踪”与“主动管理”相结合的方式获取超额收益的投资策略，就是让产品走势与指数保持一致的同时，依托管理人的管理能力，让组合的涨幅尽可能地高于指数而跌幅小于指数。
哈里·马科维茨均值方差模型	投资者需要在期初从所有可能的证券组合中选择一个最优的组合，需要同时满足尽可能高的收益率和尽可能低的不确定性风险。
海龟策略	海龟交易的具体规则是： <ol style="list-style-type: none"> 1. 当今天的收盘价大于过去 20 个交易日中的最高价时，以收盘价买入； 2. 买入后，当收盘价小于过去 10 个交易日中的最低价时，以收盘价卖出。
R^2	用来衡量线性回归的拟合度。相关性“ r ”衡量两个变量间的相关性，相关性接近 1 表示变量间具有很强的正相关性，接近-1 表示变量间具有很强的负相关性，接近 0 表示变量间没有太多的关系。 R^2 与相关性“ r ”具有很强的相关性。
夏普比率	计算投资组合每承受一单位总风险，会产生多少的超额报酬。
资本资产定价	资本资产定价模型假设所有投资者都按马克维茨的资产选择理论进行投资，对期望收益、方差和协方差等的估计完全相同，投资人可以自由借贷。基于这样的假设，资本资产定价模型研究的重点在于探求风险资产收益与风险的数量关系，即为了补偿某一特定程度的风险，投资者应该获得多少的报酬率。

套利定价理论	套利定价理论认为，套利行为是现代有效率市场（即市场均衡价格）形成的一个决定因素。如果市场未达到均衡状态的话，市场上就会存在无风险套利机会。
唐奇安通道	原始的唐奇安通道先设置一条阻力线和一条支撑线，阻力线由过去 N 天的最高价的最大值形成；支撑线由过去 N 天的最低价的最小值形成。唐奇安上阻力线：由过去 N 天的当日最高价的最大值唐奇；安下支撑线：由过去 N 天的当日最低价的最小值。
多因子模型	多因子模型是关于资产定价的模型。与资本资产定价模型和单指数模型不同，多因子模型认为证券价格并不仅仅取决于证券的风险，还取决于其他一些因素，如，投资者未来预期收入、未来消费品的相对价格及未来的投资机会等。
协方差矩阵	在统计学与概率论中，协方差矩阵的每个元素是各个向量元素之间的协方差，是从标量随机变量到高维度随机向量的自然推广。
对数收益率	连续复利情况下的收益率，即连续复利下，其净多期收益率为单期对数收益率之和。可加性使得对数收益率比算术收益率处理起来更加方便，另外和算术收益率相比，对数收益率具有更容易处理的统计性质。
IC 值	IC 具体的计算公式是全部股票在调仓周期期初的排名和本调仓周期收益排名的线性相关度。IC 均值是多个调仓周期 IC 的平均值。IC 越大，表明排名分越靠前的股票，未来收益排名也越靠前。

