

金融科技赋能投研系列之十二：

智 AI 科技 慧投未来（下）

在年报的上篇中，我们已经着重论述了商品因子体系的构建逻辑和必要的测试方法论。在本篇中，我们将重点说明更多技术细节，输出完整的测试结果，并将不同模型结果做相应对比。

投资咨询业务资格：

证监许可【2011】1289 号

研究院 量化组

陈辰

☎ 0755-23887993

✉ chenchen@htfc.com

从业资格号：F3024056

投资咨询号：Z0014257

何绪纲

☎ 0755-23887993

✉ hexugang@htfc.com

从业资格号：F3069194

高天越

☎ 0755-23887993

✉ gaotianyue@htfc.com

从业资格号：F3055799

镇谔博

☎ 0755-23887993

✉ zhenchenbo@htfc.com

从业资格号：F3080231

目录

一、商品因子应用	4
二、商品因子表现	5
三、AI 模型的应用	18
四、AI 模型预测能力探索	27
五、附录	31

图表目录

图 1: 商品因子的应用图景	5
图 2: 时序基础因子（可投资的）累计收益率	6
图 3: 展期收益率的计算原理（原则上贴水越深展期收益率越大）	7
图 4: 价值因子计算示例	8
图 5: 各块累计收益率表现	9
图 6: 2020 年板块收益率以及月均持仓金额增长情况	9
图 7: 2020 年价值因子正向暴露板块的权重	9
图 8: 2020 年价值因子反向暴露板块的权重	9
图 9: 截面因子（不可投资的）累计收益率	10
图 10: 2020 年 Curve 因子正向暴露板块的权重	10
图 11: 2020 年对 Curve 因子反向暴露板块的权重	10
图 12: 2020 年期限结构因子正向暴露板块的权重	11
图 13: 2020 年期限结构因子反向暴露板块的权重	11
图 14: 2020 年动量因子正向暴露板块的权重	11
图 15: 2020 年动量因子反向暴露板块的权重	11
图 16: 2020 年波动率因子正向暴露板块的权重	12
图 17: 2020 年波动率因子反向暴露板块的权重	12
图 18: 2020 年持仓因子正向暴露板块的权重	12
图 19: 2020 年持仓因子反向暴露板块的权重	12
图 20: 2020 年偏度因子正向暴露板块的权重	13
图 21: 2020 年偏度因子反向暴露板块的权重	13
图 22: 2020 年流动性因子正向暴露板块的权重	13
图 23: 2020 年流动性因子反向暴露板块的权重	13

图 24:	2020 年均价突破因子正向暴露板块的权重	14
图 25:	2020 年均价突破因子反向暴露板块的权重	14
图 26:	2020 年美元指数 Beta 因子正向暴露板块的权重	14
图 27:	2020 年美元指数 Beta 因子反向暴露板块的权重	14
图 28:	2020 年对美元指数因子暴露较大的品种	15
图 29:	2020 年对中国 CPI 因子暴露较大的品种	15
图 30:	2020 年 CPI Beta 因子正向暴露板块的权重	15
图 31:	2020 年 CPI Beta 因子反向暴露板块的权重	15
图 32:	时序基础风格因子收益率相关系数	16
图 33:	截面风格因子收益率相关系数	16
图 34:	全部截面因子之间保持较低相关性	17
表格 1:	华泰商品, 股票因子组的构成	17
图 35:	沪深 300 样本截面因子解释力	18
图 36:	商品全市场样本截面因子解释力	18
表格 2:	原始数据因子重要性排名及对应指标值 (回归分析)	20
表格 3:	原始数据因子重要性排名及对应指标值 (分类)	21
表格 4:	短周期数据因子重要性排名及对应指标值 (回归)	22
表格 5:	短周期数据因子重要性排名及对应指标值 (分类)	23
表格 6:	长周期数据因子重要性排名及对应指标值 (回归)	24
表格 7:	长周期数据因子重要性排名及对应指标值 (分类)	25
图 37:	不同周期因子 MSE 排名 (分类)	26
图 38:	不同周期因子 RSS 排名 (分类)	26
图 39:	不同周期因子最小分裂节点排名 (分类)	26
表格 8:	随机森林预测 T+1 周表现 (回归)	28
表格 9:	随机森林预测 T+1 周表现总结 (回归)	28
表格 10:	随机森林预测 T+2 周表现 (回归)	29
表格 11:	随机森林预测 T+2 周表现总结 (回归)	29
表格 12:	深度神经网络多步预测胜率(%)结果	30
表格 13:	华泰商品, 股票因子代码附录	31

一、商品因子应用

我们在年报上篇中详细地介绍了基础因子和截面因子这两个不同的概念，而这两套因子也有着不同的应用场景：

1) 时序基础因子（可以直接投资）

我们已经在去年推出了可以复制的动量、期限结构两套风格因子指数。指数的历史表现是可以在 wind 上直接查询到，它们在今年均有相当可观的收益表现，策略可以复现。附 Wind 代码：

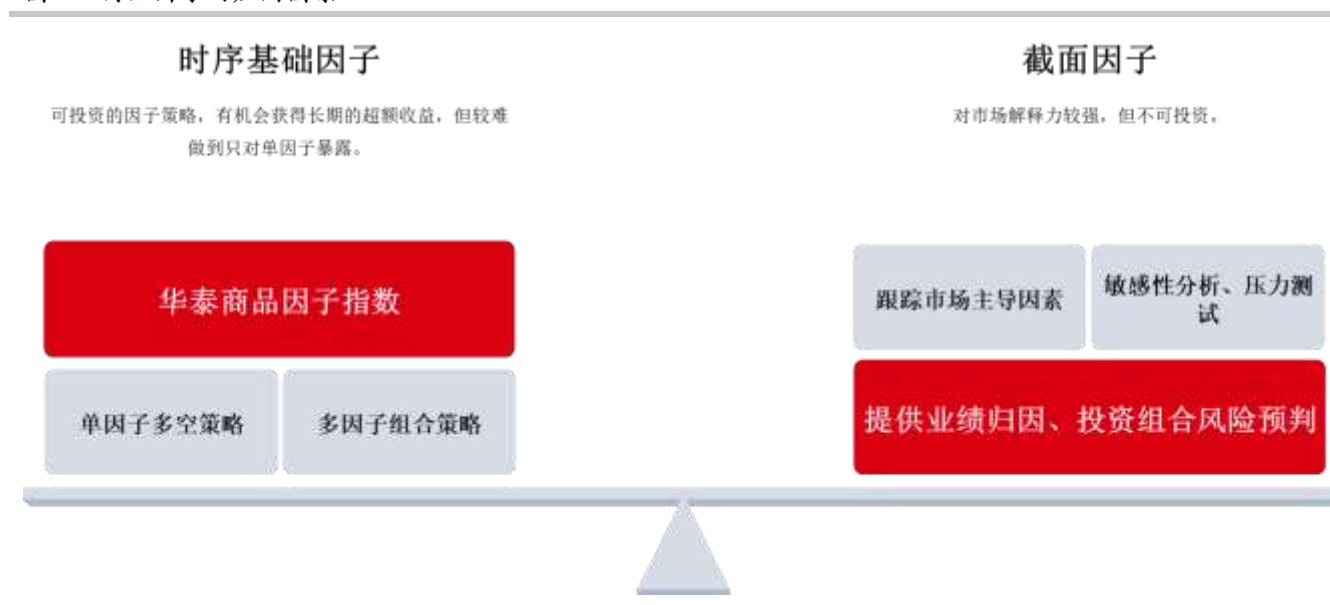
华泰商品动量 HTCI0101.WI	华泰农产品动量 HTCI0102.WI
华泰工业品动量 HTCI0103.WI	华泰能化动量 HTCI0104.WI
华泰商品期限结构 HTCI0201.WI	华泰农产品期限结构 HTCI0202.WI
华泰工业品期限结构 HTCI0203.WI	华泰能化期限结构 HTCI0204.WI

同时因子对中高频交易也具有指导作用，通过日度更新各个期货品种的风格暴露数据，每日给出时序基础因子策略的多空策略持仓，可以对投资选取商品品种提供支持；而风格因子收益率则均可达到即时跟踪。同时在基础因子的基础上，也支持我们对多因子策略、因子择时的交易策略进行深入研究，发掘更有效的多因子策略。

2) 截面因子

截面因子在计算风险矩阵、跟踪收益来源上有天然的优势。运用截面因子可以提供全市场、各个板块以及不同品种主导因素的跟踪。尽管截面因子本身不可投资，但是因为其底层的纯因子（pure factor）设计，我们可以捕捉到全市场内相互独立的风格收益来源，也可以对已有投资组合进行准确的业绩归因和组合风险预判。同时截面因子还可以为敏感性分析、压力测试等风险管理工具提供数据基础。需要特别指出，这些更高层次的精细化商品投资模式在离开截面多因子模型的情况下，都是难以实现的。

图1: 商品因子的应用图景



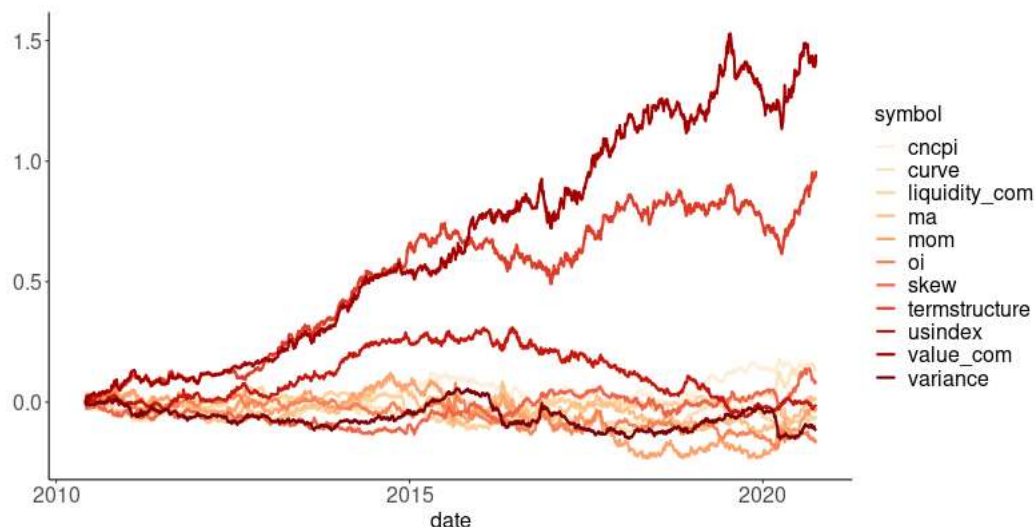
数据来源: 华泰期货研究院

回到算法开发角度, 人工智能算法是典型的非线性方法。其应用到投资领域的切入点, 并非局限于最近历史的行情表现, 做简单线性外推; 而是通过学习历史上最接近于目前市场场景时, 不同市场因素对标的物的复杂价格传导模式, 各种风格影响的强弱切换等特征, 来对当前市场标的物的未来走势做出预判。而数据是模型成功的关键。截面因子组收益率分解了市场系统性风险敞口, 其中包含的信息都具有坚实的经济学基础, 这样高质量的数据显然可以帮助算法更快的学习到更有用的信息, 从而有机会对市场预测达到更高的准确率。本文将在后面章节展示人工智能算法对商品板块重要因子影响力判断以及周度收益率预测上的表现。

二、商品因子表现

本章我们结合商品市场, 多角度对比时序基础因子与截面因子的表现。时序基础因子具有可投资性, 而在 11 个因子中, 价值因子与期限结构因子表现最好, 具有长期稳定的正向收益。

图2: 时序基础因子(可投资的)累计收益率



数据来源: Wind 天软 华泰期货研究院

为保持论述完整性, 我们不妨重新审视期限结构因子与价值因子的编制方式:

期限结构因子:

现货与期货的价格差异, 即基差 (basis), 与现货的库存水平、持有成本 (包括交通、仓储以及保险费用等)、市场利率 (购买现货的融资成本) 以及持有现货的便利收益 (convenience yield) 等因素相关。假设基差水平全部反应了库存、持有成本、利率及便利收益等因素; 升水市场中, 商品供给充足, 持有现货头寸至升水合约交割月份, 覆盖持有成本, 并卖出现货获利, 因此升水合约价格在其它条件不变的情况下, 价格下跌, 且随着时间的推移向现货价格靠拢。反之, 贴水市场中, 商品供给不足, 现货便利收益高于持有成本, 贴水期货合约在其它条件不变的情况下, 价格升高, 且随着时间的推移向现货价格靠拢。

因此, 利用展期收益来判断市场的升贴水结构, 并构建多空组合, 即买入展期收益最高的一篮子期货合约, 卖出展期收益最低的一篮子期货合约, 持有各商品的主力合约组合可获得期限结构的展期收益。

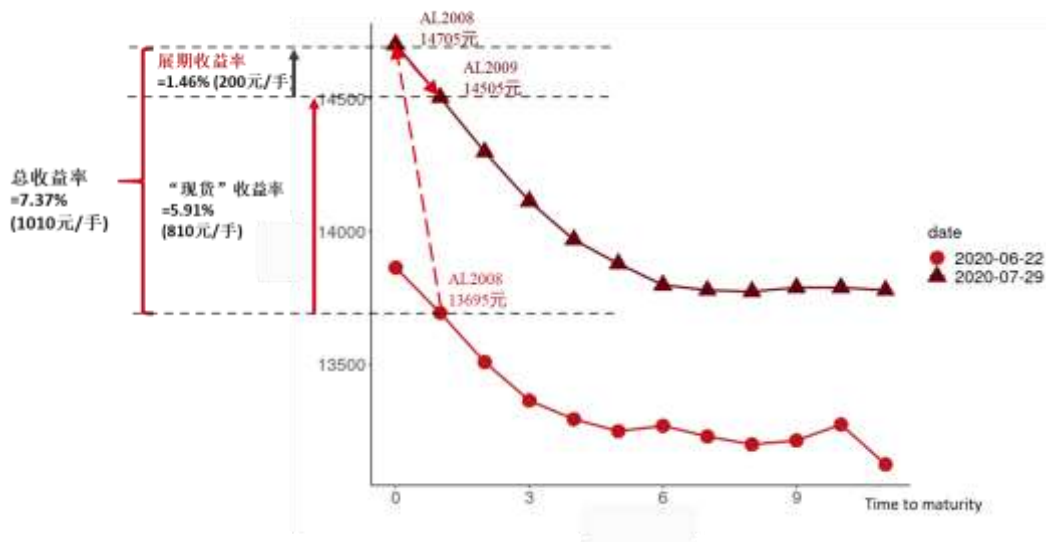
期限结构因子:

$$\text{Roll}_t^c = \ln\left(\frac{P_t^{c,\text{spot}}}{P_t^{c,\text{dom}}}\right) \times \frac{365}{t_{c,\text{dom}} - t_{c,\text{spot}}}$$

其中 $P_t^{c,\text{spot}}$ 和 $P_t^{c,\text{dom}}$ 分别为在 t 时刻期货品种 c 的近月合约价格和主力合约价格, $t_{c,\text{spot}}$ 和

$t_{c,dom}$ 分别为期货品种近月合约到期日剩余天数和期货品种 c 主力合约到期日剩余天数。

图3: 展期收益率的计算原理 (原则上贴水越深展期收益率越大)



数据来源: Wind 天软 华泰期货研究院

价值因子:

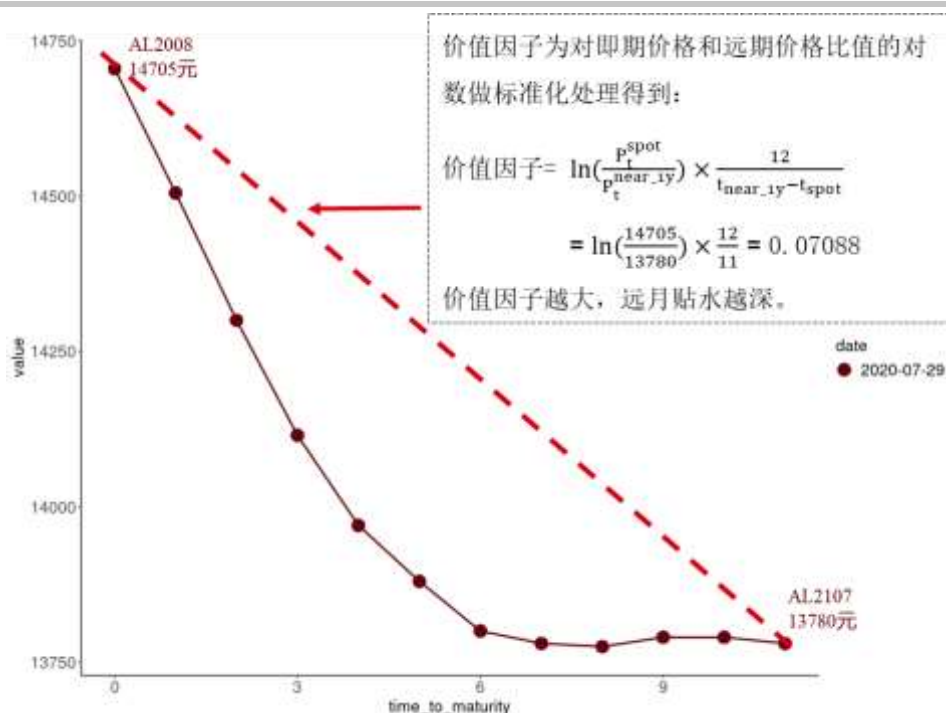
在权益市场中,著名的 Fama-French 三因子模型中使用的价值因子是提取市净率相对较低的公司对市盈率相对较高公司的风险溢价,而在商品市场中供需是评判价值的重要指标。所以,可以使用价格上升潜力较高(库存水平较低)的品种对潜力相对较低(库存水平较高)品种的风险溢价作为价值因子。但是库存数据较难及时获取,量化指标计算中可以使用期货期限结构曲线的近月合约溢价程度作为替代指标。远月贴水越深的品种,在近月的供需相对更为紧张,因此具有更大的价格上涨潜力,而远月升水越强的品种,相对上涨潜力较小,这与期限结构因子的底层逻辑一致。但是价值因子使用的合约有所不同,因为大宗商品普遍存在季节性因素(seasonal),所以对于供需状况的比较需要使用同期相比的方式来去除某些品种价格可能包含的季节性波动。最终,现货溢价程度使用期货曲线上第一个合约与未来一年合约之间的斜率衡量。

价值因子:

$$\text{value}_t^c = \ln\left(\frac{p_t^{c,\text{spot}}}{p_t^{c,\text{near}_1y}}\right) \times \frac{12}{t_{c,\text{near}_1y} - t_{c,\text{spot}}}$$

其中 $P_t^{c,spot}$ 和 $P_t^{c,near_1y}$ 分别为在 t 时刻期货品种 c 的近月合约（现货）价格和距到期最接近一年的合约价格。 $t_{c,near_1y}$ 和 $t_{c,spot}$ 为期货品种 c 距到期期限最接近一年的合约的到期日剩余月数。

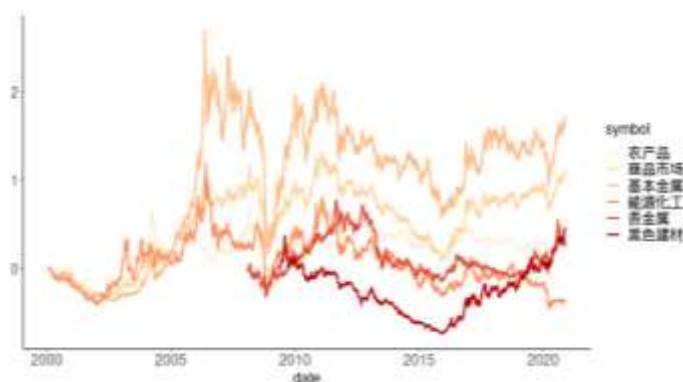
图4：价值因子计算示例



数据来源：Wind 天软 华泰期货研究院

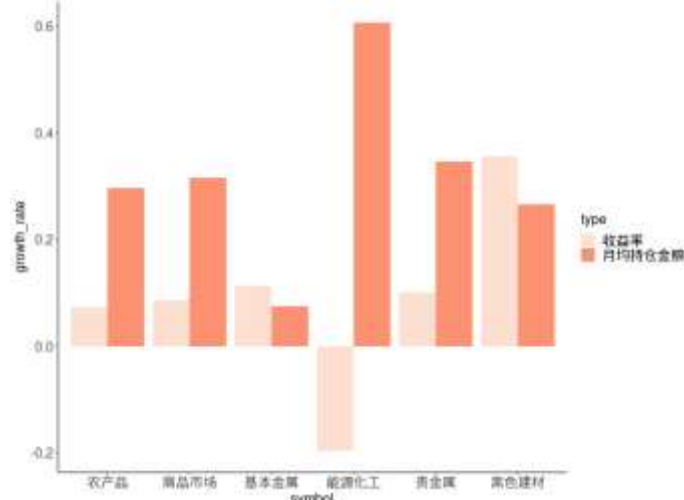
本质上，贴水越深的品种在期限结构因子与价值因子上暴露较大。而根据价值因子的时序基础因子策略每月选出的 5 个多头品种与 5 个空头品种的板块分布权重上看，黑色建材板块在今年主要正向暴露给了价值因子，农产品和能化板块反向暴露给价值因子，而对应到今年各板块的收益率表现，黑色建材板块累计收益最大，能化板块是唯一的负收益板块，所以价值因子的具有相当大的风格因子投资潜力。

