

机器学习：开拓金融量化新前沿

——机器学习系列报告之一

金融工程深度

机器学习如今无时无刻不在影响着世界，越来越多的工具和产品被发明出来以替代以往需要大量人工操作的作业，有些甚至是行业专家才能完成的工作。在棋类运动上，深蓝（DeepBlue）、AlphaGo 等程序已经陆续战胜了人类顶尖棋手，而在金融领域，世界上第一个人工智能 ETF 也在 2017 年诞生。虽然该 ETF 目前的表现较为一般，但却足够促使我们思考一个问题：机器学习或者人工智能是否会深刻地影响到金融领域？如果是的话，我们又该如何探索机器学习在金融领域的应用方式。

分析师

刘均伟（执业证书编号：S0930517040001）
 021-52523679
liujunwei@ebcn.com

联系人

胡骥聪
 021-52523683
hujicong@ebcn.com

- ◆ **金融领域尚是机器学习下一片蓝海。**21 世纪开始，尤其是最近 10 年，虽然机器学习的研究与应用成果在自然科学领域有了突飞猛进的进步，但在金融领域的应用却受到了挑战。基于对海外主流投行在机器学习应用情况的数量，我们认为机器学习在金融服务方面已经崭露头角，而在金融投资领域则有待观察。
- ◆ **量化投资结合机器学习的趋势不可阻挡。**量化投资是机器学习天然的应用场景，我们从整体环境的改变，机器学习实现难度的降低，传统量化遇到的瓶颈这三个角度论证了量化投资为何要研究机器学习。在这三个方面相互作用下，机器学习在未来量化研究中的作用将越来越大，如何找到机器学习合适的运用方式就成了当前量化研究领域最有前景的研究方向之一。
- ◆ **金融领域下机器学习的正确用法仍需探索。**许多在看似直观的应用方式下直接套用机器学习算法的做法往往并不能达到预期的效果。在金融投资领域下运用机器学习的方式可能需要更进一步的探索与尝试。我们认为可以尝试的方向包括但不限于数据处理，“白箱”化处理，过拟合程度的估计等等。我们构造三个简单的示例，在第一个示例中我们验证了可以将交易数据结构化为更适合机器学习算法的结构，成交量或成交额等分 K 线比传统时间等分 K 线有更适合的统计特征，用它们作为输入数据可以得到更好更稳定的结果；第二个示例展示通过研究算法下不同输入特征的特征重要性，可以更好地理解在算法之后是什么因素在真正起到预测作用；最后一个示例中我们通过构建生成数据研究 RSRs 择时策略的路径依赖程度，并一定程度上测试了我们所使用的开平仓阈值参数值是否为最优选择。
- ◆ **风险提示：**测试结果均基于模型和历史数据，模型存在失效的风险。

目 录

| | |
|-------------------------------|----|
| 1、 金融是机器学习下一个主战场 | 5 |
| 1.1、 自然科学领域：大放异彩 | 5 |
| 1.2、 金融领域：服务崭露头角、投资有待观察 | 6 |
| 2、 探索金融领域下机器学习的正确用法 | 11 |
| 2.1、 量化投资为什么要研究机器学习 | 11 |
| 2.2、 示例一：将数据处理成更合适的结构 | 12 |
| 2.3、 示例二：利用特征重要性打开“黑箱” | 15 |
| 2.4、 示例三：利用生成数据了解过拟程度 | 18 |
| 3、 风险提示 | 21 |

图目录

| | |
|---|----|
| 图 1：机器学习算法发展..... | 5 |
| 图 2：各大投行发展机器学习研究团队..... | 7 |
| 图 3：AI Powered Equity ETF (AIEQ) 与标普 500 指数走势比较..... | 8 |
| 图 4：海外利用机器学习算法预测股票价格的论文 | 9 |
| 图 5：机器学习在金融投资领域接受度较低 | 10 |
| 图 6：时间等分 K 线收益率分布..... | 14 |
| 图 7：Tick 等分 K 线收益率分布..... | 14 |
| 图 8：成交量等分 K 线收益率分布 | 14 |
| 图 9：成交额等分 K 线收益率分布 | 14 |
| 图 10：在随机森林算法下不同输入特征在预测 EBQC (光大综合质量) 因子收益方向上的特征重要性..... | 18 |
| 图 11：沪深 300 生成数据曲线示例 | 19 |
| 图 12：生成曲线 (1) 上测试原有参数下的 RSRS 策略 | 20 |
| 图 13：生成曲线 (2) 上测试原有参数下的 RSRS 策略 | 20 |
| 图 14：RSRS 策略在实际沪深 300 曲线上的历史回测效果..... | 20 |
| 图 15：不同参数 S 的平均排名 | 21 |

表目录

表 1：AI Powered Equity ETF (AIEQ) 与标普 500 指数统计数据..... 8

表 2：机器学习基金失败的 7 大原因 11

表 3：多种 K 线切片采样方式 13

表 4：不同 K 线切片方式下收益率数据的正态性 14

表 5：模型在各个股票上预测准确率统计 15

表 6：不同 K 线切片数据下模型预测准确率..... 15

表 7：测试因子明细表 16

表 8：择时特征变量名单..... 17

表 9：随机森林算法在不同因子上应用的特征重要性示例 17

表 10：固定参数 N，M 取值原始值，不同参数 S 取值下 RSRS 策略在图 12 与图 13 的生成数据上的最终净值 ... 20

1、金融是机器学习下一个主战场

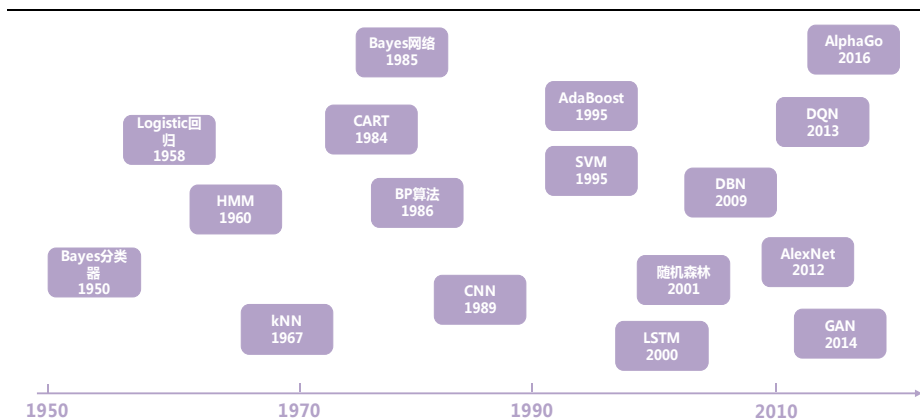
机器学习如今无时无刻不在影响着世界，越来越多的工具和产品被发明出来以替代以往需要大量人工操作的作业，有些甚至是行业专家才能完成的工作。在棋类运动上，深蓝（DeepBlue）、AlphaGo 等程序已经陆续战胜了人类顶尖棋手，而在金融领域，世界上第一个人工智能 ETF 也在 2017 年诞生。虽然首款人工智能 ETF 目前的表现较为一般，但却足够促使我们不得不思考一个问题：机器学习或者人工智能是否会深刻地影响到金融领域？

1.1、自然科学领域：大放异彩

普通大众对于机器学习的认知始于 2016 年 3 月阿尔法狗(AlphaGo)和世界顶级围棋大师李世石的人机大战。大赛以谷歌(Google)开发的机器人 AlphaGo 4:1 战胜人类结束。自此，机器学习在各个领域的应用逐渐拓展开来，例如，将双语翻译正确率提升 60%，帮助物理学家寻找新的超级材料，等等不一而足。

实际上，机器学习概念的提出与应用远比大部分人想象的要早。1980 年代相对正式的机器学习概念就已经被提出，目前比较主流的算法如神经网络、支持向量机（SVM）、树模型（CART）等也在 21 世纪之前就已经较为成熟，而相对简单的逻辑回归、kNN 等算法更可以追溯到 1950 年代。甚至目前最为流行的深度学习也在 2006 年就已经出现在正式报告里。然而虽然机器学习的概念与算法在很早时期就已经成熟，但在当时并未取得引人瞩目的成就，乃至在 20 年代末机器学习已然有些日渐式微的迹象。

图 1：机器学习算法发展



资料来源：光大证券研究所整理

但从 21 世纪开始，尤其是最近 10 年，机器学习的发展与应用成果有了突飞猛进的进步。如今机器学习在各个行业大获成功应用主要归功于三个方面：数据、算法和算力。近年来，在学术界和企业界的共同推动下，数据量和算力都有了很大的提高，算法也小有突破。而在谷歌、亚马逊、IBM、微软、阿里巴巴的推动下，高效又便宜的算力也变得越来越容易获得。在这种

背景下，机器学习在各个行业的应用前景更多的则是取决于数据在该行业的积累。

受益于大数据与大算力，我们早已无时无刻不处于机器学习环绕的世界，我们手机上的语音识别，面部认证；购物网站或视频网站针对客户消费或浏览习惯发送的推送内容。那么同样有许多数据积累的金融行业，机器学习的应用与发展又如何呢？

