

2024 年 3 月 13 日



## 因子与指数投资揭秘系列十六：铁矿石基本面量化及数据分析的研究框架

虞堪

投资咨询从业资格号：Z0002804

[yukan010359@gtjas.com](mailto:yukan010359@gtjas.com)

高宇飞（联系人）

从业资格号：F03124155

[gaoyufei028920@gtjas.com](mailto:gaoyufei028920@gtjas.com)

### 报告导读：

在之前的系列专题报告中，我们阐述了基本面量化数据选取的一些一般性的标准（参见《因子与指数投资揭秘系列十五：探究量化基本面因子品种和数据的选取方法》）。并提出，以黑色和能源化工产业链为主的较为成熟的工业品，更适合做量化基本面研究。首先，它们上下游逻辑清楚，可量化的基本面数据和指标较多。其次，供需的逻辑可以沿着产业链上下游进行传导。最后，许多工业品上市时间较早，具有较长的回测数据，统计意义更显著。

在这些品种当中，铁矿石是一个值得研究的期货品种。自 2013 年在大连商品交易所上市以来，历经十余载时间，历史数据相对充足。此外，在铁矿石的产业链条中，铁矿、螺纹钢、热卷、焦煤、焦炭等品种，互为上下游。此外近年来，锰硅、硅铁、不锈钢等品种陆续上市交易，进一步完善了这条产业链。然后，我国铁矿的供给较多依赖于外部进口，例如依赖从巴西四大矿山和澳洲等地进口，其发货量、船运、到港量等因素会对价格产生影响，这些指标较易于量化。铁矿石有较多不同品质的矿块和矿粉，这些高低品粉块之间的价差变化，通常也会对后市期货价格提供一些参考。最后，钢厂一些指标，例如钢厂铁水产量、钢厂开工率等，可以作为需求类指标，对后市期货价格的预测提供参考。

本文将首先按照类别，对铁矿石各相关指标进行梳理和介绍。然后，选择表现较好的指标并对其回测绩效进行结果展示，最后再结合结果以及近期铁矿石的行情进行相关的总结。

特别感谢实习生李泽豪对于本文撰写、数据搜集的贡献。

**风险提示：**近期铁矿石行情波动较大，注意关注政策面和情绪面的影响。

(正文)

## 1. 铁矿石产业链的介绍

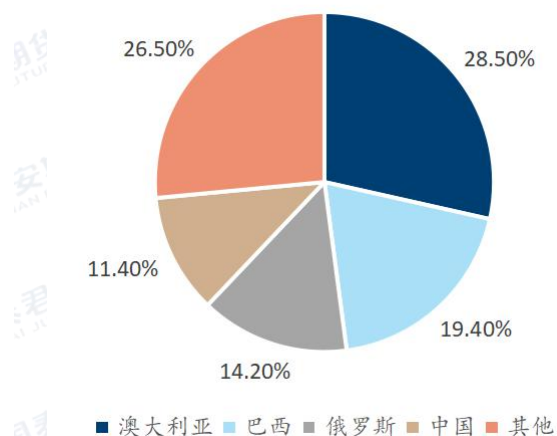
现代工业中，钢材是使用最为广泛的金属材料之一。作为生产钢材的重要原材料，铁矿石的地位举足轻重。铁矿石用途单一，98%作为钢铁生产原材料使用，1吨生铁约需要1.6吨铁矿石。因为铁矿石用途单一的这种特性，给我们在分析铁矿需求时带来了一些便利，我们并不需要像一些种类繁多的工业品那样，逐个统计下游需求量（容易产生误差并且效率较低），相反，我们只需要考虑其作为炼钢原料去生产钢材就足够了，其市场需求与钢铁行业的发展状况紧密相关。近年来，由于我国和印度等新兴经济体的蓬勃发展，钢材需求量在逐年提高，因而对于铁矿石的需求量也在日益提高。

而从供给端看，铁矿石主要分布于澳大利亚、巴西、俄罗斯和中国四个国家，占全球铁矿石总储量的70%以上。虽然我国铁矿石的储量较大，但由于矿石品质较低，高品质的铁矿石主要分布于澳大利亚和巴西，加上我国对钢材的需求也逐年增加，因此从供给端来看，我国铁矿石主要依赖于外部进口。特别是近年来，进口渠道逐渐集中于澳洲和巴西发运，而从其余国家的发运比例则逐年减少。

从矿山的发运数据看，全球铁矿石的发运主要集中于四大矿山，2023年发运量从高到低依次是：力拓(RT)，3.2亿吨；淡水河谷(Vale)，3.15亿吨；必和必拓(BHP)，2.85亿吨；福蒂斯丘，1.92亿吨。这些矿山的发运数据对于铁矿石价格走势也有一定的影响。

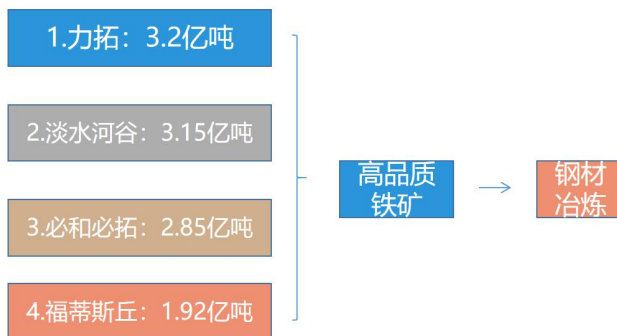
综上，我们不难看出，铁矿石的产业链上下游较为清晰明确，开采端主要依赖从澳洲和巴西四大矿山进口，需求端主要应用于钢材的制造和冶炼。这样的产业链分布也给我们提取基本面数据并进行量化信号的建模提供了便利。

