

证券研究报告



分析师:

任瞳

rentong@xyzq.com.cn

S0190511080001

麦元勋

mai yx@xyzq.com.cn

S0190514030001

股指期货交易策略系列报告之二： 基于支持向量机的日内交易策略

2014年10月10日

投资要点

- 基于数据的机器学习是一种人工智能技术，机器学习理论从系统的观测数据出发，探索其内在的规律，通过这些规律对系统未来的发展趋势加以预测。在金融市场中，金融产品的价格通常有上涨和下跌两类，此时，可以用两类支持向量机来预测。
- 本报告首次开发了一种基于支持向量机的股指期货日内交易策略，我们设定了每日开仓一次的交易模式，寻找了9个有效的输入参量，使用滚动的训练窗口，并对训练样本集和回测期进行优化。在不加杠杆情况下，该策略在回测期间实现总收益率72.71%，年化收益率39.43%，胜率为53.15%，最大回撤10.82%。

报告关键点

我们开发了一种基于支持向量机的股指期货日内交易策略。回测结果显示，在不加杠杆的情况下，该策略实现年化收益率39.43%，胜率为53.15%，最大回撤10.82%。

相关报告

《基于多项式回归的股指期货日内交易策略》2014-08-11
《华夏上证50ETF期权投资内参20140930》2014-10-09
《华安上证180ETF期现掘金锦囊20141008》2014-10-08
《华夏上证50ETF期权投资内参20141008》2014-10-08

请阅读最后一页信息披露和重要声明

敬请参阅最后一页特别声明

用使讯资得万供仅档文此

目 录

一、引言	- 3 -
二、支持向量机 (SVM) 的基本原理	- 3 -
三、策略的基本逻辑	- 4 -
四、策略回测结果	- 8 -
五、总结	- 11 -
图 1、最优超平面	- 4 -
图 2、训练样本集	- 6 -
图 3、交易流程	- 6 -
图 4、固定回测期与滚动回测期表现	- 7 -
图 5、滚动回测期示意图	- 7 -
图 6、不同参数收益率分布	- 8 -
图 7、不同回测期累计收益率	- 9 -
图 8、不同回测期的正确率与收益率	- 9 -
图 9、策略累计收益率	- 9 -
图 10、单日收益率分布	- 10 -
图 11、月度收益率分布	- 10 -
表 1、基础数据列表	- 5 -
表 2、输入参量列表	- 5 -
表 3、回测期划分	- 8 -
表 4、回测指标	- 11 -

一、引言

在本系列的第 1 篇报告——《基于多项式回归的股指期货日内交易策略》中，我们采用多项式回归方法对开盘后一段时间内价格序列进行趋势判断，该策略近期具有较好的表现。传统的日内趋势跟踪策略不仅包含日内的价格变化趋势，而且也关注上个交易日的价格信息。经典的 R-Breaker 策略就是通过日内价格与上一个交易日价格的关系做出交易判断。然而，R-Breaker 策略在近几年表现不佳，甚至出现了大的亏损，其主要原因可能在于简单的线性关系已越来越难以描述当前市场价格的波动。

作为一种机器学习算法，支持向量机已经广泛应用在模式识别、数据挖掘和时间序列预测等诸多领域。本报告尝试采用支持向量机方法设计股指期货日内交易策略。在该策略中，我们寻找了 9 个有效的输入参量，使用滚动的训练窗口，并对训练样本集和回测期进行优化，每日开平仓各一次。结果显示，该策略在回测期间实现总收益率 72.71%，年化收益率 39.43%，胜率为 53.15%，最大回撤为 10.82%。

二、支持向量机（SVM）的基本原理

基于数据的机器学习是一种人工智能技术，机器学习理论从系统的观测数据出发，探索其内在的规律，通过这些规律对系统未来的发展趋势加以预测。支持向量机（SVM, Support Vector Machine）是一种机器学习算法，该算法由 Vapnik 等人提出。支持向量机具有能够全局优化，具有较好的泛化性能，训练效率高等优点。随着支持向量机理论的逐步完善，该算法已经在模式识别、数据挖掘和时间序列预测等领域得到广泛的应用。在金融市场中，金融产品的价格通常有上涨和下跌两类，此时，可以用两类支持向量机来预测。