1.2、金融领域：服务崭露头角、投资有待观察

李开复在他的著作《人工智能》里说，“金融行业是 AI（人工智能）目前最被看好的落地领域。机器学习将会被用来模拟、识别、分析、自动化、预测、高频交易”；“未来五年，算法-量化交易策略将会取代人类，主宰资产管理领域”。其理由包括：各垂直行业相比较而言，金融是全球大数据积累最好的行业；银行、保险、证券等产业其本身业务就是基于大量数据和统计来开展的；金融行业也非常注重数据的规范化、采集自动化、存贮、集中和共享。

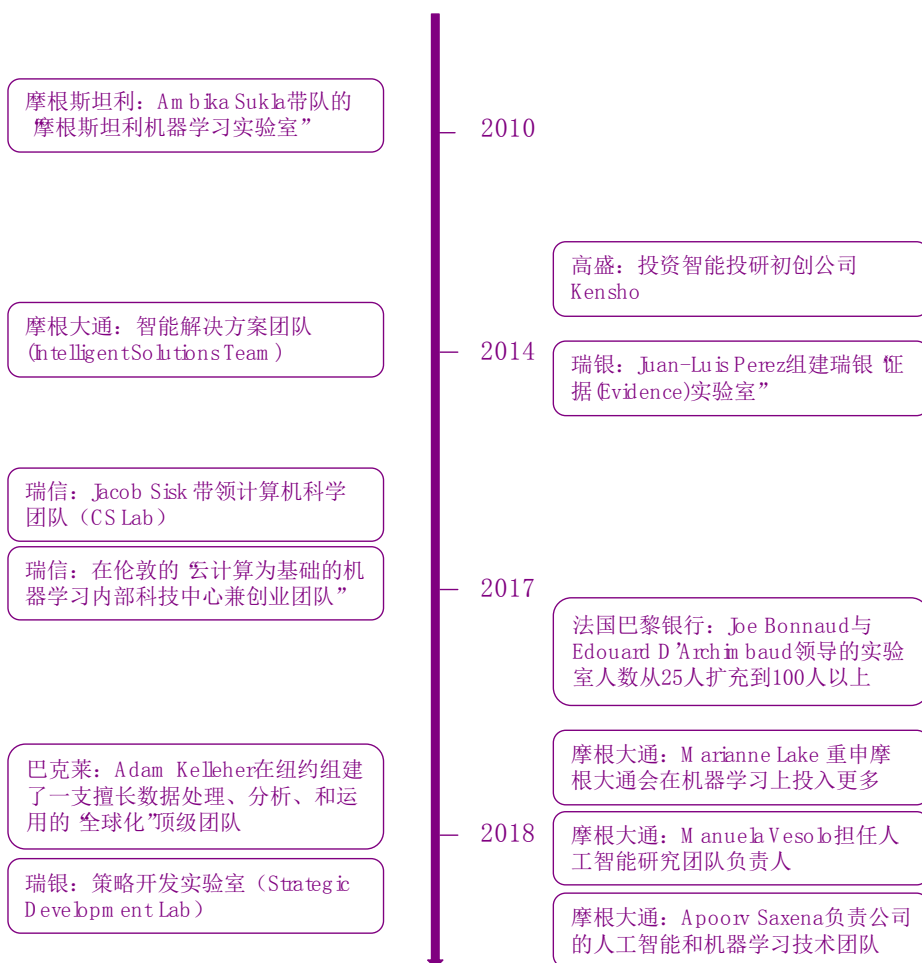
按照目前金融行业本身的业务，可以大致分为金融服务与金融投资两大类。对海外主要投行在机器学习上的布局进行梳理之后，我们认为：**机器学习在金融服务上的应用已经崭露头角，但在金融投资中的作用还有待观察。**

首先，各大投行目前在机器学习上的布局主要以搭建机器学习实验室为主，策略开发实验室的组建则较为罕见，且相对较晚。早在 2010 年，摩根斯坦利就组建了由 Ambika Sukla 带队的“摩根斯坦利机器学习实验室”。2014 年，机器学习和金融专家 Juan-Luis Perez 在瑞银组建“证据(Evidence)实验室”。2017 年，瑞信在旧金山拥有由前 Capital One 数据科学团队主管 Jacob Sisk 领导的“CS 实验室”。而在 2018 年，摩根大通、瑞银、巴克莱相继组建了偏策略开发的机器学习团队。可以看出，金融机构在机器学习领域的投入在近几年明显有加大的趋势。

其次，从各大投行基于机器学习发布的金融产品来看，同样集中在金融服务方面，以智能算法替代传统的人工业务，而聚焦在投资领域的产品则乏善可陈。比如摩根大通 2017 年开发了一款金融合同解析软件 COIN，可以在几秒钟内完成之前律师和贷款人员需要 360000 小时才能完成的工作，且错误率大大降低。又比如瑞银发布了一个机器学习系统，可以自动处理客户的建仓请求邮件和调仓请求邮件，将该工作时间从 45 分钟缩短至 2 分钟，大大提高了工作效率。

最后，即使是与投资相关的产品，也都局限在投资流程中相对简单的部分，例如算法交易、风险分析等。比如摩根大通发布的 LOXM(机器学习交易系统)在 2017 年第一季度已经在欧洲算法交易中使用，表现极佳。真正的投资决策过程，还极少直接采用机器学习的成果。

图 2：各大投行发展机器学习研究团队

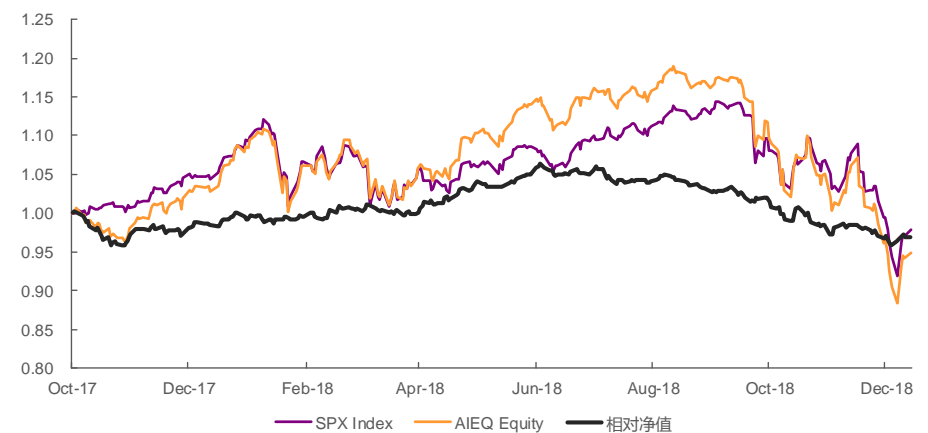


资料来源：光大证券研究所绘制

当然，无论是各大投行，还是对冲基金，甚至是大学研究机构，都一直在尝试把机器学习应用于金融投资之中。比如德意志银行在 2017 年底公开宣布的机器学习交易平台 Autobahn 2.0，能够实时的预测股票价格、交易量、和动量(momentum)，并依据实时变化调整交易策略。但从首款人工智能 ETF 的表现来看，机器学习在金融投资中的应用效果还有待观察。

作为首只在纽交所挂牌的人工智能 ETF——AI Powered Equity ETF (AIEQ)，AIEQ 背靠 IBM 的 Watson AI，从美国市场挑选 40 至 70 只股票，日频调仓，希望在同等波动水平下战胜美国大盘指数。但该 ETF 自 2017 年 10 月 17 日成立至今表现并不稳定，特别是 2018 年 10 月份至今的巨大回撤令人诧异。截至 2018 年 12 月 31 日，AIEQ 年化收益-4.34%，年化波动 17.61%，最大回撤 25.65%，夏普比率-0.17；而同期标普 500 指数收益年化-1.78%，年化波动 15.77%，最大回撤 19.78%，夏普比率-0.04。可见机器学习与人工智能在金融投资上的运用远不如我们想象中的那么容易。

图 3: AI Powered Equity ETF (AIEQ) 与标普 500 指数走势比较



资料来源: BLOOMBERG, 注: 数据截至 2018/12/31

表 1: AI Powered Equity ETF (AIEQ) 与标普 500 指数统计数据

| 标的名称 | 年化收益 | 年化波动 | 最大回撤 | 夏普比率 |
|----------------|--------|--------|---------|-------|
| S&P 500 | -1.78% | 15.77% | -19.78% | -0.04 |
| AIEQ | -4.34% | 17.61% | -25.65% | -0.17 |
| AIEQ / S&P 500 | -2.62% | 5.49% | -9.80% | -0.46 |