图5： 各块累计收益率表现



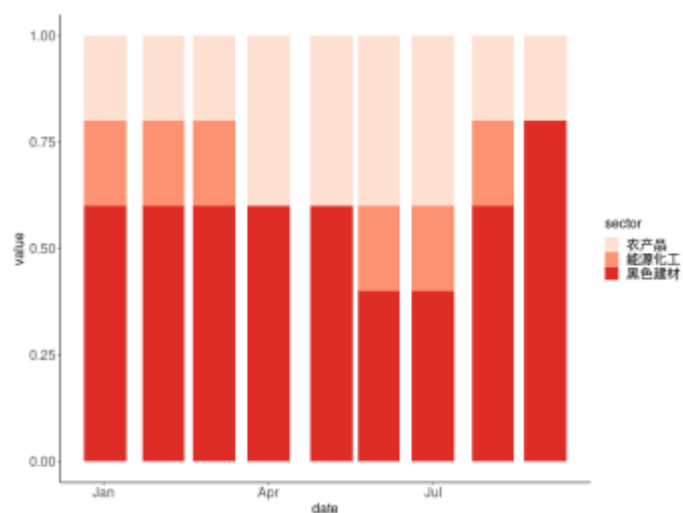
数据来源：天软 华泰期货研究院

图6： 2020 年板块收益率以及月均持仓金额增长情况



数据来源：天软 华泰期货研究院

图7： 2020 年价值因子正向暴露板块的权重



数据来源：天软 华泰期货研究院

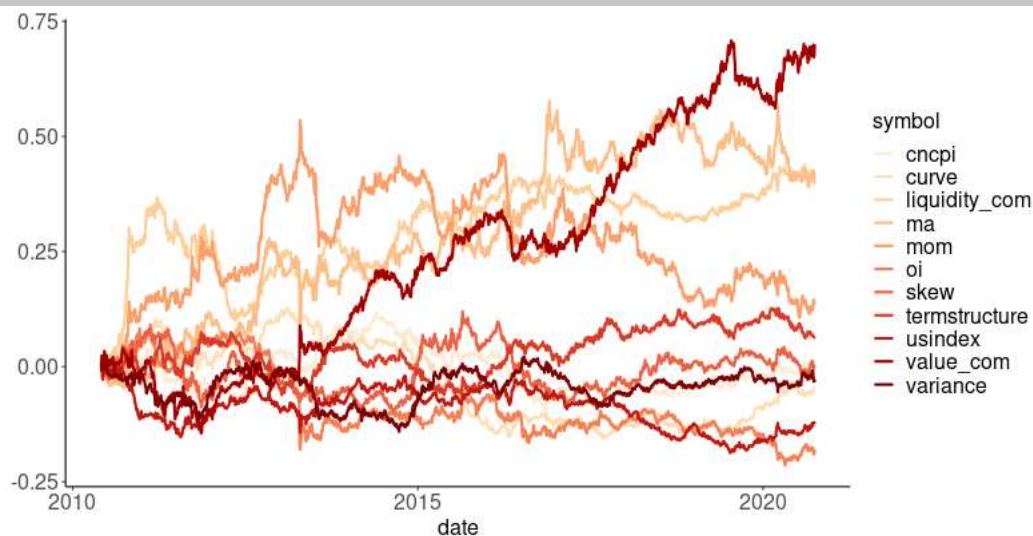
图8： 2020 年价值因子反向暴露板块的权重



数据来源：天软 华泰期货研究院

截面因子对比时序基础因子，增加了国家因子和板块因子。截面因子反映时序因子去除共线性之后得因子表现，但其收益率不可复现。

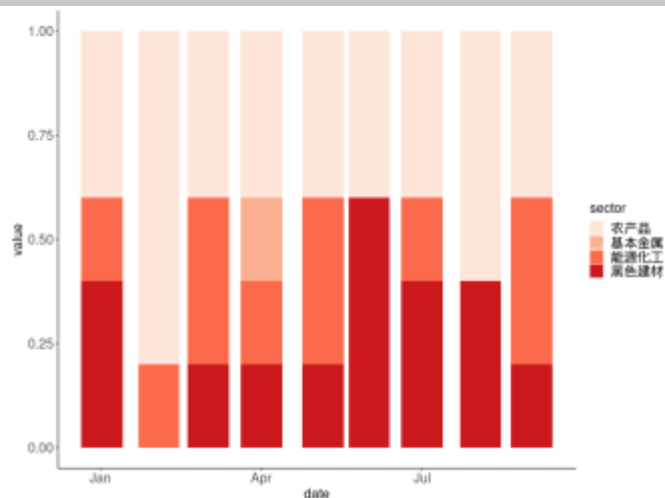
图9: 截面因子(不可投资的)累计收益率



数据来源: Wind 天软 华泰期货研究院

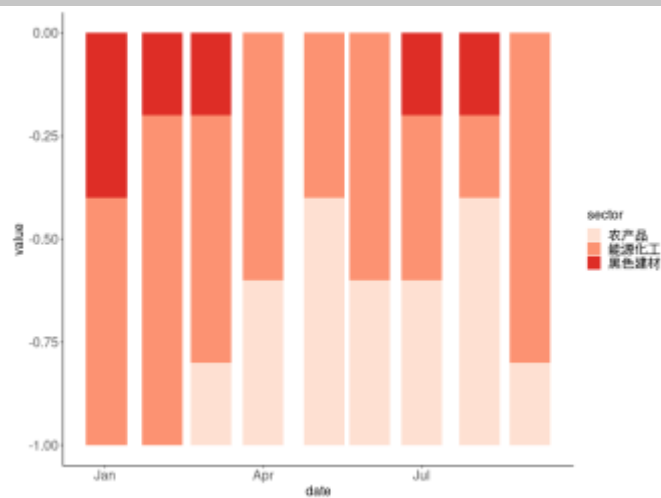
所有时序基础因子策略每月选出了5个多头品种、5个空头品种,各板块的分布权重如下:

图10: 2020年Curve因子正向暴露板块的权重



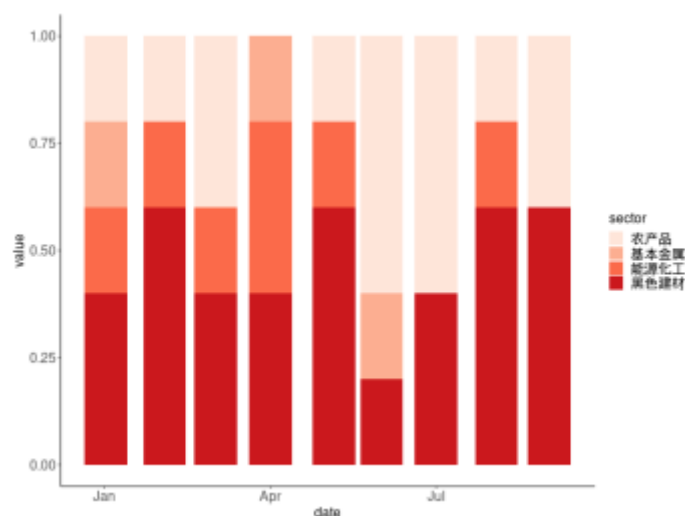
数据来源: 天软 华泰期货研究院

图11: 2020年对Curve因子反向暴露板块的权重



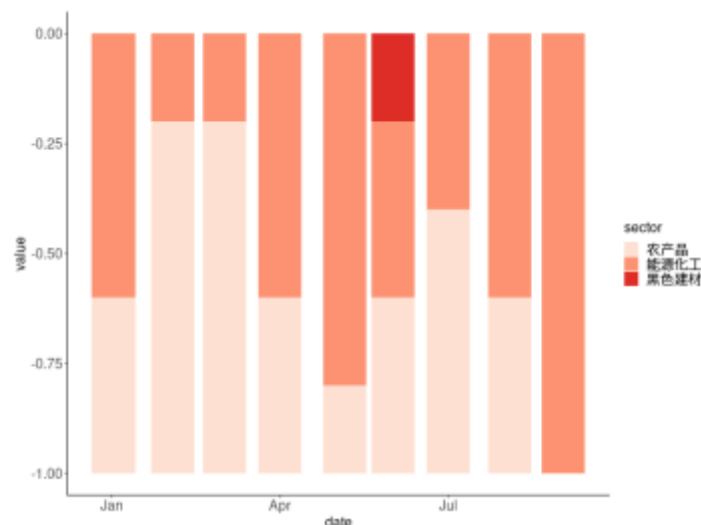
数据来源: 天软 华泰期货研究院

图12: 2020 年期限结构因子正向暴露板块的权重



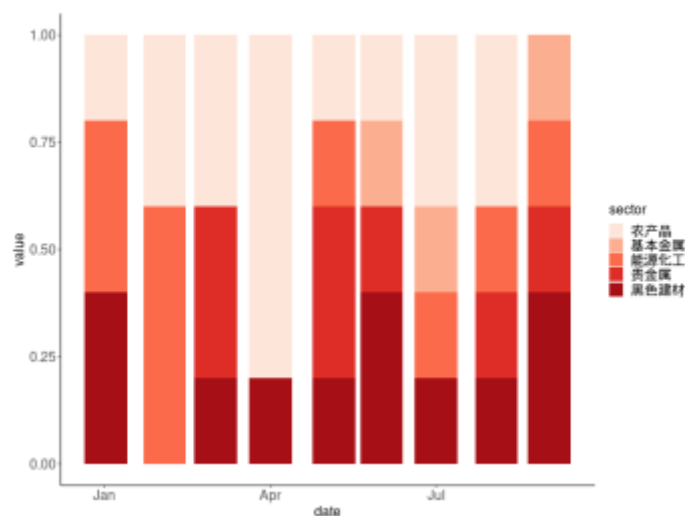
数据来源: 天软 华泰期货研究院

图13: 2020 年期限结构因子反向暴露板块的权重



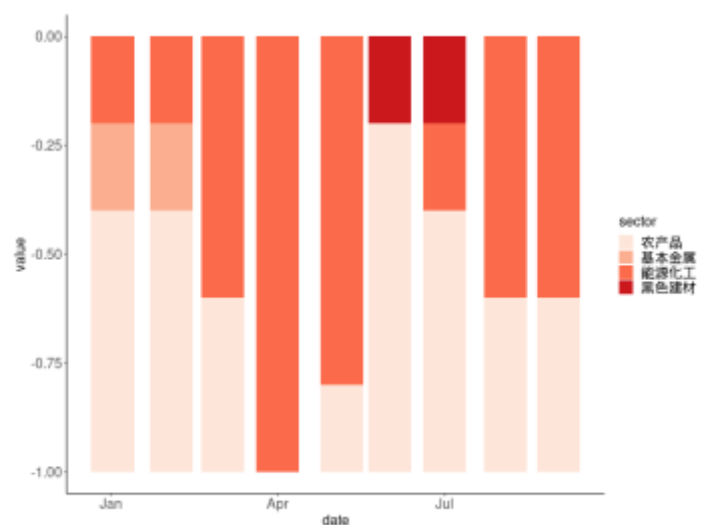
数据来源: 天软 华泰期货研究院

图14: 2020 年动量因子正向暴露板块的权重



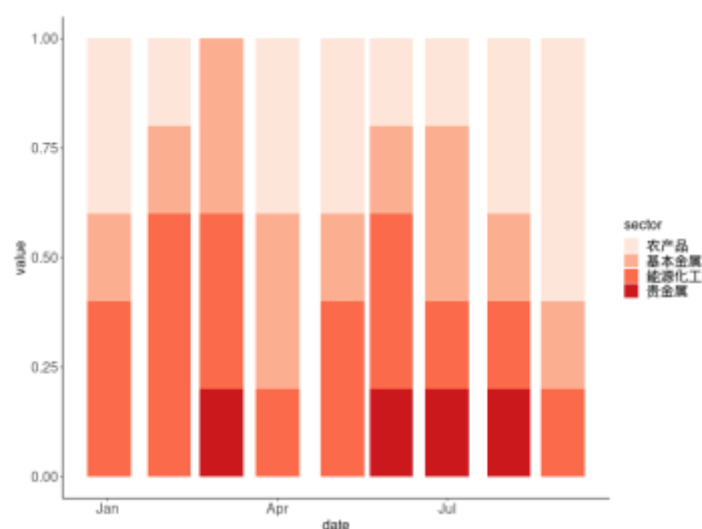
数据来源: 天软 华泰期货研究院

图15: 2020 年动量因子反向暴露板块的权重



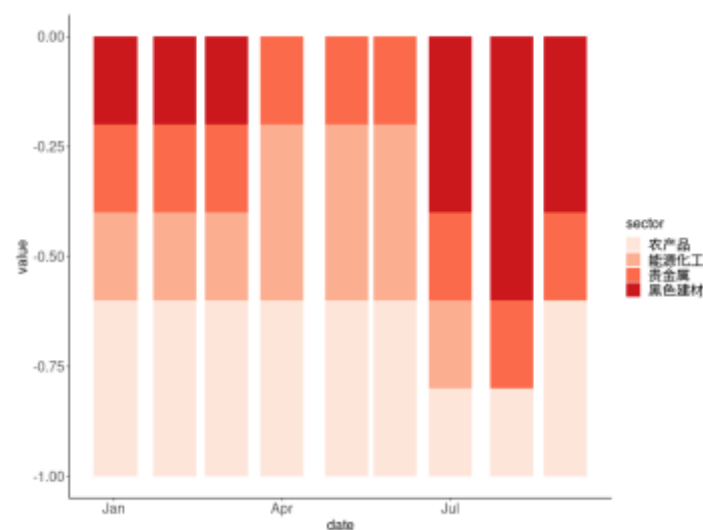
数据来源: 天软 华泰期货研究院

图16: 2020 年波动率因子正向暴露板块的权重



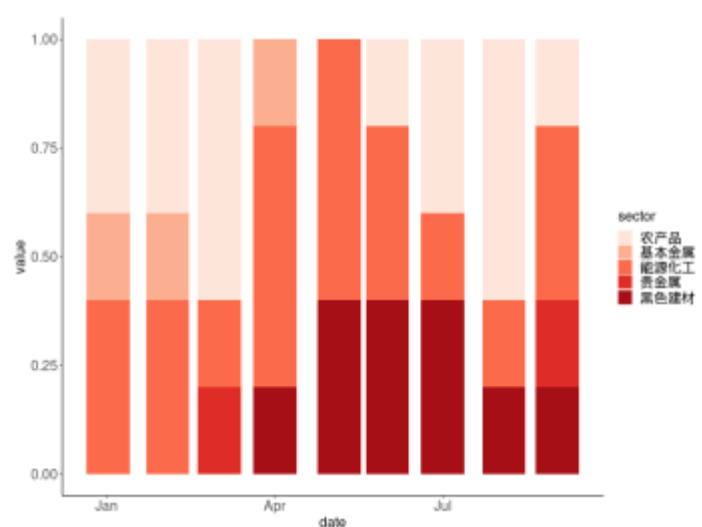
数据来源: 天软 华泰期货研究院

图17: 2020 年波动率因子反向暴露板块的权重



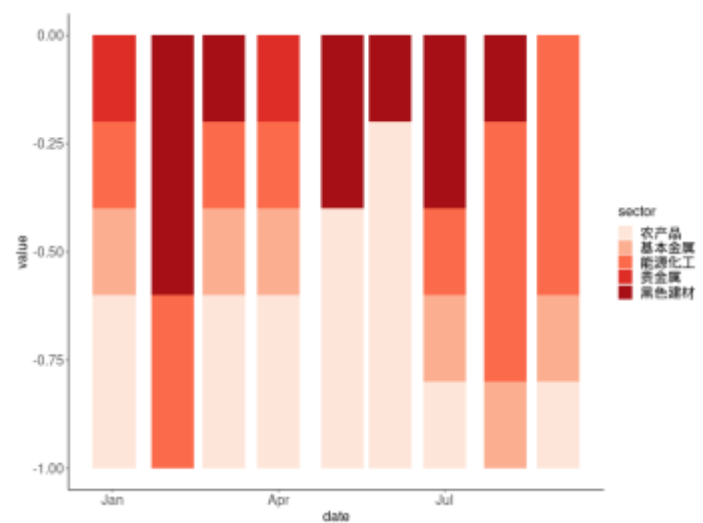
数据来源: 天软 华泰期货研究院

图18: 2020 年持仓因子正向暴露板块的权重



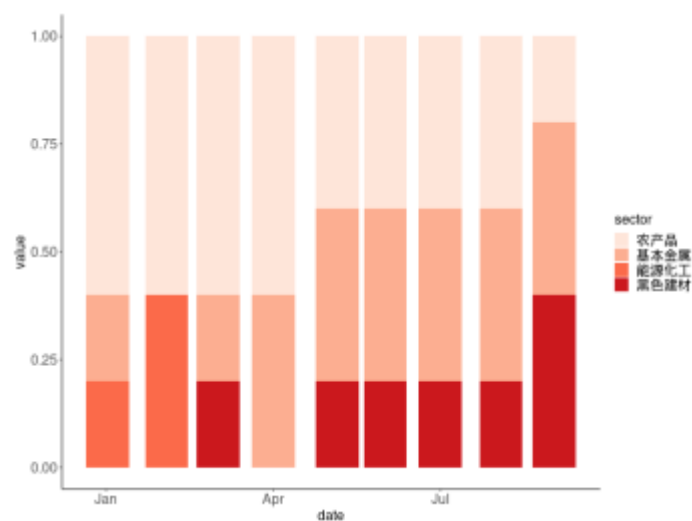
数据来源: 天软 华泰期货研究院

图19: 2020 年持仓因子反向暴露板块的权重



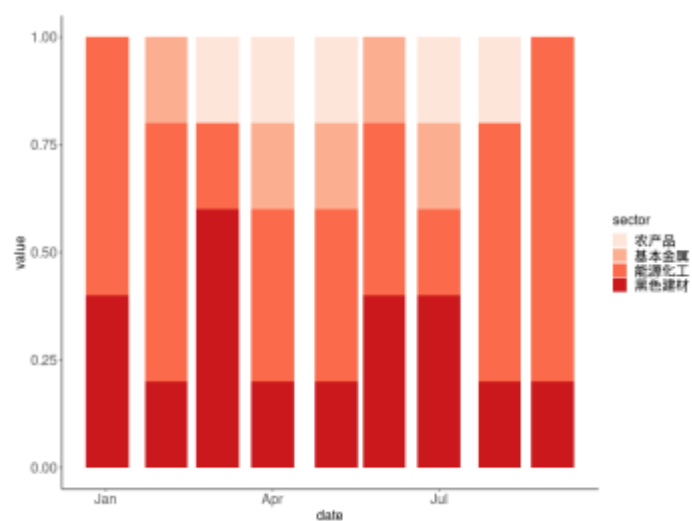
数据来源: 天软 华泰期货研究院

图20: 2020 年偏度因子正向暴露板块的权重



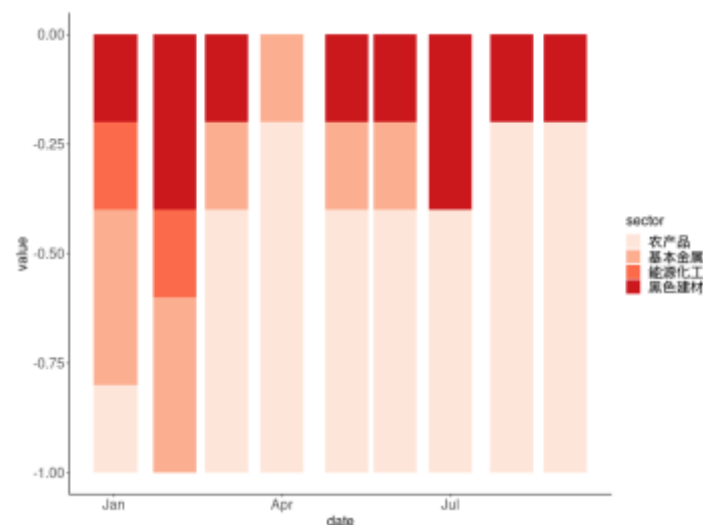
数据来源: 天软 华泰期货研究院

图22: 2020 年流动性因子正向暴露板块的权重



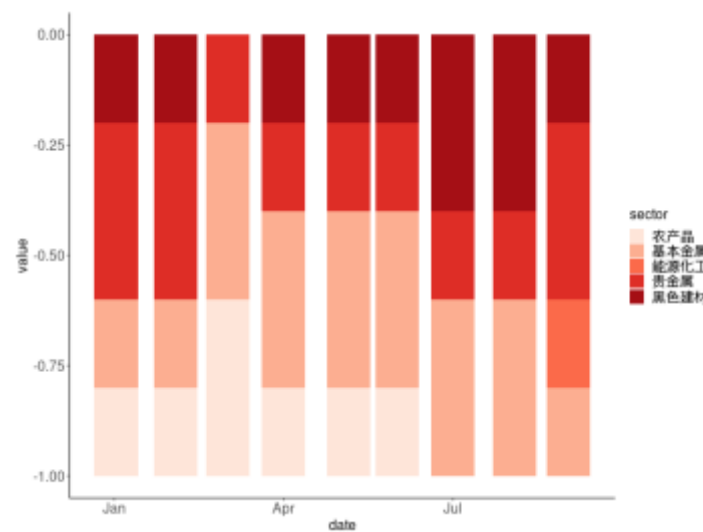
数据来源: 天软 华泰期货研究院

图21: 2020 年偏度因子反向暴露板块的权重



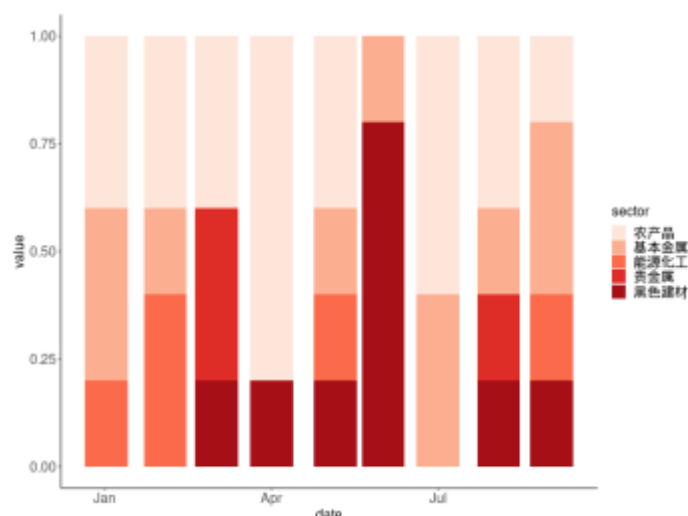
数据来源: 天软 华泰期货研究院

图23: 2020 年流动性因子反向暴露板块的权重



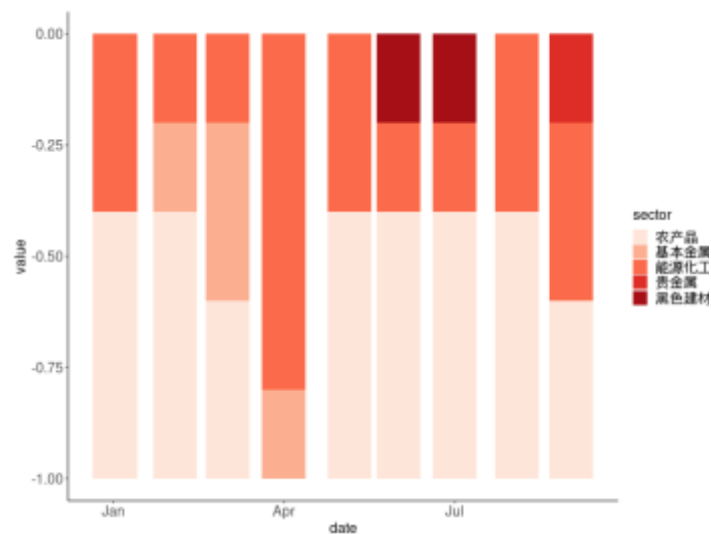
数据来源: 天软 华泰期货研究院

图24: 2020 年均价突破因子正向暴露板块的权重



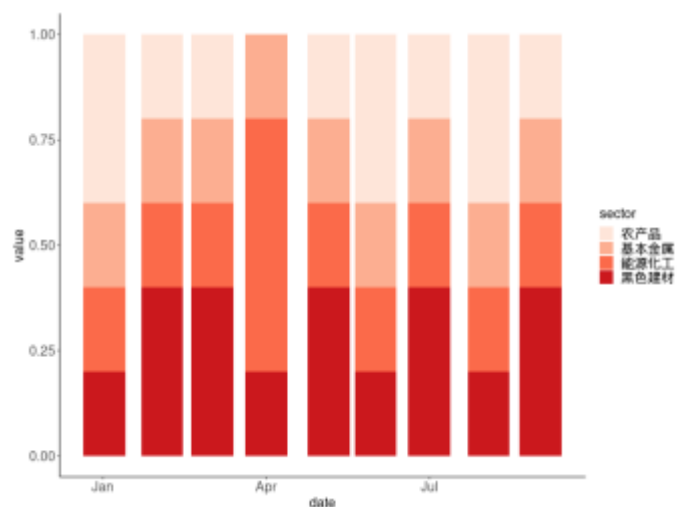
数据来源: 天软 华泰期货研究院