支持向量机在两类问题中，核心在于找到一个最优超平面，将训练样本集分成两部分（见图 1），这两部分对应输入中的两类，即“涨”与“跌”。假定大小为 l 的训练样本集 $\{(x_i, y_i), i = 1, 2, \dots, l\}$ ，由两类组成，分别对应 $y_i = 1$ 与 $y_i = -1$ 。首先，通过一个非线性函数 $\phi(x)$ 将训练集数据 x 映射到一个高维线性特征空间，寻找一个超平面 $\omega \cdot \phi(x) + b = 0$ 使得

$$\max_{\alpha} \begin{cases} L_D = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \phi(x_i) \cdot \phi(x_j) \\ = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j K(x_i, x_j) \end{cases}$$

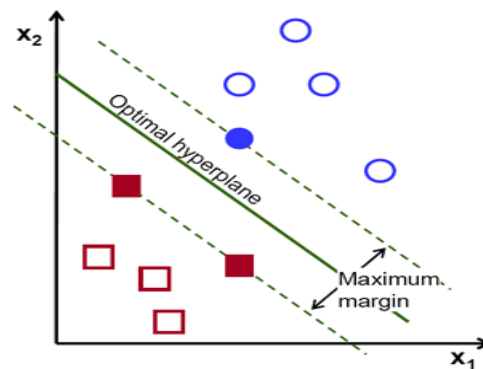
其中， $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$ 被称为核函数， α 为拉格朗日系数。然后通过决策函数便可以推断任意模式 x 对应的分类 y ，决策函数和参数 b 分别为

$$\tilde{f}(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^l y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b\right)$$

$$b = \frac{1}{N_{NSV}} \sum_{\mathbf{x}_i \in JN} (y_i - \sum_{\mathbf{x}_j \in J} \alpha_j y_j K(\mathbf{x}_j, \mathbf{x}_i))$$

上式中， $\text{sign}(\dots)$ 为符号函数， N_{NSV} 为标准支持向量数， JN 为标准支持向量集合， J 为支持向量集合。本篇报告主要使用线性核函数，对核函数不再展开讨论。支持向量机本质上就是找到一种将输入参量对应的模式分成两部分的规则，然后在未来行情对应的输入参量上应用这种规则，从而进行分类，实现未来走势“涨”和“跌”的预测。

图 1、最优超平面



数据来源：兴业证券研究所

三、策略的基本逻辑

支持向量机在股指期货上的应用主要需要解决以下几个问题：

- (1) 输入参量的选择
- (2) 训练样本集大小与策略有效期

首先我们要确定策略的输入参量。输入参量是否有效，决定着策略的“成败”。输入参量的选择以反映价与量的变化为主，本篇报告中，我们选取了上个交易日和今日开盘以来的价格、成交量和波动率指标，具体包括沪深 300 股指期货 5 分钟 K 线的 12 个基本数据（见表 1）。然而这些基础数据在不同维度上，数量级也有很大的差异，若直接作为输入参量，会导致小数淹没在大数中。另一方面，我们应当将尽可能多的有效因素包括在模型之中。于是，我们将这些基础数据进行标准化处理，最终得到 9 个输入参量（见表 2）。

表 1、基础数据列表

基础数据项	定义
昨日开盘价	上个交易日第一条 K 线的开盘价
昨日收盘价	上个交易日最后一条 K 线的收盘价
昨日最高价	上个交易日所有 K 线的最高价
昨日最低价	上个交易日所有 K 线的最低价
昨日波动率	上个交易日所有 K 线收盘价的标准差
昨日成交量	上个交易日所有 K 线的成交量之和
今日开盘价	今日第一条 K 线的开盘价
最新收盘价	今日当前 K 线的收盘价
最新最高价	今日第一条至当前 K 线的最高价
最新最低价	今日第一条至当前 K 线的最低价
最新波动率	今日第一条至当前 K 线的收盘价标准差（换算为一个交易日）
最新成交量	今日第一条至当前 K 线的累计成交量（换算为一个交易日）

数据来源：兴业证券研究所

其中，最新波动率和最新成交量换算为一个交易日的方法为

$$\text{最新波动率} = \text{今日开盘至最新 K 线收盘价标准差} \times \sqrt{\frac{\text{日总 K 线数}}{\text{开盘至当前 K 线数}}}$$

$$\text{最新成交量} = \text{今日最新累计成交量} \times \frac{\text{日总 K 线数}}{\text{开盘至当前 K 线数}}$$

