

分析师：

徐寅

xuyinsh@xyzq.com.cn

S0190514070004

研究助理：

占康萍

zhankangping@xyzq.com.cn

西学东渐--海外文献推荐系列之一百二十三

2021 年 7 月 1 日

报告关键点

本文基于动态时间规整算法 DTW 构建了标准普尔 500 行业 ETF 的模式匹配交易策略,实证表明策略在某些行业表现良好,具有较高的收益水平。

相关报告

《西学东渐--海外文献推荐系列之一百二十》

《西学东渐--海外文献推荐系列之一百二十一》

《西学东渐--海外文献推荐系列之一百二十二》

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!
- 本文基于动态时间规整算法 DTW 构建了标准普尔 500 行业 ETF 的模式匹配交易策略。具体流程如下:1) 依据动态时间归整算法 DTW 构建了完整的模式匹配交易系统,并依据训练集得到决策字典;2) 在测试数据集上应用滑动窗口函数和模式匹配算法,根据每个窗口匹配的模式总比率做出交易决策;3) 最后,系统会根据投资组合优化理论计算每个 ETF 的最优权重,从而构建策略。根据实证结果,我们指出 DTW 在标准普尔 500 指数行业 ETF 上的应用潜力巨大。本文构建的模式匹配策略在某些行业表现良好,具有较高的收益水平。

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、数据收集和准备	4 -
2.1、数据源	4 -
2.2、数据预处理	4 -
2.3、训练集、验证集和测试集	4 -
2.4、窗口切分	5 -
3、研究方法	6 -
3.1、动态时间序列归整 DTW	6 -
3.2、模式匹配交易策略	7 -
4、实证结果	9 -
4.1、初测试	9 -
4.2、改进后的模式匹配策略	13 -
4.3、总结	15 -
5、未来研究与改进方向	16 -
图表 1、Min-Max 变换算法	4 -
图表 2、基于四节点的模式集	5 -
图表 3、DTW 中算法 1 与算法 2	7 -
图表 4、最近 250 日标准普尔 500 指数行业 ETF 价格图	9 -
图表 5、最近 250 日标准普尔 500 指数板块 ETF 收益图	9 -
图表 6、固定模式匹配策略的结果展示	10 -
图表 7、动态模式匹配策略的结果展示	11 -
图表 8、200 天和 252 天动态投资组合的优化后权重	11 -
图表 9、优化投资组合在两个不同的测试期中的资产净值	11 -
图表 10、COVID-19 冲击前的 200 日动态策略的资产净值（按行业划分）	12 -
图表 11、COVID-19 冲击后的 200 日动态策略的资产净值（按行业划分）	12 -
图表 12、200 日动态模式匹配策略投资组合表现统计	13 -
图表 13、改进后的模式匹配策略表现结果	13 -
图表 14、COVID-19 冲击前的改进后策略的资产净值（按行业划分）	13 -
图表 15、COVID-19 冲击后的改进后策略的资产净值（按行业划分）	14 -
图表 16、最优参数	14 -
图表 17、COVID-19 冲击前后的最优化策略的资产净值（按行业划分）	15 -
图表 18、改进后的模式匹配策略表现	15 -

报告正文**基于 DTW 算法构建标普 500 行业 ETF 模式匹配交易策略****文献来源：**

Fleiss, A., Liu, C., Eom, G., Yu, S., & Zhang, W., Dynamic Time Warping: S&P 500 Sector ETF Pattern Matching Trading Strategy. The Journal of Financial Data Science, 2021.

推荐理由：

本文基于动态时间规整算法 DTW 构建了标准普尔 500 行业 ETF 的模式匹配交易策略。具体流程如下：1) 依据动态时间归整算法 DTW 构建了完整的模式匹配交易系统，并依据训练集得到决策字典；2) 在测试数据集上应用滑动窗口函数和模式匹配算法，根据每个窗口匹配的模式总比率做出交易决策；3) 最后，系统会根据投资组合优化理论计算每个 ETF 的最优权重，从而构建策略。根据实证结果，我们指出 DTW 在标准普尔 500 指数行业 ETF 上的应用潜力巨大。本文构建的模式匹配策略在某些行业表现良好，具有较高的收益水平。

我们的思考：

欧式距离常被用于衡量时间序列之间的相似性，但这种方法也存在某些限制。因此，本文引入动态时间归整算法 DTW 来衡量时间序列的相似性。DTW 是一种常用的语音识别算法（例如识别语音签名），可以归整语音速度上的差异。但研究发现，将其跨领域应用到金融时间序列分析方面依旧有效，这也激励投资者可将更多方法尝试应用于金融领域。

1、引言

金融时间序列相较于其他领域的数据具有较为独特的特征，其中一个关键特征为：**金融理论及其时间序列中均包含了不确定性因素**。由于不确定性的存在，统计学上的理论和方法在金融时间序列分析中发挥着重要作用。股票投资者通常会通过多种方式来使用股票价格时间序列，例如通过预测未来价格走势以帮助并指导投资决策。常用的预测价格方法有两种：1) 识别数据中隐含模式，并进行分析以便预测未来趋势，常用模式包括对称三角形、上升三角形、下降三角形和楔形。此方法的本质在于很多人相信历史一定会重演。2) 使用技术分析来观察和学习价格走势，并构建模型以预测价格。本文主要使用第一种方法进行研究。**同时为了充分捕捉时间序列中的隐含模式，我们引入一种用于衡量两组序列之间相似性的算法**。欧式距离常被用于衡量时间序列之间的相似性，但这种方法也存在某些限制。因此，本文引入动态时间归整算法 DTW 来衡量时间序列的相似性。DTW 是一种常用的语音识别算法（例如识别语音签名），可以归整语音速度上的差异。同时，一些研究发现 DTW 算法在检索类似于金融时间序列数据的多属性时间序列方面也同样有效。

本研究的目的是构建一个基于 DTW 的模式匹配交易策略。该技术将当前时间序列的价格与一组固定的、定义好的模式相匹配（参见图表 2）。其次，系统可

请务必阅读正文之后的信息披露和重要声明

以提取价格数据中的代表性模式，并在最佳进入和退出点进行交易。

本文的其余章节如下：

首先，我们在第二章中详细说明本次研究中所使用的数据，并定义策略的时间窗口。

其次，在第三章中我们详细介绍了本文所用研究方法。具体的，我们回顾 DTW、模式匹配和与其相关的交易策略的相关文献，并梳理了策略构建流程与 ETF 投资组合的优化方法。

