

SVM 模型在多因子股票池上的择时应用

——金融工程专题报告

分析师：郝惊

SAC NO: S1150517100003

2019 年 3 月 18 日

证券分析师

郝惊

022-23861600

hao_liang@bhq.com

核心观点：

● 策略逻辑

本文测试了通过传统多因子模型构建股票池，并在给定股票池的基础上，通过 SVM 模型进行择时交易的量化策略。SVM 作为传统机器学习模型中一个非常重要的分类算法，其构建方式思路为，如果数据点离分隔超平面的间隔越远，其预测效果越可信，模型的目的是找到最大支持向量到分隔超平面的距离这一最优化问题的解。

相关研究

《有色金属子行业轮动模型——有色金属行业量化专题》

《黄金、白银量化研究——有色金属行业量化专题》

《以道式理论为基础的顺势、整理体系——金融工程 CTA 策略专题报告》

● 特征集和测试流程

特征集主要使用行情数据以及基于行情数据计算出来的技术指标和统计指标。提取前十二个月股票池中股票的特征集和涨跌情况作为训练集，在训练集内完成交叉验证以确定超参，然后对每只股票逐日预测未来 N 个交易日的涨跌情况，根据预测结果进行择时交易。此后每三个月更新一次训练集，完成交叉验证和确定超参。

● 测试结果

添加日线级别择时后，年化收益率从 18.66% 增加到 22.17%，年化波动率从 34.75% 降低到 27.26%，年均换手率从 1175.23% 降低到 991.46%。添加周线级别择时后，年化收益率从 18.66% 降低到 16.07%，年化波动率从 34.75% 降低到 32.73%，年均换手率从 1175.23% 降低到 987.96%。

● 结论

1、在时间周期选择上，日频周期效果比周频周期效果要好。2、把机器学习模型应用在日频周期择时上可以提高策略表现。3、若训练样本采用滚动方式，市场行情处于顶部位置时，机器学习模型并不能提供即时的逃顶止盈，还需要和传统的止盈止损方法相结合才可以使策略有更好的表现。4、因部分股票存在一直空仓的状态，添加日线或周线择时后，年平均换手率并没有提高，反而有所降低。

● 风险提示

模型过度优化风险，模型失效风险。

目 录

| | |
|-------------------------|---|
| 1. 引言 | 4 |
| 2. 使用传统多因子模型构建股票池 | 4 |
| 3. SVM 模型介绍 | 5 |
| 3.1 理论介绍 | 5 |
| 3.2 特征集 | 6 |
| 3.3 测试流程 | 7 |
| 4. 择时效果测试 | 7 |
| 4.1 日频择时效果测试 | 8 |
| 4.2 周频择时效果测试 | 8 |
| 5. 结论 | 9 |
| 6. 未来研究方向 | 9 |
| 7. 风险提示 | 9 |

图 目 录

| | |
|--------------------------------|---|
| 图 1: 绝对收益择时策略净值强于基准走势 | 5 |
| 图 2: 二维空间分隔超平面示例 | 6 |
| 图 3: 测试流程图 | 7 |
| 图 4: SVM 日频择时+多因子股票池净值走势 | 8 |
| 图 5: SVM 周频择时+多因子股票池净值走势 | 8 |

表 目 录

| | |
|---------------------------------|---|
| 表 1: 因子列表 | 4 |
| 表 2: 绝对收益择时策略指标表现 | 5 |
| 表 3: 特征集 | 6 |
| 表 4: SVM 日频择时+传统多因子策略表现指标 | 8 |
| 表 5: SVM 周频择时+传统多因子策略表现指标 | 8 |

1. 引言

多因子模式是权益量化的主流模型。目前在多因子模型基础上的进一步研究，包括引入行业基本面、使用机器学习模型进行因子收益预测、因子择时等。本文测试了通过传统多因子模型构建股票池，并在给定股票池的基础上，通过 SVM 模型进行择时交易的方法，是否能够提高收益。

