

港股及新三板量化回测平台设计文档

刘世豪量化研究员 刘邦砥量化研究员

2018 年 8 月 28 日

目录

| | | |
|----------|--------------|----------|
| 1 | 设计背景 | 2 |
| 1.1 | 新三板 | 2 |
| 1.2 | 港股 | 2 |
| 1.3 | 港股及新三板量化交易平台 | 3 |
| 2 | 设计目标 | 4 |
| 3 | 系统环境 | 4 |
| 4 | 模块设计 | 4 |
| 4.1 | 数据设计 | 4 |
| 4.1.1 | 数据项 | 4 |
| 4.1.2 | 数据结构 | 4 |
| 4.2 | 算法描述 | 4 |
| 4.3 | 接口设计 | 5 |
| 4.3.1 | 获取数据 | 5 |
| 4.3.2 | 交易 | 6 |

1 设计背景

1.1 新三板

“新三板”市场原指中关村科技园区非上市股份有限公司进入代办股份系统进行转让试点，因挂牌企业均为高科技企业而不同于原转让系统内的退市企业及原 STAQ、NET 系统挂牌公司，故形象地称为“新三板”。新三板股票是指证券公司代办股份转让系统中关村科技园区非上市股份有限公司股份报价转让，因为挂牌企业的有别于原代办股份转让系统的企业的，所以被形象地称为“新三板”股票。新三板的推出，不仅仅是支持高新技术产业的政策落实，或者是三板市场的另一次扩容试验，其更重要的意义在于，它为建立全国统一监管下的场外交易市场实现了积极的探索，并已经取得了一定的经验积累。作为中国特色多层次资本市场的重要组成部分，新三板投资方法论研究是一个全新课题，需要我们不断实践和探索，并逐步加以总结与完善。

1.2 港股

港股市场指在香港联合交易所上市的股票，有主板市场和创业板市场之分，股票类型有：

1. 蓝筹股

指恒生指数成份股。这类股份的普遍特性是具有行业代表性、流通量高、财务状况良好、盈利稳定及派息固定。蓝筹股的固定数目是 43 支。由于这类个股的盈利能力强，大多属于基金偏爱个股（即基金长期持有或加码买入的股票），因此较少人为因素（如庄家活动）影响，股价相对稳定，适合作中长线投资。

2. 国企股

指获中国证监会批核到香港上市的中国大陆的国有企业。它亦可称为 H 股，意指在香港上市的国企股。在内地还有称为 N 股及 S 股的，是分别在美国纽约及新加坡上市的国企的简称。

3. 红筹股

指在香港上市，但由中资企业直接控制或持有三成半股权以上的上市公司股份。

港股与 A 股的区别有：

1. 比 A 股市场更成熟、理性，对世界的行情反映更灵敏。
2. 内地市场有涨跌停板制度，即涨跌波幅如超过某一百分比，有关股份即会停止交易一段指定时间；香港市场并没有此制度。
3. 香港证券市场主要以港元为交易货币；内地股市以人民币为交易货币。
4. A 股实行“T+1”交易制度，而港股市场中证券商可替投资者安排卖出当日较早前已购入的证券。
5. 香港证券市场准许进行受监管的卖空交易。

1.3 港股及新三板量化交易平台

量化交易是指以先进的数学模型替代人为的主观判断，利用计算机技术从庞大的历史数据中海选能带来超额收益的多种“大概率”事件以制定策略，极大地减少了投资者情绪波动的影响，避免在市场极度狂热或悲观的情况下作出非理性的投资决策。它是一种利用计算机技术防止人为非理性操作的投资方法。国外量化交易已经发展已经有三十多年历史，目前已经比较成熟，但我国量化交易还处于起步阶段。

量化交易平台允许投资者和研究人员在平台上编写交易策略、利用平台上的数据、选定起止时间进行回测。通过回测，可以直观地了解收益走势、收益率、年化收益率、最大回撤、夏普比率等一系列指标，是进行量化交易和量化研究的有效手段。

