

分析师：

徐寅

xuyinsh@xyzq.com.cn

S0190514070004

西学东渐--海外文献推荐系列之一百一十八

2021年5月6日

报告关键点

横截面策略成功的关键在于资产排名预测是否准确。当前资产排名常用的算法是启发式方法或者回归后再分类算法，它们在信息检索领域并不是最优的方法。为了更加准确的排序，作者用 LTR 算法构建了一个新的框架。本文以横截面动量策略作为案例，作者证明了机器学习排序模型相比于传统模型可以显著提高横截面策略的表现。

相关报告

《西学东渐--海外文献推荐系列之一百一十七》

《西学东渐--海外文献推荐系列之一百一十六》

《西学东渐--海外文献推荐系列之一百一十五》

投资要点

- 西学东渐，是指从明朝末年到近代，西方学术思想向中国传播的历史过程。西学东渐不仅推动了中国在科学技术和思想文化方面的发展，也有力地促进了社会与政治的大变革。在今天，西学东渐仍有其重要的现实意义。作为 A 股市场上以量化投资为研究方向的卖方金融工程团队，在平日的工作中，常常深感海外相关领域的研究水平之高、内容之新。而这也促使我们通过大量的材料阅读，去粗取精，将认为最有价值的海外文献呈现在您的面前！
- 横截面策略成功的关键在于资产排名预测是否准确。当前资产排名常用的算法是启发式方法或者回归后再分类算法，它们在信息检索领域并不是最优的方法。为了更加准确的排序，作者用 LTR 算法构建了一个新的框架。本文以横截面动量策略作为案例，作者证明了机器学习排序模型相比于传统模型可以显著提高横截面策略的表现。

风险提示：文献中的结果均由相应作者通过历史数据统计、建模和测算完成，在政策、市场环境发生变化时模型存在失效的风险。

请务必阅读正文之后的信息披露和重要声明

目录

1、引言	- 3 -
2、文献综述：	- 4 -
2.1、横截面动量策略	- 4 -
2.2、LTR 算法在金融领域的运用	- 4 -
3、问题定义	- 5 -
4、得分计算方法	- 6 -
4.1、经典的横截面动量	- 6 -
4.2、回归后再排序方法	- 6 -
4.3、LTR 算法	- 7 -
4.4、训练细节	- 8 -
5、表现分析	- 8 -
5.1、数据情况	- 8 -
5.2、回测和预测描述	- 8 -
5.3、模型和对比维度	- 9 -
5.4、结果与结论	- 10 -
6、结论	- 12 -
 图表 1、按目标波动率调整后的累计收益	- 9 -
图表 2、按目标波动率调整后的策略表现指标	- 10 -
图表 3、每个月的策略平均排序表现	- 11 -
图表 4、目标波动率调整后的分位数组合表现	- 11 -
图表 5、目标波动率调整后的分位数组合累计收益	- 12 -

报告正文

通过 LTR 算法构建横截面策略

文献来源：

Poh D, Lim B, Zohren S, et al. Building Cross-Sectional Systematic Strategies By Learning to Rank[J]. The Journal of Financial Data Science, 2021.

推荐理由：

横截面策略成功的关键在于资产排名预测是否准确。当前资产排名常用的算法是启发式方法或者回归后再分类算法，它们在信息检索领域并不是最优的方法。为了更加准确的排序，作者用 LTR 算法构建了一个新的框架。本文以横截面动量策略作为案例，作者证明了机器学习排序模型相比于传统模型可以显著提高横截面策略表现。

我们的思考：

横截面策略是 A 股量化选股实践中最常用的策略，本文介绍的 LTR 算法是机器学习领域较为成熟的排序算法，它能够更加全面的对资产排序结构进行学习，相比于传统方法会有更好的排序效果，进而产生更好的策略收益。A 股的投资者以纯多头的组合为主，而 LTR 算法更注重对头部组合的排序效果，对 A 股投资者更具有参考意义。

1、引言

横截面策略是一种常用的交易策略，学术文献中有很多研究(Baz et al. 2015)。与独立考虑每种资产的时间序列方法不同(Moskowitz, Ooi, and Pedersen, 2012)，横截面策略通过做多预期收益最高的资产、做空预期收益最低的资产来获取收益。比如经典的横截面动量策略(Jegadeesh and Titman, 1993)会根据过去一年的收益进行排序，假设今年也将延续同样的排名情况，做多头部资产，做空尾部资产。即便在资产具有显著相关性的前提下，横截面策略仍然可以规避市场的整体波动(Baz et al. 2015; Jusselin et al. 2017; Roncalli 2017)。

近些年来随着机器学习的兴起，很多学者们将结合横截面策略与预测模型结合起来(Kim 2019; Gu et al. 2018, 2019)，常常能够战胜传统基准。通常情况下，这些机器学习模型通过有监督的方法进行训练，目的是预期收益的均方误差最小。尽管回归模型能够精确的提供未来资产收益率的平均估计，它并不能解决横截面策略的核心问题：未来收益的排名。这样的回归模型会对策略绩效与投资决策产生负面影响。

在信息检索领域，有很多模型可以改进标准的回归模型(Pasumarthi et al. 2019)。统称为 learning to rank (LTR) 算法，现在有一些细分算法利用了深度学习、或树模型(Wang and Klabjan 2017; Li et al. 2019; Wu et al. 2010)可以显著的提高预测效果。尽管 LTR 算法在金融领域也有应用，但大多数文献都将其用于简单的股票推荐，而缺乏一般性的横截面策略框架。

请务必阅读正文之后的信息披露和重要声明

在本文中，我们以动量策略为例，展示了如何使用 LTR 算法提升传统横截面策略的效果。我们首先将选股转换为一种排名问题，这样我们就可以在选股领域运用不同的 LTR 算法。然后，我们使用美国市场的数据测试了标准的有监督学习、启发式算法与 LTR 算法的效果，证明了基于 listwise 与 pairwise 的 LTR 算法能够显著提高策略的效果。最后，算法可扩展性较强，完全可以代入较多的特征变量，本文为横截面策略提供了一个新颖的框架体系。