五、模型的建立与求解

5.1 第一问：

5.1.1 强势股的选择

- 先将收益率从最高开始排序，即：

$$\frac{\text{上一周的最后一个交易日的收盘价} - \text{上一周第一个交易日的收盘价}}{\text{上一周的第一个交易日的收盘价}}$$

赋最高收益率的股为 0, 第二高收益率的股为 1, 以此类推排出第一个数列 0-29。

- 将拟合直线得斜率按从大到小排序。即用最小二乘法公式：
设拟合直线的公式为：

$$y = k \bullet x + b$$

其中，拟合直线的斜率为：

$$k = \frac{\overline{xy} - \bar{x} \bullet \bar{y}}{x^2 - (\bar{x})^2}$$

计算出斜率后，进行排序。赋最大值 k 的股票为 0，第二大 k 值为 1，以此类推排出第二个数列 0-29。

- 将 R^2 按从大到小排序，若 R^2 平方大于 0.75，表示模型拟合度很好，可解释程度较高； R^2 小于 0.5，表示模型拟合有问题。

$$R^2 = \sqrt{\frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}}$$

计算出 R^2 后，赋最大值的一只股票为 0，第二大的股票为 1，以此类推排出第三个数列 0-29。

综合排序，将相同股票的三份数据相加，按数字从小到大取十只股票，即为每周选出的 10 支强势股票。

选择原因：对强势股的选择，首先看收益率，通过分析过去数据，考察上周的上涨程度，来分析下一周中的股票强势程度。其次是拟合直线的斜率，斜率可以反映整体股票的走势，斜率越大股票上涨程度越大。最后是 R^2 ，相关系数， R^2 值越大，说明我们拟合的直线越能反应该股票的真实走势。故三个指标综合排名前十的股票即为我们选出的下周强势股。

5.1.2 模型的建立

把一周作为一个周期，从第二周开始：

1. 计算上一周 30 支股票的收益率，拟合直线斜率及 R^2 ，通过 5.1.1 中所述方法选出上周的 10 支强势股；
2. 于第二周周一买入第一周的十支强势股，每股所用金额为本金 10%；
3. 此后于每周周一对比所持股票和上一周的强势股，若重合则保留，剩余非强势股则卖出，新的强势股则买入。若因亏损等导致剩余资产比本金少，则按数列从小到大开始，以 10% 本金每支股票的价格买入，直到现有资产不足为止。