图 1：铁矿石各国储量分布图



资料来源：国泰君安期货研究

图 2：四大矿山发运量排名



资料来源：国泰君安期货研究

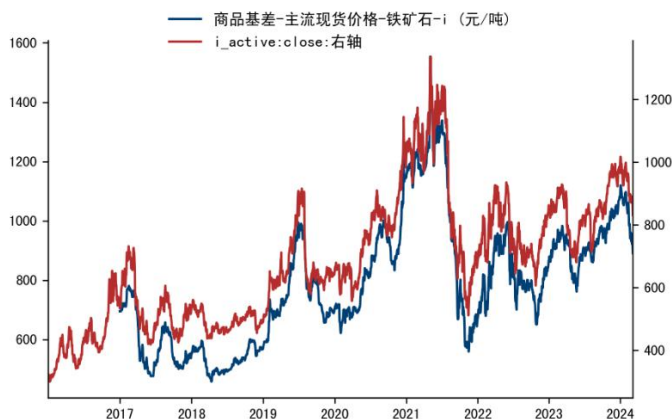
## 2. 铁矿石基本面数据举例

接下来，我们将分类列举一些基本面数据。依照《因子与指数投资揭秘系列十五：探究量化基本面因子品种和数据的选取方法》文中提到的基本面数据选择标准，所选取数据应尽可能具有较长的历史长度，较频繁的更新频率（日度或周度），以及有较强的代表性。

### 2.1 铁矿石基差

本数据选自“商品基差-主流现货基差-铁矿石-i”，数据源为国泰君安期货数据库，更新频率为日度。主要选择期货贴水较大时择时做多，因为基差走扩，期货价格往往被低估，下一步可能反转走强；反之当期货价格升水较大时择时做空，由于商品交割机制的存在，基差往往会在临近交割时收敛。

图 3：铁矿石价格和基差走势



资料来源：国泰君安期货研究

可以看出，铁矿石期货长期贴水于现货，基差始终大于零。从 2016 年到 2021 年若不择时，长期做多铁矿石也可以带来可观的收益。但需要指出，在 2022 年及之后的一段时间，随着各种事件和政策的影响，铁矿石波动加剧，趋势性不及以前好捕捉，仅做多的策略恐将不再适用。

### 2.2 铁矿石现货价格

由于不同产地、不同品质的铁矿石种类繁多，因此铁矿石现货种类价格较为繁多。我们挑选以下几种现货价格数据：

低品质：58.5%混合粉品牌价格：青岛港（日）；超特粉；

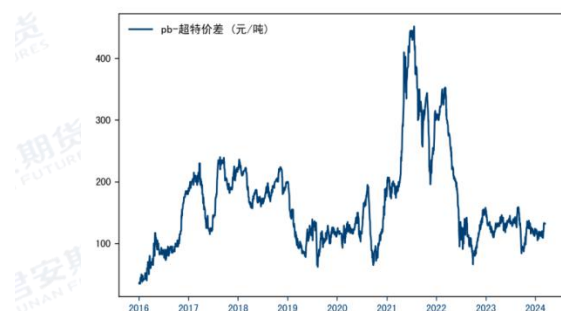
中品质：61.5%PB 粉品牌价格：青岛港（日）

高品质：65%卡拉加斯粉品牌价格：青岛港（日）；巴混粉；

以上数据均来自于钢联，每日更新。不同品质的矿粉之间有价差，例如我们可以构造 PB 粉-混合粉，PB 粉-超特粉，卡粉-超特粉（或混合粉），巴混-超特粉（或混合粉）之间的价差因子。这些价差可能会对后市价格有一定的先导预测作用。其主要逻辑是，在假定供应不变的情况下，铁矿的需求应先从价格、品味较低的矿粉开始，当价差扩大，说明需求已经延申到高品质矿粉，导致其价格走强。因此价差扩大反映需求端的走强。