图25: 2020 年均价突破因子反向暴露板块的权重



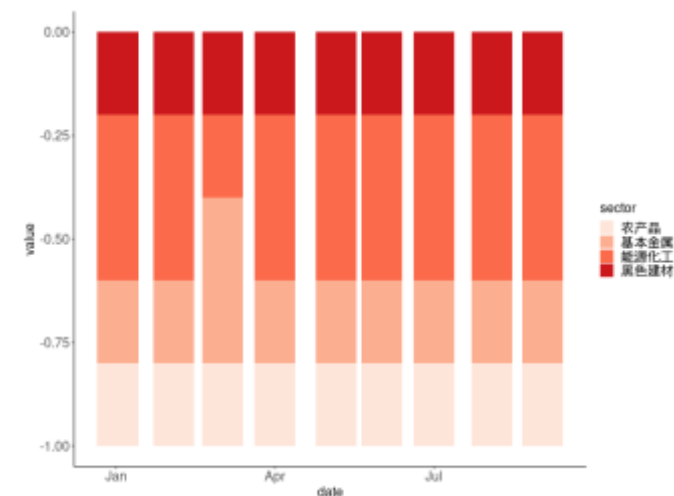
数据来源: 天软 华泰期货研究院

图26: 2020 年美元指数 Beta 因子正向暴露板块的权重



数据来源: 天软 Wind 华泰期货研究院

图27: 2020 年美元指数 Beta 因子反向暴露板块的权重



数据来源: 天软 Wind 华泰期货研究院

长期来看，在 Beta 因子上暴露较大的品种变化不大。下图是 2020 年 1-10 月，不同商品品种在相应因子策略多头组合中出现的频率。

图28: 2020 年对美元指数因子暴露较大的品种

品种	频次
天然橡胶	10 / 10
菜籽粕	10 / 10
铁矿石	10 / 10
铜	10 / 10
棉花	8 / 10

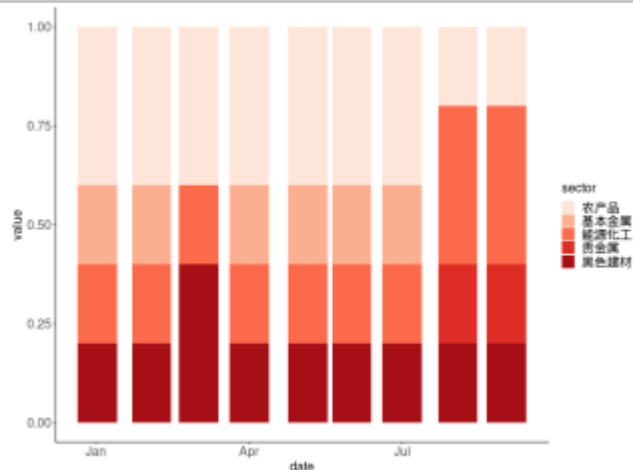
数据来源: 天软 Wind 华泰期货研究院

图29: 2020 年对中国 CPI 因子暴露较大的品种

品种	频次
天然橡胶	10 / 10
菜籽粕	10 / 10
铁矿石	10 / 10
铜	10 / 10
棉花	8 / 10

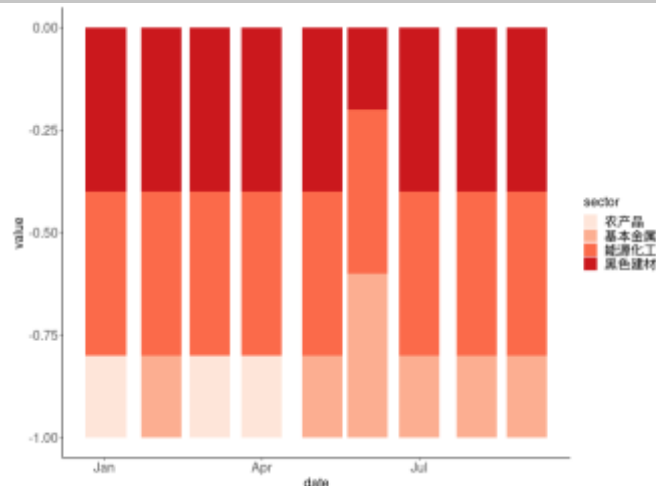
数据来源: 天软 Wind 华泰期货研究院

图30: 2020 年 CPI Beta 因子正向暴露板块的权重



数据来源: 天软 Wind 华泰期货研究院

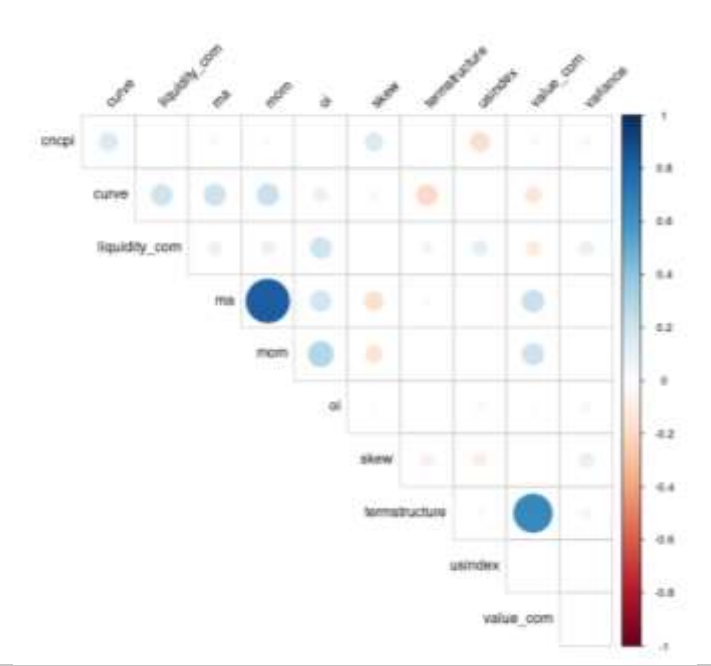
图31: 2020 年 CPI Beta 因子反向暴露板块的权重



数据来源: 天软 Wind 华泰期货研究院

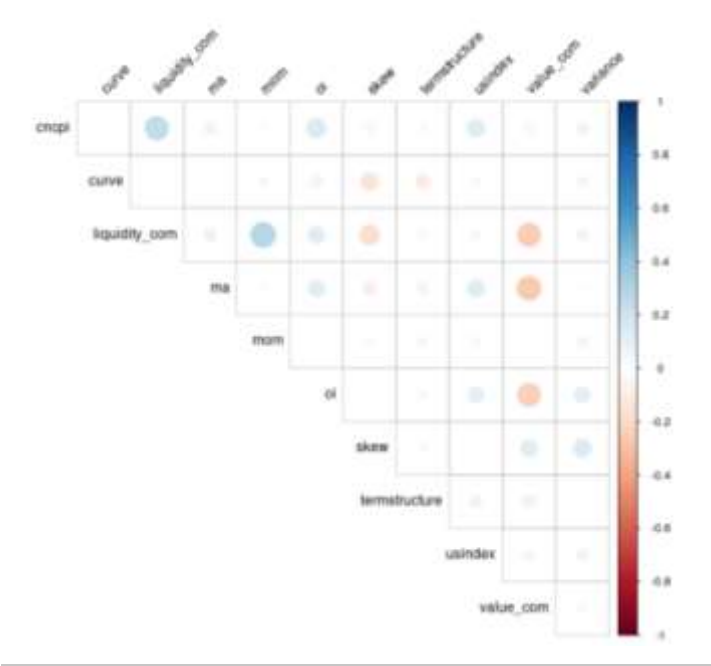
时序基础因子策略中，动量因子和均价突破因子、期限结构因子与价值因子保持了较高的相关性。截面因子策略中，风格因子间相关性基本消除，每个因子表征了相对独立的风险敞口。

图32: 时序基础风格因子收益率相关系数



数据来源: 天软 华泰期货研究院

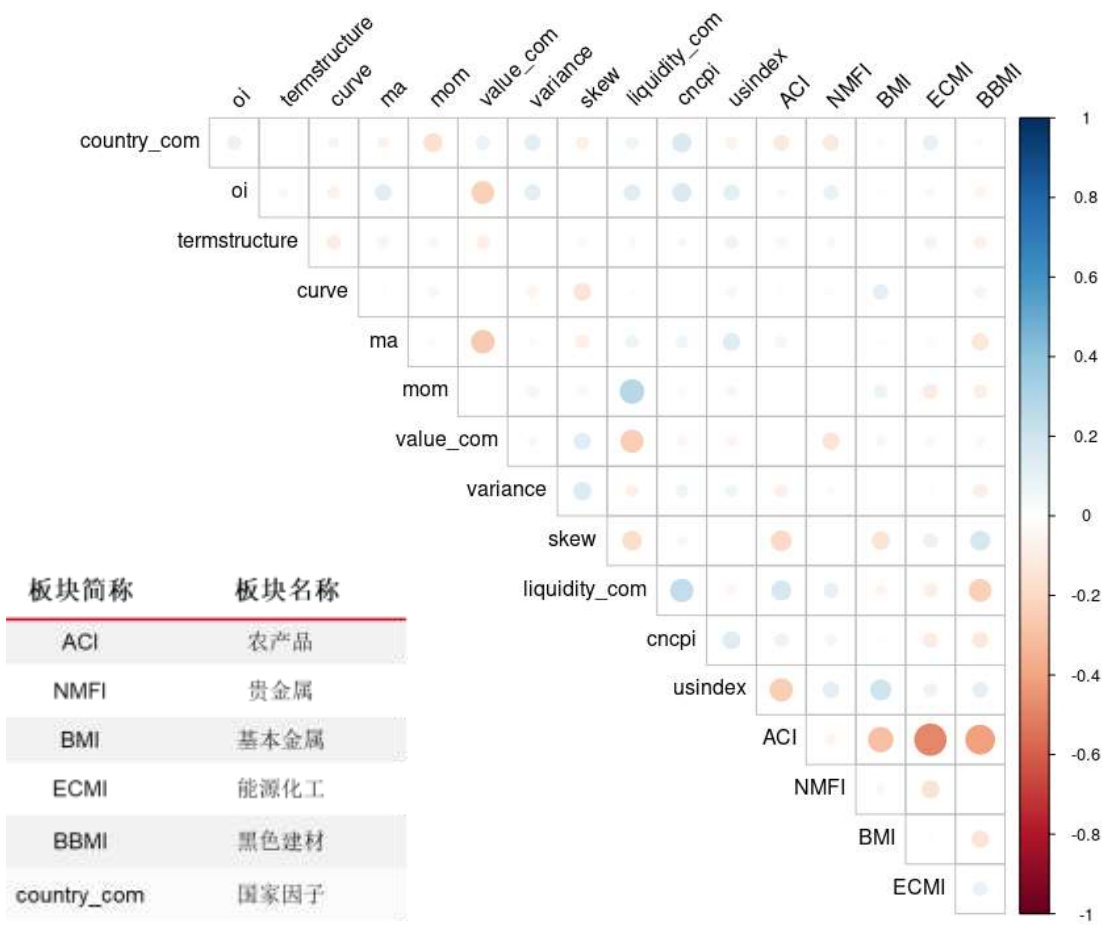
图33: 截面风格因子收益率相关系数



数据来源: 天软 华泰期货研究院

同时，截面因子中风格因子与板块因子相关性较低，说明风格因子抓住了各板块共同的风险暴露特征。全体截面因子之间相关性较小，每个因子都能刻画一类独特的风格/板块特征。农产品与其他板块都具有一定的负相关。

图34: 全部截面因子之间保持较低相关性



数据来源：Wind 天软 华泰期货研究院

接下来，我们观察截面多因子组对全品种期货市场的整体解释力。参照 barra 的风格因子算法，我们对中国股票市场也复现了一套截面因子组，样本空间选取沪深 300 的股票池，在年报上篇中我们介绍了期股联动研究，本章中我们将展示期货因子和股票因子对市场解释力的对比，下一章中我们会选用截面回归得到的商品因子以及股票因子作为特征量，研究其对商品板块指数的收益率影响力排名。