之后，我们在第四章中基于上述方法进行了相应实证，并进行总结。

最后，我们对 DTW 应用于模式匹配策略未来的可能方向以及该特定策略的应用方向进行展望。

2、数据收集和准备

2.1、数据源

本文将标普 500 指数板块 ETF 作为研究标的，具体包括 XLE、XLF、XLU、XLI、XLK、XLV、XLY、XLP 和 XLB，相对应地跟踪标准普尔 500 指数的能源、金融、公用事业、工业、信息技术、医疗保健、非必需消费品、必需消费和材料板块。由于数据上的缺失，本文研究标的不包括通信服务行业 ETF (XLC) (该 ETF 于 2018 年创建)。

ETF 交易方式与股票类似，本文中不考虑交易期权，只考虑多头和空头头寸。

2.2、数据预处理

我们收集了 2000 年 1 月 1 日至 2020 年 8 月 28 日的 ETF 价格，并使用 Min-Max 变换算法（参见图表 1）将数据转换至 -1 到 1 的范围。我们认为 Min-Max 变化可以使得价格和模式集更加接近，有利于之后的模式匹配。模式匹配具体细节将在后文叙述。

图表 1、Min-Max 变换算法

```
## Window Transform
def normalize (prices, upperbond = 1, lowerbond = -1):
    xmax = max (prices)
    xmin = min (prices)
    y = [ (upperbond - lowerbond) * (ele - xmin)/(xmax - xmin) + lowerbond
          for ele in prices]
    return y
```

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

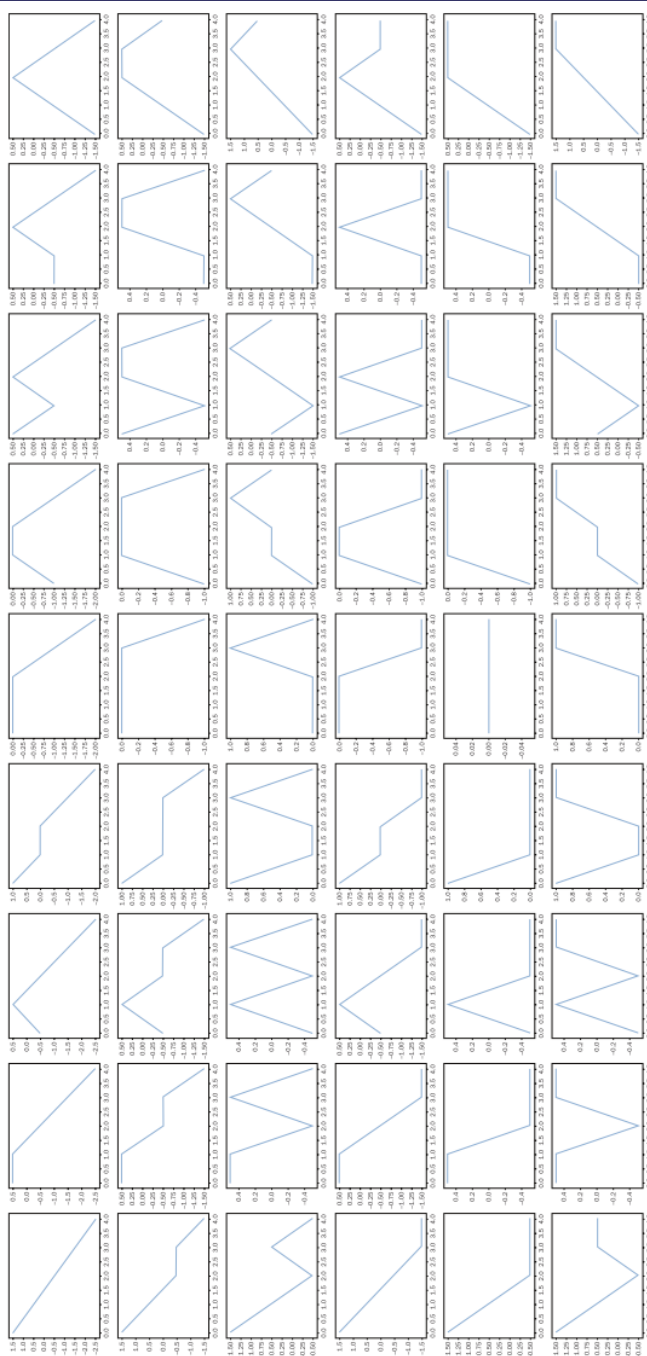
2.3、训练集、验证集和测试集

我们将原始数据分为三部分：训练集、验证集和测试集。训练集为前 14 年，从 2000 年 1 月 1 日到 2014 年 1 月 1 日。验证集为训练集后三年，即截至 2017 年 1 月 2 日。测试集则有两个版本：一个截至 2020 年 2 月 10 日，另一个截至 2020 年 8 月 28 日 (包含了 COVID-19 时期)。训练集用于得到模式匹配字典，验证集用

请务必阅读正文之后的信息披露和重要声明

于模型的验证，而测试集则用于测试样本外的表现。

图表 2、基于四节点的模式集



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

2.4、窗口切分

模式匹配交易系统需要一个“滑动窗口函数”将每个混合窗口（一组价格数据）分成两个较小的窗口：匹配窗口和决策窗口。“滑动窗口函数”将整个数据集划分为多个窗口，且窗口的数量取决于数据长度、步伐和窗口的长度。其中步伐是指每个窗口起点之间的时间长度；窗口长度是每个窗口的长度。

请务必阅读正文之后的信息披露和重要声明

3、研究方法

在本章节中，我们首先介绍动态时间序列归整算法 DTW，并提供两种与模式匹配相关的算法。其次，我们将解释如何使用 DTW 实现模式匹配交易策略，然后将这些结果用于构建最优投资组合。

3.1、动态时间序列归整 DTW

在获得多组时间序列数据后，我们通常会想要将一个序列与另一个序列进行比较。在某些领域，直接的距离度量就足够了，例如欧式距离。然而，大部分情况下两个序列彼此并不对齐（例如，跨时间）。

动态时间序列归整 DTW 是一种通过计算两个不同长度时间序列之间距离，以比较序列相似度的算法。这种算法提供了一种非线性（弹性）的对齐方式，可以由此获得更直观的相似性度量。这样，即使它们在时间轴上不尽相同，也能够检测出匹配的相似形状。该算法通过使用时间序列数据的弹性变换来识别在时序中不同模式之间的相似位置，从而最大限度地减少由于存在与时间有关的变化而导致的失真效应。

