

## 商品基本面量化系列（二）：铁矿石多因子策略探索

国投安信期货研究院

王锴 期货投资咨询号 Z0016176

在上一篇系列报告《商品量化基本面(一)：甲醇多因子探索中》中我们挖掘并梳理了甲醇产业链逻辑下的因子，通过合成时序和机器学习等方法构建了甲醇多因子择时策略，近一个季度的样本外跟踪表现较为稳定，同时模型信号与基本面判断存在印证关系。因此，我们又延续基本面量化的思路，将研究方法在新的品种上进行拓展。

铁矿石作为黑色系上市时间较早的品种，市场参与结构相对成熟完善，波动率也比较大，具有良好的交易特性。从基本面数据考虑，铁矿石产业用途集中，98%作为钢铁生产原材料使用，从而价格与下游钢铁行业连接紧密，数据的获取与更新来源较为清晰。同时，铁矿石作为黑色产业重要原材料，螺纹钢与热卷的上游，其中的价格变动也存在较强相关性。随着近两年黑色系价格波动逐渐放大，铁矿石基本面的量化研究也为黑色和商品研究和投资提供参考和借鉴。

在之前的量化基本面研究中，我们运用多种机器学习模型来进行拟合和预测，因子探索方面取得了一定的进展，但是我们仍然会以尊重主观基本逻辑为前提，以合理的量化手段，建立多维度模型加以比较和论证。本文中我们梳理了影响铁矿石价格的产业链逻辑，深度挖掘影响价格的基本面因子。我们着重分析了不同频率数据之间嵌套组合的顺序方法，并且进行了数据周期性变化的测试。

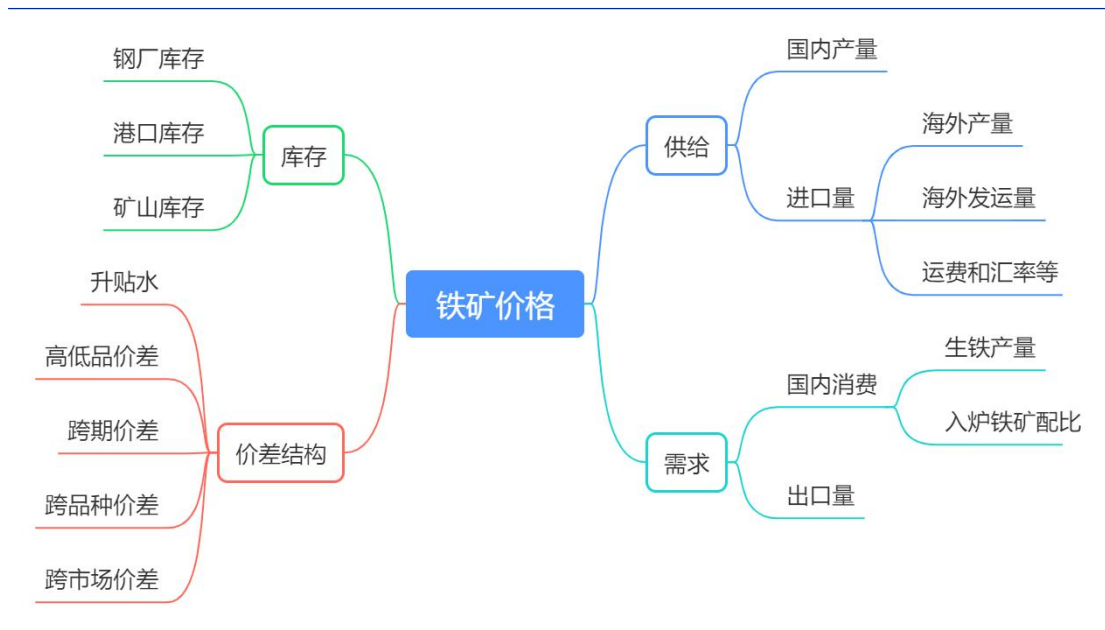
实证研究显示，周频策略相对日频策略测试表现更好。日频基本面时序策略中的供给因子策略达到夏普 1 以上，其余因子在测试区间内均有不同程度的回撤，影响了收益风险比。绝对收益方面，在铁矿石价格大幅波动区间内的 21-22 年，供给因子胜率达 60%以上，收益贡献占比较高，对于看涨行情的抓取能力较强。风险控制方面，价差因子表现较为稳健，年化开仓次数 76 次，相对较少。库存因子样本内外有一定差异，看涨区间内出现较大回撤。以窗口期胜率权重合成后风险收益比有所提升，交易频次增加，平均持仓天数 2 天。