资料来源: BLOOMBERG, 注: 数据截至 2018/12/31

我们认为, 机器学习在金融投资中的应用效果不及预期, 一方面与金融数据不稳定、信噪比偏低有关, 另一方面也与投资者积极性欠佳有关。

首先, 由于金融数据不稳定、信噪比偏低, 机器学习在金融投资领域的应用方向看似直观, 但实证效果往往不如人意。预测市场未来涨跌(分类问题), 预测股票未来价格(回归问题)都是投资决策中的核心问题, 然而绝大多数投资者在应用机器学习来达到上述目的时, 结果都令人大失所望。即使在回测过程中获得了不错的结果, 策略在样本外的有效性确往往不太令人满意。

图 4：海外利用机器学习算法预测股票价格的论文

Algorithms for stock price direction prediction used in literature.

| | Prediction method | | | | | | |
|--|-------------------|----|----|-----|----|----|----|
| | LR | NN | KN | SVM | AB | RF | KF |
| Schöneburg (1990) | | x | | | | | |
| Bessembinder and Chan (1995) | x | | | | | | |
| Brownstone (1996) | x | x | | | | | |
| Saad, Prokhorov, and Wunsch (1996) | | x | | | | | |
| Kim and Chun (1998) | | x | | | | | |
| Saad, Prokhorov, and Wunsch (1998) | | x | | | | | |
| Kim and Han (2000) | | x | | | | | |
| Kuo et al. (2001) | | x | | | | | |
| Kim (2003) | | x | | x | | | |
| Kim and Lee (2004) | | x | | | | | |
| Rodriguez and Rodriguez (2004) | x | x | | | x | x | |
| Huang et al. (2005) | | x | | x | | | |
| Kumar and Thenmozhi (2006) | x | x | | x | | x | |
| Lunga and Marwala (2006) | | | | | x | | |
| Wu et al. (2006) | | | | | | | |
| Wang and Chan (2007) | x | | | | | | |
| Huang et al. (2008) | x | x | x | x | | | |
| Senol and Ozturan (2008) | x | x | | | | | |
| Lai, Fan, and Chang (2009) | | | | | | | |
| Lee (2009) | | x | | x | | | |
| Ou and Wang (2009) | x | x | x | x | | | |
| Kara, Boyaciogly and Baykan (2011) | | x | | x | | | |
| Wei and Cheng (2011) | | x | | | | | |
| Subha and Nambi (2012) | x | | x | | | | |
| Lin, Guo, and Hu (2013) | | | | x | | | |
| De Oliveira, Nobre, and Zárate (2013) | | x | | | | | |
| Chen, Chen, Fan, and Huang (2013) | | x | | | | | |
| Rechenthin et al. (2013) | | | x | x | x | | |
| Ji, Che, and Zong (2014) | | x | | | | | |
| Bisoi and Dash (2014) | | x | | | | | |
| Zikowski (2015) | | | | x | | | |
| Hafezi, Shahrabi, and Hadavandi (2015) | | x | | | | | |
| Patel et al. (2015) | | x | | x | | x | |

资料来源：Evaluating multiple classifiers for stock price direction prediction, Expert Systems with Applications 42 (2015) 7046–7056

其次，虽然金融投资领域对机器学习的接受程度越来越大，但依然有不少业界人士对其应用前景持怀疑态度。除了金融数据的不稳定性、算法的过拟合问题、构建的新因子失效速度过快等技术难题之外，也有投资者认为是否应用机器学习并不重要，因为业界成功的投资公司中，既有应用了机器学习的公司，也有完全没有应用机器学习的公司。换言之，是否应用机器学习与投资是否成功没有必然的联系。特别是，机器学习模型相对黑箱的特点，让投资理念相对成熟的基金经理很难接受，在投资过程中也较难应用。

图 5：机器学习在金融投资领域接受度较低



资料来源：光大证券研究所绘制

综上所述，虽然目前机器学习在金融投资中的应用并不顺利，但我们认为其依然存在研究和尝试的价值，特别是在找到合理的运用方式之后，机器学习有潜力在金融投资，尤其是量化金融领域，开辟一条新的道路。目前海内外都已经有了先行者在探索这条道路，美国具有代表性的比如：WealthFront, Betterment, Jane Street, BridgeWater 等，而中国则是 AFA Finance 走在前沿。

目前，国内外金融机构中，机器学习在金融投资或策略开发中的主要应用场景集中于以下几类：

1. 资产配置方案，资产分类方式；
2. 生成新的 alpha 因子，alpha 因子结合方式；
3. 参数测试，模型优化，最优策略挑选；
4. 避免人为因素在交易策略制定中的影响等。

从上述应用场景来看，主流的应用场景都与量化投资息息相关。因此，接下来的章节，我们将进一步聚焦在量化投资领域，探讨为什么量化投资需要研究机器学习，以及应该从什么角度出发研究机器学习。

2、探索金融领域下机器学习的正确用法

在第一章节中，我们已经讨论过，许多在看似直观的应用方式下直接套用机器学习算法的做法往往并不能达到预期的效果。在金融投资领域下运用机器学习并非想象中那么容易。我们认为在算法本身已经日益强大的今天，要想利用机器学习取得金融投资的成功，需要更多投入精力与研究的不在于算法本身，而在于算法之外。在运用模型算法前应该做哪些处理，在运用算法过程中应该注意哪些细节，在运用算法后又该如何看待算法给出的结果等等。也正是基于这样的想法，我们将在 2019 年推出我们的机器学习系列报告。该系列报告的重心并不会放在利用各个不同的算法模型实现具体的策略上，更多的则是着眼于算法外合理的细节处理。

我们不少灵感与想法是源于一篇论文《The 7 Reasons Most Machine Learning Funds Fail》。在该文中作者 Marcos López de Prado 提到了造成机器学习投资失败的最主要的 7 个原因。在指出失败原因的同时其实也或直接或间接为我们的研究重心指出了方向。在我们机器学习系列报告接下来的一系列报告中，我们也将研究并展示这些大家关心的问题的一些解决方案或合理的研究方式。

表 2：机器学习基金失败的 7 大原因

| | 原因（原文） | 观点 |
|---|------------------------------|----------------|
| 1 | The Sisyphus paradigm | 团队协作影响 |
| 2 | Integer differentiation | 并不应该全部使用整数差分 |
| 3 | Inefficient sampling | 使用了不科学的采样方式 |
| 4 | Wrong labeling | 使用了不科学的打标签方式 |
| 5 | Weighting of non-IID samples | 非独立同分布样本的赋权问题 |
| 6 | Cross-validation leakage | 错误应用交叉验证导致信息泄露 |
| 7 | Backtest overfitting | 回测过拟合严重 |

资料来源：The 7 Reasons Most Machine Learning Funds Fail, Marcos López de Prado

作为该系列的开篇报告，在接下来的篇幅里，我们将首先花一点篇幅说明：做量化为什么要研究机器学习。之后则从广受机器学习研究者关注的角度展示三个示例用以表明我们认为有潜力的研究方向。虽然这些示例仅是我们后续专题报告中相对粗浅的研究结论，但已经可以从中一窥这些研究方向在金融机器学习中的巨大前景。

2.1、量化投资为什么要研究机器学习

在上一章结尾，我们已经知道在金融投资领域，有投资人士认为“有的公司利用了机器学习很成功，也有些没利用机器学习的公司同样很成功。”那么，既然不用机器学习照样可以很成功，量化投资为什么还要研究机器学习呢？

我们认为有三个原因：整体环境的改变，机器学习实现难度的降低，以及传统量化遇到的瓶颈。我们分别阐述这三个方面对于是否研究机器学习带来的影响。

整体环境的改变：市场的整体环境是不断改变的，这种改变以一种不可逆的方式向电子化发展。而随着整体环境的变化，以往常见的商业模式可能会逐步被新的方式取代。我们知道在纽约交易所电子化以前，许多经纪人（Broker，或者叫 Floor Trader）在纽交所大厅里为客户执行交易操作，并通过它来赚取佣金。在交易所刚引进电子撮合时，由于大部分人还没有完全适应新的交易环境，仍然有不少 Floor Trader 能做的很成功。然而当越来越多的经纪人开始熟悉利用更有效率的电子撮合机制，那些仍然坚持在大厅以传统方式下单的经纪人越来越难做。直至今今，很难再见到在大厅依靠手势下单的经纪人们了。虽然目前机器学习在金融投资领域的应用尚不完全成熟，但如果未来随着越来越多的人与团队参与并熟练运用研究效率更高的机器学习，那么坚持不用机器学习的研究者将面临巨大的挑战，甚至是淘汰。而正如我们在上一章所看到的，如今尝试运用机器学习的团队的确越来越多。而通过后面第二点“机器学习实现难度的降低”的剖析，我们认为市场上运用机器学习的参与者占比越来越大的趋势还将继续。换句话说，“现在不用机器学习也能成功的公司，在未来的金融环境中如果仍然不用机器学习，未必能继续成功下去了”。