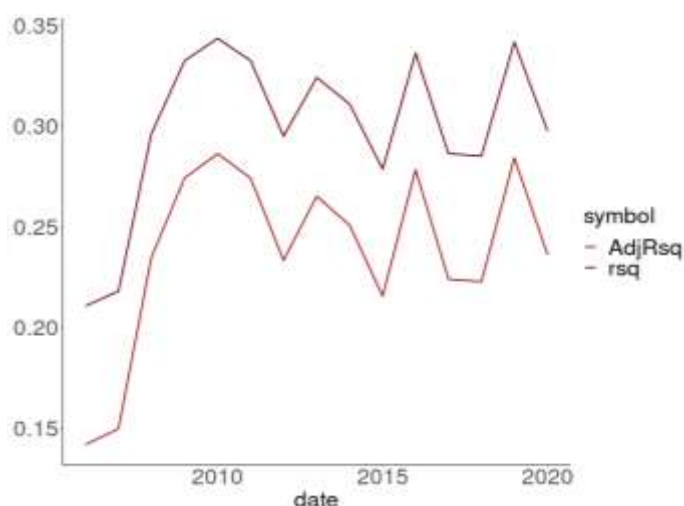
表格 1: 华泰商品，股票因子组的构成

标的物市场	国家因子	板块因子	风格因子
商品市场	1 个	5 个（华泰商品板块）	11 个（含 2 个宏观因子）
股票市场	1 个	11 个（GICS 一级）	13 个

资料来源：华泰期货研究院

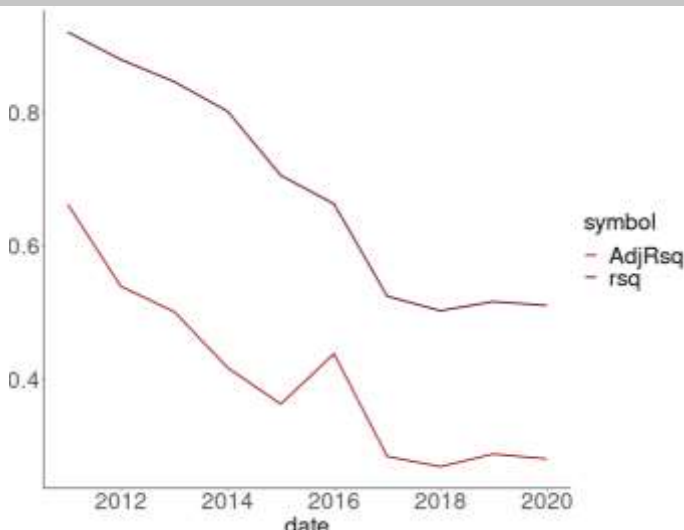
通过对比股票和商品市场因子组在历史上的解释力，我们发现多因子模型在商品市场具有更强的解释效果，Rsquare 和 AdjRsquare 明显更大。同时，商品市场的解释力随着时间的变化逐渐降低，我们认为这可能与商品期货品种的逐渐丰富有关。

图35: 沪深 300 样本截面因子解释力



数据来源：天软 华泰期货研究院

图36: 商品全市场样本截面因子解释力



数据来源：天软 华泰期货研究院

三、 AI 模型的应用

使用随机森林算法筛选重要因子，相较于传统的线性方法，可以挖掘数据间的非线性规律。随机森林算法的基本组成元素是决策树，决策树算法具有效率高，处理的信息量大等优点。但同时，决策树本身依然有着不可避免的局限性，如对缺失值不敏感，数据容易过拟合等。随机森林算法则利用多颗决策树，并进行随机的特征测试，不仅增强了模型的鲁棒性，也提高了分类的准确性。

随机森林算法分为回归和分类两种方式，他们对因子重要性的评判标准也不同。当采用回归的方式时，平均最小分裂节点和因子纯度上升是主要考量因子重要性的两个指标。简单来说，平均最小分裂节点越小，因子纯度上升越快，该因子就越重要。当采用分类的方式时，平均基尼系数降低和平均最小分裂节点这两个指标是我们衡量因子重要性的标准。类似地，平均基尼系数降低的越快以及最小分裂节点越小，该因子就越重要。下文中我们将给出回归与分类两种方法的因子重要性结果，以便于互相对比论证。

在具体数据测试过程中，我们会选用截面回归得到的**商品因子**以及**股票因子**作为特征量，使用商品板块指数作为标的物进入随机森林模型。下面我们给出针对原始数据以及多周期

分解之后的短周期、长周期数据三个口径的测试结果，为方便展示，我们统一按照缩写展示，对应中文名称可以参见附录中的商品/股票因子代码表对比查找。

随机森林算法使用回归方式。CCFI 为全市场商品指数，纵观全市场指数的主要影响因素，商品国家因子“country_com”无疑是最主要的影响因素，在主导因子排名中都位列第一，商品动量因子“mom”排名第二，而股票材料板块因子“materials”排名第三，这可能与材料板块包含大量大宗商品上下游企业有关。同时股票市场的国家因子“country”对商品市场也有较强的影响。

接着，我们再看各个板块的主要影响因素，此处我们在影响力排名时未加入板块对应的板块截面因子（若加入，影响力将排在第一位）。“country_com”无疑是最主要的影响因素。其次，商品动量因子“mom”在大多数板块中重要性排名第二，中国 CPI 因子“cncpi”对能化板块与基本金属板块有较强的影响力。而期股联动效应最明显的当属贵金属板块，股票市场价值因子“value”以及股票材料板块因子“materials”对贵金属板块均有较强的解释力。

我们的测试数据如下：

标的物：华泰商品 5 板块指数周度收益率（2010-06 至 2020-09）；

训练集：前 70%的周度标的物收益率；

测试集：近 3 年约 30%的周度标的物收益率；

特征集：所有截面因子；

树的棵树：5000 棵树（此时所有模型已经收敛）；

测试目标：测试因子重要性（使用分类/回归的方法）。

表格 2: 原始数据因子重要性排名及对应指标值 (回归分析)

指数	指标	重要因子 1	重要因子 2	重要因子 3	重要因子 4	重要因子 5
CCFI	平均最小分裂节点	country_com	mom	materials	ma	country
		1.169	3.653865	3.983694	4.434437	4.51296
CCFI	因子纯度	country_com	mom	materials	country	ma
		0.10588	0.008634	0.007573	0.006328	0.005509
BMI	平均最小分裂节点	country_com	cncpi	mom	liquidity_com	ma
		1.2954	4.021488	4.144089	4.337171	4.939333
BMI	因子纯度	country_com	mom	liquidity_com	cncpi	country
		0.114096	0.01519	0.012287	0.011038	0.009664
ACI	平均最小分裂节点	country_com	mom	value_com	skew	liquidity_com
		1.3282	4.2382	4.488833	5.183926	5.251678
ACI	因子纯度	country_com	mom	value_com	liquidity_com	skew
		0.065518	0.005787	0.005218	0.004377	0.004318
ECMI	平均最小分裂节点	country_com	cncpi	value_com	mom	country
		1.1998	3.322257	4.37822	4.500273	4.717031
ECMI	因子纯度	country_com	cncpi	value_com	mom	country
		0.170653	0.017067	0.012814	0.012477	0.012367
BBMI	平均最小分裂节点	country_com	mom	ma	materials	country
		1.1972	3.645793	4.045621	4.165094	4.3069
BBMI	因子纯度	country_com	mom	ma	country	materials
		0.178152	0.023689	0.023685	0.021146	0.020871
NMFI	平均最小分裂节点	country_com	value	materials	oi	value_com
		1.758918	3.510392	4.202797	4.362788	4.941374
NMFI	因子纯度	country_com	value	oi	materials	ma
		0.034351	0.015905	0.0151	0.013093	0.01176

资料来源: 天软 同花顺 Wind 华泰期货研究院

相比于回归分析, 分类方法明显抓住了更多期股联动的信息, 影响力排名靠前的因子中股票因子所占的席位明显增加, 股票材料板块因子“materials”对所有板块均有较大影响。例如股票电信板块因子“telecom”对商品黑色建材板块影响较大, 股票可选消费板块因子

“consumer_staples”对商品贵金属板块影响较大。所有商品板块中，贵金属板块表现出与股票市场较强联动规律，前五个重要因子中有四个来源于股票的因子组。

表格 3: 原始数据因子重要性排名及对应指标值 (分类)

指数	指标	重要因子 1	重要因子 2	重要因子 3	重要因子 4	重要因子 5
CCFI	平均最小分裂节点	country_com	materials	cncpi	country	leverage
		1.0512	3.559003	4.345607	4.54705	5.124671
CCFI	基尼平均降低	country_com	materials	country	cncpi	beta
		107.7792	8.528326	4.866498	4.025277	2.31588
BMI	平均最小分裂节点	country_com	cncpi	country	materials	telecom
		1.0698	4.232959	4.52054	4.66911	5.068748
BMI	基尼平均降低	country_com	country	cncpi	materials	mom
		71.68661	7.027792	6.198795	5.279597	3.956728
ACI	平均最小分裂节点	country_com	cncpi	materials	variance	nlsiz
		1.152	4.802959	5.322576	5.521614	5.565862
ACI	基尼平均降低	country_com	cncpi	materials	dividend	country
		56.00821	5.745984	5.089082	4.492322	4.454449
ECMI	平均最小分裂节点	country_com	cncpi	materials	country	momentum
		1.0748	3.542007	4.399244	4.971408	5.115297
ECMI	基尼平均降低	country_com	cncpi	materials	country	momentum
		74.90966	7.874616	6.625995	6.274343	3.851418
BBMI	平均最小分裂节点	country_com	country	telecom	materials	energy
		1.123225	4.087838	5.306563	5.415382	5.436793
BBMI	基尼平均降低	country_com	country	telecom	materials	termstructure
		48.58515	8.95895	5.127764	4.956116	4.61726
NMFI	平均最小分裂节点	country_com	materials	value	consumer_staples	country
		1.735491	3.311866	3.898148	5.114164	5.207934
NMFI	基尼平均降低	country_com	materials	value	consumer_staples	country
		16.61226	10.07841	7.896611	5.750435	5.487776

资料来源: 天软 同花顺 Wind 华泰期货研究院

在回归分析中，短周期中商品波动率因子“variance”对各板块影响力较大。商品 Curve 因子

对农产品板块影响力排名靠前，同时所有的板块在短周期上都受到商品国家因子“country_com”与中国 CPI 因子“cncpi”影响。

表格 4: 短周期数据因子重要性排名及对应指标值（回归）

指数	指标	重要因子 1	重要因子 2	重要因子 3	重要因子 4	重要因子 5
CCFI	平均最小分裂节点	country_com	cncpi	country	mom	variance
		1.1082	3.587051	4.220228	4.289107	4.478603
CCFI	因子纯度	country_com	cncpi	country	variance	materials
		0.000269	0.000019	0.000017	0.000012	0.000011
BMI	平均最小分裂节点	country_com	cncpi	variance	country	mom
		1.1594	3.568705	3.701848	4.163848	4.317837
BMI	因子纯度	country_com	variance	cncpi	country	mom
		0.000371	0.000039	0.000038	0.000037	0.000023
ACI	平均最小分裂节点	country_com	curve	variance	country	liquidity_com
		1.2896	4.836651	4.944178	5.18526	5.241895
ACI	因子纯度	country_com	curve	variance	liquidity_com	country
		0.000171	0.000012	0.000011	0.000011	0.00001
ECMI	平均最小分裂节点	country_com	cncpi	country	mom	variance
		1.213	3.310676	3.892736	4.531391	4.535645
ECMI	因子纯度	country_com	cncpi	country	variance	mom
		0.000441	0.000046	0.000045	0.000029	0.000027
BBMI	平均最小分裂节点	country_com	mom	cncpi	value_com	termstructure
		1.1874	4.23203	4.64945	4.668904	4.734653
BBMI	因子纯度	country_com	mom	cncpi	termstructure	value_com
		0.000372	0.000038	0.000034	0.000031	0.00003
NMFI	平均最小分裂节点	country_com	materials	variance	cncpi	ma
		1.486795	4.173994	4.466632	4.48905	4.852962
NMFI	因子纯度	country_com	materials	variance	cncpi	termstructure
		0.000127	0.000036	0.000035	0.000032	0.000028

