

# UBS 实习汇报

汇报人 许震宇 日期 2026.1.29

## 一、一个简短的自我介绍

- 许震宇
- 中国科学院大学(北京)
- 家在深圳、港籍
- 徒步、网球、滑雪...

微信：



中国香港 离岛区



扫一扫上面的二维码图案，加我为朋友。

图 1:

## 二、数据格式调研

数据格式	适用场景与优点	备注
CSV	通用、可读性强、便于快速查看与交换	体积大、读写慢、类型易丢失
Parquet	列式存储、压缩好、读取快，适合大规模因子/行情特征	生态成熟(Spark/Arrow)
Feather/Arrow	序列化快、零拷贝传输，适合进程间共享与临时缓存	适合内存分析
HDF5 (.h5)	分层结构、可存多表/多维数组、支持元数据与随机访问	适合大体量数值数据
SQLite	轻量数据库、支持 SQL 查询，适合中小规模结构化数据	单文件便携
Pickle	Python 对象持久化，开发效率高	跨语言差、版本兼容风险
JSON	可读、跨语言、适合配置与元数据	体积大、速度慢
NPZ	压缩数组打包，便于分发与存档	适合纯数值数组

表 1: 量化常见数据格式对比

## 三、数据调研

- 数个指标从 2010 年到 2024 年的分时数据
- 层级: 分时数据项 < 每日数据 < 从 2014 ~2024 的数据 < 3 个指标的从 2014 ~2024 的数据
- 每个分时数据项包含: Close High Low Open Time Date Amount Volume

## 四、构建代码框架

### 1 目录结构

```
UBS_Quant/
  assets/          # 原始数据与静态资源
    MinutesIdx.h5
  output/          # 回测输出 (gitignore)
  report/          # 报告与说明文档
  1/
    main.tex
    tex_file/head.tex
  src/             # 源码入口目录
    quant_research/
      run_experiment.py   # 主流程入口
      data/              # 数据读取与特征构建
        loader.py
        handler.py
      backtest/           # 回测与指标
        engine.py
        metrics.py
      portfolio/          # 组合与净值曲线
        #
```

```
allocator.py  
features/          # 特征基础定义  
base.py
```

## 2 主文件流程(`run_experiment.py`)

- 配置阶段: 定义数据配置、模型配置与回测配置(执行时刻、模式等)。
- 数据准备:`SimpleDataset.prepare` 读取原始数据并通过 `FeatureHandler` 构建特征/信号。
- 记录入口:`Experiment` 打开实验上下文并记录参数。
- 信号生成:`SignalRecord.generate` 输出 `signals.csv`。
- 信号分析:`SigAnaRecord.generate` 先生成收益列, 再计算 Sharpe, 输出 `sharpe_by_exec.csv` 与分析图。
- 组合回测:`PortAnaRecord.generate` 选择执行时刻, 计算收益并输出 `equity_curve.csv`。
- 结果落盘: 输出 `readouts_full.csv` 以便后续分析。

## 五、阅读 alpha101

几个印象挺深的点:

- 将 alpha 分为: 1. momentum; 2. mean-reversion;
- significant dependence on the turnover  $T_i$  here, while the average daily return  $R_i$  is strongly correlated with the daily return volatility  $\sigma_i$  and we have the scaling property (1) with  $X \approx 0.76$ .
- alphas are ephemeral(短暂的)