用 python 写入策略代码，导入附录一数据，运行得出 2011-2020 年的收益率曲线，得图表 1。



图 1

2011-2020 收益曲线

与大湾区指数进行对比：

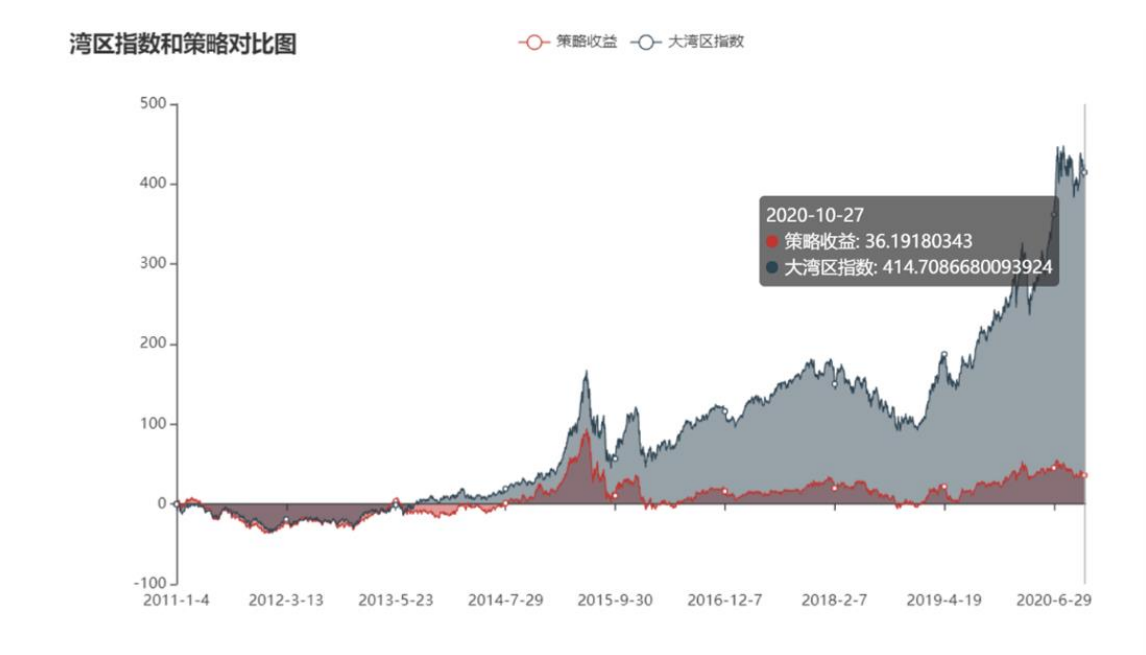


图 2
大湾区指数与策略对比图

详细信息：



图 3
业绩走势

5.1.3 模型的分析与评价

现分析投资策略成效，据图 2，策略累计收益为 36.1918%，据图 3，年化收益为 3.11%，最大回撤为 52.14%。以股票策略而言，我们认为好的量化策略必须具备以下的共性：

1、收益预测的有效性与特异性

有效性很容易理解，量化策略的收益预测模型必须是很扎实的理论依据与逻辑基础，并且是适用于所要投资的市场的一个模型，例如把美股的动量因子用到 A 股，会发现 A 股的反转效应更突出；特异性是指预测模型和市场上其他参与者的模型的相关性要低，最直观的例子就是大家如果都认可并使用的 Alpha 因子会逐渐变为 Beta 因子，同质化会影响收益预测的表现。

2、风险控制模型

好的量化模型必须有应对风险的能力，即需要一个能准确预测组合风险的风险模型。例如我们在实施量化对冲模型的时候，我们会利用优矿的风险模型去控制组合在各个风格上的暴露，控制组合与基准指数的跟踪误差，确保组合能稳定地获取超额收益；可以通过绩效归因去对收益进行分解，一个好的量化策略应该是收益绝大部分来自于选股而非风格与行业的策略，风控能力是量化策略好坏的核心之一。

3、交易模型

交易模型就是策略在实施的能最小化交易成本的模型，例如你的交易如何应对流动性的问题，如何去尽可能地减少冲击成本，做空如何处理期指的升贴水等等，这是一个策略实际运作无法避免的问题，也是策略实际运作的保证。

5.2 第二问^[1]：

5.2.1 理论基础与原理

(1) Markowitz 均值-方差模型

马科维茨把风险定义为收益率的波动率，首次将数理统计的方法应用到投资组合选择的研究中。这种方法使收益与风险的多目标优化达到最佳的平衡效果

依据以下几个假设：

- ◆ 投资者在考虑每一次投资选择时，其依据是某一持仓时间内的证券收益的概率分布。
- ◆ 投资者是根据证券的期望收益率的方差或标准差估测证券组合的风险。
- ◆ 投资者的决定仅仅是依据证券的风险和收益。

- ◆ 在一定的风险水平上，投资者期望收益最大；相对应的是在一定的收益水平上，投资者希望风险最小。

资产组合的均值和方差为：

$$\mu_p = E(r_p) = \sum_{i=1}^N x_i E(r_i) = X^T \mu$$

$$\sigma_p^2 = \sum_{i,j=1}^N x_i x_j \sigma_{ij} = \sum_{i=1}^N x_i^2 \sigma_i^2 + 2 \sum_{1 \leq i < j \leq N} x_i x_j \sigma_{ij} = X^T \Sigma X$$

(2) 优化求最优解：SLSQP 方法

用 Sequential Least Squares Programming 方法分配初始权重，给予

- ◆ 边界条件：每个权重在 0 到 1 之间；
- ◆ 约束条件：权重之和为 1。

输出最大夏普率的投资组合的权重分配。可以获得此投资组合的收益率、波动率和夏普率。

5.2.2 模型的建立与求解

(1) 初步统计数据

首先计算个股的对数收益率 (Log Return)，用当天的收盘价与前一天相比较。利用上一周数据计算年化收益率和协方差矩阵。

(2) 随机取得不同投资组合的夏普比率

每只股票都有一定权重。此资产组合的收益率（均值）为组合中个股收益率（均值）的权重之和。

$$E(r_p) = \sum w_i E(r_i)$$

使用随机数将权重随机分配给由第一题方案求出的 10 支强势股，产生大量投资组合，每一组投资组合由一组随机权重组成。将其收益率和波动率用散点图描述，并计算每个组合的夏普率。定义函数，使输入权重分配时输出该组合的收益率、波动率和夏普比率。