表 2、输入参量列表

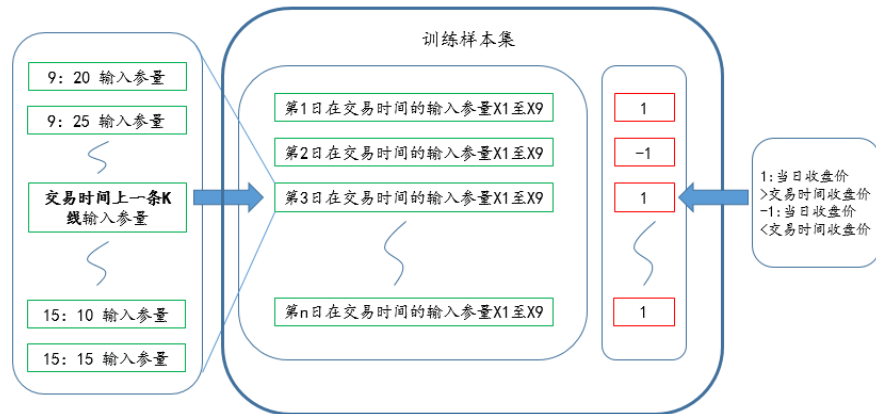
输入参量	标准化处理方法
X1	昨日开盘价/最新收盘价-1
X2	昨日收盘价/最新收盘价-1
X3	昨日最高价/最新收盘价-1
X4	昨日最低价/最新收盘价-1
X5	最新成交量/昨日成交量-1
X6	最新波动率/昨日波动率-1
X7	今日开盘价/最新收盘价-1
X8	最新最高价/最新收盘价-1
X9	最新最低价/最新收盘价-1

数据来源：兴业证券研究所

构建训练样本集时，我们应当在训练样本集和回测期上使用相同的交易规则。我们使用训练样本集内每个交易日**交易时间上一条 K 线**对应所对应的输入参量作为该交易日的训练参量。若此交易日盘尾收盘价大于**交易时间的 K 线**收盘价则认定为“涨”，并分类为 1，反之则认定为“跌”，分类为-1。这种分类方法假设以当前 K 的收盘价开仓，并以当天收盘价平仓（见图 2）。

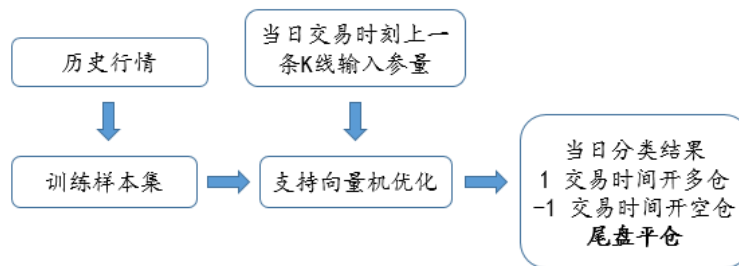
在回测期，通过优化的决策函数与交易日**交易时间上一条 K 线**对应的输入参量得到当日的分类结果，若为 1 则在交易时间开多仓，若为-1 则开空仓，尾盘平仓（见图 3）。