2、文献综述：

2.1、横截面动量策略

动量策略可以分为时间序列的动量策略（单变量）与横截面的动量策略（多变量）。在时间序列的动量策略里，资产的交易取决于其自身的历史回报，这种策略最早由 Moskowitz, Ooi, and Pedersen(2012)提出，他们测试了其在近 60 个品种过去 25 年的表现。后人有很多工作即受此启发，探讨如何进一步提高策略效果。(Baz et al. 2015; Rohrbach and Suremann 2017; Lim, Zohren, and Roberts 2019)。

横截面动量策略（CSM 策略）的思路类似，但更着重对比资产的相对表现。它的特点是首先按照某个特征（通常为历史收益率）对资产进行排序，然后做多排序靠前的资产，做空排序靠后的资产。自从 Jegadeesh and Titman (1993) 针对美国市场运用了这种策略，出现了很多文献研究这种策略在其他市场与其他资产上的效果。(LeBaron 1996; Rouwenhorst 1998; Griffin, Ji, and Martin 2003; Erb and Harvey 2006; Chui, Titman, and Wei 2010; de Groot, Pang, and Swinkels 2012)。也有人提出了改进策略的方法，Kim (2019) 用夏普比率进行排序而不是简单地历史收益率。Pirrongo (2005) 则使用波动率对历史收益率进行了标准化处理，认为可以更好的评价不同资产的表现。Baz et al. (2015) 使用了很多类似但更复杂的方法，比如使用波动率调整后的 MACD 指标作为排序变量。近期文献的共同点是他们都是用“回归后再排序”的方法(Wang and Rasheed 2018)，在构建组合前通过模型使得预测误差最小，然后再构建多空组合。通常这些文献用的损失函数都是 MSE (Kim (2019)、Gu et al. (2019))。

回顾 CSM 的相关文献，本文是第一篇考虑使用排序算法增强 CSM 策略的文章。本文运用的并不是启发式算法，也不是 pointwise 损失函数，而是用到了 LTR 算法，基于 pairwise 和 listwise 的 LTR 算法能够显著提升样本外的表现。

2.2、LTR 算法在金融领域的运用

在信息检索领域，LTR 是学者们核心的研究算法，重点是使用机器学习来训练模型，进而执行排名任务(Li 2011; Liu 2011)。现代的计算技术与技术框架都有巨大的进步（参阅 Google Cloud Cloud TPU 文档、PyTorch、TensorFlow），这些都使得机器学习的方法更容易落地，研究人员不需要从基础的 BM25 和语言模型学起(Li 2011)，也能构建相对复杂的架构(Pasumarthi et al. 2019)。

如今 LTR 算法在众多商业领域都有关键作用，如搜索引擎(Liu 2011)、电子商

请务必阅读正文之后的信息披露和重要声明

- 4 -

务(Santu, Sondhi, and Zhai 2017)和娱乐(Pereira et al. 2019)。Song, Liu, and Yang (2017)、Wang and Rasheed(2018)探索了其在金融领域的应用: Song, Liu, and Yang (2017)将 RankNet 和 ListNet 运用到 10 年的市场与新闻情绪数据中, 获取的风险调整后收益比标普 500、HFRI EMN 和纯情绪技术构建的策略都高。Wang and Rasheed(2018)证明了 LambdaMART 相比于标准神经网络在预测深交所股票日内收益的优势。

尽管他们取得了较好的效果, 但这些应用并未与传统的系统化策略进行对比, 而只是与市场的收益进行了对比, 并且他们的 LTR 算法也根据具体情况进行了改动, 难以评估 Song, Liu, and Yang (2017)、Wang and Rasheed(2018)研究的真实价值。本文明确的解决了这些问题, 将 LTR 模型纳入横截面策略框架。

3、问题定义

给定月度再平衡的组合, 在 t_m 时刻, CSM 策略可以表达如下:

$$r_{t_m, t_{m+1}}^{CSM} = \frac{1}{n_{t_m}} \sum_{i=1}^{n_{t_m}} X_{t_m}^i \frac{\sigma_{tgt}}{\sigma_{t_m}^{(i)}} r_{t_m, t_{m+1}}^i$$

其中 $t_m, t_{m+1} \in T\{1, \dots, t-1, t, t+1, \dots, T\}$, 其中 T 代表每个月的最后一个交易日。 $r_{t_m, t_{m+1}}^{CSM}$ 表示组合从 t_m 到 t_{m+1} 的实际收益。 n_{t_m} 代表组合里的股票数量。 $X_{t_m}^{(i)}$ 为-1、0、1 中的一个数, 表示 CSM 组合对于资产 i 的信号。我们选择月频调仓, 避免频率过高比如日频引起的交易成本。我们将组合的目标波动率 σ_{tgt} 设置为 15%, 除以 $\sigma_{t_m}^{(i)}$, 后者代表事前对该月波动率的预测。在本文中, 我们使用 63 天的指数移动平均计算波动率 $\sigma_{t_m}^{(i)}$, 但可以用更加复杂的方法(如 GARCH [Bollerslev 1986])。

CSM 策略的总体框架包括以下 4 个组成部分:

得分计算。在 t_m 时刻资产 i 的输入向量为 $\mathbf{u}_{t_m}^{(i)}$, 通过预测模型 f 计算得到相应的分数 $Y_{(t_m)}^{(i)}$:

$$Y_{(t_m)}^{(i)} = (f(\mathbf{u}_{t_m}^{(i)}))_{t_m}$$

对于资产数量为 N_{t_m} 的横截面, 资产对应的分数集合为 $Y_{t_m} = \{Y_{t_m}^1, \dots, Y_{t_m}^{N_{t_m}}\}$ 。

得分排序。第二步可以表示为:

$$Z_{t_m}^{(i)} = R(Y_{t_m}^{(i)})$$

其中 $Z_{t_m}^{(i)} \in \{1, \dots, N_{t_m}\}$ 为资产 i 使用 $R()$ 排序后的持仓。

资产选择。这步通常根据排名阈值构建相应的多空组合。下面的等式用头尾 10% 的分位数构建了多空组合。

$$X_{(t_m)}^i = \begin{cases} -1 & Z_{(t_m)}^{(i)} \leq \left\lfloor 0.1 \times N_{t_m} \right\rfloor \\ 1 & Z_{(t_m)}^{(i)} > \left\lfloor 0.9 \times N_{t_m} \right\rfloor \\ 0 & \text{Otherwise} \end{cases}$$