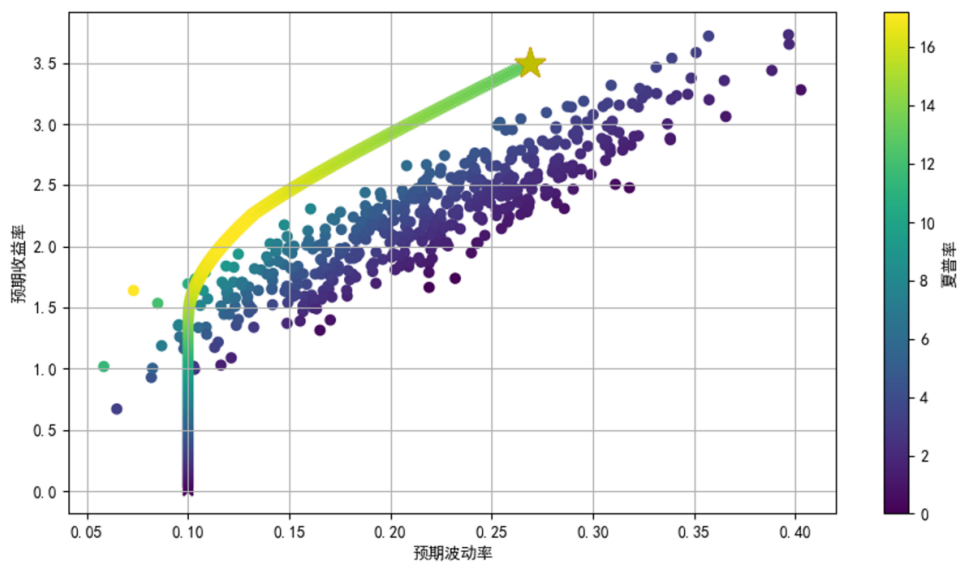


图 4

多组随机投资组合散点图

(3) 最大夏普比率与最小方差优化法

以夏普比率为例，若某投资组合的夏普比率最高，则该组合夏普比率最优，但夏普比率最优的点不一定是所产生的随机投资组合中的一个。所以需要
通过优化算法，寻找出一个恰当的权重分配输入来输出最大的夏普比率。

用最小优化算法，将最大化夏普比率问题转变为最小化负的夏普比率问题。利用 Scipy.optimize 库中的 minimize 函数，以及边界条件及约束条件，得出结果，即最大夏普比率的投资组合及最小方差的投资组合的权重分配。