从周频策略的表现来看，整体收益和风险控制能力均显著优于日频策略，价差和库存因子收益能力突出，其中库存因子风险控制能力更佳，在震荡与下行趋势内表现相对更好，夏普比达到 2 以上。表现相对较差的是需求因子，在整体开仓次数较少的情况下，在 2022 年上半年终端需求较差的行情下，单因子有显著回撤。选取表现较好的价差、库存和供给三因子合成后的策略平滑了价差和库存因子出现的回撤区间，年化开仓次数近 20 次，夏普比 2.3。

## 一、基本面相关因子

我们将铁矿价格的基本面因子分为四大类：需求因子、供给因子、库存因子以及价差因子。

图：铁矿石指标结构分类



资料来源：Wind，国投安信期货

**需求因子：**铁矿的下游需求主要来自于钢厂的高炉炼铁需求，一般来说我们以生铁的产量和铁矿的日耗来衡量铁矿的消费水平，而影响这两者的主要因素是钢厂的利润情况以及国家政策层面的影响。对于钢厂来说，利润越高其生产积极性越强，这会拉动铁矿的需求，不过在双碳目标的指导下，钢厂生产也会受到政策端的制约，特别是在秋冬季不利于污染物扩散的时候，限产执行力度会相对较强。

**供给因子：**铁矿的供给分为进口量和国内产量，而由于我国铁矿的进口依赖度比较高并且期货盘面标的又主要以海外粉矿为主，所以海外进口是我们主要关注的供给因子。海外铁矿供给我们会分为四大矿山和中小矿山，其中四大矿山由于体量较大并且数据可得性较高，所以我们会重点关注。对于海外供给我们主要从铁矿的产量和发运量来分析，而这两者又会受到海外铁矿生产成本、汇率和运费等因素的影响。国产矿这边我们也会关注铁精粉的产量和产能利用率等数据。

库存因子：铁矿的库存我们主要关注的是国内的港口库存和钢厂的库存数据。除了库存的绝对量反应了铁矿的整体供需水平外，库存的消费比也是我们需要关注的重要数据，这反应了钢厂能够维持目前生产强度多少天的情况。此外港口库存的结构、疏港、到港和压港数据也是我们判断铁矿供需水平的重要依据。

价差因子：价差因子主要包括了期限价差、跨品种价差和跨市场价差等。因为黑色产业链上下游联系紧密，所以各品种之间联动性较强。一般来说钢厂利润较高也会反应在钢材价格的走高上，而在利润较高的同时，钢厂对于原料价格也存在着较高的接受程度。此外原料之间的价差也会相互影响，当焦炭价格较高的时候，会压制铁矿的价格并影响铁矿品种的使用。最后海外掉期价格也会对国内价格产生影响。

## 二. 数据清洗

### 2.1 数据频率的处理

数据频率方面，我们选择日频和周频数据，更低频次的数据由于样本数量过少不利于拟合，且发布时间往往滞后不足以支持样本外实际跟踪。第一种方法是将周频数据映射为日频数据，但是这里并没有采用插值法来直接估算实际数据未发布前的数据，而是统一采用原数据，因为这样可以匹配更为真实客观的对应关系。第二种方法是统一将所有数据低频化处理，这主要是考虑到铁矿石所获取的基本面数中周频占比超过 80%，对应的信号变化频率也偏向周度层面的变化，因此考虑将周频模型的表现与日频模型进行比较，同时考虑周期环比和均线平滑来处理这些数据。

### 2.2 数据滞后性处理

滞后性处理：日频数据基本都会在交易日当天收盘更新，而周频数据基本在周五发布。因此我们在做特征有效性检测和建模过程中统一将这些数据延后一阶处理，即当天获得的数据生成信号在下一个交易日进行交易。

### 2.3 去极值和标准化处理

数据中的极值往往会影响在线性模型假设下的拟合，同时对于映射关系也存在错误推断的可能，因此我们将高于时序数据 3 倍标准差的数值调整为平均值 $\pm 3$  倍标准差。此外，所有的基本面因子采用 W 检验的方式处理其数据的正态关系，通过 P-value 值观

测，检验中只有少量因子存在相对正态分布关系，因此集中处理其影响模型预测精度的异常值并按照各个大类因子数据特征进行标准化处理，使得大类数据和甲醇价格数据处于同一假设关系。

整体数据选取 2017 年 1 月-2022 年 6 月年基本面数据，进行差值处理和数据填充整合，分别采用相关性、平稳性检验、回归测试单因子有效性，进而设计基本面因子指导的交易策略，