DTW 算法的目的是比较两个时间序列， $X = \{x_i\}_{i=1}^N$ 和 $Y = \{y_i\}_{i=1}^M$ （ N 不一定等于 M ），并计算它们之间的最小累积距离。DTW 在 $O(NM)$ 的时间复杂度内能够得到一个最优解，同时可以通过多尺度进一步改进。

算法首先构建一个距离矩阵 $C \in R^{N \times M}$ ，以表示两个序列 X 和 Y 之间的所有成对距离。这个矩阵也称为局部成本矩阵 C ，它是通过距离度量 $\|\cdot\|$ 定义的，如：

$$C_l \in R^{N \times M}: c_{ij} = \|x_i - y_j\|$$

生成距离矩阵后，该算法构建了汇整函数（来自两个序列两点之间对应关系），从而定义了 $x_i \in X$ 和 $y_j \in Y$ 之间的相似性。通过这个函数能够找到最小化距离的最佳点对应关系。严格地说，DTW 构建的对齐路径是一个点序列 $p = (p_1, p_2, \dots, p_K)$ ，其中 $p_l = (p_i, p_j) \in [1:N] \times [1:M]$ ，且 $l \in [1:K]$ ，同时该序列满足边界条件、单调性条件和步长条件。

与局部成本矩阵相对应的计算归整路径的代价函数被定义为：

$$C_p(X, Y) = \sum_{l=1}^L c(x_{nl}, y_{ml})$$

因此，与最小代价关联的归整路径称为最佳归整路径 P^* 。然而，确定 P^* 需要计算测试 X 和 Y 之间的每一条可能的归整路径，路径的数量级是 $O(NM)$ ，这样的计算方式是低效的。因此，我们采用动态规划的 Soft-DTW 算法：STW。

DTW 算法得到的距离可以被定义为路径的 P^* 的代价：

$$DTW(X, Y) = \min_{p \in P^{N \times M}} C_p(X, Y)$$

其中 $P^{N \times M}$ 代表所有可能的归整路径。那么，累积成本矩阵，或全局成本矩阵 D 为：

$$D_{1,j} = \sum_{k=1}^j c(x_1, y_k), \quad j = 1, \dots, M$$

$$D_{i,1} = \sum_{k=1}^i c(x_i, y_1), \quad i = 1, \dots, N$$

$$D_{i,j} = \min\{D_{i-1,j-1}, D_{i-1,j}, D_{i,j-1}\} + c(x_i, y_j)$$

图表 3 为用于计算累积成本矩阵和找出最佳归整路径的算法 1 和算法 2。

图表 3、DTW 中算法 1 与算法 2

Algorithm 1: Accumulated Cost Matrix (X, Y, C)

```

Initialization:
    • n ← |X|
    • m ← |Y|
    • dtw ← new[n × m]
    • dtw(0,0) ← 0
for i ← 1; i ≤ n; i++ do
    dtw(i, 1) ← dtw(i-1, 1) + c(i, 1)
for j ← 1; j ≤ m; j++ do
    dtw(1, j) ← dtw(1, j-1) + c(1, j)
for i ← 1; i ≤ n; i++ do
    for j ← 1; j ≤ m; j++ do
        dtw(i, j) ← c(i, j) + min{dtw(i-1, j), dtw(i, j-1), dtw(i-1, j-1)};
return: dtw

```

Algorithm 2: Optimal Warping Path (dtw)

```

Initialization:
    • path ← new array
    • i ← rows(dtw)
    • j ← columns(dtw)
while (i > 1) & (j > 1) do
    if i == 1 then
        j ← j - 1
    else if j == 1 then
        i ← i - 1
    else
        if
            dtw(i-1, j) == min{dtw(i-1, j), dtw(i, j-1), dtw(i-1, j-1)}
        then
            i ← i - 1
        else if
            dtw(i, j-1) == min{dtw(i-1, j), dtw(i, j-1), dtw(i-1, j-1)}
        then
            j ← j - 1
        else
            i ← i - 1; j ← j - 1
        path.add((i, j))
return: path

```

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

3.2、模式匹配交易策略

有了上一节的两个结果（时间窗口的定义和 DTW 算法），我们现在可以给出定义固定模式的方法并详细说明如何根据字典做出决策，从而构建策略。

3.2.1、固定模式

我们首先定义了一系列能够代表匹配期间价格趋势的模式，以用于后续的模式匹配交易系统。价格趋势的多样性取决于窗口内节点的数量。出于对时间复杂度的考虑，我们将节点的数量限制为四个，这意味着本文生成的模式均包含四个等长的时区。之后我们根据每个时区的情况分配三种趋势中的一种：持平、向上和向下。因此我们能够利用 $3^3 = 27$ 或 $3^4 = 81$ 种模式来进行训练（图表 2）。

3.2.2、决策训练

我们将整体价格数据分割成多个由匹配窗口和决策窗口组成的混合窗口后，使用 DTW 算法在每个匹配窗口上来寻找与其相似度最高的模式。为了避免出现某些匹配窗口没有特别相似模型，却被匹配至一种模式的情况，我们增加了相似度阈值以提高匹配的精度。我们将阈值设置为一个匹配窗口与根据 DTW 计算得到的它与所有模式的距离之和除以窗口长度。如果这个距离高于该阈值，则将此窗口分配给全平模式。

之后，通过比较每种模式，与它对应的匹配窗口的平均价格和相应的决策窗口的平均价格，计算这种固定模式未来上涨、下跌和持平次数。

请务必阅读正文之后的信息披露和重要声明

- 平均价格（决策窗口）> 平均价格（匹配窗口） \Rightarrow 上涨；
- 平均价格（决策窗口）< 平均价格（匹配窗口） \Rightarrow 下跌；
- 平均价格（决策窗口）= 平均价格（匹配窗口） \Rightarrow 持平；

如果系统将 100 个窗口中的 15 个识别为与第一种模式匹配，这 15 个窗口中 5 个被识别为上涨，另外 10 个被识别为下跌，则它会记录这第一个模式的比率为 $5/15 = 0.333$ 。我们在决策字典中保存这种模式与其对应的比率。但如果某些固定模式仅包含少于五个窗口（这些模式被识别的次数少于 5 次），则不会为这些模式计算比率，并将它们忽略。因此，我们的决策字典中的模式可以少于 81 个。