机器学习实现难度的降低：机器学习在未来的实现难度是在不断降低的，而且这种难度的降低是在算法、算力、数据三个维度上全面推进。算法在日新月异的更新，随着 Google 等科技巨头算法开源，世界各地的思想在 GitHub 上交汇，任何一个新的算法进展都能借由互联网在短时间传遍全世界。算力也在变得越来越强，被验证了几十年的摩尔定律依然发挥着神奇的效果，我们使用的硬件设施每间隔 1 到 2 年，就会有突破性的升级。而每时每刻在全球各个市场上发生的交易，都在扩充着我们日后可以使用到的数据信息。可以预见，随着时间的推移，运用机器学习的成本将会变得越来越低。

传统量化遇到的瓶颈：传统量化从 1950 年代 Markowitz 建立均值方差模型，到 1990 年代 alpha 因子策略的兴起，再到如今 21 世纪的广为人知。但随着对各种理论与体系的逐步完善，研究者们都逐渐开始发现传统量化的一些瓶颈，比如如何处理非线性问题。无论是资产配置，还是新因子挖掘，线性体系的完善程度已经很高，研究对于该体系的边际改善非常低。几年前大家还能基于线性逻辑开发出较多有效的新因子，如今其难度已经有了数量级上的提升，再去挖掘出这样的新因子已经是非常困难的事情。而机器学习尚且是量化研究领域的一片蓝海，而它的其中一个优势正是能找出对象之间的非线性关系。

如果说机器学习将是量化研究未来绕不开的课题，那么如何尽早找到合适的运用方式就成了当前最有前景的研究方向之一。接下来的章节，我们会展示在这一方面已经实现的部分想法与尝试。

2.2、示例一：将数据处理成更合适的结构

金融数据本身，无论是基本面财务数据，还是价量数据，都有良好的结构性。这也是为什么一些在金融领域应用机器学习的支持者看好其前景的原因之一。但正如《The 7 Reasons Most Machine Learning Funds Fail》文中提及的，这样的结构化方式以及一些处理的方式，是不是真的就适合直接运用在机器学习算法上呢？

拿最常见的量价数据来说，我们大家已经习以为常的 K 线数据，它的结构化方式（或者说取样方式）都是按时间等分切片取样。比如我们说日线，就是按每个交易日切分数据，每个 K 线里都是一个完整交易日（4 小时）的交易信息；或者我们说 5 分钟线，就是按每个 5 分钟切分数据，每个 K 线里是 5 分钟的交易信息。

时间等分切片的好处是显而易见的，人们在处理任何事情时都习惯于以时间为刻度，包括处理交易事务；同时当我们需要以图标的方式直观地观察一个价格序列的走势时，横坐标按时间等分也更方便研究者确定任何一个价格发生的时间点。

但当我们利用机器学习工具去做研究时，时间等分切片却未必是一个比较好的结构化方式。因为不少机器学习算法要求输入数据尽量满足独立同分布，或者满足同方差性等条件。而时间等分采样切片明显不太符合这样的性质，很大可能每个样本点所蕴含的信息多少是不均的，有些时刻大家更有交易意愿，波动放大；有些时候却交易情绪低迷，成交量萎靡，波动很小。总结而言，时间等分切片会面临样本点信息不均，序列自相关性，样本非同方差，收益率非正态分布等问题，因此为了更好地利用机器学习这套工具，更适合的做法是将价量数据以其它更好的采样切片方式来制作 K 线。

目前更有效的采样切片方式包括：tick 等分 K 线，成交量等分 K 线，成交额等分 K 线，信息量等分 K 线。本文我们仅做一些简单的实证，通过一些统计数据，展示时间等分、Tick 等分、成交量等分以及成交额等分的 K 线方式哪些更适合于大部分机器学习算法。信息量等分 K 线以及更详细全面的测试与分析会在我们后续的系列报告中有所展示。

表 3：多种 K 线切片采样方式

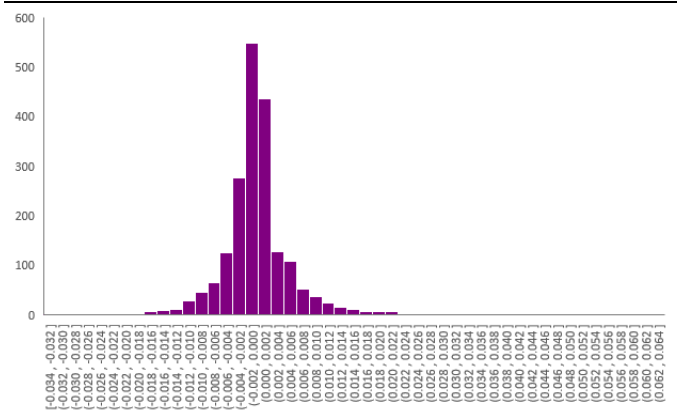
| K 线切片方式 | 切片方式解释与目的 |
|-----------|----------------|
| 时间等分切片 | 每根 K 线有相同的时间跨度 |
| Tick 等分切片 | 每根 K 线有相同的交易笔数 |
| 成交量等分切片 | 每根 K 线有相同的成交量 |
| 成交额等分切片 | 每根 K 线有相同的成交额 |
| 信息量等分切片 | 每根 K 线有相同的信息量 |

资料来源：光大证券研究所

我们以平安银行（000001.SZ）股票为例，利用其 2017 年一整年的价量数据分别构建时间等分 K 线、Tick 等分 K 线、成交量等分 K 线以及成交额等分 K 线。构建的过程中为了使其具有可比性，每种 K 线在一年内的 K 线总数在同一数量级下。

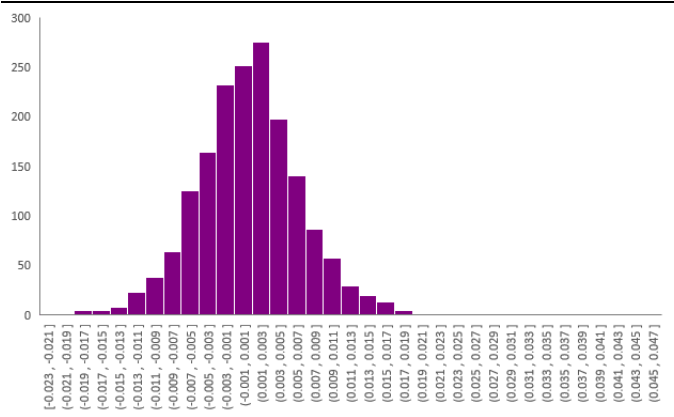
通过比较这四种不同 K 线切片方式下收益率数据的均值、方差、偏度与峰度，明显可以看出时间等分 K 线下的收益率非常明显的呈现左偏、尖峰、厚尾的统计特征，正态性很弱。而 Tick 等分 K 线、成交量等分 K 线以及成交额等分 K 线下的收益率则有较强的正态性，偏度接近 0，峰度接近 3。

图 6：时间等分 K 线收益率分布



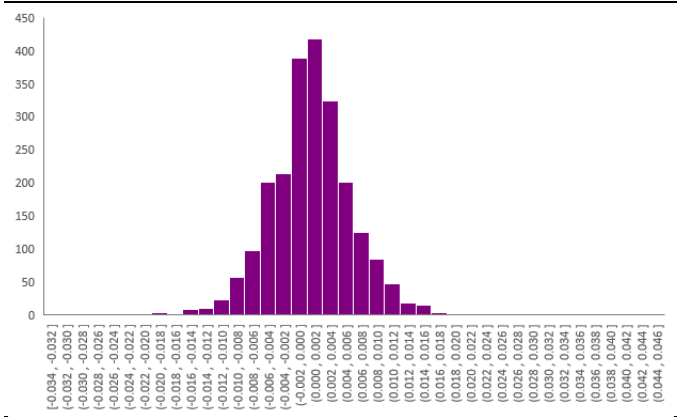
资料来源：光大证券研究所

图 7：Tick 等分 K 线收益率分布



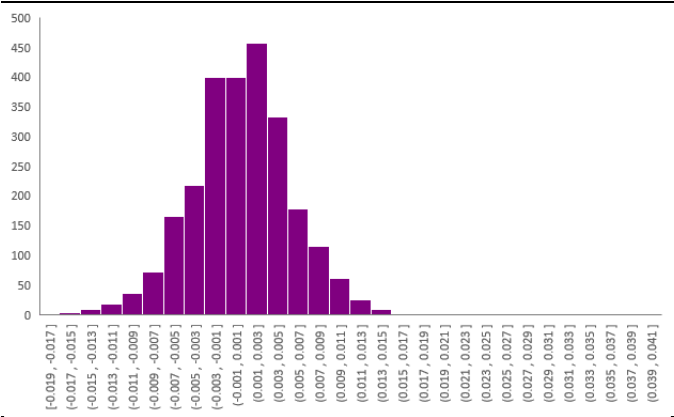
资料来源：光大证券研究所

图 8：成交量等分 K 线收益率分布



资料来源：光大证券研究所

图 9：成交额等分 K 线收益率分布



资料来源：光大证券研究所

表 4：不同 K 线切片方式下收益率数据的正态性

| K 线切片方式 | 均值 | 方差 | 偏度 | 峰度 |
|-----------|--------|--------|-------|--------|
| 时间等分切片 | 0.021% | 0.560% | 1.640 | 16.367 |
| Tick 等分切片 | 0.023% | 0.578% | 0.307 | 2.439 |
| 成交量等分切片 | 0.018% | 0.522% | 0.162 | 3.549 |
| 成交额等分切片 | 0.016% | 0.480% | 0.159 | 2.142 |