图 2、训练样本集



数据来源：兴业证券研究所

图 3、交易流程

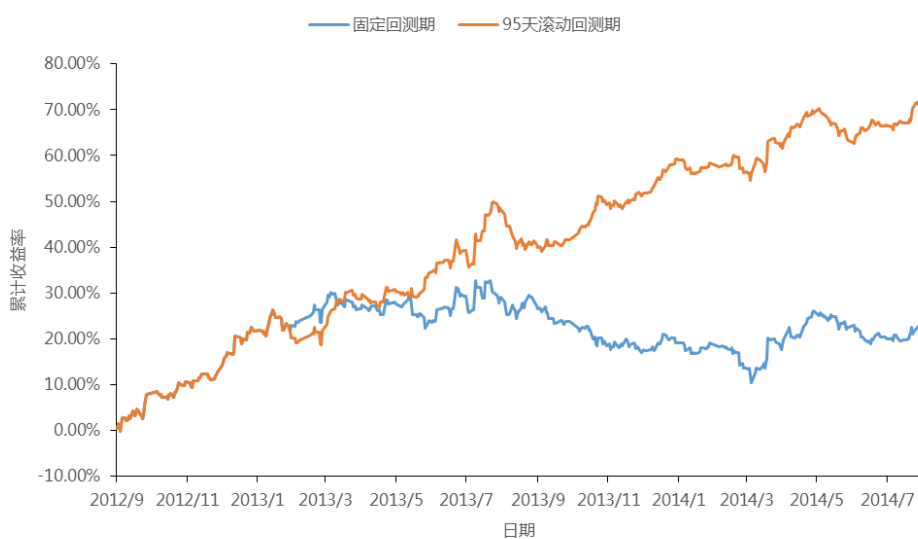


数据来源：兴业证券研究所

市场是不断演化的，过去一段时间内优化得到的规律在未来一段时间内可能有效，但很少是长期有效的。如果规律未能及时更新，策略表现可能会不如人意（见图 4）。

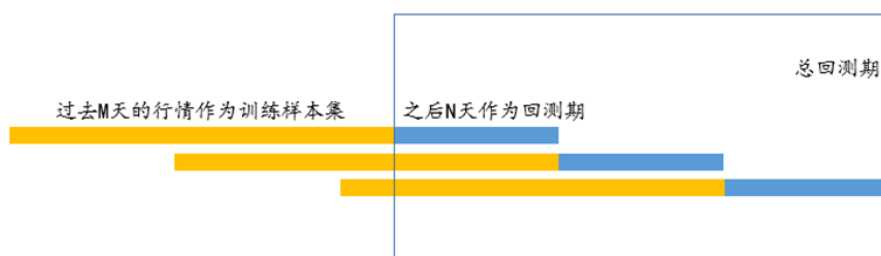
因此，我们需要不断地在优化中使用最新信息，动态优化这种规律。采用滚动的回测期和训练样本集可以很好的避免策略阶段性失效的问题。在本报告中，我们使用过去 M 天的行情作为训练样本集进行支持向量机的优化，利用优化的结果在之后 N 天回测期内进行回测（见图 5）。