3.2.3、做出决策

我们首先在测试数据集上应用滑动窗口函数和模式匹配算法，根据每个窗口匹配的模式总比率做出交易决策。我们为决策阈值设置了两个参数：多头比率阈值和空头比率阈值。如果：

1) 当期匹配出的模式对应的固定模型比率大于我们设置的多头比率阈值，系统会在匹配期的最后一个收盘日前持有多头头寸；反之若小于设置的空头比率阈值，将持有空头头寸。

2) 如果当期的比率没有超过空头或多头比率阈值，或者如果该模式的匹配次数少于五次，系统将不执行任何操作。

由于空头头寸本质上比多头头寸风险更大，我们将当前现金的 30% 分配给多头交易，将 10% 分配给空头交易。但是，为了避免由于日度交易频率而导致资金不足的情况，我们设置了一个交易阈值，这意味着如果需要进行交易的金额超过我们实时资产净值的 10%，我们选择不开仓。

在此我们举例说明交易机制，假设系统将某一天的窗口与第一种模式匹配，并且当前持有的现金为 100000 美元。如果决策字典中记录的该模式比率大于多头比率阈值，系统将持有多头头寸，分配资本为 30000 美元。相反，如果记录的比率小于空头比率阈值，它将持有一个空头头寸，分配资本为 10000 美元，并在窗口的末端平仓。

3.2.4、固定模式匹配 VS 动态模式匹配

我们在本次研究中测试了两个模型：**固定模式匹配**和**动态模式匹配**。固定模式匹配交易系统使用整个训练集构建决策字典（即固定时间窗口和相应的总比率）。相反，动态策略仅使用最近的 200 或 252 日来构建字典。其中动态模式可以更充分、更快速的对最近结构变化进行反应。

3.2.5、模型检验

在训练得到决策字典后，我们将在验证集上测试交易策略，以找到每个标准普尔 500 指数行业相应 ETF 的最佳参数和资产净值。最后，系统会根据投资组合优化方法计算每个 ETF 的最优权重。

3.2.6、投资组合优化

为每个资产适当权衡或选择最优配置以最小化方差的过程通常被称为现代投

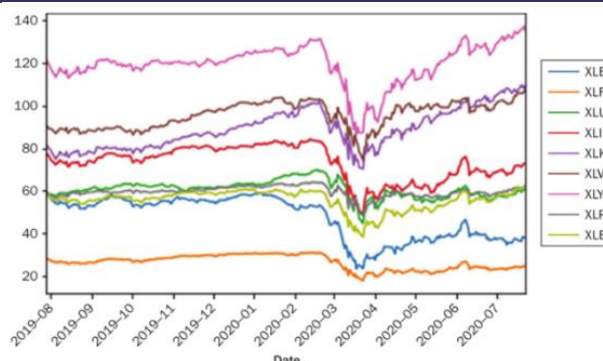
投资组合理论 (MPT)。该过程的目标通常是为了得到预期的回报、最小化风险、最大化夏普比率，或者最小化波动。我们使用 Python 包 *Pypfopt* 来计算最优权重。

系统会根据验证集期间产生的资产净值，优化由策略选择的 ETF 组成的投资组合。它将计算得出最优权重以最小化波动。然后，系统将根据产生的最优权重将初始资本分配给每个组成部分。由于时间复杂度问题以及验证期和测试期之间的连续性，我们将只执行一次优化而不是在测试集中动态优化投资组合。

4、实证结果

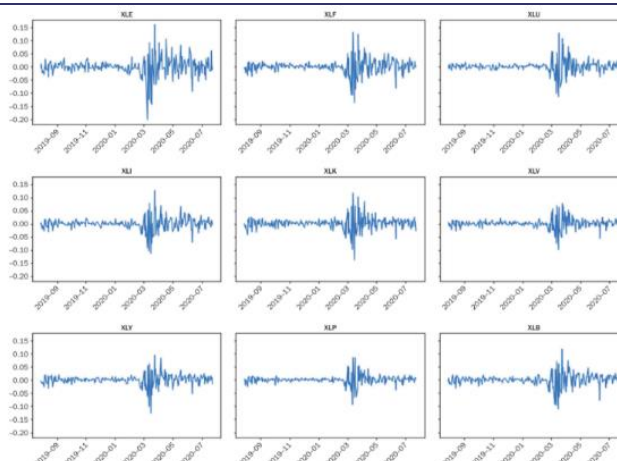
图表 4 展示了标准普尔 500 指数行业 ETF 的近期历史价格。每只 ETF 代表一个行业（例如，能源行业对应 XLE），可以看出不同行业对 COVID-19 的反应是不同的。我们还绘制了最近 250 日标准普尔 500 指数板块 ETF 收益图（图表 5）以确定 COVID-19 的相对影响。值得注意的是，能源行业(XLE)的差异最大，而其他一些行业，包括医疗保健(XLV)和必需消费品(XLP)则没有受到严重影响。

图表 4、最近 250 日标准普尔 500 指数行业 ETF 价格图



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、最近 250 日标准普尔 500 指数板块 ETF 收益图



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

4.1、初测试

我们首先使用了较短的训练周期和只有三个节点的模式集进行初测试。

请务必阅读正文之后的信息披露和重要声明

1) 固定模式匹配策略

图表 6 中展示了固定模式匹配策略结果。我们采用以下指标衡量表现：年化收益率、波动率、夏普比率、偏度、峰度、风险价值(VaR)和最大回撤(MDD)，并在验证集中优化参数（即节点数和窗口长度）以最小化波动。图表上半部分展示了每个 ETF 和基准标准普尔的策略在 COVID-19 之前的结果，下半部分则展示 COVID-19 之后的结果。

正如和预期到的市场崩盘一样，受 COVID-19 影响后的表现明显比影响前的表现差。但是，由于某些行业较少地受到 COVID-19 的影响，因此夏普比率的差异并没有那么大。以科技板块的 ETF(XLK)为例，它的夏普比率下跌了 0.21，而金融板块(XLF)则下跌了 0.67。

