

目录

第一章 遗传算法实证内容	1
1.1 策略实证	1
1.2 概述	1
1.3 策略基本设定	1
1.3.1 投资标的	1
1.3.2 数据区间	1
1.3.3 回测框架	1
1.3.4 策略参数	1
1.4 实证结果	2
1.4.1 因子形式	2
1.5 调仓情况分析	5
1.6 优化方向	6

第一章 遗传算法实证内容

1.1 策略实证

1.2 概述

现阶段挖出因子训练集收益 9.59%，测试集收益 6.67%（年化）。

1.3 策略基本设定

1.3.1 投资标的

本文实证基于 1 分钟级 IC 股指期货数据。

1.3.2 数据区间

本文采取 2019 年 11 月 4 日至 2022 年 11 月 1 日共三年数据。其中 2019 年 11 月 4 日至 2021 年 10 月 29 日为训练集，2021 年 11 月 1 日至 2022 年 11 月 1 日为测试集。

1.3.3 回测框架

本策略目前采取笔者自行搭建的回测框架。该框架将回测分为四个步骤，依次为：
1. 根据信号计算仓位、2. 根据仓位计算投资账户价值、3. 根据账户价值计算收益统计量、4. 根据账户价值绘制投资效果图。

1.3.4 策略参数

在本策略中，主要参数设定出现在 gplearn 函数包的 SymbolicRegressor 函数中。具体地，本策略主要调整的参数如表 1.1 所示。

表 1.1 遗传规划相关参数设置

参数名称	参数取值
群体数量	100
公式进化的世代数量	3
选取优质个体数	10
初始化深度	2-3

1.4 实证结果

1.4.1 因子形式

根据上述设定，得到生成因子结果如下式所示：

$$F = I\{\Delta(\min(\log X_2, X_1)) < \log(|\cos(X_4)|)\} \quad (1.1)$$

在上式中，相关变量及函数含义如表1.2所示。

表 1.2 因子指标含义说明

指标名称	指标含义
X1	时间区间内最高价
X2	时间区间内最低价
X4	时间区间内 bar 的序数（当前时刻数据是开盘后第几分钟）
delta 函数	$X_t - X_{t-1}$
lt 函数	$I_{\{X < Y\}}$
因子形式	$lt(\delta\epsilon lta(\min(\log(X2), X1)), \log(\cos(X4)))$

可以看出，由于函数最外层是指示函数，所以交易信号为单边做多。即在信号发生时进行开多仓操作，在当日收盘时平仓。

根据上述因子和投资方法，得到训练集、测试集内投资效果如表1.3所示。

表 1.3 投资结果统计量

	训练集	测试集
收益率 (年化)	9.59%	6.67%
收益率 (Buy and Hold, 参考)	20.78%	-14.73%
标准差 (年化)	0.1428	0.1798
夏普比率 (年化)	53.15%	25.94%
最大回撤水平	0.1367	0.1828
最大回撤比率	12.49%	18.01%
平均每日开仓次数	0.793	0.791
胜率	55.09%	50.26%
平均每日开多仓次数	0.793	0.791
多仓胜率	55.09%	50.26%
平均每日开空仓次数	0.0	0.0
空仓胜率	inf	inf

从表中可以看出，标的指数期货在训练集和测试集收益表现有较大差异，买入并持有策略在训练集内年化收益达到了 20% 之大，但验证集内亏损较为严重，年化收益接近-15%。

相比之下，本策略的收益表现较为稳定，在训练集和测试集内均保持正向的收益。但从最大回撤比率看，本策略的收益回撤仍处于较高的水平。

图1.1展示了训练集内账户价值变化图，其中黄线为策略累计收益，蓝线为以开盘价计算的买入并持有累计收益。从策略效果看，本策略收益在训练集内并未跑赢基准。其中在训练集内大部分时期，本策略并未呈现明显的盈利。直至2021年3月，本策略的盈利水平才开始稳步上升。



图 1.1 训练集账户价值变化

图1.2展示了训练集内策略最大回撤比率的变化情况，用 $Maxdd_t$ 表示 t 时刻最大回撤比率，其具体计算方法为

$$Maxdd_t = \frac{t \text{时刻策略累计最高点} - t \text{时刻策略账户价值}}{t \text{时刻策略累计最高点}}. \quad (1.2)$$