2. 使用传统多因子模型构建股票池

多因子选股模型中使用的因子如下表。

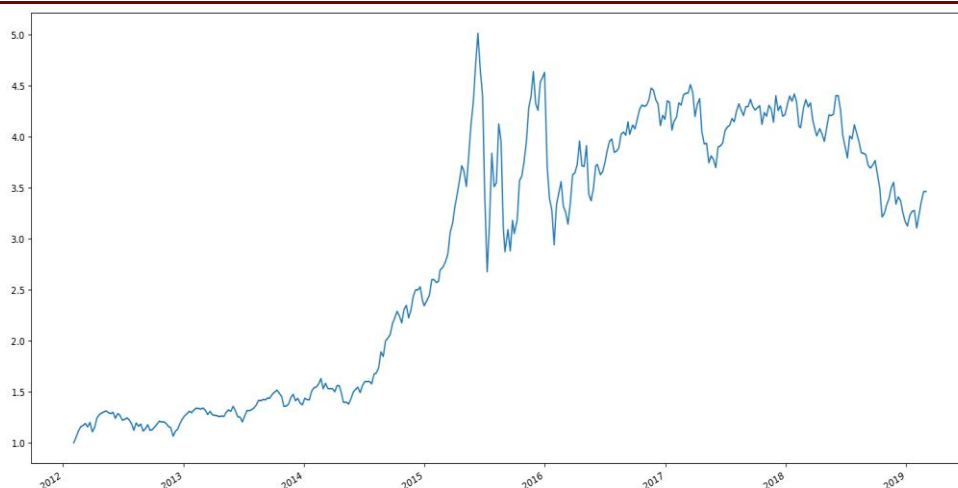
表 1: 因子列表

| 类别 | 名称 | 参数 |
|-------|------------------|-----------------------|
| 估值因子 | EPFWD_Barra | 预期 EP 值 |
| | Epcut_ttm | 1/市盈率 |
| | Bp_rel | 1/相对 PB; PB/行业 PB |
| 盈利因子 | Qfa_roe | 当季净资产收益率 |
| 成长因子 | qfa_yoysales | 单季度营业收入一年增长率 |
| | qfa_yoynetprofit | 单季度归母净利润一年增长率 |
| 波动率因子 | VOL_1 | 成交量月度波动率 |
| | VOL_3 | 成交量季度波动率 |
| 流动性因子 | STOM | 月度平均换手率 |
| | STOQ | 季度平均换手率 |
| | STOA | 年度平均换手率 |
| 动量因子 | RSTR_m12 | 相对强弱动量, T=240, 半衰期 60 |
| | RSTR_m6 | 相对强弱动量, T=120, 半衰期 30 |
| | RSTR_m3 | 相对强弱动量, T=60, 半衰期 15 |
| | RSTR_m1 | 相对强弱动量, T=20, 半衰期 5 |
| 市值因子 | 流通市值市值对数 | |

资料来源: 渤海证券研究所

每个月根据上述因子选择预期收益率最高的 100 只股票作为股票池，股票池暂未做行业中性和市值中性处理，使用移动平均法来预测因子未来收益，均线参数设为 12，设效用函数参数 U 为 1，通过风险模型确定每只股票的权重。测试时间为 2012 年 1 月份至 2019 年 1 月份。

图 1: 多因子过程净值走势



资料来源: Wind, 渤海证券研究所

表 2: 绝对收益择时策略指标表现

| 年化收益率 | 年化波动率 | 夏普比率 | 最大回撤 | 年平均换手率 |
|--------|--------|-------|--------|---------|
| 18.66% | 34.75% | 0.479 | 46.63% | 117.52% |

资料来源: Wind, 渤海证券研究所

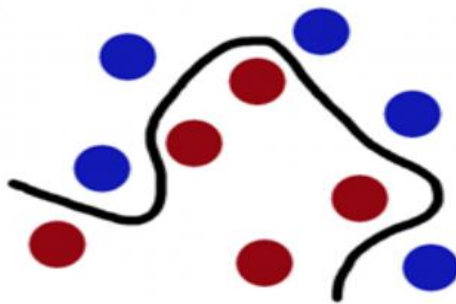
3.SVM 模型介绍

3.1 理论介绍

SVM 作为传统机器学习的一个非常重要的分类算法,它是一种通用的前馈网络类型,最早是由 Vladimir N.Vapnik 和 Alexey Ya.Chervonenkis 在 1963 年提出,目前的版本是 Corinna Cortes 和 Vapnik 在 1993 年提出,1995 年发表。深度学习出现之前,SVM 被认为是机器学习中近十几年最成功表现最好的算法。

如下图中,将红色点和蓝色点分割开来的线,在 SVM 模型中称为分隔超平面。在例子中,由于数据点在二维平面上,所以此时分隔超平面就是一条线,如果数据集是三维的,那么此时分隔超平面就是一个平面,更高维的情况依次类推。在 SVM 模型中,称数据点到分隔超平面的距离为间隔。SVM 模型构建分类器的方式为,如果数据点离分隔超平面的间隔越远,其预测效果越可信。支持向量就是离分隔超平面最近的那些点,模型的目的就是找到最大支持向量到分隔面的距离这一最优化问题的解。