组合构建。最终，我们可以根据波动率构建简单的组合。

下面的章节中，我们将用 LTR 方法与传统方法分别进行分数的计算。

4、得分计算方法

大多数 CSM 策略都是以上框架，并且后面 3 步通常类似（根据得分排序、选择资产、构架组合）。但他们的差异在于如何选择模型 f 来计算得分，有简单的启发式模型（Jegadeesh and Titman 1993），也有复杂的使用宏观变量预测的模型（Gu et al. 2018），尽管这类方法繁多，我们可以将其总结为 3 类：经典的动量、回归后再排序和本文的 LTR 方法。

4.1、经典的横截面动量

经典的横截面动量策略计算分数时方法比较简单。

Jegadeesh and Titman (1993)。作者首次提到用原始的收益率作为得分，计算窗口为过去 3 到 12 个月。

$$\text{Score Calculation: } Y_{t_m}^i = r_{t_m-252, t_m}^i$$

其中 $r_{t_m-252, t_m}^{(i)}$ 为资产 i 在过去 252 天的原始收益率。

Baz et al. (2015)。作者用到了波动率调整后的 MACD 指标，用到的时间区间也有不同。

$$Y_{t_m}^{(i)} = \xi_{t_m}^i / \text{std}(z_{t_m-252, t_m}^i)$$

$$\xi_{t_m}^i = \text{MACD}(i, t_m, S_m, L) / \text{std}(p_{t_m-63, t_m}^i)$$

$$\text{MACD}(i, t_m, S, L) = m(i, S) - m(i, L)$$

其中 $\text{std}(p_{t_m-63, t_m}^{(i)})$ 代表资产 i 在过去 63 天的滚动波动率， $m(i, S)$ 代表资产 i 的指数移动平均， S 对应的半衰期长度为 $HL = \log(0.5) / \log(1-1/S)$ ，最终信号试验了不同参数下的效果，包括 $S_k \in \{8, 16, 32\}$ ， $L_k \in \{24, 48, 96\}$ 。

$$\text{Score Calculation: } Y_{t_m}^{(i)} = \sum_{k=1}^3 \phi(Y_{t_m}^{(i)}(S_k, L_k))$$

其中 $r_{t_m-252, t_m}^{(i)}$ 为资产 i 在过去 252 天的原始收益率。

4.2、回归后再排序方法

在近年的文章常用到回归后再排序的方法，这种方法通过标准的回归过程计算得分：

请务必阅读正文之后的信息披露和重要声明

Score Calculation: $Y_{t_m}^{(i)} = f(\mathbf{u}_{t_m}^{(i)}; \boldsymbol{\theta})$

其中 f 代表输入变量为 $\mathbf{u}_{t_m}^{(i)}$ ，参数为 $\boldsymbol{\theta}$ 的预测模型，该预测模型的目标是最小化损失函数（通常为 MSE），预测变量为波动率调整后的收益：

$$\zeta(\boldsymbol{\theta}) = \frac{1}{M} \sum_{\Omega} (Y_{t_m}^{(i)} - \frac{r_{t_m, t_{m+1}}^{(i)}}{\Omega_{t_m}^{(i)}})^2$$

$$\Omega = \{(Y_{t_1}^{(1)}, r_{t_1, t_2}^{(1)} / \sigma_{t_1}^{(1)}), \dots, (Y_{t_{m-1}}^{(N_{t_{m-1}})}, r_{t_{m-1}, t_m}^{(N_{t_{m-1}})} / \sigma_{t_{m-1}}^{(N_{t_{m-1}})})\}$$

其中 Ω 表示所有可能的预测变量与被实际变量的数组集合。

4.3、LTR 算法

LTR 算法可以分为 pointwise、pairwise、和 listwise 三种。pointwise（或 pairwise）算法将排序问题转化为一个分类、回归或是有序分类的问题，listwise 算法的输入变量为排序列表，进而训练合适的排序模型。pointwise 的排序效果相对其他两种略差（Li 2011）。pairwise 与 listwise 算法下的 LTR 模型与传统的模型、回归再排序的模型有本质区别，这种区别并不只是损失函数不同。

下面我们简单展示 LTR 算法如何运用于动量策略，为了文章简洁性，正文并不展示过度细节，读者如有兴趣可以查看附录中的“Learning to Rank for Cross-Sectional Momentum”部分。

Burges et al. (2005) (RankNet)。RankNet 是第一个将神经网络技术运用到排名问题的算法。RankNet 的目标并不是最小化 MSE，而是最小化分类样本对的交叉熵误差，最优化一个样本排序高于另一个样本的概率。因为是用随机梯度下降法，RankNet 的算法复杂度随着证券数量的增长，以其数量平方的速度增长。

Burges (2010) (LambdaMART)。LambdaMART 属于 pairwise 算法，其结合了 LambdaRank 算法与 MART 算法（multiple additive regression trees）。LambdaMART 并不是直接最小化某个损失函数，而是用启发式方法进行梯度迭代（ λ 梯度）。这种特点使得算法避免陷入局部最优，规避了 LambdaRank 常见的问题（Yue and Burges 2007; Donmez, Svore, and Burges 2009）。因为使用了 λ 梯度，这种算法结合了交叉熵误差与信息检索领域的成果。另一方面，MART 是一种提升树算法，它通常能够处理好速度与效率之间的关系（这一点对搜索引擎非常重要）（Wu et al. 2010）。LambdaMART 同时结合了 LambdaRank 的最优化优势与 MART 算法的灵活性与稳健性。

Cao et al. (2007) (ListNet)。开发 ListNet 是为了解决 pairwise 在实际中遇到的问题如有些情况无法计算损失函数，有些情况下会出现算法并没有最小化整个排序误差。ListNet 运用 listwise 损失函数和基于排序的概率方法解决了以上问题。算法先训练排名最靠前的样本、然后通过 softmax 函数逐个训练排名靠后的样本，损失函数为分布的交叉熵。算法的输入为整个资产的横截面数据，因此复杂度为 $O(N_{t_m})$ ，比 RankNet 的 $O(N_{t_m}^2)$ 复杂度效率高很多。