图表 6、固定模式匹配策略的结果展示

Stats	Fixed Pattern Matching									
	SP500	XLE	XLF	XLX	XLI	XLK	XLV	XLJ	XLP	XLB
Without COVID-19 Period										
Annualized Return	0.111	-0.0432	0.0975	0.153	0.0349	0.1645	0.1039	0.0803	0.1278	-0.0618
Volatility	0.142	0.2052	0.1175	0.1061	0.1305	0.1687	0.1041	0.1427	0.1006	0.2214
Sharpe	0.7823	-0.2103	0.8296	1.4428	0.2673	0.9751	0.9984	0.5628	1.2707	-0.2791
Skewness	-0.3934	1.0871	-0.3719	-0.6043	-0.6712	-0.585	-0.559	-0.0497	-0.4139	-0.2657
Kurtosis	4.1943	14.1517	2.5663	2.6569	4.2008	2.1832	2.2914	2.8434	1.2983	6.5658
VaR	45.9025	1,717.217	1,177.148	1,226.058	1,444.500	1,680.697	1,105.555	1,600.183	1,093.325	1,855.745
MDD	0.2744	0.3088	0.2191	0.2176	0.1967	0.3028	0.1594	0.2302	0.2013	0.273
Parameters	XLE	XLF	XLX	XLI	XLK	XLV	XLJ	XLP	XLB	
Nodes	3	3	3	3	3	3	3	3	3	3
Window Length	23	20	44	23	32	17	38	38	38	38
With COVID-19 Period										
Annualized Return	0.0735	-0.2357	-0.0302	0.084	-0.0075	0.2089	0.0773	0.0682	0.0929	0.0085
Volatility	0.2631	0.3785	0.202	0.2416	0.2612	0.2742	0.2081	0.2274	0.1746	0.2513
Sharpe	0.2794	-0.6228	-0.1497	0.3476	-0.0285	0.7616	0.3714	0.2997	0.5322	0.034
Skewness	-0.952	-2.6255	-0.661	-0.2314	-0.4737	-0.6124	-0.3789	-1.0418	0.2116	-0.6794
Kurtosis	14.7161	29.2188	12.515	16.2582	11.0773	10.6778	8.9983	11.9045	9.9564	10.0765
VaR	67.22	2,331.141	1,736.078	2,055.870	2,276.406	2,985.515	1,985.740	2,111.253	1,755.911	2,065.936
MDD	0.3392	0.5985	0.2807	0.3162	0.3697	0.4268	0.2459	0.3033	0.2206	0.3233
Parameters	XLE	XLF	XLX	XLI	XLK	XLV	XLJ	XLP	XLB	
Nodes	3	3	3	3	3	3	3	3	3	3
Window Length	38	11	38	38	38	38	35	38	38	38
Without COVID-19 Period										
Annualized Return	0.111	-0.0432	0.0975	0.153	0.0349	0.1645	0.1039	0.0803	0.1278	-0.0618
Volatility	0.142	0.2052	0.1175	0.1061	0.1305	0.1687	0.1041	0.1427	0.1006	0.2214
Sharpe	0.7823	-0.2103	0.8296	1.4428	0.2673	0.9751	0.9984	0.5628	1.2707	-0.2791
Skewness	-0.3934	1.0871	-0.3719	-0.6043	-0.6712	-0.585	-0.559	-0.0497	-0.4139	-0.2657
Kurtosis	4.1943	14.1517	2.5663	2.6569	4.2008	2.1832	2.2914	2.8434	1.2983	6.5658
VaR	45.9025	1,717.217	1,177.148	1,226.058	1,444.500	1,680.697	1,105.555	1,600.183	1,093.325	1,855.745
MDD	0.2744	0.3088	0.2191	0.2176	0.1967	0.3028	0.1594	0.2302	0.2013	0.273
Parameters	XLE	XLF	XLX	XLI	XLK	XLV	XLJ	XLP	XLB	
Nodes	3	3	3	3	3	3	3	3	3	3
Window Length	23	20	44	23	32	17	38	38	38	38
With COVID-19 Period										
Annualized Return	0.0735	-0.2357	-0.0302	0.084	-0.0075	0.2089	0.0773	0.0682	0.0929	0.0085
Volatility	0.2631	0.3785	0.202	0.2416	0.2612	0.2742	0.2081	0.2274	0.1746	0.2513
Sharpe	0.2794	-0.6228	-0.1497	0.3476	-0.0285	0.7616	0.3714	0.2997	0.5322	0.034
Skewness	-0.952	-2.6255	-0.661	-0.2314	-0.4737	-0.6124	-0.3789	-1.0418	0.2116	-0.6794
Kurtosis	14.7161	29.2188	12.515	16.2582	11.0773	10.6778	8.9983	11.9045	9.9564	10.0765
VaR	67.22	2,331.141	1,736.078	2,055.870	2,276.406	2,985.515	1,985.740	2,111.253	1,755.911	2,065.936
MDD	0.3392	0.5985	0.2807	0.3162	0.3697	0.4268	0.2459	0.3033	0.2206	0.3233
Parameters	XLE	XLF	XLX	XLI	XLK	XLV	XLJ	XLP	XLB	
Nodes	3	3	3	3	3	3	3	3	3	3
Window Length	38	11	38	38	38	38	35	38	38	38

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

2) 动态模式匹配策略

动态策略的目的是仅使用最新数据计算得到的决策字典，以捕获任何最近的结构变化。因此，我们预计动态策略在 COVID-19 影响后的平均表现要优于相对应的固定策略。图表 7 给出了动态策略在 COVID-19 之前和之后的结果。

对于 XLE、XLF、XLI 和 XLV 而言，该策略在 COVID-19 之前的夏普比率更高。然而，动态策略的所有夏普比率都是正的，而固定策略的方差则更大。同样，对于 XLE、XLF 和 XLV，COVID-19 后的夏普比率更高了。有趣的是，动态策略在疫情影响后方差更大，且平均表现并不比固定策略好。这可能是因为使用的决策字典时间窗口（200 日）的长度设置问题。这个长度导致更新速度太慢，而无法准确适应最新的结构变化。然而，窗口太小又会导致夏普比率减小。