图 2：二维空间分隔超平面示例



资料来源：Wind，渤海证券研究所

SVM 模型中核函数的主要作用就是提供一种从低维空间到高维空间的映射，从而解决部分低维空间中线性不可分的问题，常用的核函数包括线性核函数、多项式核函数、高斯核函数或称为径向基核函数（RBF）、Sigmoid 核函数等等。其中径向基核函数是非线性分类 SVM 的主流核函数。

3.2 特征集

特征集主要使用行情数据以及基于行情数据进一步计算出来的技术指标和统计指标。技术指标的计算可使用 TA-Lib 库。

表 3：特征集

| 类别 | 名称 | 参数 |
|----------|------------|----|
| 价格行情 | Open | 无 |
| | High | |
| | Low | |
| | Close | |
| 均线类 | MA1 | 10 |
| | MA2 | 20 |
| | MA3 | 30 |
| | MA4 | 40 |
| | MA5 | 60 |
| 动量类 | ADX | 20 |
| 技术指标波动率类 | ATR | 20 |
| 截距项 | Intercept1 | 10 |
| | Intercept2 | 20 |
| | Intercept3 | 30 |
| 斜率 | Slope1 | 10 |
| | Slope2 | 20 |
| | Slope3 | 30 |

| | | |
|-------|---------|----|
| | Stddev1 | 10 |
| 统计波动率 | Stddev2 | 20 |
| | Stddev3 | 30 |

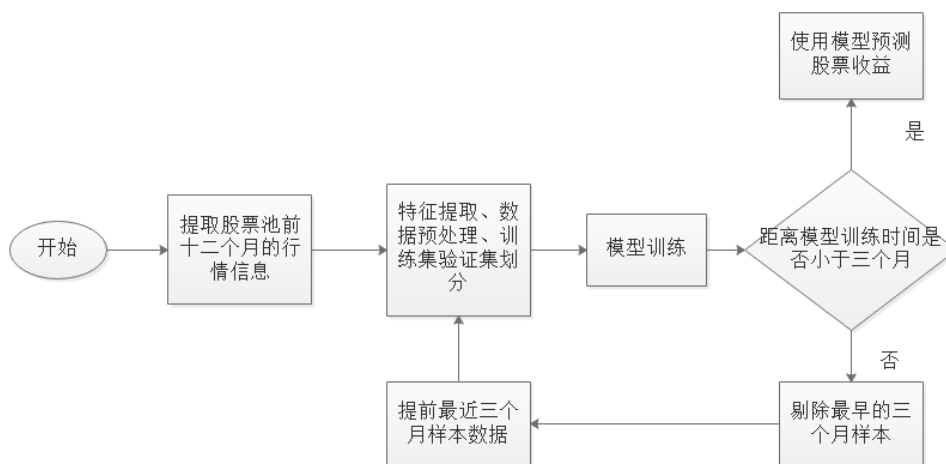
资料来源：渤海证券研究所

使用上述指标作为特征集，即模型中的 X 。使用股票未来 N 个时间周期的上涨或下跌状态为样本值，设上涨状态为 1，下跌状态为 0，及模型中的 Y ，这样就把股票涨跌转化为经典的二分类问题。设定 N 为 3。

3.3 测试流程

提取前十二个月股票池中股票的特征集和涨跌情况作为训练集，在训练集内完成交叉验证以确定超参，然后预测未来一个月的股票涨跌情况，根据预测结果进行择时交易。此后每三个月更新一次训练集，完成交叉验证和确定超参，具体流程如下图：