Xia et al. (2008) (ListMLE)。为了给 listwise 的算法表现与理论进行结合，Xia et al. (2008) 提出了具有一致性、完整性和线性复杂度的似然损失函数。似然函

请务必阅读正文之后的信息披露和重要声明

数连续、可微且是凸函数(Boyd and Vandenberghe 2004)。ListMLE 将排序问题的损失函数定义为最小化损失概率或者说最大化似然函数。Xia et al. 2008 证实 ListMLE 在同等复杂度的前提下，在基准数据集上表现超过其他的 listwise 算法。

经典的横截面动量策略计算分数时方法比较简单。后文中我们也发现，ListMLE 与 ListNet 的线性复杂度更适合大规模数据集。

4.4、训练细节

RankNet, ListNet, 和 ListMLE 运用 Adam 优化器执行对应的排序损失函数最小化，最多反向传播 100 层，我们每次使用 90% 的数据进行训练，10% 的数据进行验证。出于实际考虑，我们的目标是未来 21 个交易日的收益而非未来一个月的收益率。对于用到神经网络的 LTR 模型，我们使用 2 层隐藏层，每层的宽度是一个可调节的超参数。如果 25 次迭代损失函数没有显著变化，我们就会提前停止训练，防止过拟合。另外，我们还使用 dropout 正则化防止过拟合，其中 dropout rate 是一个超参数。所有模型中的超参数进行 50 次迭代以达到最优超参数(Bergstra et al. 2015)。关于超参数的设置细节可见附录中“Additional Training Details”部分。

5、表现分析

5.1、数据情况

本文使用 CRSP 2019 的数据构建了月频调仓的组合。我们的样本包括 1980 年到 2019 年在 NYSE 交易的，且 CRSP 的股票代码为 10 和 11 的所有股票。每次调仓间隔，我们只交易股票价格在 1 美金之上的股票。另外，我们只考虑过去一年有活跃成交的股票，以下所有的价格都是收盘价。

5.2、回测和预测描述

除了两种运用启发式排名的经典策略外，所有模型的超参数 5 年调整一次，策略在每个月的最后一个交易日调仓。每次我们分别做多和做空 100 只股票，100 只股票大概占比所有可交易股票的 10%，我们用经典的得分计算方法：

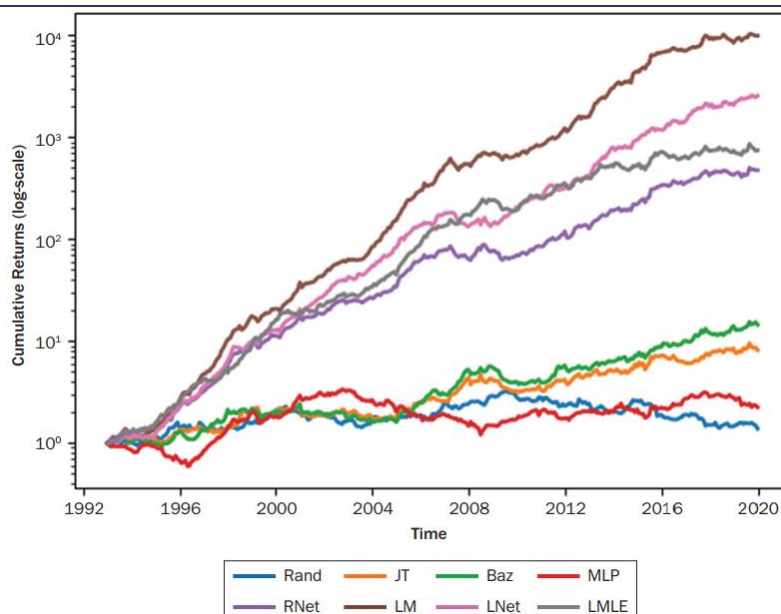
1. 计算原始收益率：参考 Jegadeesh and Titman (1993)，计算过去 3、6、12 个月的收益率；
2. 标准化收益：分别计算 3、6、12 个月窗口的日频波动率，然后根据时间维度进行调整；
3. MACD 指标：我们参考 Baz et al.(2015)的方法构建了最终的信号。我们把原始的中间信号 $Y_{t_m}^{(i)}(S_k, L_k)$ 加入到预测指标集中，其中 $k=1,2,3$ ；分别计算过去 1、2、6、12 个月的信号值，因此一共有 16 个预测信号。

5.3、模型和对比维度

LTR 和基准模型（缩写见后面括号）如下：

1. Random (Rand): 随机选择股票构建组合，提供一个完全随机的基准。
2. Random return(JT): 参考 Jegadeesh and Titman (1993) 的基于启发式排序方法，最早的 CSM 策略。
3. Volatility normalized MACD (Baz): 参考 Baz et al. (2015) 构建的更复杂的启发式排序算法。
4. Multilayer perceptron (MLP): 典型的回归后再排序的技术。
5. RankNet (RNet): Pairwise LTR 模型，参考 Burges et al. (2005)。
6. LambdaMART (LM): Pairwise LTR 模型，参考 Burges (2010)。
7. ListNet (LNet): Listwise LTR 模型，参考 Cao et al. (2007)。
8. ListMLE (LMLE): Listwise LTR 模型，参考 Xia et al. (2008)。

图表 1、按目标波动率调整后的累计收益



资料来源：The Journal of Financial Data Science，兴业证券经济与金融研究院整理

我们按照以下两类标准评价不同算法的表现：第一类是金融领域通用的方法（以下 1 到 3）、第二类是信息检索和排序领域的方法（以下 4）：

1. 收益：预期收益和样本外组合正收益的占比。
2. 风险：月频波动率，最大回撤、下行偏离度。
3. 金融指标：夏普率（预期收益率比波动率）、索提诺比率（预期收益比最大回撤）、卡玛比率（预期收益比下行偏离度）。此外，我们还用到了平均利润比平均亏损作为评价指标。
4. 排序表现：Kendall's Tau 指标，还有 NDCG@k，参考 Järvelin and Kekäläinen 2000，适合评价多头组合的收益（Wu et al. 2010）。其中 k 本文设定为 100。