图 4、固定回测期与滚动回测期表现



数据来源：兴业证券研究所

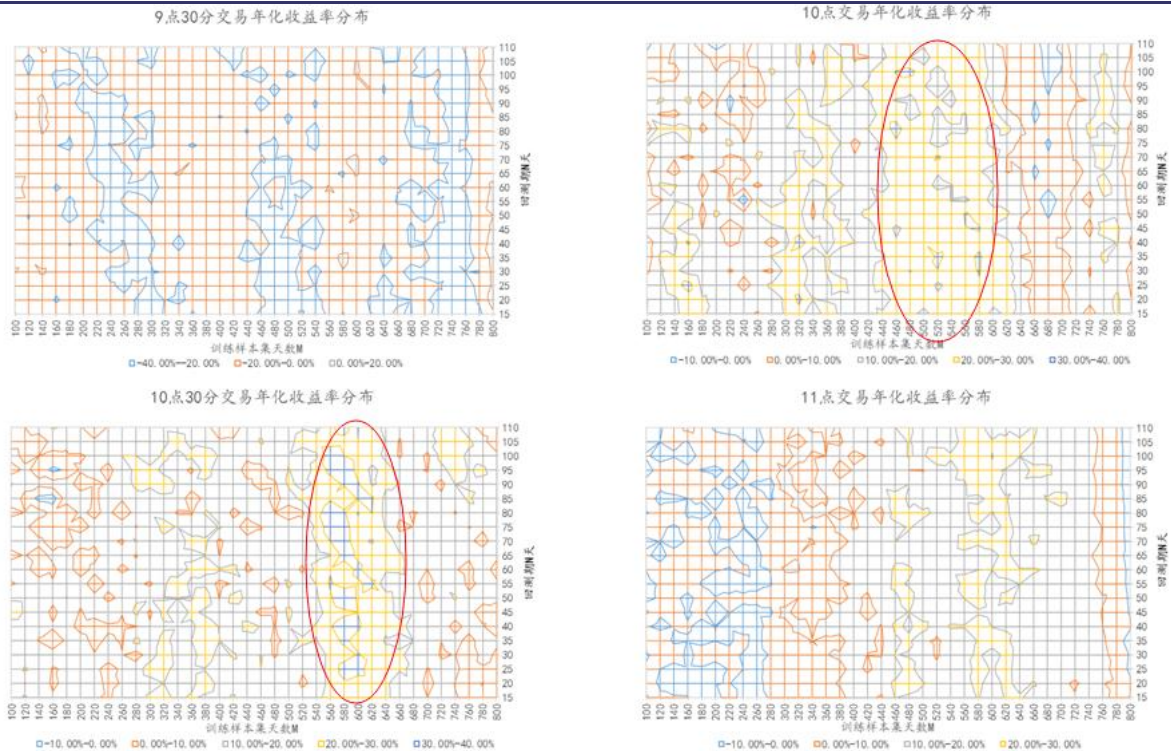
图 5、滚动回测期示意图



数据来源：兴业证券研究所

训练样本集天数 M 和回测期天数 N 是两个很重要的参量。训练样本集过小，会使训练结果陷入某些局部最优值，而过大又不能准确反映近期的规律。此外，我们分别测试了不同交易时间的效果，发现在 10 点至 10 点半之间进行交易，并且使用 560 至 660 天作为训练样本集天数 M ，具有较为稳定和优异的表现。策略对回测期天数 N 反映出策略有效的天数，最终我们选择 95 天。（见图 6，其中横坐标为训练样本集的天数 M ，纵坐标为回测期天数 N ，不同的颜色区域对应不同的年化收益率范围。）

图 6、不同参数收益率分布



数据来源：兴业证券研究所

四、策略回测结果

我们用沪深 300 股指期货主力合约作为标的，K 线的周期取 5 分钟；每日交易时间设置为当日 10 点 30 分，成交价为收盘价，尾盘平仓，手续费为万分之 0.4；采用滚动的训练样本集和回测期，训练样本集为 580 天，回测期为 95 天；支持向量机采用线性核函数；总样本为 