资料来源：光大证券研究所

考虑到不同交易者进行一笔交易的成交量可能大不相同，以及一些高送转股会有拆股行为，或者一些公司的股票增发与回购行为，成交额等分 K 线相比 Tick 等分 K 线与成交量等分 K 线往往更为稳定。

我们也构造了一个极为简单的模型，来对不同 K 线切片方式下的算法表现作一定展示。模型的思路是基于 SVM 分类算法，利用历史最近固定数量 K 线的收益率序列来尝试预测下一根 K 线的走势（涨、平、跌）。这里我们分别拿时间等分切片 K 线与成交量等分切片 K 线作为数据进行比较。

选取了目前上证 50 中交易历史足够久的 40 只股票作预测，与我们预期的一样，像这样简单粗暴的算法套用并没有实质上的预测效果。但通过对比

不同 K 线上的结果，我们依然可以发现不同 K 线下的效果有所差异，成交量等分 K 线数据下的模型效果整体表现更为稳定。在测试集上，40 只股票中有 29 只在成交量切片 K 线数据上的预测准确性更高。无论在训练集与测试集，成交量切片 K 线数据上的模型准确度均值都高出时间切片 K 线约 1 个百分点；同时观察不同股票模型准确率标准差，也能发现算法在成交量切片数据上的表现更为稳定。

表 5：模型在各个股票上预测准确率统计

| | 时间等分切片 K 线 | | 成交量等分切片 K 线 | |
|--------|------------|--------|-------------|--------|
| | 训练集 | 测试集 | 训练集 | 测试集 |
| 准确率均值 | 45.71% | 45.44% | 47.29% | 46.25% |
| 准确率标准差 | 4.15% | 4.30% | 3.21% | 4.08% |

资料来源：光大证券研究所，注：数据样本区间为 2010/1/1 至 2017/12/31

表 6：不同 K 线切片数据下模型预测准确率

| 股票代码 | 训练集准确率 | | 测试集准确率 | | 股票代码 | 训练集准确率 | | 测试集准确率 | |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | 时间切片 | 成交量切片 | 时间切片 | 成交量切片 | | 时间切片 | 成交量切片 | 时间切片 | 成交量切片 |
| 600000 | 47.27% | 47.75% | 47.66% | 48.32% | 600887 | 48.95% | 48.19% | 45.89% | 46.45% |
| 600016 | 43.70% | 46.89% | 41.47% | 46.14% | 601006 | 43.98% | 45.97% | 43.25% | 44.42% |
| 600019 | 41.13% | 45.46% | 45.28% | 44.26% | 601088 | 49.04% | 50.21% | 46.85% | 48.17% |
| 600028 | 42.00% | 43.90% | 39.80% | 41.07% | 601166 | 48.07% | 48.30% | 47.66% | 47.77% |
| 600029 | 43.77% | 46.53% | 45.03% | 46.40% | 601169 | 45.23% | 47.71% | 41.47% | 44.62% |
| 600030 | 48.95% | 49.41% | 49.24% | 50.15% | 601186 | 44.23% | 47.54% | 49.44% | 50.25% |
| 600036 | 47.40% | 48.76% | 48.88% | 49.54% | 601288 | 38.16% | 36.91% | 36.35% | 34.26% |
| 600048 | 47.82% | 49.21% | 47.26% | 46.45% | 601318 | 50.70% | 51.01% | 48.83% | 48.88% |
| 600050 | 41.07% | 44.92% | 45.79% | 49.34% | 601328 | 40.28% | 44.02% | 37.51% | 40.96% |
| 600104 | 48.88% | 50.50% | 47.82% | 48.48% | 601390 | 40.45% | 45.88% | 45.74% | 48.22% |
| 600196 | 48.18% | 49.41% | 48.48% | 46.29% | 601398 | 39.78% | 41.23% | 39.95% | 40.36% |
| 600276 | 49.24% | 49.25% | 48.98% | 48.07% | 601601 | 49.37% | 50.66% | 48.48% | 47.97% |
| 600309 | 49.25% | 50.61% | 49.59% | 49.70% | 601628 | 49.37% | 48.87% | 49.29% | 47.82% |
| 600340 | 48.25% | 48.80% | 49.80% | 51.88% | 601688 | 49.20% | 50.57% | 51.17% | 48.53% |
| 600519 | 50.74% | 52.76% | 47.46% | 51.47% | 601818 | 37.30% | 42.58% | 34.97% | 38.53% |
| 600547 | 49.65% | 49.51% | 48.48% | 51.07% | 601857 | 42.97% | 47.11% | 42.28% | 44.42% |
| 600585 | 49.47% | 49.03% | 47.06% | 47.72% | 601888 | 50.53% | 49.18% | 49.95% | 47.51% |
| 600606 | 46.22% | 45.76% | 45.43% | 42.44% | 601939 | 37.77% | 42.97% | 39.14% | 42.34% |
| 600690 | 48.16% | 47.99% | 46.09% | 47.06% | 601988 | 37.73% | 40.70% | 36.14% | 35.58% |
| 600703 | 48.18% | 48.41% | 47.92% | 48.12% | 601989 | 45.97% | 47.19% | 45.89% | 48.88% |

资料来源：光大证券研究所，注：数据样本区间为 2010/1/1 至 2017/12/31

2.3、示例二：利用特征重要性打开“黑箱”

机器学习不为大部分投资者接受的一个广为人知的原因在于其不可解释性。尤其是如今机器学习算法的“王者”们——深度神经网络算法及其延伸算法。人们在运用这些算法得到一个看上去挺美好的结果后，往往很难解

释到底是什么特征，在以什么样的方式导致了这个策略的有效性。这就使得在使用这些机器学习策略时很难评估信心程度。

会出现这样的现象的原因在于人们仅仅把机器学习算法当成预测工具。实际上更多时候，机器学习更适合当成一个研究工具而非预测工具。机器学习算法给我们提供了一种寻找那些从经典逻辑或计量经济学逻辑出发难以理解的非线性预测信息。通过机器学习我们有了一个对这样非线性解释性现象的研究入口；至于当我们理解了这些非线性预测能力的逻辑与思路，如何利用它们构建我们的投资策略，那就是很多业界投资者们最擅长的事情了。

打开机器学习“黑箱”的一个有效方式就是研究每一个预测模型中的特征重要性（Feature Importance）。不同的机器学习算法的“黑箱”程度不同，比如树模型(CART)则是一种相对“白箱”的算法，而支持向量机(SVM)，神经网络(NN)则非常“黑箱”。这就使得探索不同算法的特征重要性的难度不一。在本篇报告中，我们仅拿相对简单的树模型做示例，相对全面的研究方法我们将在后续的系列报告中剖析。

我们在 2018 年上半年发表了一篇利用机器学习算法来进行因子择时的深度报告《因子正交与择时：基于分类模型的动态权重配置》中，报告中我们测试分别用不同的分类算法来对因子未来一个月的收益方向进行预测，从样本内及样本外跟踪的效果上看都很优秀。预测模型里有十几项涉及到不同类型的输入变量，那么到底是哪些变量在真正提供预测能力，我们尝试做一些初步的分析。

因子择时模型需要预测的是表 4 中的 14 个常用因子未来一个月的收益方向，而择时变量则选取了表 5 中所示的宏观经济环境，货币政策，市场状态变量，以及因子自身收益与波动情况数据。（具体的因子择时策略构建与最终表现参考报告《因子正交与择时：基于分类模型的动态权重配置》）

表 7：测试因子明细表

| 因子名称 | 因子描述 |
|--------------|-----------------|
| BP_LR | 账面市值比(最近报告期) |
| EP_TTM | 盈利市值比(TTM) |
| Ln_MC | 市值对数 |
| Momentum_1M | 动量(1 个月) |
| Momentum_24M | 动量(24 个月) |
| STD_1M | 波动(1 个月) |
| TURNOVER_1M | 换手率(1 个月) |
| VSTD_3M | 流动性(3 个月) |
| DP_TTM | 股息率(TTM) |
| EEP | 一致预期 EP |
| EEChange_3M | 一致预期盈利增长率(3 个月) |
| TAG_TTM | 总资产增长率(TTM) |
| ROE_TTM | 净资产回报率(TTM) |
| Debt_Asset | 资产负债比 |