请务必阅读正文之后的信息披露和重要声明

5.4、结果与结论

图表 1 展示了样本外的各策略表现，图表 2 展示了它们的金融指标。为了更好地比较各策略，我们将每个组合的波动率都调整为 15%，所有计算均为考虑交易成本。无论是净值还是指标表现，我们的 LTR 算法都明显强于基准策略，LambdaMART 策略在较多维度上均排首位。

图表 2、按目标波动率调整后的策略表现指标

	Benchmarks				LRT models			
	Rand	JT	Baz	MLP	RNet	LM	LNet	LMLE
E[returns]	0.024	0.092	0.112	0.044	0.243	0.359	0.306	0.260
Volatility	0.156	0.167	0.161	0.165	0.162	0.166	0.155	0.162
Sharpe	0.155	0.551	0.696	0.265	1.502	2.156	1.970	1.611
Downside dev.	0.106	0.106	0.097	0.112	0.081	0.067	0.068	0.071
MDD	0.584	0.328	0.337	0.641	0.294	0.231	0.274	0.236
Sortino	0.228	0.872	1.157	0.389	3.012	5.321	4.470	3.647
Calmar	0.042	0.281	0.333	0.068	0.828	1.555	1.115	1.102
% Positive Returns	0.545	0.582	0.591	0.551	0.693	0.762	0.715	0.681
Avg. P/Avg. L	0.947	1.114	1.184	1.001	1.407	1.594	1.679	1.534

资料来源：The Journal of Financial Data Science，兴业证券经济与金融研究院整理

收益维度上，排序算法在预期收益与胜率表现上均有明显改善，最差的 LTR 算法也能显著超越表现最好的基准算法。尽管所有策略均调整为同一个风险水平，但 LTR 算法在波动率与下行风险的控制上均表现更好。在其他维度上，表现最差的 LTR 模型也仍然强于表现最好的基准模型，并且表现最好 LTR 模型在各个维度表现很强，清楚的说明了 LTR 算法可以显著的提高动量策略的表现。

进一步分析每个组中策略的相对表现，我们首先发现，基于 listwise 的 LTR 策略相比基于 pairwise 的 LTR 策略没有明显优势。有人可能认为基于 listwise 的 LTR 策略学习了更广泛的结构因此其表现理应更好，但结果并不支持这种说法。可以解释为：金融数据中的信噪比较低，另外 listwise 用到的数据量是每年 $12 \times N_{\text{average}}$ ，而 pairwise 用到的数据量是 $12 \times N_{\text{average}}^2$ 每年，其中 N_{average} 是每个月横截面平均样本数。

在基准策略中，随机模型表现最差，只有 MLP 策略表现略好一些。由于数据量有限且噪音较多，这有可能是过拟合的结果。另外，MLP 做收益预测时只用到了价格数据，Naccarato, Pierini, and Ferraro (2019) 认为这可能是模型的隐患。（参见“Dataset Overview in the PerformanceEvaluation”部分）。

NDCG 通过秩相关性和调整后的权重来对模型效果进行评价，适合评价头部排序的准确性（Wang et al.2013）因此在信息检索领域得到广泛运用（Li 2011）。我们认为 NDCG 特别适合评价 CSM 策略的表现，因为它能衡量头部资产的排序效果，这与最终的多空组合表现直接挂钩。基于此我们用 NDCG 评价所有策略的表现，设置 k 为 100。从排序的表现看，在 NDCG@100 的评价体系下，LTR 算法的表现全面超越基准策略的表现，说明其在排序能力方面表现更好。Kendall's Tau（秩相关系数）上的规律也完全相同，说明 LTR 算法对全体样本的排序也更准确。

图表 3、每个月的策略平均排序表现

	Benchmarks				LTR Models			
	Rand	JT	Baz	MLP	RNet	LM	LNet	LMLE
Kendall's Tau	0.000	0.016	0.013	0.008	0.032	0.032	0.033	0.020
NDCG@100 (longs)	0.549	0.555	0.562	0.550	0.576	0.576	0.578	0.565
NDCG@100 (shorts)	0.552	0.562	0.555	0.564	0.575	0.585	0.579	0.567

资料来源：The Journal of Financial Data Science，兴业证券经济与金融研究院整理

为了进一步评价样本外的排序表现，我们构建了一下多空组合：每个月根据模型的信号，等权重构建组合。比如多头组合即为 MLP 打分最高的 10%，收益表现与之前类似，但分位数指标 $D_{t_m}^{(i)}$ 替换了原来的 $X_{t_m}^{(i)}$ ：

$$r_{t_m, t_{m+1}}^{DEC} = \frac{1}{n_{t_m}} \sum_{i=1}^{n_{t_m}} D_{t_m}^{(i)} \frac{\sigma_{tgt}}{\sigma_{t_m}^{(i)}} r_{t_m, t_{m+1}}^i$$

其中 $D_{t_m}^{(i)} = \{0, 1\}$ ，对于分为组合的股票为 1，否则为 0。

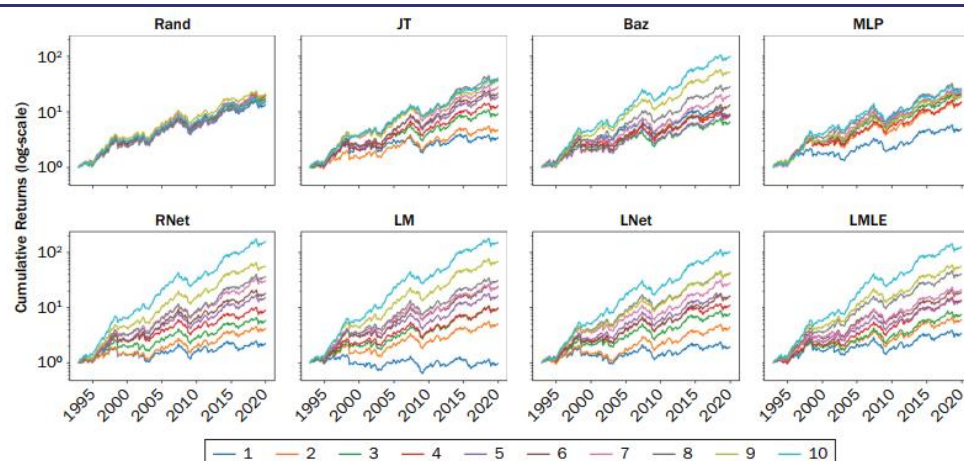
我们还在组合层面进行了额外的调整，从图表 4 的表现汇总可以看到：从分位数组合 1 到组合 10，除了随机模型，其余模型的收益与夏普比都是大致单调的，这说明动量效应确实存在。尤其是 LTR 算法对应的分位数组合，不同组的差异更加显著。图表 5 绘制了不同策略的分位数组合收益，进一步说明了算法的排序能力与组合的最终分位数组合收益的离散度有关。LTR 模型中，二者的关系最为显著，验证了我们的假设：资产排序能力越好，策略表现越好。