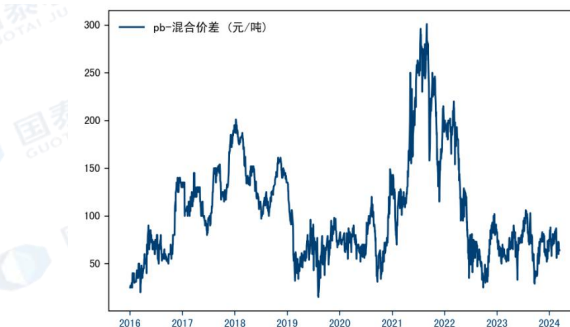


图 4：PB-超特粉价差



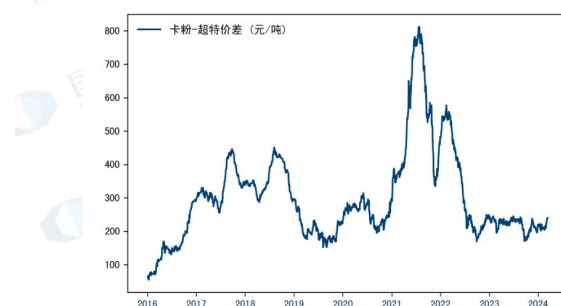
资料来源：国泰君安期货研究

图 5：PB-混合粉价差



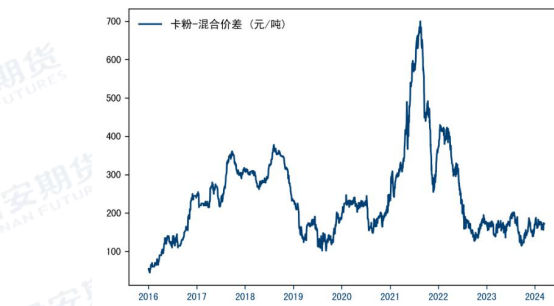
资料来源：国泰君安期货研究

图 6：卡粉-超特粉价差



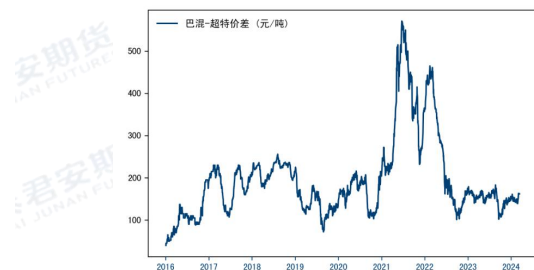
资料来源：国泰君安期货研究

图 7：卡粉-混合粉价差



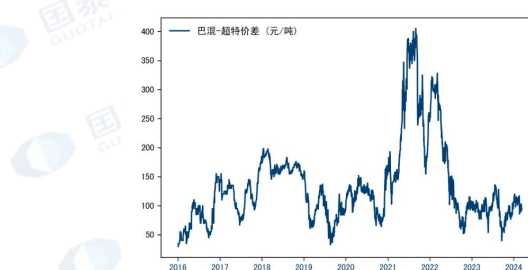
资料来源：国泰君安期货研究

图 8：巴混-超特粉价差



资料来源：国泰君安期货研究

图 9：巴混-混合粉价差



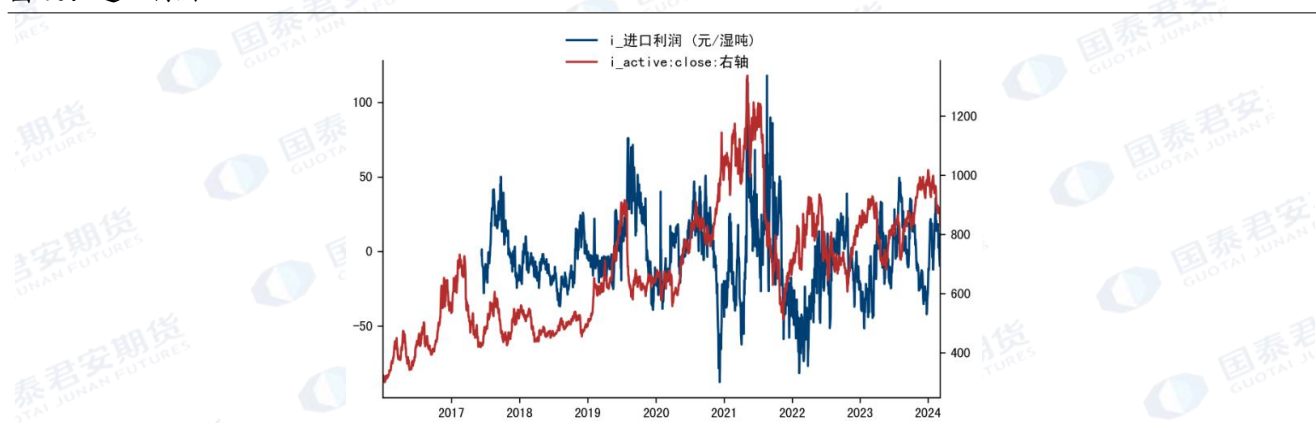
资料来源：国泰君安期货研究

由于这些价差因子相关性较高，在具体使用时可以选择表现最稳定的一个，或多个价差因子取等权平均值，从而降低单个因子预测的风险。

## 2.3 进口利润

如前文所述，我国铁矿石较多依赖从海外进口。进口利润是一个关键的指标，它的主要逻辑是，假定需求不变的情况下，当进口利润增加时，更倾向于从海外进口铁矿石，从而影响铁矿的供给增加。从下图中我们也可以看到，进入 2021 年之后，期货价格和进口利润呈现较为明显的反向变动关系。

