**实习汇总文档**

崔晏菲

1. 主要的工作进展
   1. 调出了xgboost最优的参数范围（, ）
   2. 建立了一套目前最优的模型工作流
      1. 将同code\_id的因子用线性模型聚合
      2. 将聚合好的数据用于gbtree模型的训练
         * 1. 先对称，对称的效果要远好于不对称
           2. 再将三倍标准差之外的真实收益率clip
      3. 将模型输出的预测值送入OPT，OPT参数为默认参数，除了hconstrain=0.5
   3. 建立了一些较为有效的统计指标（计算代码已发给清源）
      1. IC最好在0.04以上，0.03以上也能接受
      2. R2小于-0.005，结果一定会很差。正常应该在0附近（±0.0001）。当R2>0.001时，结果大概率较好
      3. <-3bp Sharp，指的是模型的预测值小于-3bp的那一部分，其对应的真实收益率的均值除以标准差。这个值应当小于0，一般来说效益-0.06算是比较不错的。
      4. >3bp Sharp，指的是模型的预测值大于3bp的那一部分，其对应的真实收益率的均值除以标准差。这个值越大越好，若低于0.03则说明预测失效，若大于0.04则说明效果很好。
      5. >5bp Sharp，指的是模型的预测值大于5bp的那一部分，其对应的真实收益率的均值除以标准差。这个值越大越好，一般要求大于0.05
      6. >8bp Sharp，指的是模型的预测值大于8bp的那一部分，其对应的真实收益率的均值除以标准差。这个值越大越好，一般要求大于0.07
   4. 建立了筛因子的方法（已废弃）
   5. 建立了因子融合的流程（已交接给清源）
      1. 详见[2024.01.19 - 崔晏菲 - Confluence (pinghu.tech)](https://confluence.pinghu.tech/display/~yfcui1/2024.01.19)
   6. 初步证实了加入品种信息能改善模型表现
      1. 详见[2024.01.23 - 崔晏菲 - Confluence (pinghu.tech)](https://confluence.pinghu.tech/display/~yfcui1/2024.01.23)