图表 4、目标波动率调整后的分位数组合表现

	Decile										L-S
	1	2	3	4	5	6	7	8	9	10	
Rand											
E[returns]	0.110	0.120	0.118	0.124	0.114	0.123	0.126	0.115	0.126	0.115	0.028
Volatility	0.162	0.163	0.164	0.162	0.164	0.162	0.163	0.161	0.163	0.164	0.156
Sharpe	0.675	0.737	0.721	0.769	0.697	0.763	0.772	0.710	0.772	0.702	0.177
JT											
E[returns]	0.059	0.071	0.096	0.108	0.122	0.127	0.137	0.151	0.146	0.146	0.094
Volatility	0.165	0.164	0.165	0.164	0.163	0.163	0.163	0.162	0.160	0.156	0.167
Sharpe	0.360	0.435	0.581	0.661	0.746	0.780	0.840	0.928	0.910	0.938	0.565
Baz											
E[returns]	0.095	0.094	0.083	0.094	0.092	0.109	0.124	0.137	0.160	0.185	0.107
Volatility	0.163	0.163	0.162	0.162	0.163	0.162	0.162	0.162	0.163	0.163	0.161
Sharpe	0.582	0.573	0.510	0.579	0.566	0.671	0.765	0.849	0.984	1.130	0.664
MLP											
E[returns]	0.072	0.112	0.122	0.114	0.127	0.126	0.130	0.135	0.124	0.132	0.097
Volatility	0.163	0.161	0.163	0.162	0.163	0.162	0.163	0.163	0.163	0.164	0.168
Sharpe	0.443	0.697	0.751	0.703	0.780	0.779	0.800	0.825	0.758	0.806	0.578
RNet											
E[returns]	0.043	0.067	0.079	0.095	0.115	0.121	0.140	0.147	0.163	0.202	0.246
Volatility	0.164	0.164	0.164	0.164	0.164	0.163	0.161	0.162	0.161	0.163	0.161
Sharpe	0.263	0.405	0.480	0.580	0.698	0.742	0.870	0.906	1.014	1.238	1.527
LM											
E[returns]	0.012	0.074	0.097	0.098	0.117	0.132	0.131	0.141	0.171	0.201	0.349
Volatility	0.161	0.165	0.162	0.163	0.164	0.162	0.164	0.162	0.163	0.162	0.155
Sharpe	0.075	0.449	0.599	0.606	0.716	0.806	0.800	0.868	1.053	1.232	2.107
LNet											
E[returns]	0.037	0.069	0.089	0.102	0.116	0.117	0.137	0.152	0.151	0.186	0.296
Volatility	0.161	0.165	0.162	0.163	0.164	0.162	0.164	0.162	0.163	0.162	0.155
Sharpe	0.232	0.416	0.549	0.628	0.711	0.720	0.837	0.938	0.929	1.148	1.911
LMLE											
E[returns]	0.059	0.080	0.088	0.110	0.109	0.123	0.126	0.151	0.163	0.193	0.244
Volatility	0.165	0.164	0.164	0.165	0.161	0.164	0.162	0.161	0.160	0.163	0.160
Sharpe	0.360	0.489	0.537	0.671	0.677	0.750	0.782	0.937	1.016	1.186	1.530

资料来源：The Journal of Financial Data Science，兴业证券经济与金融研究院整理

图表 5、目标波动率调整后的分位数组合累计收益



资料来源：The Journal of Financial Data Science，兴业证券经济与金融研究院整理

6、结论

我们针对 CSM 策略中重要的排序步骤，提出了一个新颖的算法——LTR 算法。LTR 算法框架可以灵活的调整输入变量，是一种扩展性很强的算法。通过学习样本排序的关系（pairwise 或是 listwise），我们可以获得更精确的排序结果，而普通的启发式算法、回归后再排序算法都无法获得精确的结果。相比传统的方法，LTR 算法对应的排序结果更准确，策略效果也更好。