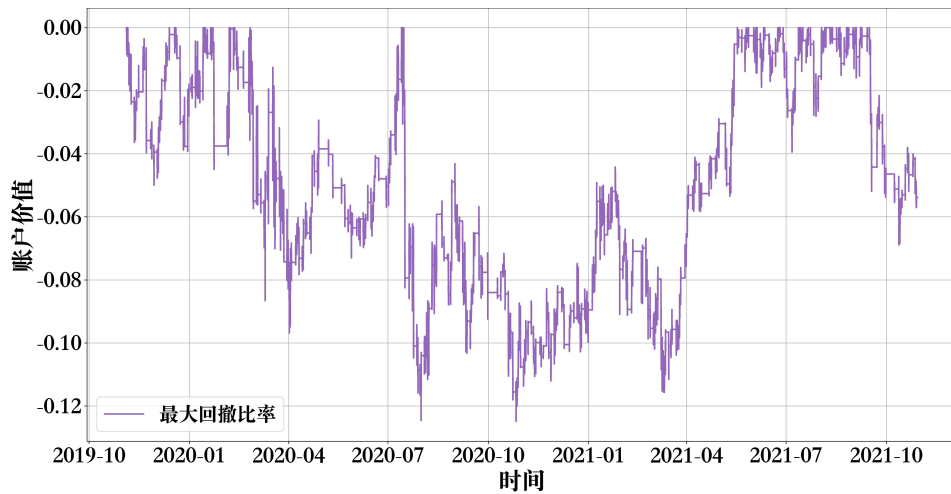


图 1.2 训练集最大回撤比率变化

从图中可以看出，相比策略收益情况，策略的最大回撤水平处于较高的水平。

图1.3展示了训练集内策略逐笔收益频率直方图。从中可以看出，逐笔收益值有较为明显的右偏形态。虽然收益整体表现为正，但较厚的左尾仍表明策略有较大的损失风险。

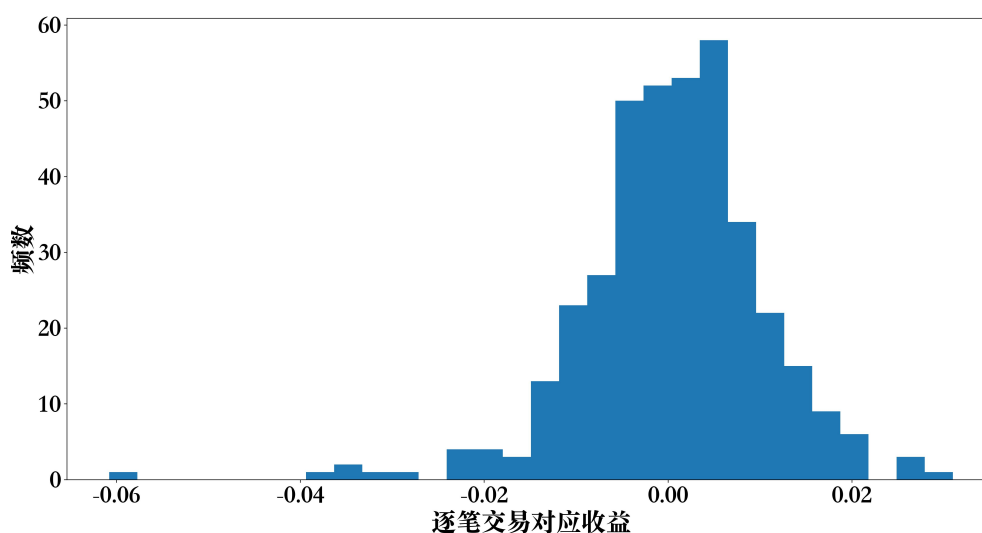


图 1.3 训练集逐笔收益

接下来分析测试集内投资效果。图1.4展示了训练集内账户价值变化图，其中黄线为策略累计收益，蓝线为以开盘价计算的买入并持有累计收益。从策略效果看，本策略收益在指数下跌时多为不开仓交易的状态，在2022年上半年初始表现较为平稳，后随指数一同下跌。在2022年下半年后，本策略逐步开始实现收益的稳步上升，最终获得年化6%的收益，高出基准近20%。