资料来源：天软 同花顺 Wind 华泰期货研究院

将数据分周期后，可以发现更多有意思的数据特征，包含大量可供挖掘的信息，特别是商

品资产定价的复杂底层逻辑，我们将在后续的报告中进行深入细致的研究。此处我们以贵金属板块为例，短周期数据显示，商品流动性因子“liquidity_com”与股票流动性因子“liquidity”对贵金属板块均有较强的影响力。贵金属板块在长周期，受中国 CPI 因子“cncpi”影响力较强，这恰好与贵金属板块抗通胀的属性相呼应。

表格 5: 短周期数据因子重要性排名及对应指标值（分类）

指数	指标	重要因子 1	重要因子 2	重要因子 3	重要因子 4	重要因子 5
CCFI	平均最小分裂	country_com	cncpi	country	mom	residualvolatility
	节点	1.0106	3.445618	3.519236	3.895744	4.43413
CCFI	基尼平均降低	country_com	country	cncpi	mom	nsize
		98.95552	6.225542	4.130717	2.873652	2.436741
BMI	平均最小分裂	country_com	country	mom	ma	skew
	节点	1.0266	2.709482	2.829331	3.744401	3.807335
BMI	基尼平均降低	country_com	country	mom	ma	cncpi
		55.62655	7.075812	5.712964	4.385223	4.13649
ACI	平均最小分裂	country_com	cncpi	dividend	country	skew
	节点	1.0516	3.333327	3.462538	3.64037	3.79076
ACI	基尼平均降低	country_com	cncpi	dividend	skew	mom
		41.6892	4.957643	4.706505	4.357007	4.354272
ECMI	平均最小分裂	country_com	cncpi	country	telecom	mom
	节点	1.0544	2.465798	2.557709	4.051334	4.074616
ECMI	基尼平均降低	country_com	country	cncpi	mom	skew
		58.74126	8.184023	7.336258	3.76324	3.735209
BBMI	平均最小分裂	country_com	country	mom	cncpi	consumer_discreti onary
	节点	1.0366	3.539	3.583139	3.655866	4.009854
BBMI	基尼平均降低	country_com	mom	country	cncpi	bab
		37.51239	4.937976	4.66333	4.566867	4.559108
NMFI	平均最小分裂	country_com	liquidity_com	liquidity	materials	variance
	节点	1.172877	2.112799	3.398114	3.419959	3.509487
NMFI	基尼平均降低	country_com	liquidity_com	liquidity	variance	materials
		20.40677	9.370527	5.991875	5.760703	5.614699

资料来源：天软 同花顺 Wind 华泰期货研究院

表格 6: 长周期数据因子重要性排名及对应指标值 (回归)

指数	指标	重要因子 1	重要因子 2	重要因子 3	重要因子 4	重要因子 5
CCFI	平均最小分裂节点	country_com	mom	materials	ma	cncpi
		1.2298	3.811534	4.456733	4.786381	4.82986
CCFI	因子纯度	country_com	mom	materials	cncpi	ma
		0.001347	0.000114	0.000072	0.00007	0.00007
BMI	平均最小分裂节点	country_com	cncpi	mom	liquidity_com	value_com
		1.3282	3.605905	3.784024	4.457076	5.474335
BMI	因子纯度	country_com	cncpi	mom	liquidity_com	oi
		0.001624	0.000244	0.000227	0.000156	0.000104
ACI	平均最小分裂节点	country_com	mom	oi	cncpi	ma
		1.283	4.228776	4.614476	4.815815	5.179705
ACI	因子纯度	country_com	mom	oi	cncpi	country
		0.000843	0.00007	0.000058	0.00005	0.000046
ECMI	平均最小分裂节点	country_com	cncpi	mom	country	value_com
		1.2748	3.654053	4.988058	5.166443	5.3396
ECMI	因子纯度	country_com	cncpi	country	mom	curve
		0.002244	0.000246	0.00016	0.000137	0.00012
BBMI	平均最小分裂节点	country_com	ma	mom	value_com	materials
		1.28	3.403548	4.407436	4.781245	4.938124
BBMI	因子纯度	country_com	ma	mom	value_com	materials
		0.00201	0.00039	0.000242	0.000207	0.000192
NMFI	平均最小分裂节点	country_com	materials	oi	value	mom
		1.915916	3.706798	3.930021	4.951271	5.23786
NMFI	因子纯度	country_com	oi	materials	value_com	ma
		0.000458	0.00023	0.000213	0.000167	0.000144

资料来源: 天软 同花顺 Wind 华泰期货研究院

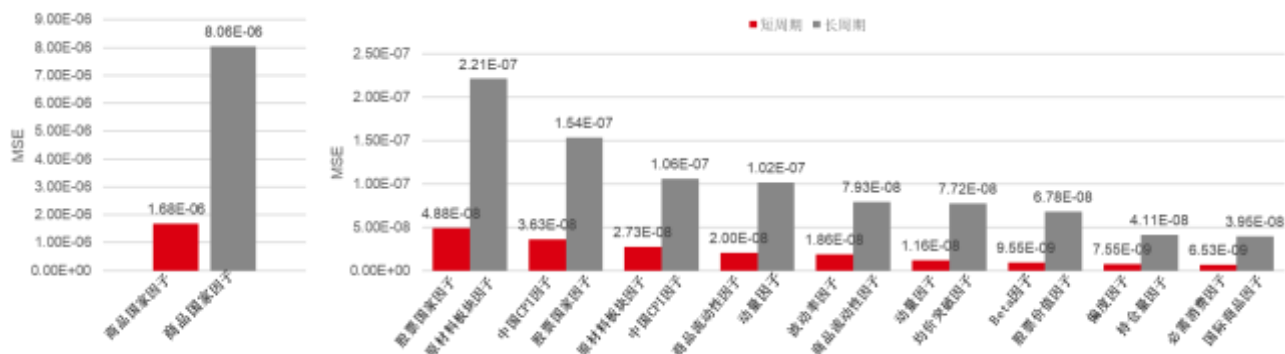
表格 7: 长周期数据因子重要性排名及对应指标值 (分类)

指数	指标	重要因子 1	重要因子 2	重要因子 3	重要因子 4	重要因子 5
CCFI	平均最小分裂节点	country_com	materials	value_com	value	country
		1.0332	2.937904	3.884672	3.980352	4.174608
CCFI	基尼平均降低	country_com	materials	value_com	country	energy
		88.51427	4.956135	3.58894	3.395316	3.162585
BMI	平均最小分裂节点	country_com	mom	cncpi	variance	size
		1.0348	2.725488	2.89872	3.606848	3.740856
BMI	基尼平均降低	country_com	mom	cncpi	variance	earnings
		35.56541	6.870832	6.131003	5.082203	4.774999
ACI	平均最小分裂节点	country_com	growth	nlsiz	value_com	consumer_staples
		1.0414	3.719851	3.812336	3.943396	4.145293
ACI	基尼平均降低	country_com	nlsiz	growth	value_com	termstructure
		55.83628	4.426981	4.187505	3.970684	3.838619
ECMI	平均最小分裂节点	country_com	country	energy	cncpi	momentum
		1.026	3.242244	3.443338	3.502569	3.824138
ECMI	基尼平均降低	country_com	country	cncpi	energy	ma
		56.53563	5.635919	4.607681	4.498467	4.211922
BBMI	平均最小分裂节点	country_com	materials	rici	nlsiz	country
		1.0754	2.821994	3.390504	3.509515	3.788427
BBMI	基尼平均降低	country_com	materials	nlsiz	rici	oi
		32.1719	6.019482	5.271676	5.154959	4.527522
NMFI	平均最小分裂节点	country_com	value	rici	beta	materials
		1.671897	3.011211	3.201967	3.602805	3.769973
NMFI	基尼平均降低	country_com	value	rici	cncpi	residualvolatility
		9.634981	6.396121	5.676417	5.133926	5.100377

资料来源: 天软 同花顺 Wind 华泰期货研究院

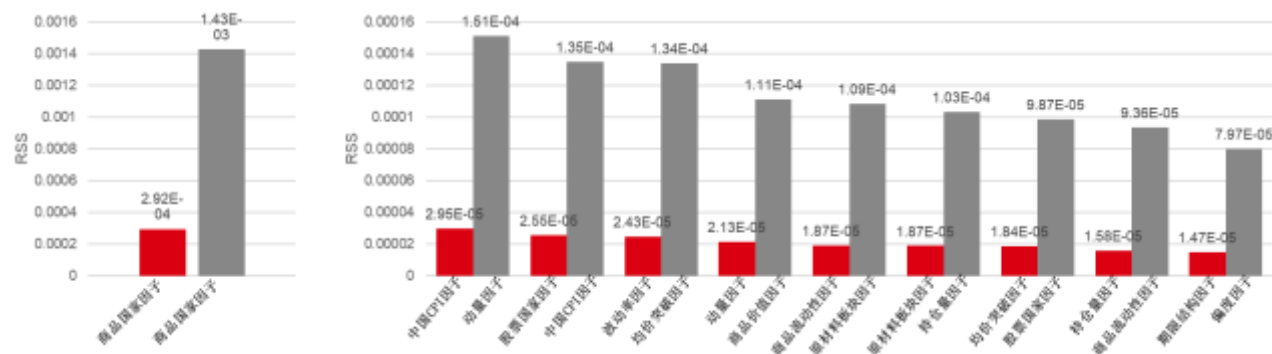
基于前述结果, 商品国家因子对各个板块、各个尺度的影响力均位列第一, 我们从统计的角度, 通过因子间的对比, 评估因子对全市场的重要性中影响作用的相对强弱, 具体方法是计算各因子在不同周期针对五大板块的重要性指标之和然后进行排序, 根据选取不同特征指标得到如下结果:

图37: 不同周期因子 MSE 排名 (分类)



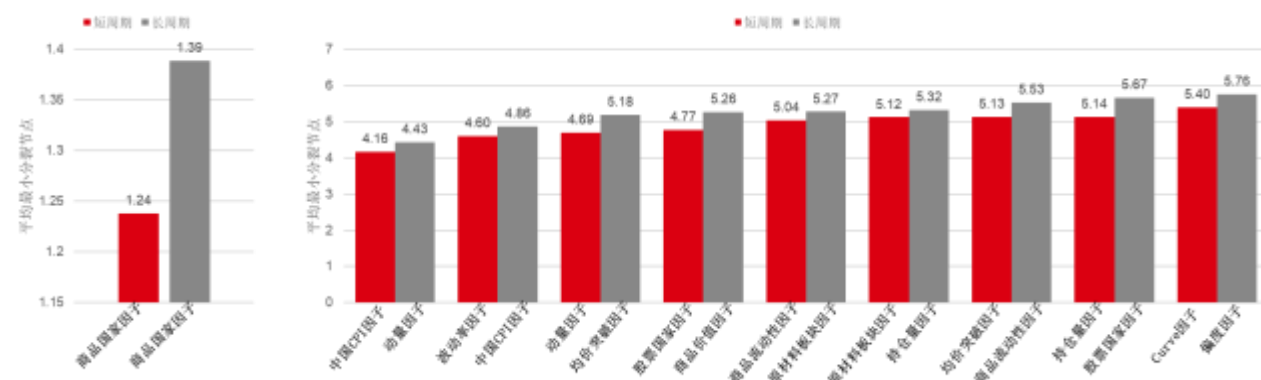
数据来源: 天软 同花顺 Wind 华泰期货研究院

图38: 不同周期因子 RSS 排名 (分类)



数据来源: 天软 同花顺 Wind 华泰期货研究院

图39: 不同周期因子最小分裂节点排名 (分类)



数据来源: 天软 同花顺 Wind 华泰期货研究院

四、 AI 模型预测能力探索

截面因子能够将抽象的因子概念，通过行情数据具象化，并且能够帮助投资人更清晰理解具有系统性特征的收益/风险来源。所以，对于投资市场风格和商品板块的投资策略来说，截面因子是不可或缺的风险溢价定量分析基础。基于截面因子在投资策略中的优势，我们在本章对截面因子的预测能力进行进一步的探索。