图 10：进口利润



资料来源：国泰君安期货研究

## 2.4 铁矿石发运量

我们考虑从澳洲、巴西，以及四大矿山的一些发运指标：

巴西铁矿石发货量：14 个港口合计（周）；

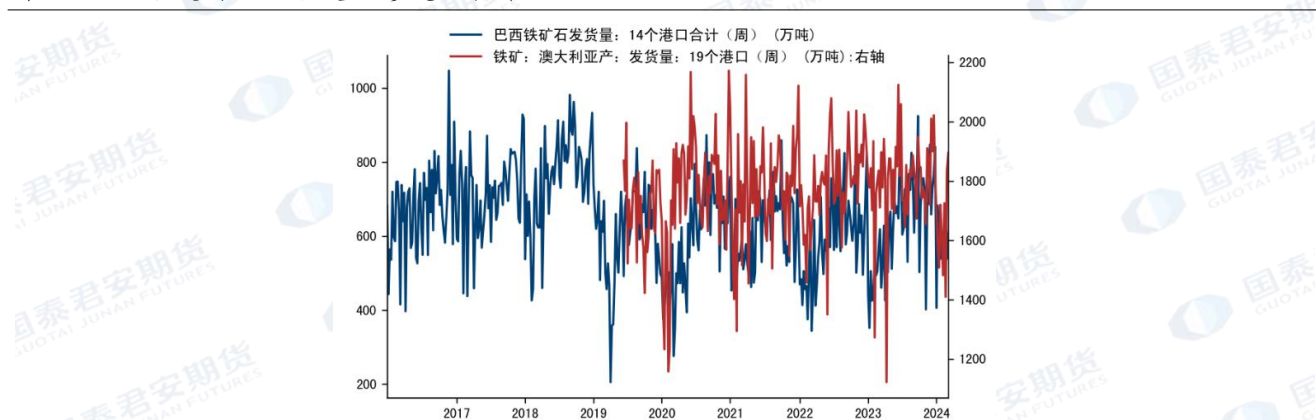
福蒂斯丘（FMG）铁矿石发货量：至中国（周）；

铁矿：到港量：26 个港口（周）；

铁矿：澳大利亚产：发货量：19 个港口（周）。

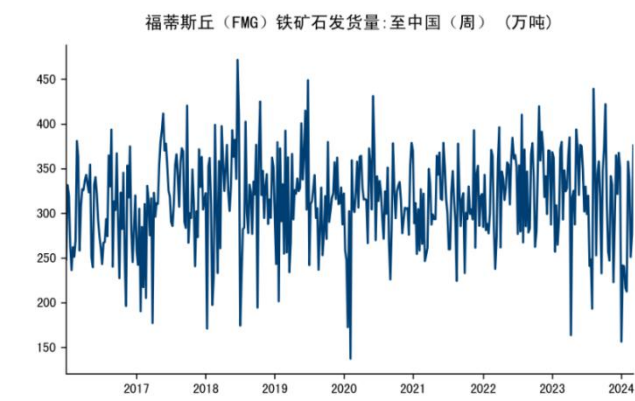
关于铁矿的发运和供应指标还有许多，读者可以根据需要自行选择合适的指标进行测试。从下图可以看出，发运指标的变化较为剧烈，无明显趋势性，可以考虑使用移动平均等方法对曲线的光滑度进行处理。