图：特征筛选排序

			IC	RankIC	param	pvalue	tvalue	adjR2
供给	周度	铁矿：发货量：巴西：淡水河谷	-0.019	-0.028	0.000	0.458	-0.742	0.000
		澳洲&巴西铁矿石：发货量合计 14 个港口	-0.019	-0.036	0.000	0.460	-0.740	0.000
		铁矿：到港量：北方港口	-0.010	0.011	0.000	0.692	-0.396	-0.001
		126 家矿山企业：产能利用率	-0.018	0.006	0.000	0.501	-0.673	0.000
		126 家矿山企业：铁精粉：产量：中国	-0.020	0.001	-0.001	0.436	-0.779	0.000
		247 家钢铁企业：高炉开工率：中国	0.019	0.026	0.001	0.473	0.717	0.000
		247 家钢铁企业：高炉炼铁产能利用率：中国	0.006	0.016	0.000	0.820	0.227	-0.001
		唐山钢厂产能利用率	0.004	0.004	0.000	0.865	0.170	-0.001
		247 家钢铁企业：盈利率：中国	-0.039	-0.027	-0.001	0.140	-1.477	0.001
	日度	铁矿石运价巴西图巴朗-青岛	-0.055	-0.048	-0.002	0.036	-2.099	0.002
		铁矿石运价西澳-青岛	-0.062	-0.046	-0.002	0.018	-2.369	0.003
		即期汇率美元兑人民币	0.008	0.002	0.000	0.748	0.322	-0.001
		汇率	0.010	0.002	0.000	0.703	0.382	-0.001
需求	周度	进口铁矿：日均消耗量：247 家钢铁企业	0.051	0.047	0.002	0.049	1.971	0.002
		国产烧结粉：日均消耗量：64 家钢厂	-0.007	0.018	0.000	0.775	-0.286	-0.001
		进口烧结粉：日均消耗量：64 家钢厂	0.053	0.047	0.001	0.043	2.024	0.002
		247 家钢铁企业：铁水：日均产量：中国	0.020	0.027	0.001	0.439	0.774	0.000
		247 家钢铁企业：高炉开工率：中国	0.019	0.026	0.001	0.473	0.717	0.000
		247 家钢铁企业：高炉炼铁产能利用率：中国	0.006	0.016	0.000	0.820	0.227	-0.001
		唐山钢厂产能利用率	0.004	0.004	0.000	0.865	0.170	-0.001
		247 家钢铁企业：盈利率：中国	-0.039	-0.027	-0.001	0.140	-1.477	0.001
	日度	生产利润	-0.025	-0.050	-0.001	0.346	-0.944	0.000
		螺纹利润	-0.003	-0.023	0.000	0.912	-0.111	-0.001
		钢坯利润	-0.021	-0.032	-0.001	0.417	-0.812	0.000
		热轧利润	-0.018	-0.031	-0.001	0.501	-0.673	0.000
		冷轧利润	-0.032	-0.042	-0.001	0.228	-1.206	0.000
		生产成本	-0.009	0.016	0.000	0.722	-0.355	-0.001
价差	日度	.5%PB 粉品牌价格：日照港	-0.013	0.016	0.000	0.626	-0.488	-0.001
		DCE：铁矿：主力合约：收盘价	0.030	0.050	0.001	0.245	1.164	0.000
		DCE：焦炭：主力合约：收盘价	-0.007	0.011	0.000	0.800	-0.254	-0.001

		SHFE: 螺纹钢: 主力合约: 收盘价	-0.013	0.012	0.000	0.628	-0.484	-0.001
		DCE: 铁矿: 合约: 收盘价	0.021	0.040	0.001	0.421	0.805	0.000
		DCE: 铁矿: 05 合约: 收盘价	0.022	0.036	0.001	0.390	0.861	0.000
		DCE: 铁矿: 09 合约: 收盘价	0.008	0.032	0.000	0.751	0.317	-0.001
		DCE: 焦炭: 合约: 收盘价	-0.002	0.013	0.000	0.948	-0.065	-0.001
		DCE: 焦炭: 05 合约: 收盘价	-0.012	0.003	0.000	0.639	-0.470	-0.001
		DCE: 焦炭: 09 合约: 收盘价	-0.019	0.007	-0.001	0.475	-0.715	0.000
		SHFE: 螺纹钢: 合约: 收盘价	-0.015	0.005	0.000	0.554	-0.592	0.000
		SHFE: 螺纹钢: 05 合约: 收盘价	-0.022	0.002	-0.001	0.394	-0.853	0.000
		SHFE: 螺纹钢 0 合约: 收盘价	-0.018	0.001	0.000	0.485	-0.699	0.000
		普氏铁矿石价格指数 62%FeCFR 中国北方	0.003	0.031	0.000	0.920	0.101	-0.001
		市场价废钢 6-8mm 唐山	-0.030	-0.011	-0.001	0.257	-1.135	0.000
		价格螺纹钢 HRB400 20mm 石家庄	-0.039	-0.021	-0.001	0.138	-1.483	0.001
		价格螺纹钢 HRB400 20mm 上海	-0.033	-0.011	-0.001	0.205	-1.268	0.000
		期货收盘价螺纹钢	-0.014	0.008	0.000	0.605	-0.517	-0.001
		价格方坯 Q235 唐山	-0.037	-0.015	-0.001	0.160	-1.405	0.001
		价格热轧板卷 4.75mm 天津	-0.028	-0.016	-0.001	0.277	-1.087	0.000
		价格中板普 20mm 上海	-0.037	-0.024	-0.001	0.160	-1.407	0.001
		铁矿石普氏	0.002	0.029	0.000	0.938	0.077	-0.001
		焦炭	-0.040	-0.028	-0.002	0.129	-1.521	0.001
		废钢	-0.029	-0.011	-0.001	0.261	-1.125	0.000
		螺纹钢现货	-0.032	-0.011	-0.001	0.221	-1.224	0.000
		螺纹钢期货	-0.014	0.007	0.000	0.583	-0.549	0.000
库存	周度	全国主要港口铁矿石库存总计	-0.025	-0.041	-0.001	0.330	-0.974	0.000
		126 家矿山企业: 铁精矿: 库存: 中国	0.009	-0.019	0.000	0.726	0.350	-0.001
		铁矿: 进口: 库存: 45 个港口	-0.026	-0.041	-0.001	0.325	-0.984	0.000
		进口铁矿石平均库存可用天数国内大中型钢厂	-0.001	0.015	0.000	0.956	-0.055	-0.001
	日度	铁矿: 进口: 库存: 45 个港口	-0.026	-0.041	-0.001	0.325	-0.984	0.000