运行该策略代码，得出：



图 5

最大夏普比率优化法业绩走势

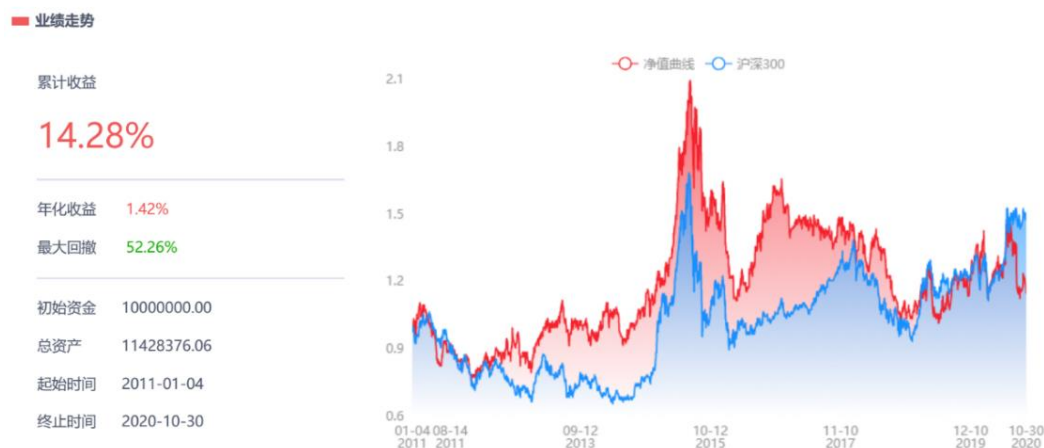


图 6

最小方差优化法业绩走势

5.2.3 模型的优化与分析

对比可知，最大夏普比率优化法相较于最小方差优化法，虽然回撤率稍高出 13 个百分点，但累计收益及年化收益是后者的 7 倍。所以对于仓位配置，宜采用最大夏普比率优化法。

5.3 第三问：

5.3.1 理论基础与原理

本题目中谈到调仓时间不变，单只股票的投资额为本金 10%，需要选择自己的策略。我们组选用经典多因子选股理论基础，先通过分析规模类：总市值因子（PE）；估值类：净市率（BP）；每股指标类：权益回报率（ROE）、资产回报率（ROA）；流动性类：换手率相对波动率（volatility）和动量类：6 日乖离率(bias 6)这六个因子，之后分别分析六个因子，其次加权每支股票六个因子的 IC 值，演算出的后对其进行排序，选出前十支股票。此即为我们的策略。

（1）经典的多因子选股理论基础

多因子模型是关于资产定价的模型。与资本资产定价模型和单指数模型不同，多因子模型认为股票价格并不仅仅取决于股票的风险，还取决于其他一些因素。

5.3.2 模型的建立与求解

为了验证所选因子的有效性，我将这六个因子分别进行因子回测，最后算它们的周期平均收益率和 IC 值。最后发现，他们的表现如图 8 所示，根据以 5 天为周期的分析，资产回报率因子的收益率最高，换手率相对波动率的因子收益率最低，同时换手率相对波动率的这个因子在 IC 值的表现也不算很好，为倒数第二名在 IC 分析中。但总的来说，这六个因子与行情吻合度很高，可以用来进行做多因子的分析。

	A	B	C	D	E
收益率	10D	1D	5D	五日排名	
资产回报率	0.00195624	0.000181608	0.000986481		1
权益回报率	0.000784515	9.66E-05	0.000409697		2
净市率	0.00058766	8.26E-05	0.000392974		3
总市值	-0.001607239	-0.000137232	-0.000801251		4
6日乖离率	-0.001941265	3.79E-05	-0.001945905		5
换手率相对波动率	-0.004086023	-0.000433857	-0.002488556		6
IC	10D	1D	5D	五日排名	
资产回报率	0.006816932	-0.00323533	0.007738411		1
净市率	0.006583166	0.014076296	0.006451795		2
权益回报率	0.005641954	-0.002934425	0.004795639		3
总市值	-0.01977787	-0.012935181	-0.016627752		4
换手率相对波动率	-0.025383925	-0.012341717	-0.026562376		5
6日乖离率	-0.033071624	-0.042389328	-0.052285211		6

图 7

六因子的回测 IC 和收益率排名图

将分析出的 IC 值进行排序，得到排名前 10 支的股票，进行策略的演算，得到如下收益图。



图 8

多因子策略回测数据图

5.3.4 模型的优化分析

可以看到的是，模型的收益率很高的同时，夏普比率并不算很高，同时最大回撤也才近 30%，最后来算盈亏比的话，盈亏比达到了 20:3 这样的盈亏比值值得我们使用这个交易策略去交易。从策略的收益率，夏普率和盈亏比来看，多因子策略绝对是一个非常优秀的一个策略。但同时也存在着一些问题，比如因子库中的因子繁多复杂如何从中挑出有用的那些因子来进行交易。在这一步中我们使用了因子回测的这一个方法来挑选出对行情敏感的因子。在后续的可能优化中，我们可以适当增加因子的个数，以及对因子之间的相关性进行一个分析，剔除掉剔除冗余因子。还可以对因子的组合上采用最大化 IC_IR 方法来进行因子的组合。