图 11：巴西和澳洲铁矿发运量（多港口合计）



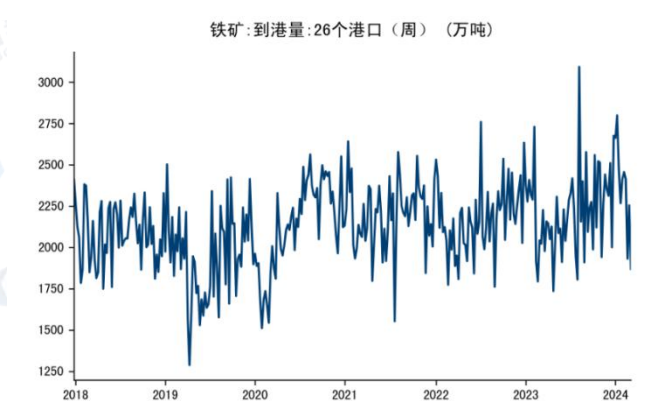
资料来源：国泰君安期货研究、钢联

图 12：福蒂斯丘（FMG）铁矿石发货量：至中国（周）



资料来源：国泰君安期货研究、钢联

图 13：铁矿：到港量：26 个港口（周）



资料来源：国泰君安期货研究、钢联

## 2.5 铁矿石库存、仓单指标

库存和仓单是商品的重要基本面特征。我们选取以下库存指标：

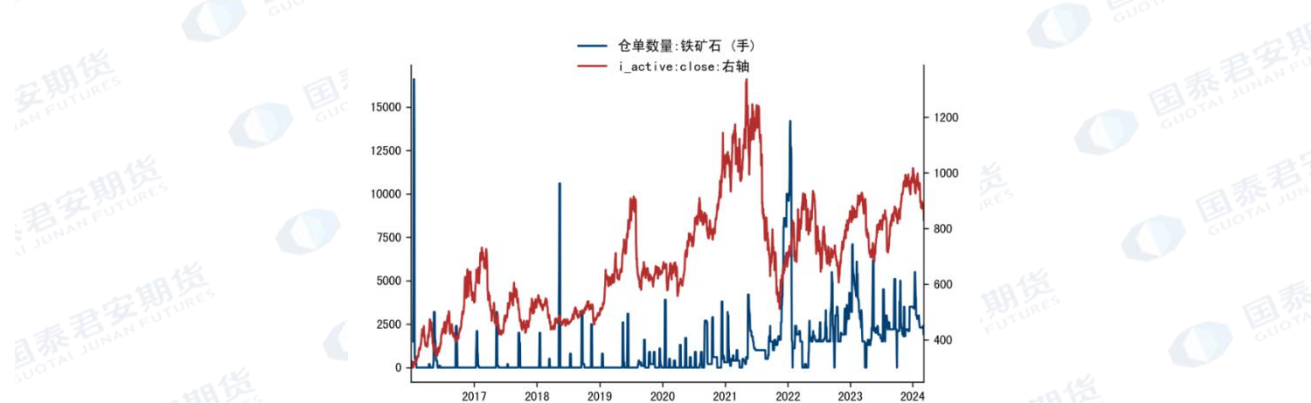
仓单数量：铁矿石（交易所每日盘后公布）

铁矿：进口：库存：247 家钢铁企业（周）

进口铁矿石港口总库存：45 个港口总计（周）

钢铁企业库存属于铁矿的下游库存。企业库存增加，反映企业生产有利润，会继续生产导致需求旺盛，因此倾向于做多。而港口库存则是相反的逻辑，港口库存增加，说明现货成交减少。它往往和价格呈现呈反向变动关系。仓单的逻辑和铁矿的中游库存类似，属于反映供应端的增减，和价格往往负相关。需要注意的是，仓单数量在 2021 年之后才具有比较明显的变化规律，而 2021 年之前则有大量的无效数据。

图 14：铁矿石仓单与期货价格

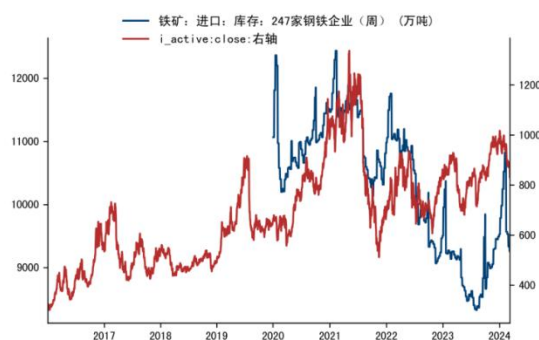


资料来源：国泰君安期货研究、同花顺

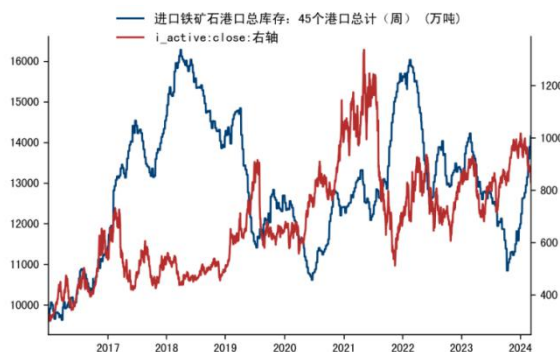
图 15：铁矿：进口：库存：247 家钢铁企业（周）

图 16：进口铁矿石港口总库存：45 个港口总计（周）





资料来源：国泰君安期货研究、钢联



资料来源：国泰君安期货研究、钢联

## 2.6 铁矿石下游需求端指标

从钢铁冶炼端出发，我们大致可以选取如下的一些需求指标：

Mysteel 全国 247 家钢厂日均铁水产量（周）

Mysteel 全国 247 家钢厂高炉开工率（周）

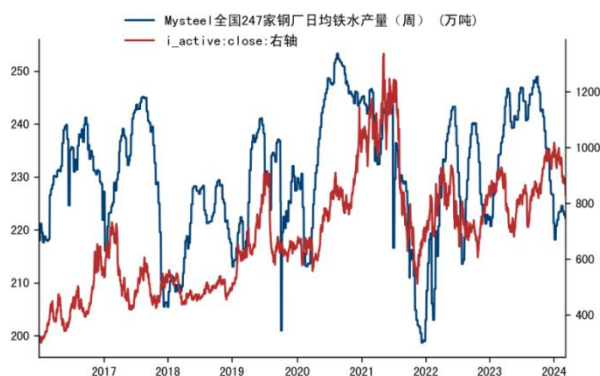
247 家钢铁企业：高炉产能利用率：中国（周）

进口铁矿石港口总库存：日均疏港量：45 个港口总计（周）

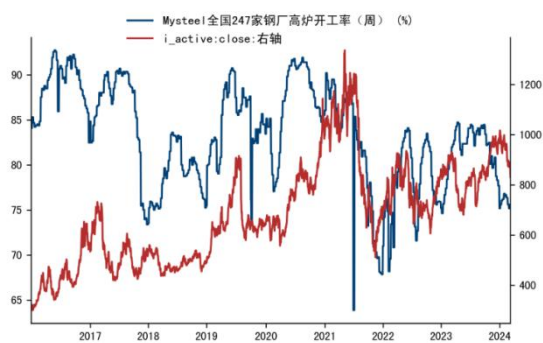
铁水是铁矿石下游重要应用，是铁矿石需求的最直接体现。产量增加时做多。钢厂开工率反映钢厂对于铁矿石的需求，为正向变动关系。可与钢厂高炉利用率交叉对比，反映当前的生产利润水平。钢厂产能利用率反映钢厂的利润水平，高的利润水平能够与铁矿石的需求形成正反馈。日均疏港量能够较好反映港口材料堆积程度，从而反映需求程度。当疏港量增多时做多，疏港量减少时做空。