图表 7、动态模式匹配策略的结果展示

Dynamic Pattern Matching									
Stats	SP500	XLE	XLF	XLU	XLI	XLK	XLV	XLX	XLB
Without COVID-19 Period									
Annualized Return	0.111	0.1755	0.1352	0.0146	0.0516	0.0387	2.0666	0.0234	0.0734
Volatility	0.142	0.207	0.1047	0.35	0.0957	0.1527	10.498	0.0848	0.1297
Sharpe	0.7823	0.8779	1.2915	0.0417	0.539	0.2534	1.9689	0.2761	0.2089
Skewness	-0.3934	2.9101	0.7556	0.4927	-1.0766	-0.7727	4.7133	-1.4137	-0.1751
Kurtosis	4.1943	24.637	4.5392	47.5818	4.8695	5.0468	83.7353	7.6546	3.5979
VaR	45.9025	1,691,066	1,098,090	1,391,777	1,034,698	1493,716	10,344,734	830,055	1,030,335
MDD	0.2744	0.4106	0.2235	0.3794	0.1614	0.2574	2.7773	0.1331	0.1802
Parameters		XLE	XLF	XLU	XLI	XLK	XLV	XLX	XLB
Nodes		3	3	3	3	3	3	3	3
Window Length		23	20	35	17	29	23	20	38
With COVID-19 Period									
Annualized Return	0.0735	0.0216	0.0142	-0.0466	-0.0153	0.0785	0.1742	-0.4953	-0.0109
Volatility	0.2631	0.1737	0.7789	0.3842	0.1848	0.1777	0.1803	6.1078	0.1225
Sharpe	0.2794	0.1245	0.0182	-0.1212	-0.0829	0.442	0.9666	-0.0811	-0.0891
Skewness	-0.952	-1.3263	0.5038	4.5302	-0.6548	-0.1967	0.5476	-15.1809	0.1631
Kurtosis	14.7161	32.2239	9.9956	86.7388	12.3117	21.2782	24.6287	321.9376	10.1511
VaR	67.22	1,275,109	5,157,895	1,894,221	1,615,600	1,662,607	1,569,015	2,653,783	1173,650
MDD	0.3392	0.2813	0.6147	0.4041	0.2162	0.2477	0.2792	1.1824	0.1598
Parameters		XLE	XLF	XLU	XLI	XLK	XLV	XLX	XLB
Nodes		3	3	3	3	3	3	3	3
Window Length		11	38	38	17	11	14	23	23

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

3) 模式匹配策略

基于 200 日和 252 日的 ETF 动态模式匹配策略，我们可以构建一个优化的均值-方差投资组合。权重仅限于多头头寸。

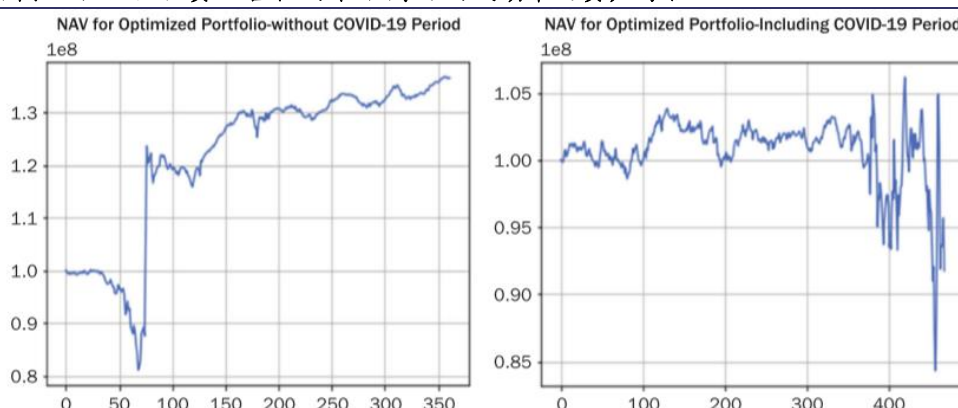
图表 8 总结了通过优化夏普比率而计算出的权重。使用这个优化投资组合，我们使用相同的统计数据以衡量 200 日的动态模式匹配策略。图表 9 中绘制了我们的最优投资组合在 COVID-19 时期前后的净值曲线，以及图表 10 和 11 中我们展示了两个时期按行业划分的资产净值图。

图表 8、200 天和 252 天动态投资组合的优化后权重

252-Day	XLE	XLF	XLU	XLI	XLK	XLV	XLX	XLZ	XLB
Weight	14.45%	15.10%	18.66%	1.09%	11.75%	20.10%	18.85%	0.00%	0.00%
200-Day	XLE	XLF	XLU	XLI	XLK	XLV	XLX	XLZ	XLB
Weight	9.58%	6.82%	13.22%	12.59%	19.36%	6.72%	17.02%	7.11%	7.59%

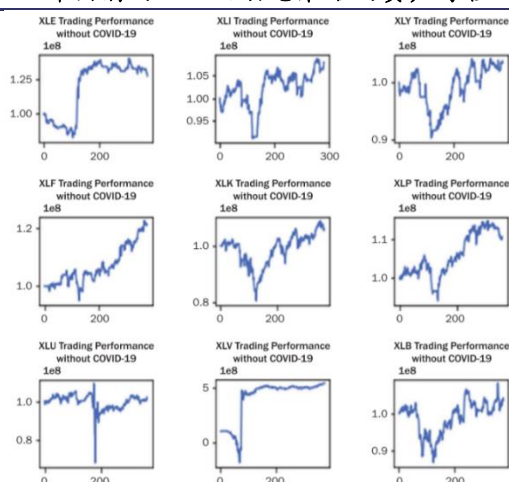
资料来源：The Journal of Financial Data Science，兴业证券经济与金融研究院整理

图表 9、优化投资组合在两个不同的测试期中的资产净值



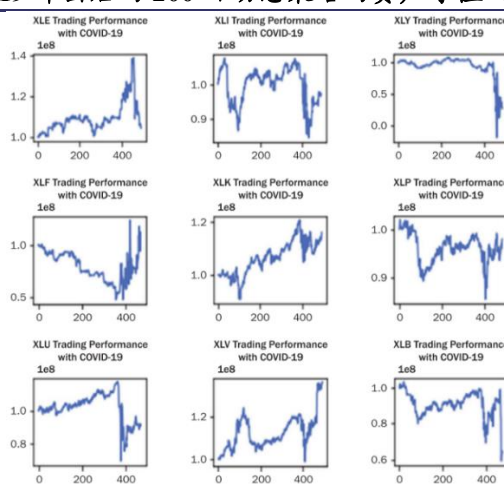
资料来源：The Journal of Financial Data Science，兴业证券经济与金融研究院整理

图表 10、COVID-19 冲击前的 200 日动态策略的资产净值（按行业划分）



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 11、COVID-19 冲击后的 200 日动态策略的资产净值（按行业划分）



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 12 中给出了统计数据。最终的优化后投资组合在 COVID-19 之前表现良好，夏普比率为 1.028。

值得注意的是，该投资组合的表现并不是很具有吸引力。我们可以在图表 12 中观察到一些明显的问题。首先，在 COVID-19 后投资组合的回报为负数，最大回撤为 20%。其次，尽管 COVID-19 之前的投资组合产生正回报，但波动率很大。此外，投资组合的最大回撤为 40%。这表明我们需要做更多的研究来进一步优化其表现。