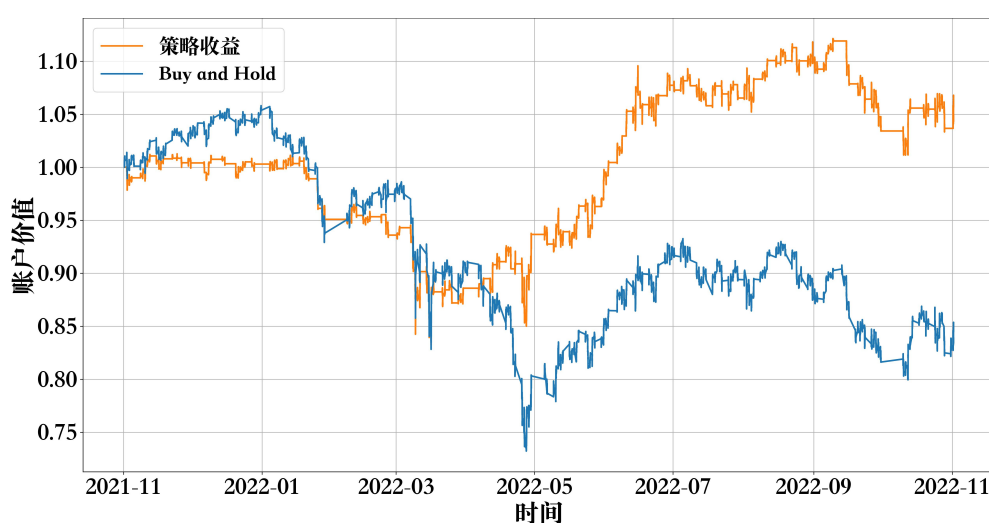


图 1.4 测试集账户价值变化

图1.5展示了训练集内策略最大回撤比率的变化情况，同训练集一样，策略的最大

回撤水平相对较高。

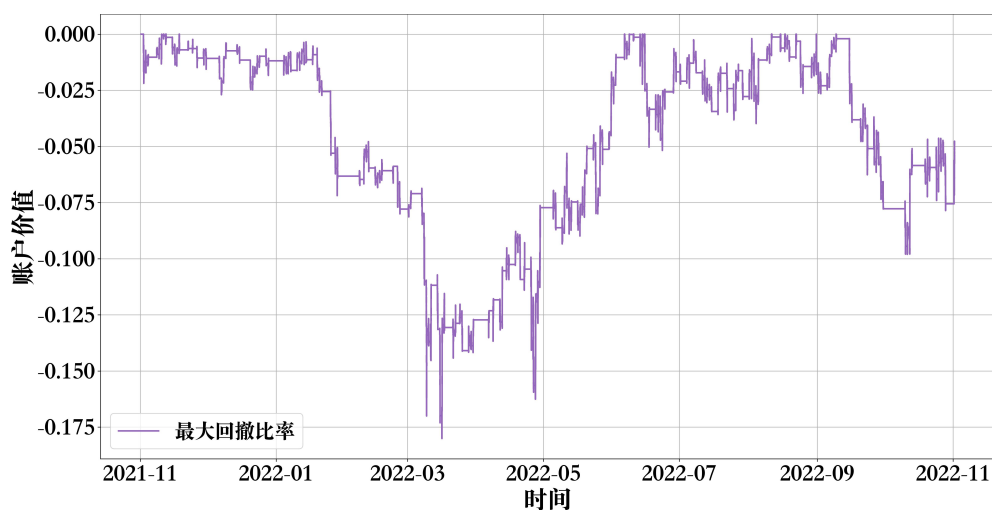


图 1.5 测试集最大回撤比率变化

图1.6进一步展示了训练集内策略逐笔收益散点图。从图中可以看出，测试集内每笔策略收益更为集中，峰度有所下降。收益极端值出现的更为频繁。

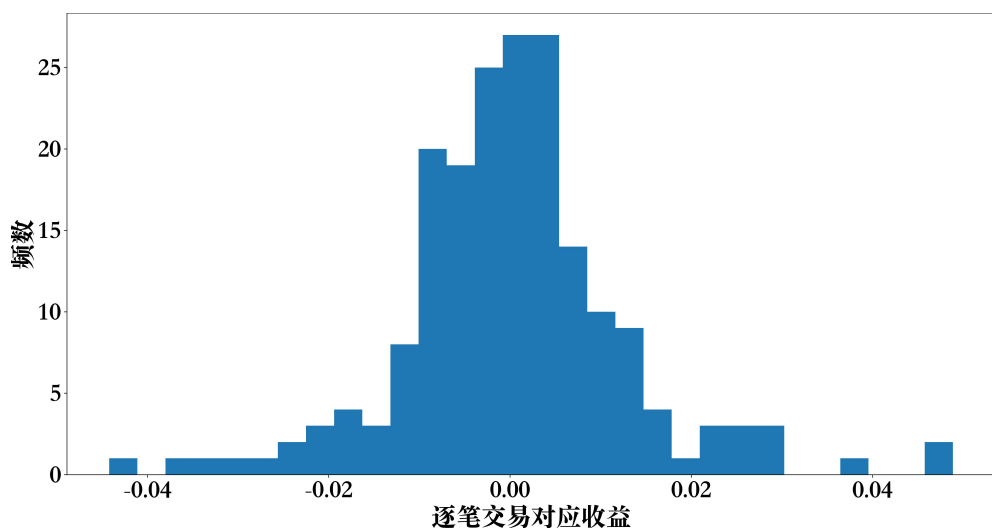


图 1.6 测试集逐笔收益

1.5 调仓情况分析

接下来分析在调仓中一个比较有趣的现象。在观察调仓过程时，我们发现调仓集中在特定的几个时点。图1.7，1.8展示了调仓时点的频数统计直方图。

从图中可以看出，交易大多集中在 9:52，并且频率随时间呈现递减的趋势。进一步观察分钟序数，发现依次为 22，44，66 等。

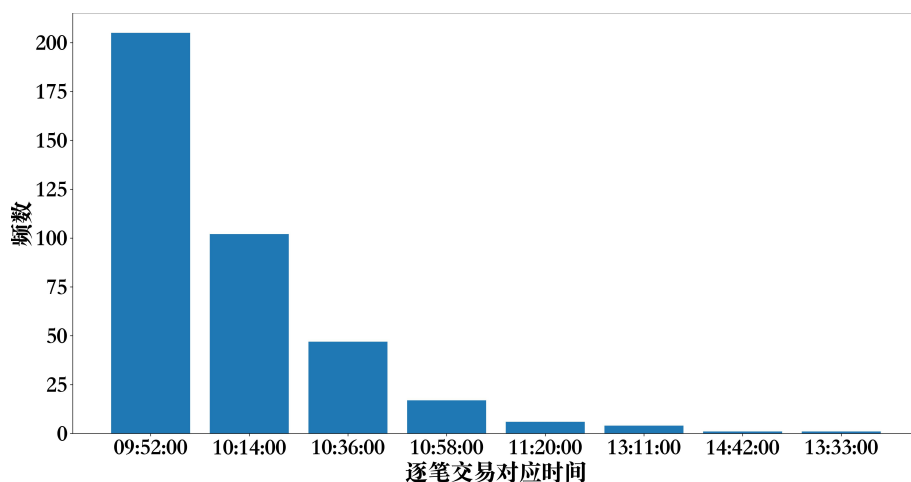