资料来源: Wind, 国投安信期货

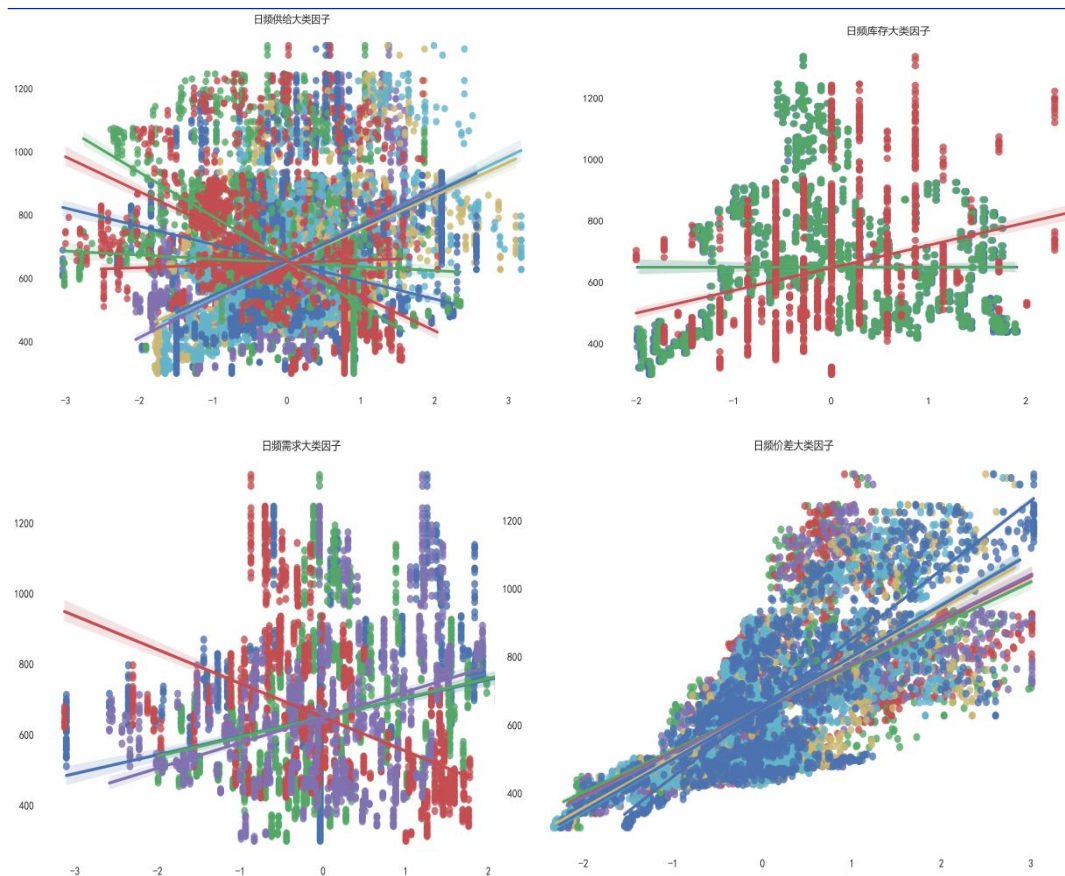
供给因子中澳洲巴西 14 港口发货量胜率较高。我国铁矿石进口主要来自于澳洲和巴西, 两者合计占我国总进口量的 8 成以上, 而我国铁矿的进口依赖度也达到 8 成左右, 所以澳巴铁矿的发货量对于我国铁矿的短期供给存在重要影响。需求因子中进口烧结粉的日均消耗量胜率较高。我国铁矿主要依赖进口, 而期货盘面标的又以进口粉矿为主, 而粉矿的日均消耗量能直观的显示目前钢厂的生产强度, 所以该因子相对更加重要并有效。