资料来源：光大证券研究所

表 8：择时特征变量名单

| | 指标代码 | 指标名称 | 备注 |
|----------|------------|----------------|------------------------------|
| 货币政策变量 | Tbill_3M | 3 个月国债收益率 | - |
| | M1 | M1 货币供应量同比增长率 | - |
| 经济环境变量 | CPI | CPI 同比增长率 | - |
| | PPI | PPI 同比增长率 | - |
| | IND | 规模以上工业增加值同比增长率 | - |
| 市场状态变量 | TS | 期限利差 | 10 年国债到期收益率 - 1 年国债到期收益率 |
| | CS | 信用利差 | 1 年中债中短期票据到期收益率 - 1 年国债到期收益率 |
| | 300_RET | 沪深 300 收益率 | 月度 |
| | 1000_RET | 中证 1000 收益率 | 月度 |
| | 300_STD | 沪深 300 波动率 | 月度 |
| | 1000_STD | 中证 1000 波动率 | 月度 |
| | RET_Spread | 大小盘收益差值 | 300_RET - 1000_RET |
| | STD_Spread | 大小盘波动差值 | 300_STD - 1000_STD |
| 因子收益衍生变量 | Ret_Factor | 因子收益 | 6 个月加权移动平均 |
| | Std_Factor | 因子波动 | 6 个月加权移动平均 |

资料来源：光大证券研究所，Wind

这里我们研究使用的随机森林 (Random Forest) 算法进行因子择时时不同特征的特征重要性，像树模型这样本身分类方式较为直观的，相对“白箱”的算法，其特征重要性也较容易计算。这里我们使用“平均不纯度减少”方法 (MDI) 来测算不同输入变量的特征重要性，MDI 是通过计算每个变量对分类树每个节点上观测值的异质性的影响，从而比较变量的重要性。MDI 值越大表示该变量的重要性越大，其蕴含的预测能力越强。

可以看到在预测不同的因子时，真正提供有效信息的输入变量并不一样，有些变量实际上并没有真正起到预测作用，反而可能更多地制造了干扰预测效果的噪音。例如 PPI 数据对预测市值因子 (Ln_MC) 卓有贡献，但几乎对动量因子 (Momentum) 没有预测作用。

当然，MDI 仅仅只是让我们对哪些变量有解释作用有个初步的了解，其算法也天然有些不足之处，例如替代效应等，这就需要配合其它的计算一起考虑。同时这些变量以什么样的逻辑链解释预测能力，以及之后如何利用这个结果来改进策略也都有待更深的研究。

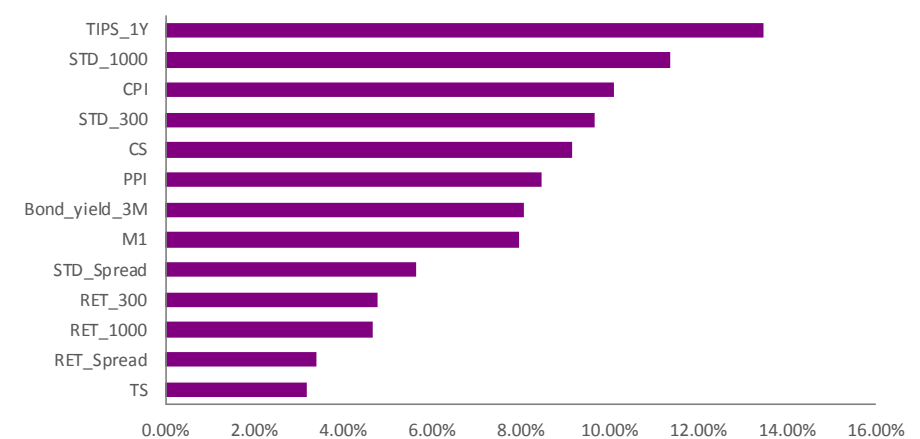
表 9：随机森林算法在不同因子上应用的特征重要性示例

| | EBQC | BP_LR | Momentum_1M | STD_3M | Ln_MC | TURNOVER_1M |
|-----|--------|-------|-------------|--------|--------|-------------|
| TS | 3.20% | 2.68% | 4.62% | 4.15% | 7.60% | 3.63% |
| CS | 9.16% | 2.84% | 3.91% | 8.41% | 4.49% | 3.22% |
| CPI | 10.12% | 8.85% | 5.09% | 5.87% | 10.89% | 11.79% |

| | | | | | | |
|----------------------|--------|--------|--------|--------|--------|--------|
| PPI | 8.46% | 8.78% | 3.98% | 4.83% | 12.58% | 8.24% |
| STD_Spread | 5.63% | 5.92% | 16.64% | 13.70% | 3.16% | 19.43% |
| RET_Spread | 3.40% | 7.18% | 10.84% | 4.47% | 8.65% | 8.73% |
| M1 | 7.98% | 2.40% | 15.69% | 15.18% | 2.86% | 8.34% |
| Bond_yield_3M | 8.07% | 16.58% | 6.53% | 2.82% | 6.39% | 6.09% |
| TIPS_1Y | 13.49% | 16.03% | 5.24% | 5.89% | 6.63% | 3.44% |
| STD_1000 | 11.39% | 6.71% | 11.30% | 5.38% | 11.65% | 4.10% |
| STD_300 | 9.66% | 6.76% | 4.03% | 5.88% | 10.73% | 2.14% |
| RET_1000 | 4.66% | 7.54% | 9.97% | 16.27% | 9.96% | 9.33% |
| RET_300 | 4.79% | 7.72% | 2.16% | 7.16% | 4.40% | 11.50% |

资料来源：光大证券研究所，注：测试区间为 2008 - 2018

图 10：在随机森林算法下不同输入特征在预测 EBQC（光大综合质量）因子收益方向上的特征重要性



资料来源：光大证券研究所，注：测试区间为 2008 - 2018

2.4、示例三：利用生成数据了解过拟合程度

另一个机器学习被人诟病的方面就是过拟合。机器学习在找寻数据关联性的强大能力使其极其容易出现过拟合的现象，这就使得研究者能较容易在样本内得到非常好看的策略表现与净值曲线，但一到样本外策略的表现就大大不如样本内。

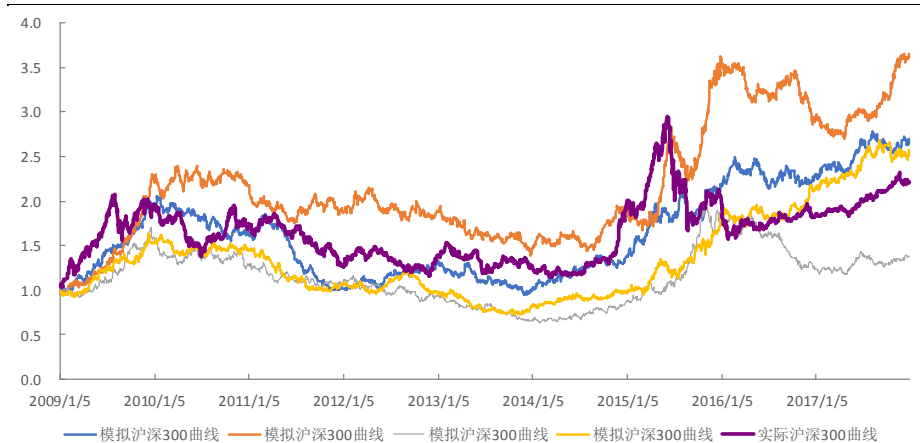
许多不当的回测方式都会造成过拟合的发生，大家比较熟知的比如幸存者偏差 (Survivorship Bias)，未来信息 (Look-ahead Bias)，过分数据挖掘 (Data Mining)，路径依赖 (Path Dependence) 等。甚至在正确地进行回测时，也不可避免地会有过拟合的风险，重复在同一块数据上回测会增大研究者发现“看似有效策略”的概率。当然，这些“看似有效”的策略实际上完全不可信，依靠它们交易往往会以损失惨重收尾。

因此处理过拟合问题是机器学习绕不开的问题，也是大部分策略开发需要考虑的一点。如何应对评估策略的过拟合程度以及如何合理的采取尽可能减少过拟合概率的回测方法，也将在我们后续的一系列报告中探讨。我们在本文中通过一个示例展示一些基本的评估过拟合程度的方式。

A 股中很多研究者都在研究择时策略，希望能一定程度上预测市场未来的涨跌。有些利用技术指标，有些统计模型。我们在 2017 年初构建的 RSRS 择时模型就是运用了最简单的统计模型——线性回归，来构建了一个指数择时模型。在样本内以及近 2 年的样本外跟踪下效果都很优秀，受到业界的肯定。但实际上这个策略是真的有效，还是只是这两年运气好使得策略“看上去有效”呢？包括大部分其它研报上的择时策略，在回测时都是基于 A 股历史走势来测试效果的，包括一些参数最优化的选取也是基于同样一条历史走势。这就使得策略很容易产生路径依赖的过拟合现象。为了评估策略参数是否过分优化以及什么样的参数真正合适的取值是多少，一个简单的方式是将策略在生成数据（Synthetic Data）而不是历史数据（Historical Data）上回测。这里生成数据是指基于股票或指数历史量价数据的统计特征而生成的价格走势数据。这样我们就不再只在唯一的一条历史走势曲线的回测，而是得以在很多模拟的走势曲线上进行测试。采用 Monte Carlo Simulation，我们可以初步判断 RSRS 现有的参数是否严重过拟合，以及最适合的参数可能取值多少。