图 17: Mysteel 全国 247 家钢厂日均铁水产量（周）

图 18: Mysteel 全国 247 家钢厂高炉开工率（周）

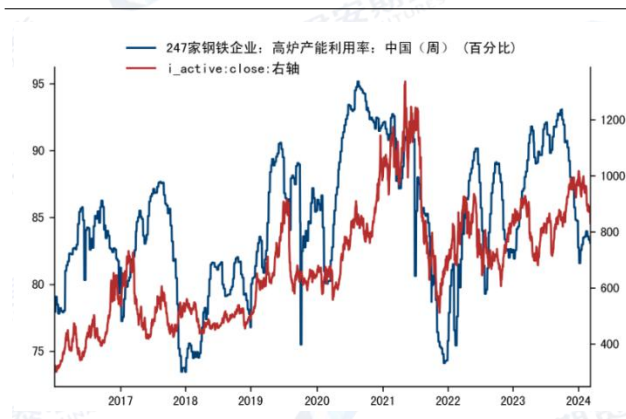


资料来源：国泰君安期货研究、钢联



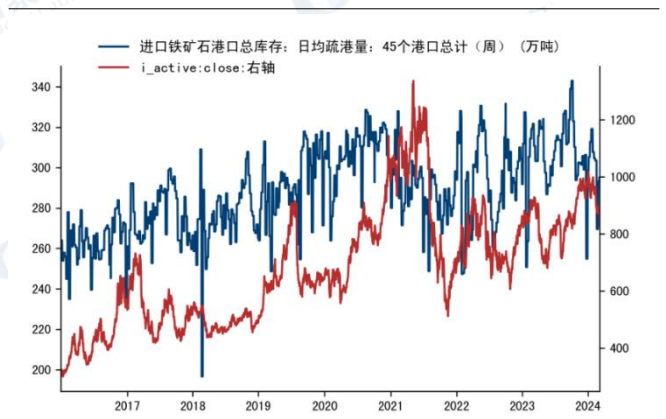
资料来源：国泰君安期货研究、钢联

图 19：247 家钢铁企业：高炉产能利用率：中国（周）



资料来源：国泰君安期货研究、钢联

图 20：进口铁矿石港口总库存：日均疏港量：45 个港口总计



资料来源：国泰君安期货研究、钢联

### 3. 因子的构造与模型回测

#### 3.1 数据的处理和因子构造方法

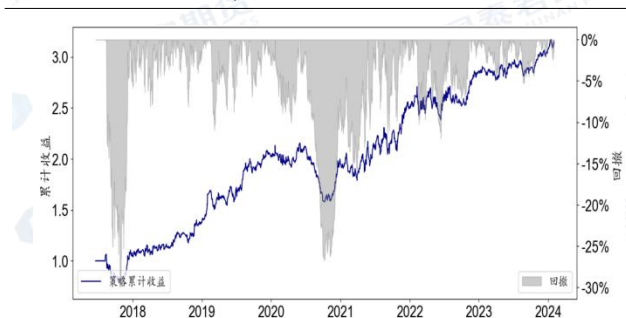
我们提取单条时间序列数据，周度的数据向后填充到交易日中，例如某数据周五更新，则在更新之前均使用上周五的数据作为最新值。同时对于有缺失值的数据，也向后填充补足。我们对于得到的时间序列将进行一系列处理，最终将其映射为 1（做多）、0（不持有合约，进行现金管理）、-1（做空）。当信号由 0 变为 1 或者 -1 时，则选择开仓，从 1 或 -1 变为 0 时，则平当前仓位。由 1 变成 -1 或由 -1 变成 1 时，则先平当前仓位，再进行反向开仓。

对于基本面因子的处理方法通常有以下几种：1. 取 N 周期的同环比，以此来衡量数据的增减变化情况。2. 取 N 周期的差值，通常用在库存、开工率等数值常为正的情况。3. 取 N 周期的移动平均，通常适用与数据变化剧烈或不够光滑的情况。4. 设定阈值，例如当数据偏离 N 个标准差时触发信号，反之则不触发交易信号。5. 构造特定规则，例如分别求较短的移动均线和较长移动均线，当短均线突破长均线触发信号，长均线突破短均线触发相反的信号。