库存因子中全国港口库存胜率较高。铁矿港口库存的数据综合反应了钢厂和贸易商等多个市场参与者的库存水平, 是能直观反应国内铁矿总供需的重要数据。价差因子中



钢材和焦炭的价格胜率较高。黑色产业链的利润分配主要集中在上游原材料中的铁矿和焦炭以及下游的成材之间。当钢材价格较高且利润偏强的时候，钢厂对于原材料价格的接受程度也高，焦炭和铁矿的走势也会偏强并同向波动。

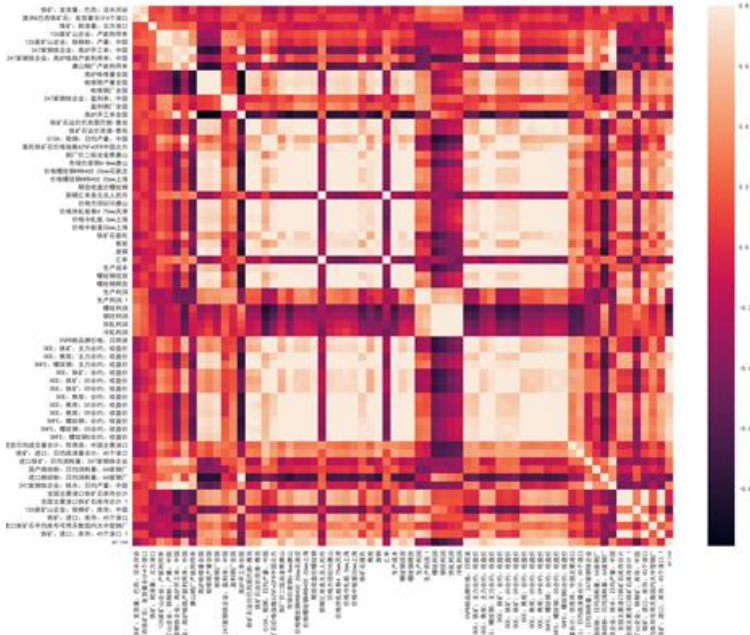
图：因子一元线性拟合效果



资料来源：Wind，国投安信期货

通过计算信息系数 IC，我们可以了解因子值对于下期铁矿石收益率的预测能力。从整个单因子线性回归结果看，需求因子和价差因子对于价格影响较大，相关性也更为显著，同时日频数据数据量大于周度数据整体相关性水平更高。结合特征筛选指标，我们可以得到 32 个重要因子。

图：IC 相关性



资料来源：Wind，国投安信期货

### 三. 基本多空策略周期信号的嵌套和比较

#### 3.1 时序多空模型

我们将筛选得到的供给、需求、库存和价差四大类中的每个因子，从时间序列维度出发，根据 IC 正负关系，遍历参数得到单因子开仓阈值，随后我们将这些单因子开仓阈值在每个大类中进行加权合成，从而产生大类因子的开仓信号。

$$Class_j = c_1 \dots c_i \dots c_n$$

$$c_i = \begin{cases} 1 & c_i > percentile_a \quad a \in [lower, upper] \\ 0 & other \\ -1 & c_i < percentile_b \quad b \in [lower, upper] \end{cases}$$

$$c_i = \begin{cases} 1 & c_i > Class\_len_a \quad a \in [lower, upper] \\ 0 & other \\ -1 & c_i < Class\_len_b \quad b \in [lower, upper] \end{cases}$$

