

DFQ-FactorVAE：融合变分自编码器和概率动态因子模型的 alpha 预测方案

——因子选股系列之一〇三

研究结论

FactorVAE 模型架构

- FactorVAE 模型，来自国际人工智能会议 AAAI 2022。原论文中提出了一种新的基于变分自编码器的概率动态因子模型，以弥合噪声数据与有效因子之间的差距。本文在原论文基础上进行修改优化，所得选股因子在沪深 300 指数增强组合中表现优异。
- FactorVAE 模型融合了变分自编码器与概率动态因子模型的思想，建立股票收益率预测模型，学习输入特征和标签之间的关系：（1）采用变分自编码器的编码器-解码器架构，有助于模型学习到数据分布的结构，并且可以在潜在空间中生成新样本；（2）采用概率动态因子模型的思想，将因子作为 VAE 模型中的潜在变量，提取动态公共因子的分布，从而起到降维降噪的作用。（3）采用一种“前验-后验”的学习方法，将预测股票收益率的问题转化成预测有效因子，使用标签收益率指导模型提取有效因子。

DFQ-FactorVAE 模型优势

- DFQ-FactorVAE 模型所得因子的稳定性、在沪深 300 股票池中的多头表现突出：
 - （1）在中证全指股票池中，DFQ-FactorVAE 模型所得到因子的稳定性明显最强，ICIR、RANKICIR、多头日超额收益夏普比均为最高。测试集（2020.01.01-2024.03.31）上 rankic 达到 15%，rankicir 达到 1.38，20 分组多头日度超额年化收益率达到 31.75%，多头日超额收益夏普比 3.52，多头日度超额收益最大回撤 8.28%，多头月度胜率 88%，月均单边换手 79%。分组单调性好。
 - （2）在沪深 300 股票池中，DFQ-FactorVAE 模型所得到因子的多头表现明显最强。测试集上 rankic 达到 10.6%，rankicir 达到 0.6，5 分组多头日度超额年化收益率达到 14.47%，多头日超额收益夏普比 1.72，多头日度超额收益最大回撤 6.74%，月均单边换手 54%。分组单调性好。分年表现上未出现明显衰减。
- DFQ-FactorVAE 模型所得因子稳定性突出，与模型的 VAE 架构和概率框架有关。VAE 架构有助于模型学习到数据分布的结构，样本外泛化能力更强。概率框架考虑到了风险建模，更适合含有噪声的股票收益率预测。
- DFQ-FactorVAE 模型所得因子在沪深 300 股票池中的多头表现突出，与模型的因子模型架构有关。先验因子对沪深 300 股票的解释度最高，平均能达到 36.5%，其次为中证 500 成分股的 27.95%，中证 1000 成分股的 25.45%，全市场解释度最低，为 20.78%。

DFQ-FactorVAE 模型在沪深 300 指增组合中的表现

- DFQ-FactorVAE 模型所得到的合成因子在沪深 300 指增组合中表现十分突出：
 - （1）整体表现：2020 年以来年化信息比达到 2.55，年化对冲收益 13.53%，年化跟踪误差 5.03%，超额收益最大回撤仅为 5.35%，单边年换手 7.67 倍。
 - （2）分年表现：2020-2023 每年取得 10%以上的正超额，2023 年超额收益达 16%。2024 年前三个月超额收益为-0.36%。
 - （3）风格暴露：相对基准沪深 300 指数，组合在市值、信息确定性、成长维度具有明显的负向暴露，但在 BETA、流动性、波动率、估值等维度都没有明显暴露。
 - （4）成分股约束：设置 100%成分内选股增强，可以明显提高组合业绩表现稳定性。跟踪误差可以降低到 4.37%，超额收益最大回撤降低到 3.31%，最大回撤恢复仅需 30 天，并且仍然可以获得超过 10%的年化超额。

风险提示

- 量化模型失效风险。
- 极端市场环境对模型的影响。

报告发布日期

2024 年 05 月 14 日

证券分析师

杨怡玲 yangyiling@orientsec.com.cn
执业证书编号：S0860523040002

刘静涵 021-63325888*3211
liujinghan@orientsec.com.cn
执业证书编号：S0860520080003
香港证监会牌照：BSX840

相关报告

基本面因子的重构：——因子选股系列之一〇二 2024-03-21

自适应时空图网络周频 alpha 模型：——因子选股系列之一〇一 2024-02-28

DFQ-HIST：添加图信息的选股因子挖掘系统：——因子选股系列之一百 2024-02-07

基于异构图神经网络的股票关联因子挖掘：——因子选股系列之九十九 2024-01-02

DFQ-TRA：多交易模式学习因子挖掘系统：——因子选股系列之九十七 2023-11-14

DFQ 强化学习因子组合挖掘系统：——因子选股系列之九十五 2023-08-17

DFQ 遗传规划价值因子挖掘系统：——因子选股系列之九十 2023-05-28

目录

一、模型概述	5
1.1 自编码器 (Auto Encoder, AE)	5
1.2 变分自编码器 (Variational Auto-Encoder, VAE)	5
1.3 因子模型 (Factor Model)	6
1.4 概率动态因子模型 (Probabilistic Dynamic Factor Model)	6
1.5 FactorVAE	6
二、模型架构	7
2.1 模型训练和预测过程	7
2.2 特征提取器	8
2.3 因子编码器	9
2.4 因子解码器	10
2.5 因子预测器	11
三、模型说明	12
3.1 数据说明	12
3.2 模型输入	12
3.3 模型参数	13
四、模型结果	14
4.1 运算用时	14
4.2 因子绩效表现	14
4.3 随机种子的影响	21
4.4 与其他量价模型相关性	21
4.5 中性化因子表现	23
五、沪深 300 指数增强组合	24
5.1 指数增强组合构建说明	24
5.2 指数增强组合的业绩表现	24
5.3 指数增强组合的风格暴露	25
5.4 组合优化约束对组合业绩的影响	26
六、总结	28
参考文献	29
风险提示	29