图 3：测试流程图



资料来源：Wind，渤海证券研究所

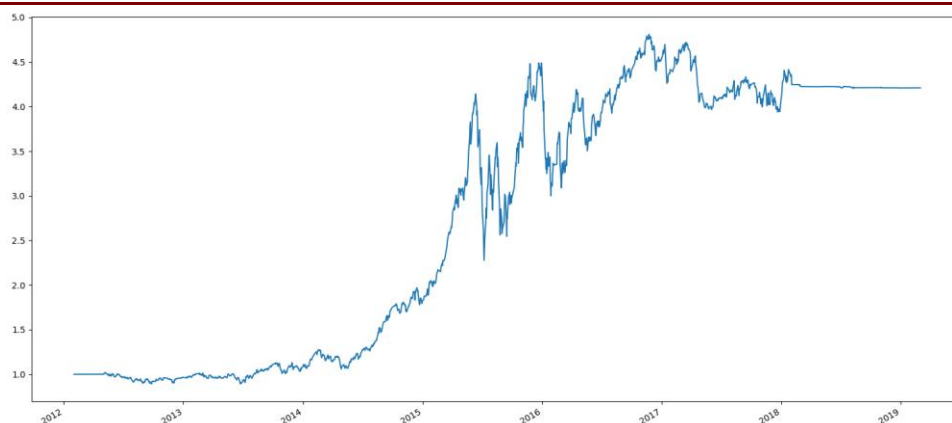
数据预处理过程中，对训练集做正则化处理，预测前对预测集使用训练集的参数进行正则化处理，避免训练过程中引入预测集数据导致未来函数问题。SVM 的核函数选择径向基核函数（RBF），其超参 C 值的待选范围为 1、5、10。

4.择时效果测试

4.1 日频择时效果测试

股票池由第二部分中的多因子选股模型确定，根据 SVM 模型的预测效果，在给定股票池的基础上做择时操作。测试时间与第二部分中相同。

图 4：SVM 日频择时+多因子股票池净值走势



资料来源：Wind，渤海证券研究所

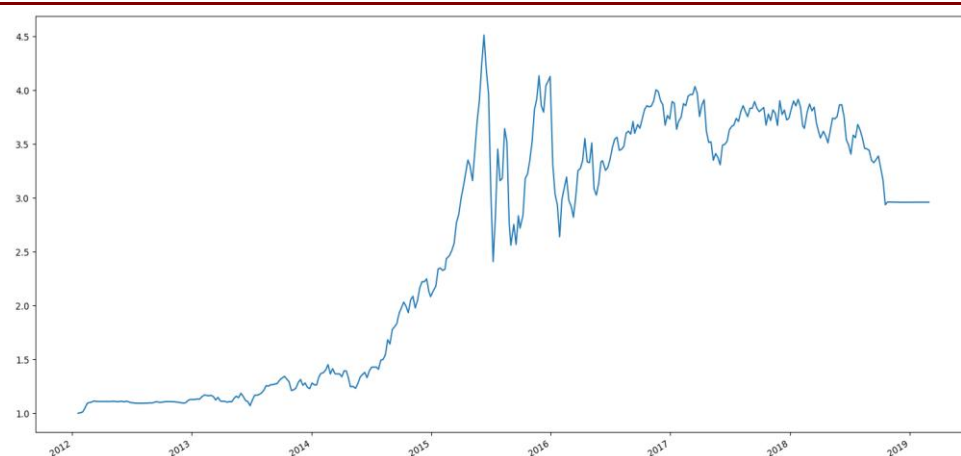
表 4：SVM 日频择时+传统多因子策略表现指标

| 年化收益率 | 年化波动率 | 夏普比率 | 最大回撤 | 年平均换手率 |
|--------|--------|------|--------|--------|
| 22.17% | 27.26% | 0.74 | 44.96% | 991.4% |

资料来源：Wind，渤海证券研究所

4.2 周频择时效果测试

图 5：SVM 周频择时+多因子股票池净值走势



资料来源：Wind，渤海证券研究所

表 5: SVM 周频择时+传统多因子策略表现指标

| 年化收益率 | 年化波动率 | 夏普比率 | 最大回撤 | 年平均换手率 |
|--------|--------|-------|--------|---------|
| 16.07% | 32.73% | 0.430 | 46.63% | 987.96% |

资料来源: Wind, 渤海证券研究所

5.结论

- 1、在时间周期选择上,日频周期效果比周频周期效果要好。
- 2、把机器学习模型应用在日频周期择时上可以提高策略表现。
- 3、若训练样本采用滚动方式,市场行情处于顶部位置时,机器学习模型并不能提供即时的逃顶止盈,还需要和传统的止盈止损方法相结合才可以使策略有更好的表现。
- 4、因部分股票存在一直空仓的状态,添加日线或周线择时后,年平均换手率并没有提高,反而有所降低。

6.未来研究方向

在进一步的研究中,我们将测试更多的机器学习模型,选择出较好的机器学习器;同时测试更多的特征,在特征选择上进一步深化;测试机器学习模型在 CTA 策略上的应用效果。此外,在股票池构建过程中加深对因子择时的研究。