5.4 第四问：

5.4.1 海龟交易策略

海龟交易法则定义了一整套非常严谨的仓位控制、止盈止损的规则。对于币市的大涨大跌行情，海龟交易法则正是应付这种极端行情的利器。

加仓：海龟法则的加仓原则是定义好一个小单位（Unit），使得该仓位的预期价值波动与总净资产的 1% 对应。也就是说，如果买入了这 1 个小单位的资产，那当天该仓位的市值变动幅度不会超过总净资产的 1%。

建仓：建仓的动作来自于趋势突破信号的产生。如果当前价格冲破上轨，就产生了一个买的建仓信号，如果当前价格跌破下轨，就产生了一个卖空的建仓信号。

止损：用户可根据需求自定义动态止损方案。

5.4.2 模型的建立与求解

对于本题我们利用多因子策略选股，以一天为周期换仓；利用海龟交易策略进行仓位配置、止盈止损等。

1. 首先，利用问题三中的多因子策略对 30 支股进行排名，选出排名前 15 的股票，对每支股票建立唐奇安隧道。
2. 对历史的价格波动进行统计，预估 Unit 带来的价值波动：

真实波动 = $\max(\text{最高价}-\text{最低价}, \text{最高价}-\text{昨日收盘价}, \text{昨日收盘价}-\text{最低价})$

$$N = \frac{\text{前日的 N 值之和} + \text{当时的真实波动}}{\text{总天数}}$$

得到该 Unit 的值。

3. 对于所选的每支股票，进行如下操作：
 - ◆ 当单支股票的 K 线穿过唐奇安通道的上轨道线时则建仓，投入资金值为一个 unit；
 - ◆ 当该股票每上涨一个 ATR，则增加投入 0.5 个 unit；
 - ◆ 当该股 K 线穿过唐奇安通道下轨道线，则平仓；当上涨了 6 个 ATR 则平仓；当亏损超过 2ATR，则平仓。

运行代码得出最终的业绩走势：



图 9

该方法虽然收益率较低，但回撤率也很低，不失为一种稳定收益的好策略。

六、参考文献

- [1]许维鸿, 基于经典四因子模型的因子选股策略, 西南证券, 2013, 1~14
- [2]路来政, 量化投资策略的应用效果研究, 暨南大学, 2012, 4~17
- [3]黄鲲, 量化交易及相应的软件系统开发, 天津大学, 2011, 1~31
- [4]丁鹏, 量化投资—策略与技术, 电子工业出版社, 2012, 112~126
- [5]国泰安开发团队, 国泰安量化投资研究平台简介, 2011, 1~2
- [6]张晓斌, 金字塔决策交易系统—初级教程(2013 修订版), 2013, 1~118
- [7]忻海, 解读量化投资, 机械工业出版社, 2012, 1~33
- [8] 曾长兴, 量化投资方法与量化产品设计, 安信证券, 2009, 1~14

七、附录

第一到第四题的代码:

在 question_code 文件夹内

real_one.py

real_two.py

real_three.py

real_four.py

2020 年“大湾区杯”粤港澳金融数学建模竞赛承诺书

第一题:

策略收益率 1.html

绩效报告.pdf

湾区指数和策略对比图 1.html

第二题:

策略收益率(最大夏普率).html

策略收益率(最小方差).html

绩效报告(最大夏普率).pdf

绩效报告(最小方差).pdf

湾区指数和策略对比图(最大夏普率).html

湾区指数和策略对比图(最小方差).html

第三题

六个因子的分析报告

策略对比图.jpg

第三题绩效报告.pdf

收益率图.jpg

第四题

绩效报告.pdf