由于 RSRS 策略需要的数据是每天的最高价与最低价，因此仅仅拟合生成日收益数据是不够的。我们基于实际的 2009 年 1 月 1 日至 2017 年 12 月 31 日内的沪深 300 指数 1 分钟价格数据来制造同时间段内的生成数据。这样我们可以基于生成的 1 分钟数据序列得到生成数据中每日的开盘价、最高价、最低价与收盘价。我们共生成了 1000 条模拟沪深 300 指数，在图 11 中我们随机抽取了 4 条生成曲线作为展示。

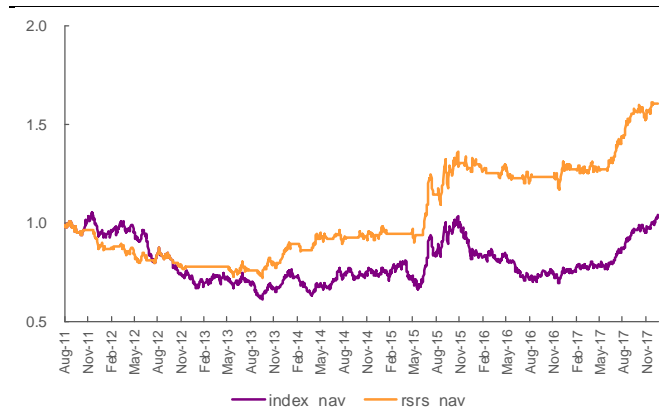
图 11：沪深 300 生成数据曲线示例



资料来源：光大证券研究所

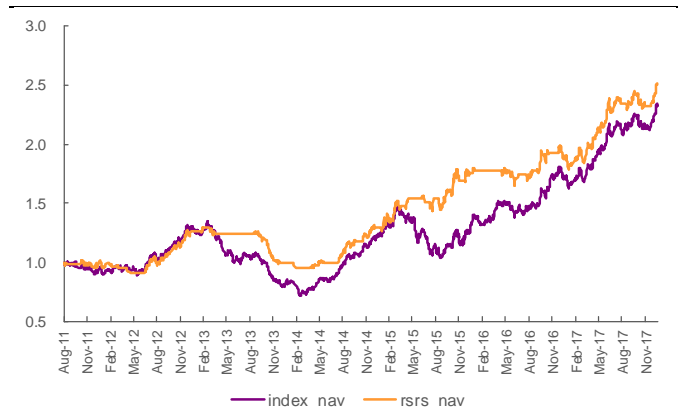
以原有 RSRS 沪深 300 择时策略的参数，我们在每个生成曲线上测试了 RSRS 择时策略的效果，整体来看，RSRS 择时策略都较为有效，尤其在规避大幅下跌以及长期震荡下跌的效果上较好。但整体回撤与效果明显不如在实际沪深 300 指数上那么优秀，同时我们也发现如果生成指数长期处于类似于美股指数那样稳定上升的趋势里，策略表现往往不如指数本身。在图 13 中展示的生成曲线上可以明显看出，在后半段的上升趋势里，RSRS 策略有间歇性的踏空现象。

图 12: 生成曲线 (1) 上测试原有参数下的 RSRs 策略



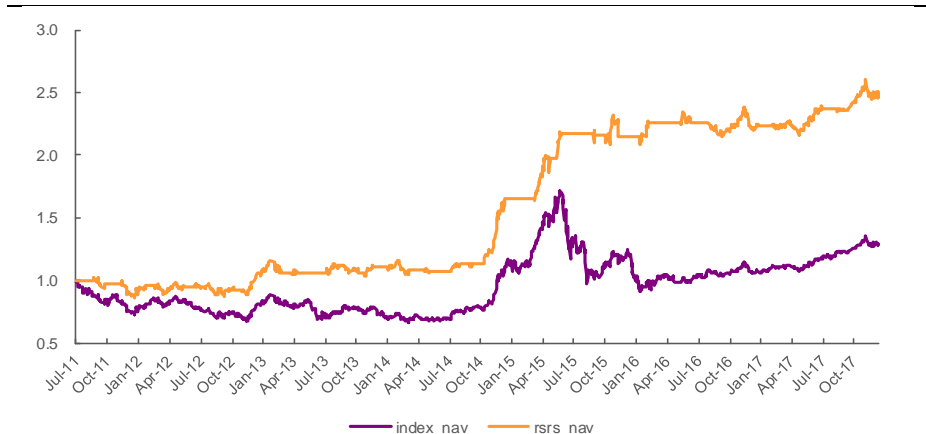
资料来源：光大证券研究所

图 13: 生成曲线 (2) 上测试原有参数下的 RSRs 策略



资料来源：光大证券研究所

图 14: RSRs 策略在实际沪深 300 曲线上的历史回测效果



资料来源：光大证券研究所

在上述研究中我们可以发现 RSRs 策略在生成数据上也有效果，那么参数是否有明显的路径依赖现象呢，最优参数到底取多少更合适？RSRs 策略共有 3 个参数：回归期数 N ，标准化窗宽 M ，开平仓阈值 S 。这里我们仅就最后的参数 S 展开测试与讨论， N 与 M 的取值就取实际中运用的参数 18 与 600。我们测试每个生成指数上 RSRs 策略在参数 S 取值 0.5、0.6、...、1.5 这 11 个不同情况下的表现。在图 12 与图 13 的生成指数上，参数 S 默认取值 0.7 虽然表现不错但并不是最优参数，图 12 的生成指数上最优参数是 0.9，而图 13 的生成指数最优参数是 0.6。

表 10: 固定参数 N ， M 取值原始值，不同参数 S 取值下 RSRs 策略在图 12 与图 13 的生成数据上的最终净值

| 阈值参数 S | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1 | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 |
|----------|------|------|------|------|------|------|------|------|------|------|------|
| 生成指数 (1) | 1.48 | 1.76 | 1.61 | 1.83 | 1.94 | 1.80 | 1.37 | 1.36 | 1.36 | 1.00 | 0.83 |
| 生成指数 (2) | 2.31 | 2.56 | 2.50 | 2.32 | 2.26 | 2.36 | 2.07 | 2.21 | 2.11 | 2.08 | 1.91 |

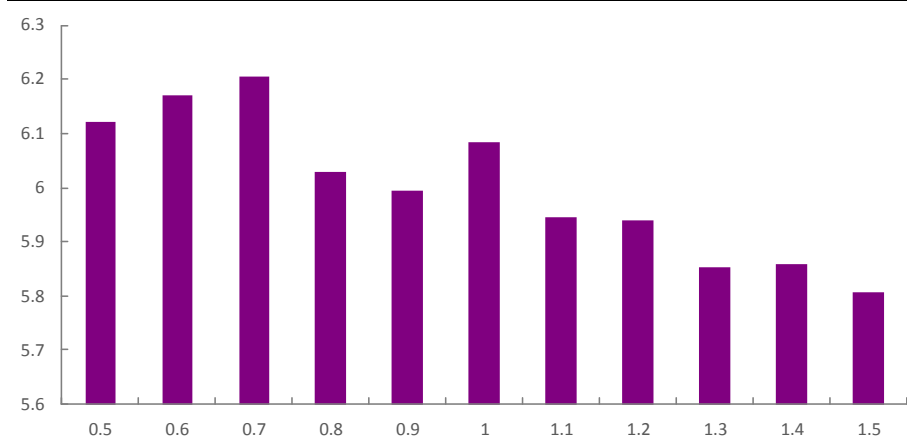
资料来源：光大证券研究所，注： $S=0.7$ 为原始策略参数取值

为了更直观的体现哪个参数大概率是最稳定优秀的参数，我们构建了参数平均排名指标。其目的是希望能将不同生成指数上的参数表现一起考虑进来。参数平均排名的计算方式如下：

1. 计算每个生成指数上策略净值在不同参数下的排名,最小的排名1。
2. 计算每个参数在所有生成指数上的净值排名的均值。

在参数 S 的平均排名指标计算中，排名可取值数值范围为 1 到 11 的正整数,从下图可以看出,参数 S 的参数平均排名指标随着参数从 0.5 升至 1.5,有一个较明显的先升后降的下凹抛物线的形状，顶点在 $S=0.7$ 处达到。可见在参数 N 与 M 取值默认参数下，阈值 S 取值 0.7 的确是更为稳定合理的最优参数。

图 15：不同参数 S 的平均排名



资料来源：光大证券研究所

在上述研究中，所有的结论的可靠性实际也是基于生成数据的可靠性基础上的。模拟出来的生成数据到底多大程度上能表征沪深 300 指数实际上真正的特性，就取决于模拟数据的生成方式。能够更好更精准地制作有效生成数据将会大大提高我们对策略参数过拟程度的认知判断，我们将在后续的研究系列报告里更深入地探讨研究这一课题。