图表 12、200 日动态模式匹配策略投资组合表现统计

Stats	
With COVID-19	
Annualized Return	-4.5%
Volatility	19.6%
Sharpe	-0.231
Skewness	1.6843
Kurtosis	36.5143
VaR	1,396,285
MDD	0.206
Without COVID-19	
Annualized Return	24.2%
Volatility	23.6%
Sharpe	1.028
Skewness	8.9683
Kurtosis	108.1863
VaR	959,980
MDD	0.407

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

4.2、改进后的模式匹配策略

由于初测试采用动态模式匹配方法构建策略，因此一种显而易见的改进是使用固定模式匹配策略。动态策略仅考虑最近一段时间（即 200 或 252 日）的模式，固定模式可以增加更多的数据样本。同时初测试是在只有三个节点的模式集上训练的。我们在此使用有四个节点的模式集合（即总共 34 个模式；参见图表 13）。

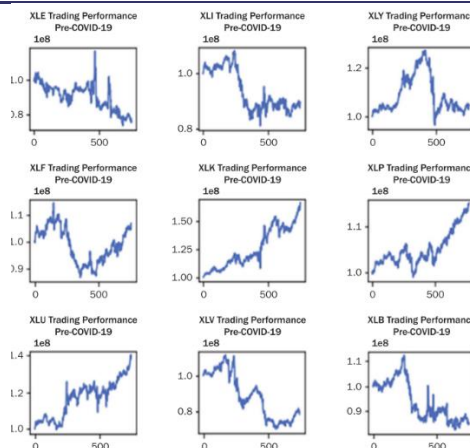
图表 13、改进后的模式匹配策略表现结果

Stats		Sector ETF Testing Result and Optimization Weights									
		SP500	XLE	XLF	XLU	XLI	XLK	XLV	XLX	XLZ	XLB
Pre-COVID-19 January 3, 2017-February 10, 2020											
Annualized Return	13.70%	-8.71%	2.14%	11.98%	-3.65%	18.31%	-8.03%	1.56%	4.71%	-5.20%	
Volatility	0.1276	0.2302	0.1116	0.1356	0.1274	0.1574	0.1359	0.1118	0.0485	0.1353	
Sharpe	1.074	-0.3784	0.1914	0.8834	-0.2862	1.1629	-0.5905	0.1395	0.9704	-0.3842	
Skewness	-0.7088	-0.5314	-0.534	0.1695	-0.2891	0.605	-1.4779	-2.3794	-0.9396	-1.1147	
Kurtosis	5.3647	9.8308	6.5599	12.4276	7.4472	20.0827	10.0922	22.5103	8.7183	11.5051	
VaR	37.5069	1,909.77	1,034.45	1,347.24	1,181.00	1,633.34	1,251.02	1,088.75	454.104	1,270.73	
Max Drawdown	0.3332	0.3724	0.2398	0.2912	0.2534	0.3961	0.3754	0.2431	0.1424	0.263	
Post-COVID-19 January 3, 2017-August 28, 2020											
Annualized Return	12.81%	3.01%	0.10%	3.97%	1.12%	5.75%	-12.29%	6.11%	3.46%	-5.45%	
Volatility	0.2078	0.5221	0.2096	0.1974	0.2439	0.336	0.1836	0.1809	0.0696	0.1599	
Sharpe	0.6166	0.0577	0.0047	0.2012	0.046	0.171	-0.6694	0.3376	0.4973	-0.3406	
Skewness	-1.1726	-0.2948	-1.5086	-0.1945	-1.754	-4.0251	-1.0638	-1.0784	-0.0604	-1.6398	
Kurtosis	22.2189	40.1426	30.9161	13.4083	55.1764	66.5949	11.3759	59.6102	15.5014	13.7381	
VaR	51.4086	2,629.35	1,658.15	2,144.35	1,415.52	2,241.97	1,316.03	1,244.59	589.215	1,337.84	
Max Drawdown	0.3622	0.7319	0.3033	0.2994	0.3685	0.4834	0.4979	0.3614	0.137	0.3805	

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

以下是固定模式匹配策略的汇总结果，包括 COVID-19 前后以及标普基准。图表 14 和 15 展示出了按行业划分的资产净值图。

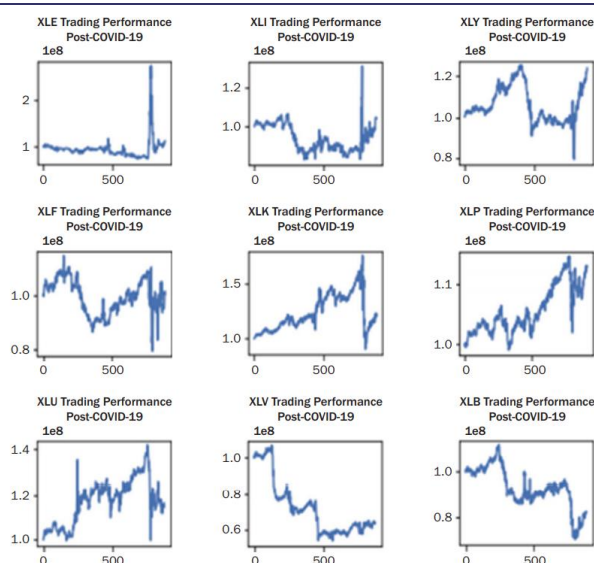
图表 14、COVID-19 冲击前的改进后策略的资产净值（按行业划分）



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

请务必阅读正文之后的信息披露和重要声明

图表 15、COVID-19 冲击后的改进后策略的资产净值（按行业划分）



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

结合图表 13 的结果和图表 16 的最优权重，我们可以分析在有四个节点的更大的模式集上训练的策略表现。

图表 16、最优参数

Parameters	Window_len	Weights
XLE	33	0.01
XLF	33	0.01752
XLU	44	0.14521
XLI	33	0.3892
XLK	33	0.15391
XLV	44	0.05559
XLY	33	0.06708
XLP	11	0.5
XLB	33	0.01178