我们以商品板块指数作为标的，通过使用周期分解之后短周期和长周期上股票、商品的历史因子数据，运用随机森林的方法对标的进行预测，分别得到短周期和长周期的预测结果，然后再将两个周期的预测结果叠加得到最终的预测结果。主要参数选取如下：以周度作为时间单位，通过选取不同历史长度的因子数据，对各商品板块 T+1 周以及 T+2 周的收益率进行预测，RMSE 越小、胜率越高说明预测效果越好。

从使用的历史数据长度上看，大部分商品板块用**历史两周**的数据对未来一周数据进行预测效果最好。从不同板块的预测效果上看，商品农产品板块、黑色建材板块的预测效果最好，贵金属板块的预测效果最差。同时，随着选取的历史时间变长，RMSE 与胜率的标准差也逐渐增大。

我们的测试数据如下：

标的物：华泰商品 5 板块指数周度收益率（2010-06 至 2020-09）；

训练集：前 70%的周度标的物收益率；

测试集：近 3 年约 30%的周度标的物收益率；

特征集：所有截面因子；

树的棵树：15000 棵（为了保证使用因子历史数据长度为 T-6 周至今时模型也能收敛）

预测目标：测试随机森林的回归能力--下一周标的物收益率；或预测下两周的标的物收益率。

表格 8: 随机森林预测 T+1 周表现 (回归)

选取因子历史数据长度		T 周	T-1 周至今	T-2 周至今	T-3 周至今	T-4 周至今	T-5 周至今	T-6 周至今
ACI	RMSE	0.013535	0.013458	0.013475	0.013480	0.013461	0.013514	0.013514
	胜率 (%)	48.00	58.00	53.02	53.69	57.72	54.05	53.38
BBMI	RMSE	0.032893	0.032828	0.032990	0.033007	0.033005	0.032727	0.032739
	胜率 (%)	48.67	54.00	53.02	45.64	46.31	47.30	45.27
BMI	RMSE	0.023324	0.023244	0.023312	0.023301	0.023307	0.023051	0.023062
	胜率 (%)	51.33	53.33	48.99	44.30	38.93	39.86	38.51
CCFI	RMSE	0.017043	0.016993	0.017070	0.017073	0.017053	0.016927	0.016938
	胜率 (%)	49.33	52.67	46.98	44.30	45.64	49.32	48.65
ECMI	RMSE	0.029140	0.029098	0.029209	0.029252	0.029227	0.029221	0.029227
	胜率 (%)	49.33	48.67	52.35	48.32	50.34	52.70	46.62
NMFI	RMSE	0.028048	0.028079	0.028143	0.028174	0.028136	0.028223	0.028219
	胜率 (%)	45.33	46.00	44.97	44.30	43.62	49.32	45.95

资料来源: 天软 同花顺 Wind 华泰期货研究院

表格 9: 随机森林预测 T+1 周表现总结 (回归)

选取因子历史数据长度	RMSE 均值	胜率均值(%)	RMSE 标准差	胜率标准差(%)
T 周	0.023997	48.67	0.006835	1.81
T-1 周至今	0.023950	52.11	0.006848	3.86
T-2 周至今	0.024033	49.89	0.006885	3.14
T-3 周至今	0.024048	46.76	0.006895	3.41
T-4 周至今	0.024031	47.09	0.006896	5.84
T-5 周至今	0.023944	48.76	0.006857	4.57
T-6 周至今	0.023950	46.40	0.006858	4.42

资料来源: 天软 同花顺 Wind 华泰期货研究院

我们再来观察对 T+2 周标的物收益率的预测效果。从使用的历史数据长度上看, 大部分商品板块用**历史一周**的数据对 T+2 周数据进行预测效果最好。从不同板块的预测效果上看, 商品农产品板块、能源化工板块的预测效果最好, 贵金属板块的预测效果最差。同时, 随着选取历史时间变长, RMSE 逐渐增大, 胜率的标准差在使用两周历史数据时达到最小。

通过上述结果，T-1 周因子数据能有效增强 T+1 周预测效果，T 周因子数据能有效增强 T+2 周预测效果，似乎相隔的两周之间具有重要的关联。

表格 10: 随机森林预测 T+2 周表现 (回归)

选取因子历史数据长度		T 周	T-1 周至今	T-2 周至今	T-3 周至今	T-4 周至今	T-5 周至今	T-6 周至今
ACI	RMSE	0.013445	0.013472	0.013478	0.013458	0.013506	0.013511	0.013514
	胜率 (%)	59.33	48.99	46.31	52.35	52.70	51.35	50.68
BBMI	RMSE	0.032837	0.033018	0.033020	0.033027	0.032755	0.032766	0.032753
	胜率 (%)	52.67	48.32	50.34	48.99	50.68	46.62	46.62
BMI	RMSE	0.023111	0.023208	0.023231	0.023242	0.022992	0.023006	0.022983
	胜率 (%)	52.67	46.31	44.30	44.97	43.24	41.89	45.27
CCFI	RMSE	0.016988	0.017074	0.017084	0.017058	0.016932	0.016939	0.016922
	胜率 (%)	51.33	46.98	39.60	41.61	45.27	50.00	50.68
ECMI	RMSE	0.029088	0.029187	0.029261	0.029233	0.029228	0.029235	0.029223
	胜率 (%)	54.00	52.35	46.98	50.34	48.65	46.62	47.97
NMFI	RMSE	0.028093	0.028134	0.028153	0.028112	0.028203	0.028202	0.028201
	胜率 (%)	44.00	51.01	45.64	50.34	50.00	45.95	44.59

资料来源: 天软 同花顺 Wind 华泰期货研究院

表格 11: 随机森林预测 T+2 周表现总结 (回归)

选取因子历史数据长度	RMSE 均值	胜率均值(%)	RMSE 标准差	胜率标准差(%)
T 周	0.023927	52.33	0.006857	4.52
T-1 周至今	0.024015	48.99	0.006889	2.12
T-2 周至今	0.024038	45.53	0.006897	3.23
T-3 周至今	0.024022	48.10	0.006900	3.67
T-4 周至今	0.023936	48.42	0.006864	3.23
T-5 周至今	0.023943	47.07	0.006865	3.04
T-6 周至今	0.023933	47.64	0.006863	2.40

资料来源: 天软 同花顺 Wind 华泰期货研究院

接下来，我们对比深度学习模型的结果。针对每一个商品板块指数，我们都挑选对标的物影响力最强的前 10 个因子作为测试特征。特征重要性判断使用随机森林模型的最小分裂

节点方法，因为该方法并不依赖于随机森林预测结果的好坏，而只和随机森林模型的拓扑结构以及采用的节点分裂方法有关。所以，即使目前使用的因子组数据不具备全面描述标的物价格规律的信息，也可以用该方法从因子组中挑选出相对而言影响力最强的因子。

我们的测试数据如下：

标的物：华泰商品 5 板块指数周度收益率（2010-06 至 2020-09）；

训练集：前 70% 的周度标的物收益率；

测试集：近 3 年约 30% 的周度标的物收益率；

特征集：标的物长/短周期尺度上对应最强影响力的 10 个截面因子；

预测目标：测试深度学习模型的分类能力--下一周标的物涨跌；或同时预测下两周、下三周的标的物涨跌

表格 12：深度神经网络多步预测胜率(%)结果

	预测一步		预测两步		预测三步	
	step 1	step 1	step 2	step 1	step 2	step 3
农产品指数	56	54	53	53	54	53
基本金属指数	58	54	54	54	53	53
黑色建材指数	55	57	58	57	56	58
能源化工指数	56	NAN	NAN	NAN	NAN	NAN
贵金属指数	57	54	55	55	54	55

资料来源：天软 同花顺 Wind 华泰期货研究院

从表中可以看出多步预测效果并未明显降低，这是一般线性模型难以做到的。实际上，我们知道市场波动率随时间延迟而迅速增大，所以金融信号中的高噪音一般会迫使模型的预测效能迅速降低。多步预测时的第一步预测胜率相对只做一步预测的效果一般要有所下降，但黑色建材例外。另外，能源化工多步预测无法获得有效值，可能是由于数据蕴含的预测信息并不充分，未使用更广泛的海外数据或技术指标类型的量价数据等原因。

对比表格 8&10，我们发现随机森林模型无法对贵金属做出有效的涨跌预测，主要原因是贵金属受到海外市场的影响很大，而且交易者结构较复杂，难以用浅度 AI 模型识别价格传导模式。尽管深度学习模型在多步预测的测试中胜率也有所下降，但是依然可以做出一

定程度的涨跌有效预测，我们认为这是深度学习模型能够从数据中挖掘到领先信息，并能识别复杂价格传导模式的重要佐证。

五、 附录

表格 13: 华泰商品，股票因子代码附录

所属板块	因子代码	因子名称
商品宏观因子	cncpi	CPI 因子
	usindex	美元指数因子
商品国家因子	country_com	国家因子
	oi	持仓因子
商品风格因子	termstructure	期限结构因子
	curve	curve 因子
	ma	均价突破因子
	mom	动量因子
	value_com	价值因子
	variance	波动率因子
	skew	偏度因子
	liquidity_com	流动性因子
商品板块因子	ACI	农产品板块因子
	NMFI	贵金属板块因子
	BMI	基本金属板块因子
	ECMI	能源化工板块因子
	BBMI	黑色建材板块因子
股票国家因子	country	国家因子
股票风格因子	beta	beta 因子
	value	价值因子
	growth	成长因子
	momentum	动量因子
	size	规模因子
	leverage	杠杆因子

股票板块因子	earnings	收益因子
	liquidity	流动性因子
	nsize	非线性规模因子
	dividend	分红因子
	rici	国际商品因子
	bab	低波因子
	residualvolatility	残差波动率因子
	energy	能源板块因子
	materials	材料板块因子
	industrials	工业板块因子
	consumer_discretionary	可选消费板块因子
	consumer_staples	必须消费板块因子
	health_care	健康医疗板块因子
	financials	金融板块因子
	information	信息板块因子
	telecom	电信板块因子
	utilities	公共事业板块因子
	real_estate	房地产板块因子

资料来源：华泰期货研究院

● 免责声明

此报告并非针对或意图送发给或为任何就送发、发布、可得到或使用此报告而使华泰期货有限公司违反当地的法律或法规或可致使华泰期货有限公司受制于的法律或法规的任何地区、国家或其它管辖区域的公民或居民。除非另有显示，否则所有此报告中的材料的版权均属华泰期货有限公司。未经华泰期货有限公司事先书面授权下，不得更改或以任何方式发送、复印此报告的材料、内容或其复印本予任何其它人。所有于此报告中使用的商标、服务标记及标记均为华泰期货有限公司的商标、服务标记及标记。

此报告所载的资料、工具及材料只提供给阁下作查照之用。此报告的内容并不构成对任何人的投资建议，而华泰期货有限公司不会因接收人收到此报告而视他们为其客户。

此报告所载资料的来源及观点的出处皆被华泰期货有限公司认为可靠，但华泰期货有限公司不能担保其准确性或完整性，而华泰期货有限公司不对因使用此报告的材料而引致的损失而负任何责任。并不能依靠此报告以取代行使独立判断。华泰期货有限公司可发出其它与本报告所载资料不一致及有不同结论的报告。本报告及该等报告反映编写分析员的不同设想、见解及分析方法。为免生疑，本报告所载的观点并不代表华泰期货有限公司，或任何其附属或联营公司的立场。

此报告中所指的投资及服务可能不适合阁下，我们建议阁下如有任何疑问应咨询独立投资顾问。此报告并不构成投资、法律、会计或税务建议或担保任何投资或策略适合或切合阁下个别情况。此报告并不构成给予阁下私人咨询建议。

华泰期货有限公司2019版权所有。保留一切权利。

● 公司总部

地址：广东省广州市越秀区东风东路761号丽丰大厦20层

电话：400-6280-888

网址：www.htfc.com