3、风险提示

本报告中的测试结果均基于模型和历史数据，历史数据存在不被重复验证的可能，模型存在失效的风险。

行业及公司评级体系

| 评级 | 说明 |
|-----|--|
| 买入 | 未来 6-12 个月的投资收益率领先市场基准指数 15% 以上； |
| 增持 | 未来 6-12 个月的投资收益率领先市场基准指数 5% 至 15%； |
| 中性 | 未来 6-12 个月的投资收益率与市场基准指数的变动幅度相差 -5% 至 5%； |
| 减持 | 未来 6-12 个月的投资收益率落后市场基准指数 5% 至 15%； |
| 卖出 | 未来 6-12 个月的投资收益率落后市场基准指数 15% 以上； |
| 无评级 | 因无法获取必要的资料，或者公司面临无法预见结果的重大不确定性事件，或者其他原因，致使无法给出明确的投资评级。 |

基准指数说明：A 股主板基准为沪深 300 指数；中小盘基准为中小板指；创业板基准为创业板指；新三板基准为新三板指数；港股基准指数为恒生指数。

分析、估值方法的局限性说明

本报告所包含的分析基于各种假设，不同假设可能导致分析结果出现重大不同。本报告采用的各种估值方法及模型均有其局限性，估值结果不保证所涉及证券能够在该价格交易。

分析师声明

本报告署名分析师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、专业审慎的研究方法，使用合法合规的信息，独立、客观地出具本报告，并对本报告的内容和观点负责。负责准备本报告以及撰写本报告的所有研究分析师或工作人员在此保证，本研究报告中关于任何发行商或证券所发表的观点均如实反映分析人员的个人观点。负责准备本报告的分析师获取报酬的评判因素包括研究的质量和准确性、客户的反馈、竞争性因素以及光大证券股份有限公司的整体收益。所有研究分析师或工作人员保证他们报酬的任何一部分不曾与、不与、也将不会与本报告中的具体的推荐意见或观点有直接或间接的联系。

特别声明

光大证券股份有限公司（以下简称“本公司”）创建于 1996 年，系由中国光大（集团）总公司投资控股的全国性综合类股份制证券公司，是中国证监会批准的首批三家创新试点公司之一。根据中国证监会核发的经营证券期货业务许可，光大证券股份有限公司的经营范围包括证券投资咨询业务。

本公司经营范围：证券经纪；证券投资咨询；与证券交易、证券投资活动有关的财务顾问；证券承销与保荐；证券自营；为期货公司提供中间介绍业务；证券投资基金代销；融资融券业务；中国证监会批准的其他业务。此外，公司还通过全资或控股子公司开展资产管理、直接投资、期货、基金管理以及香港证券业务。

本证券研究报告由光大证券股份有限公司研究所（以下简称“光大证券研究所”）编写，以合法获得的我们相信为可靠、准确、完整的信息为基础，但不保证我们所获得的原始信息以及报告所载信息之准确性和完整性。光大证券研究所可能将不时补充、修订或更新有关信息，但不保证及时发布该等更新。

本报告中的资料、意见、预测均反映报告初次发布时光大证券研究所的判断，可能需随时进行调整且不予通知。报告中的信息或所表达的意见不构成任何投资、法律、会计或税务方面的最终操作建议，本公司不就任何人依据报告中的内容而最终操作建议做出任何形式的保证和承诺。在任何情况下，本报告中的信息或所表达的意见并不构成对任何人的投资建议。客户应自主作出投资决策并自行承担投资风险。本报告中的信息或所表达的意见并未考虑到个别投资者的具体投资目的、财务状况以及特定需求。投资者应当充分考虑自身特定状况，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。对依据或者使用本报告所造成的一切后果，本公司及作者均不承担任何法律责任。

不同时期，本公司可能会撰写并发布与本报告所载信息、建议及预测不一致的报告。本公司的销售人员、交易人员和其他专业人员可能会向客户提供与本报告中所载观点不同的口头或书面评论或交易策略。本公司的资产管理部、自营部门以及其他投资业务部门可能会独立做出与本报告的意见或建议不相一致的投资决策。本公司提醒投资者注意并理解投资证券及投资产品存在的风险，在做出投资决策前，建议投资者务必向专业人士咨询并谨慎抉择。

在法律允许的情况下，本公司及其附属机构可能持有报告中提及的公司所发行证券的头寸并进行交易，也可能为这些公司提供或正在争取提供投资银行、财务顾问或金融产品等相关服务。投资者应当充分考虑本公司及本公司附属机构就报告内容可能存在的利益冲突，勿将本报告作为投资决策的唯一信赖依据。

本报告根据中华人民共和国法律在中华人民共和国境内分发，仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为客户。本报告仅向特定客户传送，未经本公司书面授权，本研究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。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。如欲引用或转载本文内容，务必联络本公司并获得许可，并需注明出处为光大证券研究所，且不得对本文进行有悖原意的引用和删改。

光大证券股份有限公司

上海市新闻路 1508 号静安国际广场 3 楼 邮编 200040

总机：021-22169999 传真：021-22169114、22169134

| 机构业务总部 | 姓名 | 办公电话 | 手机 | 电子邮件 |
|--------|-----|---------------|-------------------------|-------------------------|
| 上海 | 徐硕 | 021-52523543 | 13817283600 | shuoxu@ebsecn.com |
| | 李文渊 | | 18217788607 | liwenyuan@ebsecn.com |
| | 李强 | 021-52523547 | 18621590998 | liqiang88@ebsecn.com |
| | 罗德锦 | 021-52523578 | 13661875949/13609618940 | luodj@ebsecn.com |
| | 张弓 | 021-52523558 | 13918550549 | zhanggong@ebsecn.com |
| | 黄素青 | 021-22169130 | 13162521110 | huangsuqing@ebsecn.com |
| | 邢可 | 021-22167108 | 15618296961 | xingk@ebsecn.com |
| | 李晓琳 | 021-52523559 | 13918461216 | lixiaolin@ebsecn.com |
| | 郎珈艺 | 021-52523557 | 18801762801 | dingdian@ebsecn.com |
| | 余鹏 | 021-52523565 | 17702167366 | yupeng88@ebsecn.com |
| 北京 | 丁点 | 021-52523577 | 18221129383 | dingdian@ebsecn.com |
| | 郭永佳 | | 13190020865 | guoyongjia@ebsecn.com |
| | 郝辉 | 010-58452028 | 13511017986 | haohui@ebsecn.com |
| | 梁晨 | 010-58452025 | 13901184256 | liangchen@ebsecn.com |
| | 吕凌 | 010-58452035 | 15811398181 | lvling@ebsecn.com |
| | 郭晓远 | 010-58452029 | 15120072716 | guoxiaoyuan@ebsecn.com |
| | 张彦斌 | 010-58452026 | 15135130865 | zhangyanbin@ebsecn.com |
| | 庞舒然 | 010-58452040 | 18810659385 | pangsr@ebsecn.com |
| | 黎晓宇 | 0755-83553559 | 13823771340 | lix1@ebsecn.com |
| | 张亦潇 | 0755-23996409 | 13725559855 | zhangyx@ebsecn.com |
| 深圳 | 王渊锋 | 0755-83551458 | 18576778603 | wangyuanfeng@ebsecn.com |
| | 张靖雯 | 0755-83553249 | 18589058561 | zhangjingwen@ebsecn.com |
| | 苏一耘 | | 13828709460 | su1@ebsecn.com |
| | 常密密 | | 15626455220 | changmm@ebsecn.com |
| | 陶奕 | 021-52523546 | 18018609199 | taoyi@ebsecn.com |
| | 梁超 | 021-52523562 | 15158266108 | liangc@ebsecn.com |
| | 金英光 | | 13311088991 | jinyg@ebsecn.com |
| | 王佳 | 021-22169095 | 13761696184 | wangjia1@ebsecn.com |
| | 郑锐 | 021-22169080 | 18616663030 | zhui@ebsecn.com |
| | 凌贺鹏 | 021-22169093 | 13003155285 | linghp@ebsecn.com |
| 国际业务 | 周梦颖 | 021-52523550 | 15618752262 | zhoumengying@ebsecn.com |
| | 戚德文 | 021-52523708 | 18101889111 | qidw@ebsecn.com |
| | 安玲娴 | 021-52523708 | 15821276905 | anlx@ebsecn.com |
| | 张浩东 | 021-52523709 | 18516161380 | zhanghd@ebsecn.com |
| | 吴冕 | 0755-23617467 | 18682306302 | wumian@ebsecn.com |
| | 吴琦 | 021-52523706 | 13761057445 | wuqi@ebsecn.com |
| | 王舒 | 021-22169419 | 15869111599 | wangshu@ebsecn.com |
| | 傅裕 | 021-52523702 | 13564655558 | fuyu@ebsecn.com |
| | 王婧 | 021-22169359 | 18217302895 | wangjing@ebsecn.com |
| | 陈潞 | 021-22169146 | 18701777950 | chenlu@ebsecn.com |
| 私募业务部 | 王涵洲 | | 18601076781 | wanghanzhou@ebsecn.com |