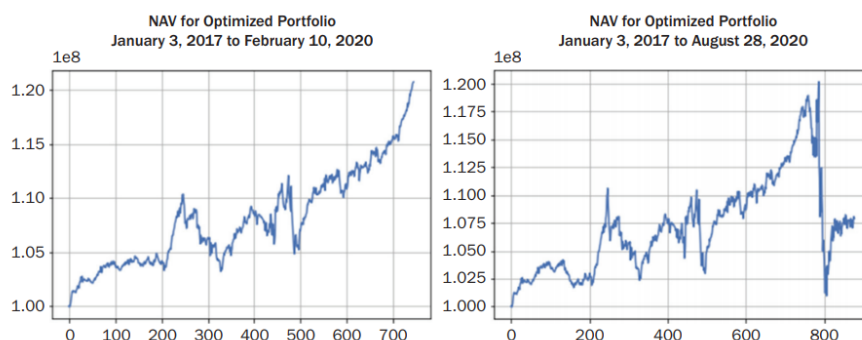
资料来源：The Journal of Financial Data Science，兴业证券经济与金融研究院整理

改进后的投资组合给出的结果优于之前的表现。虽然投资组合的收益比之前的收益小，但波动明显更小，从而提升了夏普比率。VaR 也从大约的一百万减少为一半。最大回撤从 41% 下降到改进后的 17%。同时，统计数据上更高的矩也说明其更具有吸引力：偏度从 9 左右下降到 0.04，峰度从 108 下降到 8.6。

尽管 COVID-19 后的表现并不那么诱人，但也相对更好。之前的投资组合在 COVID-19 后亏损，而这个投资组合则产生了正收益。波动也显著下降，从约 20% 降至 8%。这也自然会得出了更好的夏普比率，从 -0.23 提升到 +0.26。与 COVID-19 冲击之前的结果一样，在使用了较大的模式集后，最大回撤减少了几近一半；VaR 也从 140 万下降到 70 万。

整体而言，我们可以观察到改进后策略的表现优于改进前。这意味着，尽管当前的策略可能不会跑赢基准标准普尔，但我们能在下一节中讨论，目前的策略仍有改进的潜力。

图表 17、COVID-19 冲击前后的最优化策略的资产净值（按行业划分）



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 18、改进后的模式匹配策略表现

Stats	Portfolio
Optimized Portfolio Pre-COVID-19	
Annualized Return	6.57%
Volatility	0.0506
Sharpe	1.2979
Skew	0.0358
Kurt	8.6479
VaR	527,053.3618
MDD	0.172
Optimized Portfolio Post-COVID-19	
Annualized Return	2.21%
Volatility	0.0846
Sharpe	0.2613
Skew	-1.8855
Kurt	37.2994
VaR	720,197.5437
MDD	0.1684

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

4.3、总结

DTW 算法在金融时间序列数据上的应用使我们能够捕捉潜在趋势并可以对比我们构建的一系列模式集来找出最匹配的一种模式。使用匹配模式，我们可以计算比率并根据波动的相对频率来做出决策。

我们首先构建了两种策略：固定模式匹配策略和动态模式匹配策略。固定策略用整个训练集上训练得出决策字典。相比之下，动态字典每天都会重新计算字典，从今天起回溯 200 日以捕捉最新的市场变化，并且只考虑最近的数据。在应用了固定模式匹配策略和动态策略来获得我们的决策字典之后，我们使用标准普尔 500 指数行业 ETF 的数据，并且优化每个 ETF 的策略参数以最小化波动。

然而，初始测试的表现（参见图表 12）并不是特别令人满意。因此，我们通过扩展数据集（包含 2008 年市场崩盘）并使用固定模式训练字典以改进策略。这个微小的改进却能够展示了这个策略的灵活性：所有的统计指标都有所提高。这意味着该技术可以与现有策略结合使用，以获得更具吸引力的结果。

本研究的分析表明，DTW 在标准普尔 500 指数行业 ETF 上的应用潜力巨大。本文中构建的模式匹配策略在某些行业表现良好。基于这些策略的马科维茨均值-方差优化投资组合能够获得正回报。通过使用这些策略我们发现，具有重要趋势的时间序列是在交易中是最有作用的。同时，我们可以通过修改交易决策以捕捉任何板块或行业的特定趋势。

请务必阅读正文之后的信息披露和重要声明

5、未来研究与改进方向

目前，我们想到几种方法可以改进现在的策略。例如，我们可以考虑一个简单的参数化决策策略，该策略假设价格过程遵循几何布朗运动并由此计算出最准确的漂移和波动率。然后我们可以基于这些估计值做出决策。同时强化学习算法也可能可以优化当前的 DTW 模式匹配策略。更具体地说，状态空间可以由匹配模式、DTW 距离、价格和其他定量因素（例如移动平均线、动量、波动率等）来定义。

除此之外，我们也给出三个具体的改进方向：

首先，出于对时间复杂性的考量，当前策略仅根据移动平均价格是否上涨来做出交易决策。然而，人们可以期望基通过更广泛的信号做出决策，并使得交易系统表现更好。例如，该系统可以应用于不同的定量信号，如布林带、动量、变化率或波动率，以优化表现。

其次，由于数据的可用性有限，除了计算要求外，该算法还可以使用日内数据。有关价格或其他信号的日内数据应该有助于提升模式匹配策略的准确性。

第三，该策略目前仅根据窗口长度验证表现，如“分割时间窗口”部分所述。我们认为，增加验证参数的复杂性可能会提升表现。

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。

参考文献

- 【1】 Das, G., D. Gunopulos, and H. Mannila. “Finding Similar Time Series.” In European Symposium on Principles of Data Mining and Knowledge Discovery. Berlin, Heidelberg: Springer, 1997. Ge, X. “Pattern Matching in Financial Time Series Data.” Final project report for ICS 278, 1998.
- 【2】 Keogh, E. J., and M. J. Pazzani. “Derivative Dynamic Time Warping.” In Proceedings of the 2001 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics. Philadelphia, PA: Siam, 2001.
- 【3】 Kim, S. H. et al. 2018. “Pattern Matching Trading System Based on the Dynamic Time Warping Algorithm.” Sustainability 10 (12): 4641.
- 【4】 Pierre-Rouanet. “DTW–Dynamic Time Warping.” GitHub Repository, 2019, <https://github.com/pierre-rouanet/dtw>.
- 【5】 Senin, P. “Dynamic Time Warping Algorithm Review.” Information and Computer Science Department, University of Hawaii at Manoa Honolulu, 855.1-23: 40, 2008. 【14】 Chan, L. K., J.
- 【6】 Karceski, and J. Lakonishok. 2000. “New Paradigm or Same Old Hype in Equity Investing?” Financial Analysts Journal 56 (4): 23–36.

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