$$j \in \{supply, demand, stock, index\}$$

$$OR = \frac{p_1 / (1 - p_1)}{p_2 / (1 - p_2)}$$



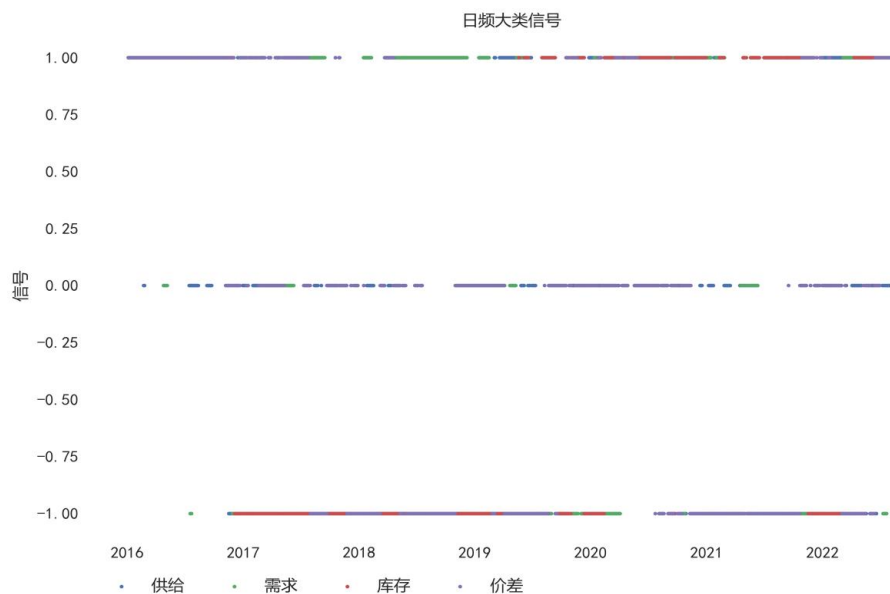
### 3.1.1 日频信号

铁矿石日内价格波动较大，因此从信号生成和调仓频率角度出发，我们优先考虑将周度数据高频化处理，与日度数据合成日频信号。由于周度数据在整体筛选后的指标中占比较大，需要同时观察高频化后信号的变化和调仓频率，以及对于周度数据运用插值法进行处理的合理性。在比较了样本内运用前值填充和梯度线性插值方法后，我们发现尽管前值填充使信号变化频率降低，梯度线性填充并不能很好地降低信息完整性的缺失。因此，我们仍然采用前值填充的方法。

在进行分类后选取 IC 值存在有效性的因子做阈值筛选，并按照 IC 的正负相关性方向作为信号的变化方向，当与标的处于正相关关系的时候同方向变化，反之信号反方向变动。我们使用 Odds Ratio 作为筛选指标对每个大类下的因子进行遍历。我们通过排序选取 OR，从而增加合成开仓信号的胜率大小，发现最终开仓阈值的结果处于 0.9-1.3 之间。对于不同的因子的最高 OR 筛选的阈值进行合成后折算成不同因子相应的胜率，并按照胜率大小对不同大类下的合成因子进行加权。通过单因子净值的统计指标分析，我们删去卡玛比率小于 0.3 的大类因子，供给大类、需求大类、库存大类以及价差大类单因子的平均胜率分别为 0.24，0.27，0.36 以及 0.34 左右，同时各个大类的平均年化收益集中在 47.22%、48.54%、24.98% 以及 38.30% 左右。从大类单因子入手观察对于各个大类的收益变化影响，我们可以看到，对于供给大类来说澳洲和巴西铁矿石以及高炉开工率对于策略收益有最为显著的提振作用。从需求角度来看，其中收益贡献最大的是进口烧结粉，而国产烧结粉因子对于需求大类的收益存在扰动。库存角度出发，其中国内以及进口铁矿石对于净值存在相似的影响幅度，

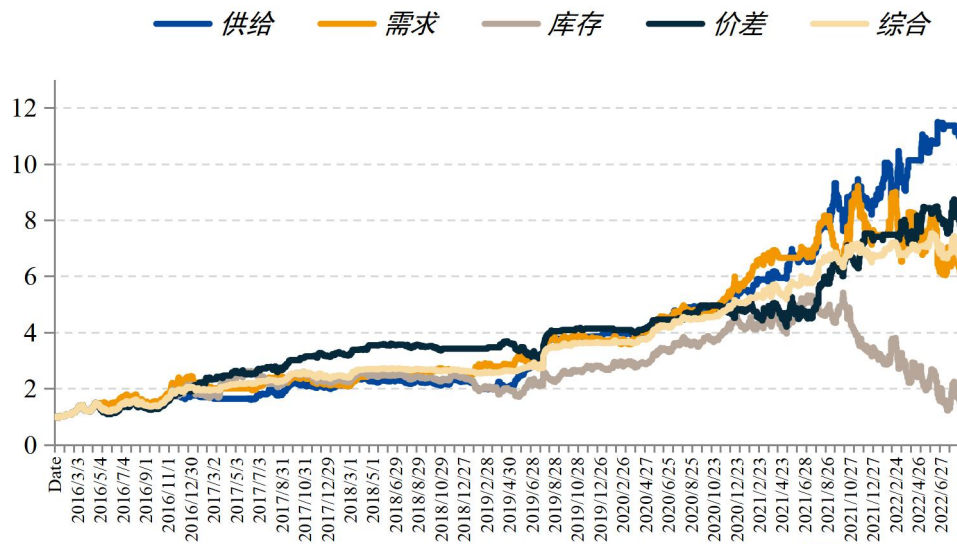
库存大类中的因子均在 2019 年到 2020 上半年未达到历史开仓阈值，净值趋于平稳。价差大类在四个大类中波动最大，从训练集角度出发，波动变化集中在 2020 年下半年以及 2021 年上半年，各个价差因子均出现相同程度的扰动，拉低整体的价差大类因子的净值。从大类因子角度出发，2021 年 7 月日频策略净值进入高波动阶段，库存因子净值开始下滑，下拉综合策略的整体收益。从历史数据观察，2021 年 7 月开始铁矿石价格从波动底部开始回暖，但是市场库存量开始累积，成交量持续下滑，库存因子信号偏空头磨损部分收益。同一阶段需求因子净值震荡走强，并出现大幅回撤，供给因子收益相对表现良好，提振综合收益效果。