#### 3.2 因子回测绩效展示

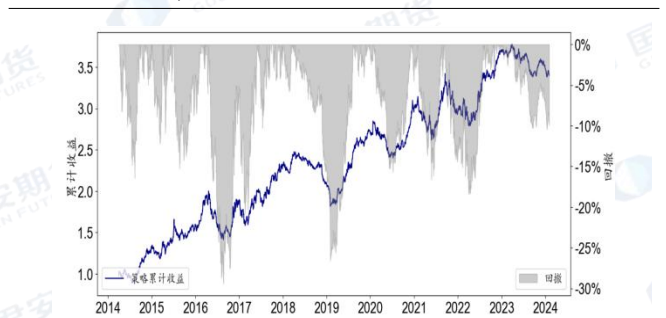
下面选取一些回测相对较好的因子进行展示，手续费设为双边万三。

图 21：进口利润指标，夏普比率 1.55，卡玛 1.11，其中 2021 年后夏普 1.22，卡玛 1.65



资料来源：国泰君安期货研究

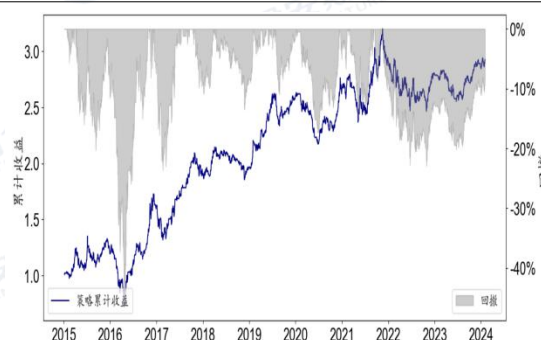
图 22：PB-混合价差，夏普比率 1.38，卡玛 0.85，2021 年后夏普 0.55，卡玛 0.5