图表目录

图 1: FactorVAE 模型示意图	7
图 2: FactorVAE 模型架构	8
图 3: 不同特征提取器的效果对比 (2020.01.01-2024.2.28)	9
图 4: 因子编码器结构	9
图 5: 因子编码器计算公式	9
图 6: 不同投资组合数量的效果对比 (2020.01.01-2024.2.28)	10
图 7: 因子解码器结构	10
图 8: 因子解码器计算公式	10
图 9: 预测收益率分布 VS 预测收益率均值效果对比 (2020.01.01-2024.2.28)	11
图 10: 因子预测器结构	11
图 11: 因子预测器计算公式	11
图 12: 预测标签 Y 不同处理方式下的模型效果对比 (2020.01.01-2024.2.28)	12
图 13: 不同输入下的模型效果对比 (2020.01.01-2024.1.16)	13
图 14: DFQ-FactorVAE 模型参数列表	13
图 15: 不同 batchsize 下的模型效果对比 (2020.01.01-2024.2.28)	13
图 16: 训练集、验证集、测试集中 IC 变化	14
图 17: 训练集、验证集、测试集中 rankIC 变化	14
图 18: 中证全指股票池各模型因子绩效表现 (2020.1.1-2024.3.31)	15
图 19: 中证全指股票池各模型分组年化超额收益 (2020.1.1-2024.3.31)	15
图 20: 中证全指股票池各模型多头组超额收益净值&回撤 (2020.1.1-2024.3.31)	15
图 21: 中证全指股票池各模型分年绩效表现 (2020.1.1-2023.12.31)	16
图 22: 沪深 300 股票池各模型因子绩效表现 (2020.1.1-2024.3.31)	16
图 23: 沪深 300 股票池各模型分组年化超额收益 (2020.1.1-2024.3.31)	16
图 24: 沪深 300 股票池各模型多头组超额收益净值&回撤 (2020.1.1-2024.3.31)	17
图 25: 沪深 300 股票池各模型分年绩效表现 (2020.1.1-2024.3.31)	17
图 26: 中证 500 股票池各模型因子绩效表现 (2020.1.1-2024.3.31)	18
图 27: 中证 500 股票池各模型分组年化超额收益 (2020.1.1-2024.3.31)	18
图 28: 中证 500 股票池各模型多头组超额收益净值&回撤 (2020.1.1-2024.3.31)	18
图 29: 中证 500 股票池各模型分年绩效表现 (2020.1.1-2024.3.31)	19
图 30: 中证 1000 股票池各模型因子绩效表现 (2020.1.1-2024.3.31)	19
图 31: 中证 1000 股票池各模型分组年化超额收益 (2020.1.1-2024.3.31)	19
图 32: 中证 1000 股票池各模型多头组超额收益净值&回撤 (2020.1.1-2024.3.31)	20
图 33: 中证 1000 股票池各模型分年绩效表现 (2020.1.1-2024.3.31)	20
图 34: 中证全指股票池 6 随机种子得到的因子值相关系数 (2020.01.01-2024.2.28)	21

图 35：中证全指股票池 6 随机种子得到的 rankIC 相关系数（2020.01.01-2024.2.28）	21
图 36：中证全指股票池中各模型相关性（2020.01.01-2024.3.31）	21
图 37：沪深 300 股票池中各模型相关性（2020.01.01-2024.3.31）	22
图 38：中证 500 股票池中各模型相关性（2020.01.01-2024.3.31）	22
图 39：中证 1000 股票池中各模型相关性（2020.01.01-2024.3.31）	22
图 40：中证全指股票池各模型中性化因子绩效表现（2020.1.1-2024.3.31）	23
图 41：沪深 300 股票池各模型中性化因子绩效表现（2020.1.1-2024.3.31）	23
图 42：中证 500 股票池各模型中性化因子绩效表现（2020.1.1-2024.3.31）	23
图 43：中证 1000 股票池各模型中性化因子绩效表现（2020.1.1-2024.3.31）	23
图 44：沪深 300 股票池指数增强组合绩效表现（2020.1.1-2024.3.31）	24
图 45：沪深 300 股票池指数增强组合超额净值与回撤（2020.1.1-2024.3.31）	25
图 46：沪深 300 股票池指数增强组合相对基准的平均风格暴露（2020.1.1-2024.3.31）	25
图 47：沪深 300 股票池指数增强组合相对基准的风格暴露变化（2020.1.1-2024.3.31）	25
图 48：不同成分股约束下，沪深 300 股票池指数增强组合的绩效表现（2020.1.1-2024.3.31）	26
图 49：不同成分股约束下，沪深 300 股票池指数增强组合超额净值与回撤（2020.1.1-2024.3.31）	26
图 50：不同风险暴露约束下，沪深 300 股票池指数增强组合的绩效表现（2020.1.1-2024.3.31）	27

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/776022102222010131>