2010 年 4 月 16 日至 2014 年 7 月 31 日，但由于需要 580 天的训练样本集，所以总回测期为 2012 年 9 月 3 日至 2014 年 7 月 31 日，其中，可分为 5 个回测期，如下表：

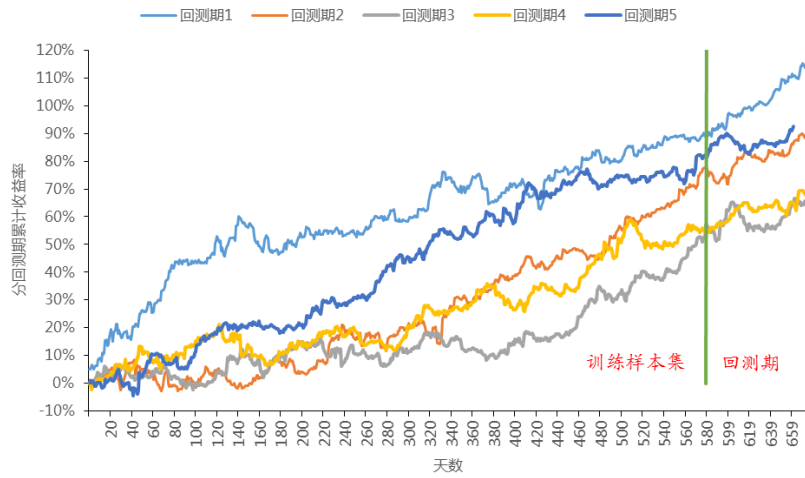
表 3、回测期划分

回测期	时间段
回测期 1	2012 年 9 月 3 日至 2013 年 1 月 23 日
回测期 2	2013 年 1 月 24 日至 2013 年 6 月 24 日
回测期 3	2013 年 6 月 25 日至 2013 年 11 月 13 日
回测期 4	2013 年 11 月 14 日至 2014 年 4 月 3 日
回测期 5	2014 年 4 月 4 日至 2014 年 7 月 31 日

数据来源：兴业证券研究所

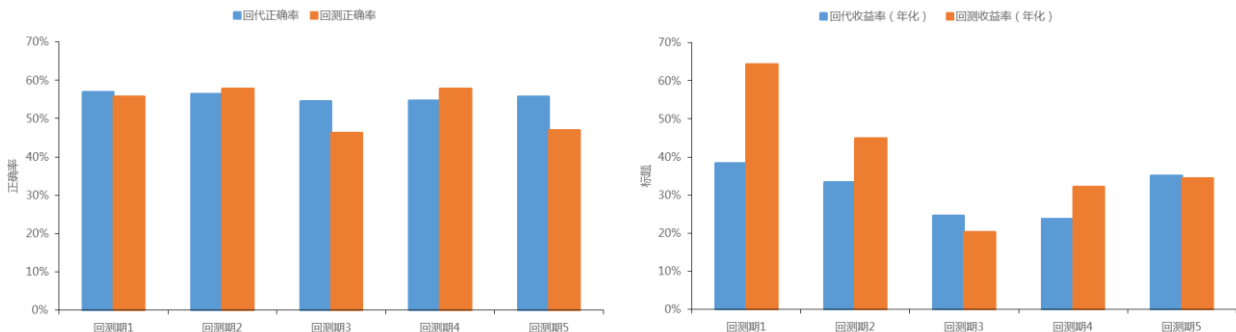
在 5 个回测期中，训练集内和回测期策略的效果具有较好的持续性（见图 7），而回代正确率与回测正确率基本都在 50% 以上（见图 8）。由此可见，通过设计这种交易方法，可以很好的维持策略的有效性。