资料来源：国泰君安期货研究

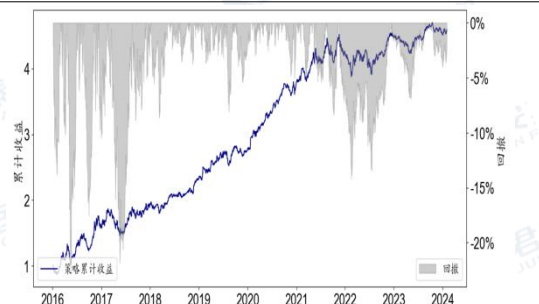


图 23：高炉产能利用率，夏普 1.04，卡玛 0.47，2021 年后夏普 0.19，卡玛 0.17



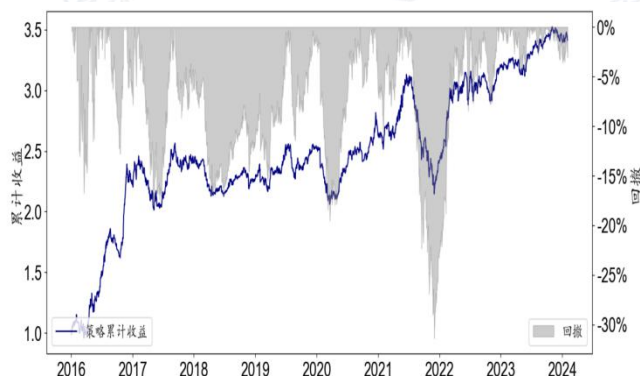
资料来源：国泰君安期货研究

图 24：仓单数量指标，夏普 2.51，卡玛 1.99，2021 年后夏普 0.8，卡玛 0.72



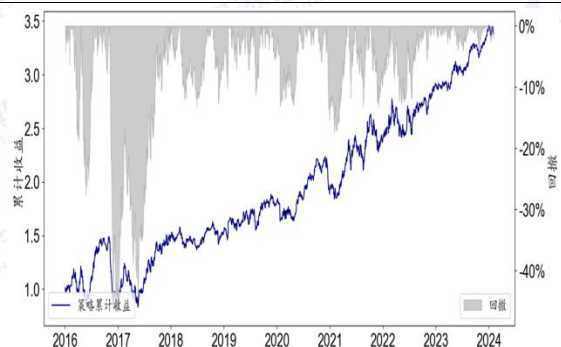
资料来源：国泰君安期货研究

图 25：巴西铁矿发货量，夏普 1.69，卡玛 0.98，2021 年后夏普 0.64，卡玛 0.39



资料来源：国泰君安期货研究

图 26：福蒂斯丘发货量，夏普 1.17，卡玛 0.65，2021 年后夏普 1.66，卡玛 2.26



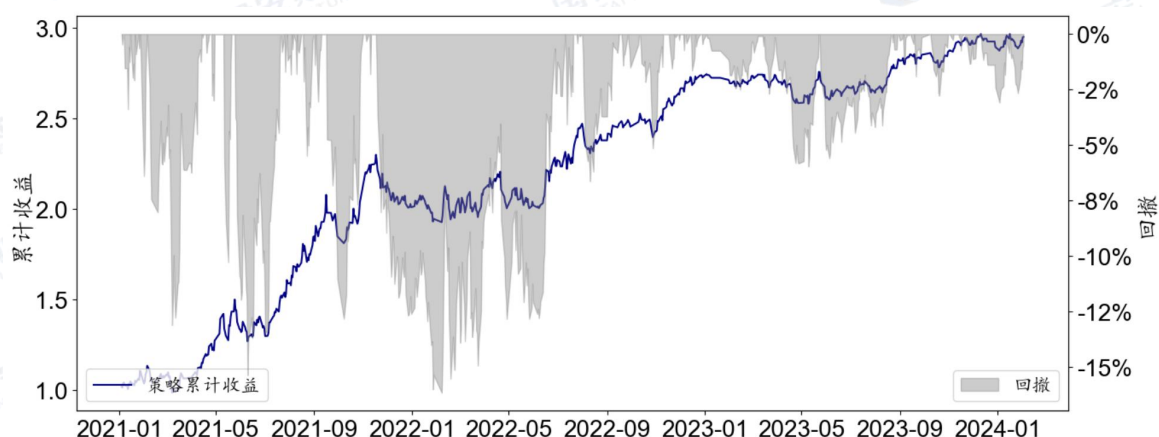
资料来源：国泰君安期货研究

### 3.3 因子组合

下面我们将简单叙述一些因子组合的方法。我们首先研究较为简单的两因子组合情形。例如我们考虑库存和基差的两因子组合。我们设定库存阈值为  $X$ ，若当期指标小于（大于）阈值  $X$ ，我们则在下一期做多（做空）标的期货。我们设定基差阈值为  $Y$ ，若当期指标大于（小于）阈值  $Y$ ，我们则在下一期做多（做空）标的期货。总而言之，当我们认为只有两因子均满足开仓条件才会开仓，其余情况下均不开仓。我们可以检验组合因子的绩效是否提升了单因子的绩效，从而检验组合是否有效。

对于多个因子，我们依然可以依此法进行组合，但随着因子数量的提升，条件限制越来越严格，开仓点将会变少，从而存在过拟合的风险。因此，可以采用多因子打分模型，例如当超过一半权重的因子满足开仓条件就进行开仓。也可以采用阶梯强度信号代表具体需要做多和做空的手数，当然在此情况下，多数情况下都非满仓操作。例如我们采用进口利润、价差、高炉产能利用率（代表需求）、仓单数量、发运量因子进行五因子组合，这些因子彼此相关性较低。

图 27：进口利润、价差、高炉产能利用率（代表需求）、仓单数量、发运量五因子等权组合表现，2021 年后夏普 2.7，卡玛 4.01



资料来源：国泰君安期货研究

#### 4. 近期行情总结及量化因子展望

近期铁矿石行情反复，波动较大。05 主力合约在十个交易日内已跌去 8.74%。从供应方面解读，铁矿供应偏强。巴西、澳洲发运量维持在高位水平，且除了主流发运外，非主流发运也有所增加。需求方面，各地钢厂时有出现检修，减产，停产现象，符合负反馈传导逻辑（钢厂减产导致需求被压制，进一步对铁矿石价格形成压力）。铁水产量处于 220-230 区间，为一年以来的最低位，较去年同期减少 14.22 万吨，从而可见需求端目前是偏弱势的状态。下游成材方面，热卷与螺纹的累库速度已处于同期中高位。铁矿中游库存方面，全国主要库存合计已连续三个月累库。对于量化因子来说，因为铁矿石在之前长期趋势上行，因此近期的剧烈波动可能会使得中长周期趋势跟踪面临一些损失。短周期量价因子的表现会稍好。对于基本面因子来说，主要因子较符合基本面底层传导逻辑，但仍然需要有以下值得注意的因素：1. 若发运量持续上升，可能需要注意主流发运（澳巴发运）因子影响的边际递减，非主流发运的影响可能还会边际递增。2. 部分事件性因素，例如停产、检修等，可能无法及时反映在基本面数据中。3. 注意政策面和情绪面的影响。



## 本公司具有中国证监会核准的期货交易咨询业务资格

本内容的观点和信息仅供国泰君安期货的专业投资者参考。本内容难以设置访问权限，若给您造成不便，敬请谅解。若您并非国泰君安期货客户中的专业投资者，请勿阅读、订阅或接收任何相关信息。本内容不构成具体业务或产品的推介，亦不应被视为相应金融衍生品的投资建议。请您根据自身的风险承受能力自行做出投资决定并自主承担投资风险，不应凭借本内容进行具体操作。

## 分析师声明

作者具有中国期货业协会授予的期货投资咨询执业资格或相当的专业胜任能力，保证报告所采用的数据均来自合规渠道，分析逻辑基于作者的职业理解，本报告清晰准确地反映了作者的研究观点，力求独立、客观和公正，结论不受任何第三方的授意或影响，特此声明。

## 免责声明

本报告的信息来源于已公开的资料，本公司对该等信息的准确性、完整性或可靠性不作任何保证。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，本报告所指的期货标的的价格可升可跌，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。

本报告中所指的研究服务可能不适合个别客户，不构成客户私人咨询建议。在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议。在任何情况下，本公司、本公司员工或者关联机构不承诺投资者一定获利，不与投资者分享投资收益，也不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。投资者务必注意，其据此做出的任何投资决策与本公司、本公司员工或者关联机构无关。

市场有风险，投资需谨慎。投资者不应将本报告作为做出投资决策的唯一参考因素，亦不应认为本报告可以取代自己的判断。在决定投资前，如有需要，投资者务必向专业人士咨询并谨慎决策。

## 版权声明

本报告版权仅为本公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制、发表或引用。如征得本公司同意进行引用、刊发的，需在允许的范围内使用，并注明出处为“国泰君安期货研究”，且不得对本报告进行任何有悖原意的引用、删节和修改。

若本公司以外的其他机构（以下简称“该机构”）发送本报告，则由该机构独自为此发送行为负责。通过此途径获得本报告的投资者应自行联系该机构以要求获悉更详细信息或进而交易本报告中提及的期货品种。本报告不构成本公司向该机构之客户提供的投资建议，本公司、本公司员工或者关联机构亦不为该机构之客户因使用本报告或报告所载内容引起的任何损失承担任何责任。