目前，国内几大量化交易平台支持的研究数据为：聚宽（joinquant）提供沪深 A 股行情数据，上市公司财务数据，场内基金数据，指数数据，期货数据以及宏观经济数据；优矿（uqer）提供沪深交易所股票的基本信息以及日/分钟级别的股票行情、财务报表、公司行为、基金、期货等信息；米筐（ricequant）提供日级别和分钟级别的 A 股、ETF、LOF、分级基金、商品期货、股指期货和国债期货的市场数据和 400+ 指标的财务数据。可见，国内量化平台的股票版块主要集中在 A 股的量化研究，并不包括逐渐成为中国资本市场重要组成部分的新三板及港股。

2 设计目标

本平台为投资者提供新三板股票量化平台，给用户在新三板市场筛选股票提供极大的便利。在新三板量化平台上，用户可以选定筛选标准、自定义筛选出的股票数量。本平台与我司自主研发的新三板数据库 `ssdata` 相结合，用户可以选择回测区间进行回测。通过回测，用户可以直观了解策略收益情况。新三板企业众多、情况复杂。通过使用量化方法合理筛选，用户既可以直接以选出的股票为标的买卖；也可以将选出股票作为考察范围，对其中企业进行更加深入的调查研究。

总之，本平台旨在用量化方法为新三板投资者提供投资方案，将调研分析的决策成本降到最低。

3 系统环境

该量化平台在 `python3` 下运行，需要安装 `ssdata`, `pandas`, `numpy`, `matplotlib` 等包，可用 `pip` 安装。

4 模块设计

4.1 数据设计

4.1.1 数据项

该量化平台数据项有：股票代码、成交量、成交额、是否有做市商。

4.1.2 数据结构

目前数据结构为股票代码与成交量、股票代码与成交额、股票代码与是否有做市商的一对一映射关系。具体则是三个 `csv` 文件。

4.2 算法描述

该量化平台的回测算法主要分为三步：

1. 从 ssdata 上获取股票数据

需要调用 ssdata 库的 `get_data` 函数。

2. 利用数据进行交易

需要在 `handle_data` 中设计具体的交易策略，必要时可以增加辅助函数，比如选股函数等。

3. 将交易过程中的收益、回撤等输出为图、表。

4.3 接口设计

该量化平台提供了两类接口：获取数据接口和交易接口。

4.3.1 获取数据

获取数据目前有 ssdata 的 `get_data` 函数，该函数具体语法如下：

```
data = ssdata.get_data(secid=stock,
                        start_date=start_date,
                        end_date=end_date,
                        field='open,yoyop')
```

其中，`secid` 参数为证券代码，`start_date` 参数为起始日期，`end_date` 参数为终止日期，`field` 参数为想要获取的数据名称，可以是多个数据，中间用“，”隔开，目前支持的数据有：

- open: 开盘价
- avgprice: 均价
- close: 收盘价
- pb: 市净率
- roediluted: 净资产收益率 ROE（摊薄）
- yoyor: 营业收入同比增长率

- yoyop: 营业利润同比增长率
- yoypni: 归属母公司股东的净利润同比增长率
- yoyni: 净利润同比增长率

ssdata.get_data 将返回一个 DataFrame，索引为日期，列名为数据名称。

注：返回的 DataFrame 中可能含有 NaN 数据，这是因为某些时间过早的（比如 2014、2015 年）数据（比如 yoyop 等）不存在，需要用 dropna() 函数将这些数据剔除：

```
data.dropna()
```

除了上述这些数据，还有成交量、成交额、是否有做市商的数据，这三类数据存储在本地的 csv 文件中，可用 pandas 包的 read_csv 读取。

4.3.2 交易

目前交易有两个接口函数：order_to 和 order_pct_to。

1. order_to：下单到一定股数

该函数语法如下：

```
order_to(target)
```

其中，target 为一个 Series 数据结构，该 Series 的索引为股票代码，值为目标股数。利用该函数，可以下单到目标股数。

2. order_pct_to：下单到一定百分比仓位

该函数语法如下：

```
order_pct_to(pct_target)
```

其中，pct_target 为一个 Series 数据结构，该 Series 的索引为股票代码，值为目标百分比仓位。利用该函数，可以使下单后的仓位达到一定百分比。