图 7、不同回测期累计收益率



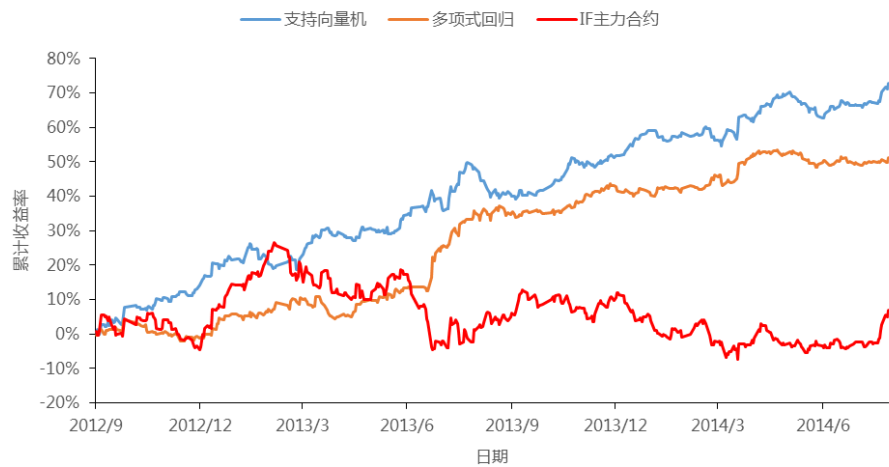
数据来源：兴业证券研究所

图 8、不同回测期的正确率与收益率



数据来源：兴业证券研究所

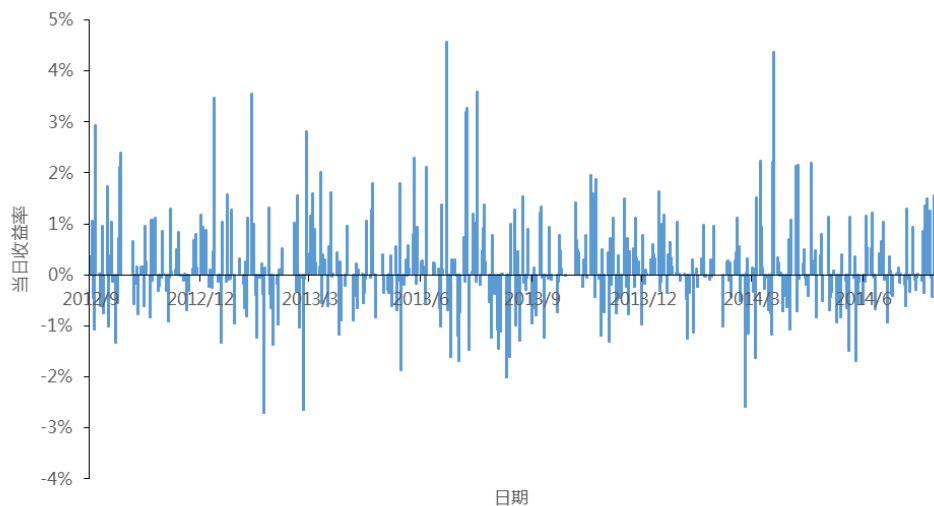
图 9、策略累计收益率



数据来源：兴业证券研究所

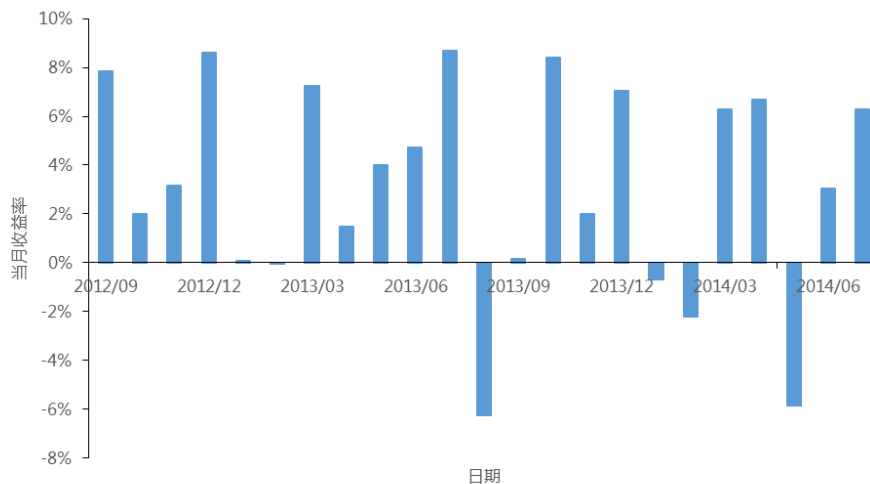
策略在 2012 年 9 月 3 日至 2014 年 7 月 31 日期间的累计收益率为 72.71%，年化收益率为 39.43%，胜率为 53.15%，最大回撤 10.82%。夏普比率为 2.32。在所有回测期的 23 个月中，78% 的月份收益为正，月收益率平均为 3.16%。

图 10、单日收益率分布



数据来源：兴业证券研究所

图 11、月度收益率分布



数据来源：兴业证券研究所

表 4、回测指标

指标	数值
累计收益率	72.71%
年化收益率	39.43%
交易次数	461
盈利次数	245
亏损次数	216
胜率	53.15%
平均收益	0.16%
平均单笔盈利	0.77%
平均单笔亏损	-0.54%
盈亏比	1.43
总的毛盈利	189.83%
总的毛亏损	-117.12%
最大单笔盈利	4.57%
最大单笔亏损	-2.70%
最大连续盈利次数	10
最大连续亏损次数	6
平均连续盈利次数	2.12
平均连续亏损次数	1.88
最大回撤	-10.82%
最长再创新高周期	58
夏普比率	2.32

数据来源：兴业证券研究所

五、总结

支持向量机是一种机器学习算法，已经广泛应用在模式识别、数据挖掘和时间序列预测等多个领域。本报告首次开发了一种基于支持向量机的股指期货日内交易策略，我们设定了每日开仓一次的交易模式，寻找了 9 个有效的输入参量，使用滚动的训练窗口，并对训练样本集和回测期进行优化。策略在回测期间实现总收益率 72.71%，年化收益率 39.43%，胜率为 53.15%，最大回撤 10.82%。

支持向量机结合我们设置的交易规则，可以很好的预测股指期货日内走势，同时，我们也看到许多可以进一步研究的地方，比如，这种二类的分类器不能考虑不开仓这种情况，在某些时候，当日内价格变化幅度低于手续费时，应选择不开仓，这需要用多个支持向量机进行过滤。另外，输入参量的选择也决定着策略的有效性。总之，我们会对支持向量机在股指期货上的应用进行后续研究。

投资评级说明

行业评级 报告发布日后的 12 个月内行业股票指数的涨跌幅度相对同期上证综指/深圳成指的涨跌幅为基准,投资建议的评级标准为:

- 推 荐: 相对表现优于市场;
中 性: 相对表现与市场持平
回 避: 相对表现弱于市场

公司评级 报告发布日后的 12 个月内公司的涨跌幅度相对同期上证综指/深圳成指的涨跌幅为基准,投资建议的评级标准为:

- 买 入: 相对大盘涨幅大于 15% ;
增 持: 相对大盘涨幅在 5% ~ 15%之间
中 性: 相对大盘涨幅在-5% ~ 5%;
减 持: 相对大盘涨幅小于-5%