图：大类因子开仓信号分布



资料来源：Wind，国投安信期货

图：基本面因子净值曲线



数据来源：wind，国投安信期货

图：策略收益表现对比

	供给	需求	库存	价差	大类合成
累积收益%	990.88	512.53	46.27	681.32	557.75
年化收益%	144.53	74.76	6.75	99.38	81.35
最大回撤%	170.18	317.43	419.70	96.34	95.20
夏普比率	1.05	0.46	0.06	0.93	1.14
卡玛比率	0.85	0.24	0.02	1.03	0.86
年化开仓次数	87.51	106.18	112.75	76.28	116.25

资料来源：Wind，国投安信期货

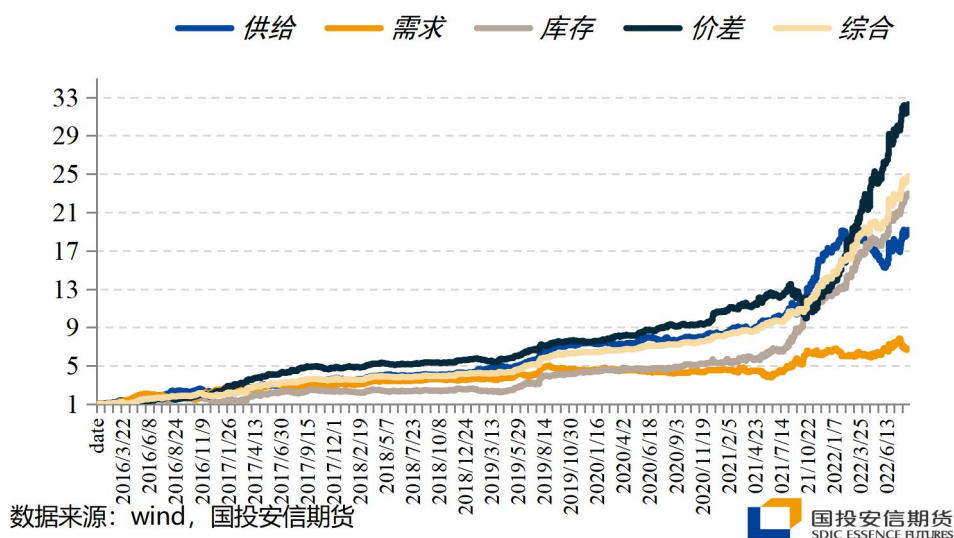
### 3.1.2 周频信号

将周频因子进行高频化处理必然会造成信息缺失，如果以降低调仓频率为成本，将日频进行低频化处理，可以更好地保留信息的完整性。同时，在周度层面上也更便于检验分析数据在同比、环比等时间周期方面的有效性。

周度数据整体预测表现效果从稳定性以及收益情况来说优于日频策略，尤其是在因子回撤的控制上。从整体数据角度出发，铁矿石周度数据量占比显著超过日度数据，尤其体现在需求大类。相比于日度数据，周度数据的 OR 值相对跨度更大，位于 0.5-3.6 之间。从年化收益的角度出发，周度数据的供给大类、需求大类、库存大类以及价差大类单因子的平均收益集中在 37.50%、25.22%、62.72%、65.75% 左右。对比周度和日度数据，库存因子周频平均年化收益是日频 3 倍，而价差大类因子日频是周频策略的 2 倍。从单因子层面出发，供给因子中收益贡献最高的是普氏铁矿石价格指数，而在 2020 年期间供给因子出现冲高时唐山废钢和方坯进入平台期。需求大类方面，提供最多收益贡献的是进口铁矿的日均消耗量，而进口烧结粉则在 2021 年拉动总体收益大幅下滑。库存因子中铁精矿库存因子净值在 2021 年同样伴随价格出现波动性回撤，而铁矿石平均