7.风险提示

模型过度优化风险,模型失效风险。

投资评级说明

| 项目名称 | 投资评级 | 评级说明 |
|--------|------|-------------------------------------|
| 公司评级标准 | 买入 | 未来 6 个月内相对沪深 300 指数涨幅超过 20% |
| | 增持 | 未来 6 个月内相对沪深 300 指数涨幅介于 10%~20%之间 |
| | 中性 | 未来 6 个月内相对沪深 300 指数涨幅介于-10%~10%之间 |
| | 减持 | 未来 6 个月内相对沪深 300 指数跌幅超过 10% |
| 行业评级标准 | 看好 | 未来 12 个月内相对于沪深 300 指数涨幅超过 10% |
| | 中性 | 未来 12 个月内相对于沪深 300 指数涨幅介于-10%-10%之间 |
| | 看淡 | 未来 12 个月内相对于沪深 300 指数跌幅超过 10% |

免责声明：本报告中的信息均来源于已公开的资料，我公司对这些信息的准确性和完整性不作任何保证，不保证该信息未经任何更新，也不保证本公司做出的任何建议不会发生任何变更。在任何情况下，报告中的信息或所表达的意见并不构成所述证券买卖的出价或询价。在任何情况下，我公司不就本报告中的任何内容对任何投资做出任何形式的担保，投资者自主作出投资决策并自行承担投资风险，任何形式的分享证券投资收益或者分担证券投资损失书面或口头承诺均为无效。我公司及其关联机构可能会持有报告中提到的公司所发行的证券并进行交易，还可能为这些公司提供或争取提供投资银行或财务顾问服务。我公司的关联机构或个人可能在本报告公开发表之前已经使用或了解其中的信息。本报告的版权归渤海证券股份有限公司所有，未获得渤海证券股份有限公司事先书面授权，任何人不得对本报告进行任何形式的发布、复制。如引用、刊发，需注明出处为“渤海证券股份有限公司”，也不得对本报告进行有悖原意的删节和修改。

渤海证券股份有限公司研究所
所长&金融行业研究

张继袖

+86 22 2845 1845

副所长&产品研发部经理

崔健

+86 22 2845 1618

计算机行业研究小组

王洪磊（部门经理）

+86 22 2845 1975

张源

+86 22 2383 9067

汽车行业研究小组

郑连声

+86 22 2845 1904

陈兰芳

+86 22 2383 9069

食品饮料行业研究

刘瑞

+86 22 2386 1670

电力设备与新能源行业研究

张冬明

+86 22 2845 1857

刘秀峰

+86 10 6810 4658

滕飞

+86 10 6810 4686

医药行业研究小组

赵波

+86 22 2845 1632

甘英健

+86 22 2383 9063

陈晨

+86 22 2383 9062

通信行业研究小组

徐勇

+86 10 6810 4602

公用事业行业研究

刘蕾

+86 10 6810 4662

餐饮旅游行业研究

刘瑞

+86 22 2386 1670

杨旭

+86 22 2845 1879

非银金融行业研究

洪程程

+86 10 6810 4609

中小盘行业研究

徐中华

+86 10 6810 4898

机械行业研究

张冬明

+86 22 2845 1857

传媒行业研究

姚磊

+86 22 2383 9065

电子行业研究

王磊

+86 22 2845 1802

固定收益研究

冯振

+86 22 2845 1605

夏捷

+86 22 2386 1355

朱林宁

+86 22 2387 3123

李元玮

+86 22 2387 3121

金融工程研究

宋旻

+86 22 2845 1131

李莘泰

+86 22 2387 3122

张世良

+86 22 2383 9061

金融工程研究

祝涛

+86 22 2845 1653

郝惊

+86 22 2386 1600

流动性、战略研究&部门经理

周喜

+86 22 2845 1972

策略研究

宋亦威

+86 22 2386 1608

严佩佩

+86 22 2383 9070

宏观研究

宋亦威

+86 22 2386 1608

孟凡迪

+86 22 2383 9071

博士后工作站

张佳佳 资产配置

+86 22 2383 9072

张一帆 公用事业、信用评级

+86 22 2383 9073

综合管理&部门经理

齐艳莉

+86 22 2845 1625

机构销售•投资顾问

朱艳君

+86 22 2845 1995

刘璐

合规管理&部门经理

任宪功

+86 10 6810 4615

风控专员

白骐玮

+86 22 2845 1659

渤海证券研究所

天津

天津市南开区宾水西道 8 号

邮政编码: 300381

电话: (022) 28451888

传真: (022) 28451615

北京

北京市西城区西直门外大街甲 143 号 凯旋大厦 A 座 2 层

邮政编码: 100086

电话: (010) 68104192

传真: (010) 68104192

渤海证券研究所网址: www.ewwww.com.cn