机构销售经理联系方式					
机构销售负责人		邓亚萍	021-38565916	dengyp@xyzq.com.cn	
上海地区销售经理					
姓名	办公电话	邮 箱	姓名	办公电话	邮 箱
罗龙飞	021-38565795	luolf@xyzq.com.cn	盛英君	021-38565938	shengyj@xyzq.com.cn
杨 忱	021-38565915	yangchen@xyzq.com.cn	王 政	021-38565966	wangz@xyzq.com.cn
冯 诚	021-38565411	fengcheng@xyzq.com.cn	王 溪	021-20370618	wangxi@xyzq.com.cn
顾 超	021-20370627	guchao@xyzq.com.cn	李远帆	021-20370716	liyuanfan@xyzq.com.cn
地址: 上海市浦东新区民生路 1199 弄证大五道口广场 1 号楼 20 层 (200135) 传真: 021-38565955					
北京地区销售经理					
姓名	办公电话	邮 箱	姓名	办公电话	邮 箱
朱圣诞	010-66290197	zhusd@xyzq.com.cn	李 丹	010-66290223	lidan@xyzq.com.cn
肖 霞	010-66290195	xiaoxia@xyzq.com.cn	郑小平	010-66290223	zhengxiaoping@xyzq.com.cn
刘晓浏	010-66290220	liuxiaoliu@xyzq.com.cn	吴 磊	010-66290190	wulei@xyzq.com.cn
何嘉	010-66290195	hejia@xyzq.com.cn			
地址: 北京市西城区武定侯街 2 号泰康国际大厦 6 层 609 (100033) 传真: 010-66290200					
深圳地区销售经理					
姓名	办公电话	邮 箱	姓名	办公电话	邮 箱
朱元贱	0755-82796036	zhuyy@xyzq.com.cn	李 昇	0755-82790526	lisheng@xyzq.com.cn
杨 剑	0755-82797217	yangjian@xyzq.com.cn	邵景丽	0755-23836027	shaojingli@xyzq.com.cn
王维宇	0755-23826029	wangweiyu@xyzq.com.cn			
地址: 福田区中心四路一号嘉里建设广场第一座 701 (518035) 传真: 0755-23826017					
海外销售经理					
姓名	办公电话	邮 箱	姓名	办公电话	邮 箱
刘易容	021-38565452	liuyirong@xyzq.com.cn	徐 皓	021-38565450	xuhao@xyzq.com.cn
张珍岚	021-20370633	zhangzhenlan@xyzq.com.cn	陈志云	021-38565439	chanchiwan@xyzq.com.cn
地址: 上海市浦东新区民生路 1199 弄证大五道口广场 1 号楼 20 层 (200135) 传真: 021-38565955					
私募销售经理					
姓名	办公电话	邮 箱	姓名	办公电话	邮 箱
徐 瑞	021-38565811	xur@xyzq.com.cn	杨雪婷	021-38565470	yangxueting@xyzq.com.cn
地址: 上海市浦东新区民生路 1199 弄证大五道口广场 1 号楼 20 层 (200135) 传真: 021-38565955					

【信息披露】

本公司在知晓的范围内履行信息披露义务。客户可登录 www.xyzq.com.cn 内幕交易防控栏内查询静默期安排和关联公司持股情况。

【分析师声明】

本人具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，以勤勉的职业态度，独立、客观地出具本报告。本报告清晰准确地反映了本人的研究观点。本人不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接收到任何形式的补偿。

【法律声明】

兴业证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本报告仅供兴业证券股份有限公司（以下简称“本公司”）的客户使用。本公司不会因接收人收到本报告而视其为客户。客户应当认识到有关本报告的短信提示、电话推荐等只是研究观点的简要沟通，需以本公司 <http://www.xyzq.com.cn> 网站刊载的完整报告为准，本公司接受客户的后续问询。

本公司的销售人员、交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。本公司没有将此意见及建议向报告所有接收者进行更新的义务。

本公司的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。

本公司系列报告的信息均来源于公开资料，我们对这些信息的准确性和完整性不作任何保证，也不保证所包含的信息和建议不会发生任何变更。我们已力求报告内容的客观、公正，但文中的观点、结论和建议仅供参考，报告中的信息或意见并不构成所述证券的买卖出价或征价，投资者据此做出的任何投资决策与本公司和作者无关。

在法律许可的情况下，兴业证券股份有限公司可能会持有本报告中提及公司所发行的证券头寸并进行交易，也可能为这些公司提供或争取提供投资银行业务服务。因此，投资者应当考虑到兴业证券股份有限公司及/或其相关人员可能存在影响本报告观点客观性的潜在利益冲突。投资者请勿将本报告视为投资或其他决定的唯一信赖依据。

若本报告的接收人非本公司的客户，应在基于本报告作出任何投资决定或就本报告要求任何解释前咨询独立投资顾问。

本报告的版权归本公司所有。本公司对本报告保留一切权利。除非另有书面显示，否则本报告中的所有材料的版权均属本公司。未经本公司事先书面授权，本报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。未经授权的转载，本公司不承担任何转载责任。