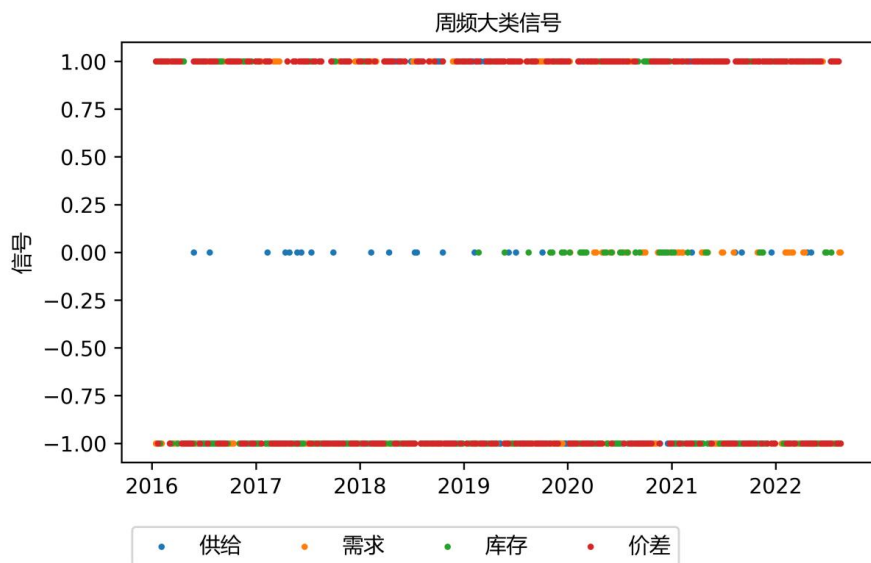
可用天数因子净值走势相对平滑，提振大类因子整体收益表现。价差因子在周度策略上均表现较佳，价差各因子平均回撤小于其他因子，除去焦煤 09 合约因子在 2021 年末大幅回调外，各价差因子净值走势平稳。从总体大类因子的角度观察，价差因子年化收益最高拉动大类合成因子的攀升，而需求相对表现较弱净值位于低位区间震荡。在 2022 年中旬，供给因子出现大幅回撤同时库存因子收益出现小幅下滑，但总体表现而言，库存因子净值相对波动较小上行趋势显著。

图：基本面因子净值



资料来源：Wind，国投安信期货

图：大类因子开仓信号分布



资料来源: Wind, 国投安信期货

图：策略收益表现对比

	供给	需求	库存	价差	大类合成
累积收益	1797.76	577.77	2183.61	3103.17	2361.51
年化收益	264.69	85.07	321.50	456.89	347.90
最大回撤	402.00	122.87	109.99	356.42	75.37
夏普比率	1.33	0.83	2.00	1.78	2.30
卡玛比率	0.66	0.69	2.92	1.28	4.62
年化开仓次数	13.57	10.63	10.51	17.30	18.09

资料来源: Wind, 国投安信期货

## 四、实证分析

1. 从短周期的日频基本面时序策略表现来看，供给因子策略达到夏普 1 以上，其余因子在测试区间内均有过不同程度的大幅回撤，影响了收益风险比。从绝对收益角度看，在铁矿石波动较大的区间内，供给因子胜率达 60%以上，收益贡献占比较高，对于看涨行情的抓取能力较强。从风险控制角度看，价差因子表现较为稳健，年化开仓次数 76，相对较少。库存因子样本内外有一定差异，看涨区间内出现较大回撤。以窗口期胜率权重合成后风险收益比有所提升，交易频次增加，平均持仓天数 2 天。
2. 从周频策略的表现来看，整体收益和风险控制能力均显著优于日频策略，价差和库存因子收益能力突出，其中库存因子风险控制能力更佳，在震荡与下行趋势内表现相对更好，夏普比达到 2 以上。表现相对较差的是需求因子，在整体开仓次数较少的情况下，在 2022 年上半年终端需求较差的行情下，单因子有显著回撤。选取表现较好的价差、库存和供给三因子合成后的策略平滑了价差和库存因子出现的回撤区间，年化开仓次数近 20 次，夏普 2.3。



## 免责声明

本研究报告由国投安信期货有限公司撰写,研究报告中所提供的信息仅供参考。报告根据国际和行业通行的准则,以合法渠道获得这些信息,尽可能保证可靠、准确和完整,但并不保证报告所述信息的准确性和完整性。本报告不能作为投资研究决策的依据,不能作为道义的、责任的和法律的依据或者凭证,无论是否已经明示或者暗示。国投安信期货有限公司将随时补充、更正和修订有关信息,但不保证及时发布。对于本报告所提供信息所导致的任何直接的或者间接的投资盈亏后果不承担任何责任。

本报告版权仅为国投安信期货有限公司所有,未经书面许可,任何机构和个人不得以任何形式翻版、复制和发布。如引用发布,需注明出处为国投安信期货有限公司,且不得对本报告进行有悖原意的引用、删节和修改。国投安信期货有限公司对于本免责声明条款具有修改权和最终解释权。