我们认为本文未来的改进方向包括对模型的结构进行创新，提升对高频数据（如订单簿数据）或是资产类别方面的排序效率。

参考文献

- 【1】Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. “TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems.” 2015, tensorflow.org.
- 【2】Baz, J., N. M. Granger, C. R. Harvey, N. Le Roux, and S. Rattray. 2015. “Dissecting Investment Strategies in the Cross Section and Time Series.” SSRN Electronic Journal. Available at SSRN: <https://ssrn.com/abstract=2695101> or <http://dx.doi.org/10.2139/ssrn.2695101>.
- 【3】Bergstra, J., B. Komer, C. Eliasmith, D. Yamins, and D. D. Cox. 2015. “Hyperopt: A Python Library for Model Selection and Hyperparameter Optimization.” Computational Science & Discovery 8 (1): 014008.
- 【4】Bollerslev, T. 1986. “Generalized Autoregressive Conditional Heteroskedasticity.” Journal of Econometrics 31 (3): 307 – 327.
- 【5】Boyd, S., and L. Vandenberghe. 2004. Convex Optimization. USA: Cambridge University Press.
- 【6】Bruch, S. 2020. “An Alternative Cross Entropy Loss for Learning-to-Rank.” arXiv:1911.09798, 2020.
- 【7】Burges, C. “From RankNet to LambdaRank to LambdaMART: An Overview.” Microsoft Research, Technical report MSR-TR-2010-82, 2010.
- 【8】Burges, C., T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender. “Learning to Rank Using Gradient Descent.” Proceedings of the 22nd International Conference on Machine Learning. ICML ’05, pp. 89 – 96. New York, NY: Association for Computing Machinery, 2005.
- 【9】Burges, C., R. Ragno, and Q. Le. 2006. “Learning to Rank with Nonsmooth Cost Functions.” Proceedings of the 19th International Conference on Neural Information Processing Systems. NIPS ’06, pp. 193 – 200. Cambridge, MA: MIT Press, 2006.
- 【10】Cao, Z., T. Qin, T.-Y. Liu, M.-F. Tsai, and H. Li. 2007. “Learning to Rank: From Pairwise Approach to Listwise Approach.” Proceedings of the 24th International Conference on Machine Learning—ICML ’07, pp. 129 – 136. Corvallis, OR: ACM Press, 2007.
- 【11】Center for Research in Security Prices (CRSP), The University of Chicago Booth School of Business. “NYSE Equity Data from 1980 to 2019.” Calculated (or derived) based on data from CRSP Daily Stock, 2019.
- 【12】Chen, T., and C. Guestrin. “XGBoost: A Scalable Tree Boosting System.” Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD ’16, pp. 785 – 794. New York, NY: Association for Computing Machinery, 2016.
- 【13】Chui, A. C. W., S. Titman, and K. C. J. Wei. 2010. “Individualism and Momentum around the World.” The Journal of Finance 65 (1): 361 – 392.
- 【14】de Groot, W., J. Pang, and L. Swinkels. 2012. “The Cross-Section of Stock Returns in Frontier Emerging Markets.” Journal of Empirical Finance 19 (5): 796 – 818.
- 【15】Donmez, P., K. M. Svore, and C. Burges. “On the Local Optimality of LambdaRank.” Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval—SIGIR ’09, pp. 460. Boston, MA: ACM Press, 2009.
- 【16】Erb, C. B., and C. R. Harvey. 2006. “The Strategic and Tactical Value of Commodity Futures.” Financial Analysts Journal 62 (2): 69 – 97.
- 【17】Griffin, J. M., X. Ji, and J. S. Martin. 2003. “Momentum Investing and Business Cycle Risk: Evidence from Pole to Pole.” The Journal of Finance 58 (6): 2515 – 2547.
- 【18】Gu, S., B. Kelly, and D. Xiu. Empirical Asset Pricing via Machine Learning. Technical report w25398, National Bureau of Economic Research, 2018.
- 【19】——. “Autoencoder Asset Pricing Models.” Yale ICF Working Paper No. 2019-04, Chicago Booth Research Paper No. 19-24, 2019, <https://ssrn.com/abstract=3335536>.
- 【20】Järvelin, K., and J. Kekäläinen. “IR Evaluation Methods for Retrieving Highly Relevant Documents.” Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval—SIGIR ’00, pp. 41–48. Athens, Greece: ACM Press, 2000.
- 【21】Jegadeesh, N., and S. Titman. 1993. “Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency.” The Journal of Finance 48 (1): 65 – 91.
- 【22】Jusselin, P., E. Lezmi, H. Malongo, C. Masselin, T. Roncalli, and T.-L. Dao. “Understanding the Momentum Risk Premium: An In-Depth Journey through Trend-Following Strategies.” SSRN Electronic Journal, 2017, <https://ssrn.com/abstract=3042173>.

- 【23】Kim, S. 2019. “Enhancing the Momentum Strategy through Deep Regression.” *Quantitative Finance* 19 (7): 1121 – 1133.
- 【24】LeBaron, B. Technical Trading Rule Profitability and Foreign Exchange Intervention. Technical report w5505, National Bureau of Economic Research, 1996.
- 【25】Li, H. 2011. “Learning to Rank for Information Retrieval and Natural Language Processing.” *Syn- thesis Lectures on Human Language Technologies* 4 (1): 1 – 113.
- 【26】Li, P., Z. Qin, X. Wang, and D. Metzler. “Combining Decision Trees and Neural Networks for Learning-to-Rank in Personal Search.” *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2032 – 2040. Anchorage, AK: Association for Computing Machinery, 2019.
- 【27】Lim, B., S. Zohren, and S. Roberts. “Enhancing Time Series Momentum Strategies Using Deep Neural Networks.” *SSRN Electronic Journal*, 2019, <https://ssrn.com/abstract/3369195>.
- 【28】Liu, T.-Y. *Learning to Rank for Information Retrieval*. Berlin, Heidelberg: Springer, 2011.
- 【29】Moskowitz, T. J., Y. H. Ooi, and L. H. Pedersen. 2012. “Time Series Momentum.” *Journal of Financial Economics* 104 (2): 228 – 250.
- 【30】Naccarato, A., A. Pierini, and G. Ferraro. 2019. “Markowitz Portfolio Optimization through Pairs Trading Cointegrated Strategy in Long-Term Investment.” *Annals of Operations Research* 276 (1 – 2): 1 – 19.
- 【31】Nguyen, P., J. Wang, and A. Kalousis. 2016. “Factorizing LambdaMART for Cold Start Recommendations.” *Machine Learning* 104 (2): 223 – 242.
- 【32】Pasumarthi, R. K., S. Bruch, X. Wang, C. Li, M. Bendersky, M. Najork, J. Pfeifer, N. Golbandi, R. Anil, and S. Wolf. “TF-Ranking: Scalable TensorFlow Library for Learning-to-Rank.” *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2970–2978. Anchorage, AK: Association for Computing Machinery, 2019.
- 【33】Paszke, A., S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. “PyTorch: An Imperative Style, High-Performance Deep Learning Library.” *Advances in Neural Information Processing Systems* 32, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché-Buc, E. Fox, and R. Garnett, pp. 8024–8035. Curran Associates, 2019.
- 【34】Pereira, B. L., A. Ueda, G. Penha, R. L. T. Santos, and N. Ziviani. 2019. “Online Learning to Rank for Sequential Music Recommendation.” *Proceedings of the 13th ACM Conference on Recommender Systems*, pp. 237 – 245. Copenhagen, Denmark: Association for Computing Machinery.
- 【35】Pirrion, C. 2005. “Momentum in Futures Markets.” *SSRN Electronic Journal*, 2005.
- Rohrbach, J., and S. Suremann. 2017. “Momentum in Traditional and Cryptocurrencies Made Simple.” *SSRN Electronic Journal*, 2017.
- 【36】Roncalli, T. 2017. “Keep Up the Momentum.” *SSRN Electronic Journal*, 2017, <https://ssrn.com/abstract/3751012>.
- 【37】Rouwenhorst, K. G. 1998. “International Momentum Strategies.” *The Journal of Finance* 53 (1): 267 – 284.
- 【38】Santu, S. K. K., P. Sondhi, and C.-X. Zhai. 2017. “On Application of Learning to Rank for E-Commerce Search.” *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 475 – 484.
- 【39】Song, Q., A. Liu, and S. Y. Yang. 2017. “Stock Portfolio Selection Using Learning-to-Rank Algorithms with News Sentiment.” *Neurocomputing* 264 (November): 20 – 28.
- 【40】Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. 2014. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting.” *Journal of Machine Learning Research* 15 (56): 1929 – 1958.
- 【41】Wang, B., and D. Klabjan. 2017. “An Attention-Based Deep Net for Learning to Rank.” 2017, arXiv:1702.06106 [cs].
- 【42】Wang, L., and K. Rasheed. “Stock Ranking with Market Microstructure, Technical Indicator and News.” *Proceedings on the International Conference on Artificial Intelligence (ICAI)*, pp. 322 – 328. New York: Springer, 2018.
- 【43】Wang, P., C. Liu, Y. Yang, and S. Huang. “A Robo-Advisor Design Using Multiobjective Rank- nets with Gated Neural Network Structure.” *2019 IEEE International Conference on Agents (ICA)*, pp. 77 – 78. IEEE, 2019.
- 【44】Wang, Y., L. Wang, Y. Li, D. He, and T.-Y. Liu. “A Theoretical Analysis of NDCG Type Ranking Measures.” *26th Annual Conference on Learning Theory. Proceedings of Machine Learning Research*, vol. 30, edited by S. Shalev-Shwartz and I. Steinwart, pp. 25 – 54. Princeton, NJ:

请务必阅读正文之后的信息披露和重要声明

- 14 -

PMLR, 2013.

【45】Wu, Q., C. Burges, K. M. Svore, and J. Gao. 2010. “Adapting Boosting for Information Retrieval Measures.” *Information Retrieval* 13 (3): 254 – 270.

【46】Xia, F., T.-Y. Liu, J. Wang, W. Zhang, and H. Li. 2008. “Listwise Approach to Learning to Rank: Theory and Algorithm.” *Proceedings of the 25th International Conference on Machine Learning. ICML ’08*, pp. 1192 – 1199. New York: Association for Computing Machinery.

【46】Yue, Y., and C. Burges. “On Using Simultaneous Perturbation Stochastic Approximation for Learning to Rank, and the Empirical Optimality of LambdaRank.” Microsoft Research, Technical report MSR-TR-2007-115, 2007.

风险提示：文献中的结果均由相应作者通过历史数据统计、建模和测算完成，在政策、市场环境发生变化时模型存在失效的风险。

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。

投资评级说明

投资建议的评级标准	类别	评级	说明
报告中投资建议所涉及的评级分为股票评级和行业评级(另有说明的除外)。评级标准为报告发布日后的12个月内公司股价(或行业指数)相对同期相关证券市场代表性指数的涨跌幅。其中：A股市场以上证综指或深圳成指为基准，香港市场以恒生指数为基准；美国市场以标普500或纳斯达克综合指数为基准。	股票评级	买入	相对同期相关证券市场代表性指数涨幅大于15%
		审慎增持	相对同期相关证券市场代表性指数涨幅在5%~15%之间
		中性	相对同期相关证券市场代表性指数涨幅在-5%~5%之间
		减持	相对同期相关证券市场代表性指数涨幅小于-5%
		无评级	由于我们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，致使我们无法给出明确的投资评级
	行业评级	推荐	相对表现优于同期相关证券市场代表性指数
		中性	相对表现与同期相关证券市场代表性指数持平
		回避	相对表现弱于同期相关证券市场代表性指数

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。

使用本研究报告的风险提示及法律声明

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。

，本公司不会因接收人收到本报告而视其为客户。本报

告中的信息、意见等均仅供客户参考，不构成所述证券买卖的出价或征价邀请或要约。该等信息、意见并未考虑到获取本报告人员的具体投资目的、财务状况以及特定需求，在任何时候均不构成对任何人的个人推荐。客户应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求，必要时就法律、商业、财务、税收等方面咨询专家的意见。对依据或者使用本报告所造成的一切后果，本公司及/或其关联人员均不承担任何法律责任。

本报告所载资料的来源被认为是可靠的，但本公司不保证其准确性或完整性，也不保证所包含的信息和建议不会发生任何变更。本公司并不对使用本报告所包含的材料产生的任何直接或间接损失或与此相关的其他任何损失承担任何责任。

本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可升可跌，过往表现不应作为日后的表现依据；在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告；本公司不保证本报告所含信息保持在最新状态。同时，本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

除非另行说明，本报告中所引用的关于业绩的数据代表过往表现。过往的业绩表现亦不应作为日后回报的预示。我们不承诺也不保证，任何所预示的回报会得以实现。分析中所做的回报预测可能是基于相应的假设。任何假设的变化可能会显著地影响所预测的回报。

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。

本报告并非针对或意图发送予或为任何就发送、发布、可得到或使用此报告而使兴业证券股份有限公司及其关联子公司等违反当地的法律或法规或可致使兴业证券股份有限公司受制于相关法律或法规的任何地区、国家或其他管辖区域的公民或居民，包括但不限于美国及美国公民（1934年美国《证券交易所》第15a-6条例定义为本「主要美国机构投资者」除外）。

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。

特别声明

在法律许可的情况下，兴业证券股份有限公司可能会利差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。

兴业证券研究

上海	北京	深圳
地址：上海浦东新区长柳路36号兴业证券大厦15层	地址：北京西城区锦什坊街35号北楼601-605	地址：深圳市福田区皇岗路5001号深业上城T2座52楼
邮编：200135	邮编：100033	邮编：518035
邮箱：research@xyzq.com.cn	邮箱：research@xyzq.com.cn	邮箱：research@xyzq.com.cn