图 1.7 训练集调仓时点情况

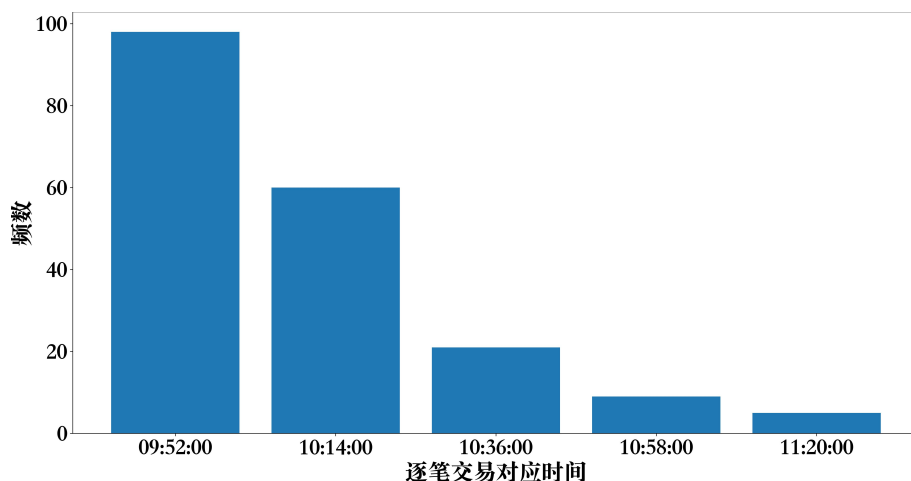


图 1.8 测试集调仓时点情况

分析原因不难发现，在因子公式1.1中，指示函数在内部的不等式左边小于右边时取值为 1，否则取为 0。但看等式右侧， $\log(\cos X_4)$ 外部的 $\log(\cdot)$ 为增函数，而 $|\cos(\cdot)|$ 仅在内部接近 π 时才接近最大值 1，因此交易也多发生在分钟数接近 π 的整数倍时。如 $22 \approx 7.0028\pi$, $44 \approx 14.0056\pi$, $66 \approx 21.0084\pi$ 。又因为信号的触发仅在第一次发生时被考虑，所以不同分钟数对应触发频数也随开盘时间的增加而下降。

1.6 优化方向

1. 数据集用滤波器，SMA 等方法进行扩充；函数加入滤波方法等
2. 看遗传算法底层逻辑和参数含义
3. sttc2plot 的收益里面分开画多仓和空仓收益，堆起来
4. 补充交易量数据，补充股票指数数据，扩充数据集。