

# 铅基本面多因子量化策略探讨

国投安信期货研究院

铅宏观属性弱，主要围绕基本面展开交易。消费领域集中，基本面逻辑相对简单，适合进行基本面量化探索。

结合基本面研究经验，本文通过量化手段，构建了四个大类因子，借助阈值开仓方法生成开仓信号，并最终以等权方式得到大类合成因子的开仓信号，在测试集和训练集上取得了预期效果。

## 1 基本面相关因子分类

我们将基本面因子划分四大类：需求、供给、库存和价差因子。

**需求因子：**我国 85% 的铅用于生产铅酸蓄电池，需求类因子主要包括铅锭及铅下游产品出口、铅酸蓄电池企业开工、电池企业和经销商成品库存、汽车、通信基站等终端品产量，主要电池型号价格。

**供给因子：**原、再生铅炼厂的原料供应，开工，产量都会影响供应。所以供应因子主要包括铅精矿供应（产量+进口）、TC、银价、银精矿进口、废电瓶价格、原/再生铅炼厂利润、原/再生铅炼厂开工等数据。

**库存因子：**库存是供需平衡的最终表现，重要性不言而喻。主要关注炼厂原料和成品库存，铅锭交易所和社会库存，港口库存。

**价差因子：**价差反应不同期限、不同品种之间的供求差异，主要包括了期限价差、现货升贴水、精废价差。铅和其他有色品种关联度低，不关注跨品种价差。

图：铅指标结构分类



资料来源：国投安信期货

## 2 基本面因子的预处理和大类因子的搭建

先构建基本面因子，在通过一定方式加权求和，形成 4 个大类因子的时间序列，然后根据这 4 个大类因子时间序列形成最后的交易策略。

### 2.1 数据采集

原始基本面因子 80 个，数据从 2012 年初开始，一直截止到最新日期，频率为日度、周度和月度。部分数据早期尚未开始搜集，各因子除频率不同外，时序长度也有差异。日度数据当日收盘后可得，连贯性强；周度数据周五收盘后可得，其中 SMM 铅锭社库每周一、周五更新；月度数据如产量、产成品库存、进出口发布滞后，所含信息量较少。遇节假日可能会略有调整。

### 2.2 基本面因子处理

#### 2.2.1 数据预处理：空值、极值

原始因子序列时间戳都是日度的，周度和月度数据有大量空值，统一用前值补全，得到全部因子的日度序列。

数据中的极值往往会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。

### 2.2.2 频率转换

铅价从 2020 年以来的变化幅度明显缩窄，波动变动频率接近周度。从信号生成和调仓频率角度出发，统一取因子每周的均值作为该周周五的值，最后只保留周五的这个值，从而实现了所有因子统一向周度转换。再将周度数据合成周度的大类因子序列。

对于交易标的，取沪铅主力合约每周五的收盘价，同样调整为周度。这也意味着基本面策略将以周度为单位进行。

### 2.2.3 标准化处理

基本面因子和沪铅价格全部转变成周度数据以后，考虑不同因子的量级和量纲差异较大，量级较大的指标可能导致量级较小的指标被忽略，当量级差别很大时，消除量纲影响也有必要。用因子的当期值减去一定窗口期的均值来降低因子量级的影响，同时通过计算因子窗口期内环比变化抹除量纲影响，从而实现了基本面因子的标准化。

具体构造方式如下：

单个基本面因子  $i$  的构造

$$F_{i,t} = f_{i,t} - \text{Mean}(f_{i,t}, f_{i,t-1}, f_{i,t-2}, \dots, f_{i,t-\text{windows}})$$

$$\text{Factor}_{i,t} = \frac{F_{i,t}}{F_{i,t-\text{windows}}} - 1$$

沪铅周度收益率序列的构造

$$R_{pb,t} = \frac{P_{pb,t}}{P_{pb,t-1}} - 1$$

其中：

1.  $f_{i,t}$  表示基本面因子  $i$  在  $t$  期的原始值；
2.  $\text{windows}$  表示窗口期，由于前期已经将频率转为周度， $\text{windows} = 3$  则代表窗口期为 3 周；
3.  $\text{Mean}()$  表示过去窗口期内因子值的平均数；
4.  $F_{i,t}$  表示某因子的  $t$  期的原始值减去窗口期均值；
5.  $\text{Factor}_{i,t}$  表示基本面因子  $i$  经过标准化以后， $t$  期的值；
6.  $P_{pb,t}$  表示沪铅在  $t$  期的收盘价；
7.  $R_{pb,t}$  表示沪铅的周度收益率；

这样就得到了不同基本面因子、沪铅收益率的时间序列。沪铅收益率就是需要预测的量，数学上也称被解释变量；构造的基本面因子可以理解成基本面上的已知信息，属于解释变量。

#### 2.2.4 滞后性处理

对所有因子统一进行延后一阶处理，即用  $t-1$  期的基本面因子来预测  $t$  期的沪铅收益率，以避免使用未来数据。

#### 2.2.5 最优窗口期确定

每个因子的时间序列开始时间不完全相同，为了保证训练集宽度足以覆盖窗口期，固定住训练集的结束时间为 2020 年 12 月 25 日，测试集则使用 2021-2023 年的。

遍历 1-52 周的窗口期，构造该因子对应的不同时间序列，然后在训练集上计算该因子对应信息系数 IC、胜率、盈亏比，以此作为筛选标准，选择最优窗口期，并最终确定对应的因子序列。其中：

信息系数 IC：基本面因子和沪铅收益率的 pearson 相关系数。如果  $IC > 0$ ，意味着因子和沪铅收益率正相关，因子值越大，对应下一期沪铅收益率可能越高；而因子值越小，对应下一期的沪铅收益率可能越低，甚至收益为负。反之亦然。

为了减少开仓次数，采用阈值开仓法，只有价格突破阈值上沿或者下沿才会出现开仓信号；因子未突破阈值的情况下，对应平仓或不开仓。

如果  $IC > 0$ ：

$$\text{开仓信号} \begin{cases} 1 & F_{i,t} > upper \\ 0 & upper > F_{i,t} > lower \\ -1 & F_{i,t} < lower \end{cases}$$

如果  $IC < 0$ ：

$$\text{开仓信号} \begin{cases} -1 & F_{i,t} > upper \\ 0 & upper > F_{i,t} > lower \\ 1 & F_{i,t} < lower \end{cases}$$

其中：

1 是买入信号，-1 是卖出信号，0 是平仓或不开仓信号； $upper$  和  $lower$  是阈值的上沿和下沿。

通过开平仓信号，计算因子训练集上的胜率和盈亏比。

胜率：

$$\text{Win\_ratio} = \frac{\text{number of profitable trades}}{\text{total number of traders}}$$

盈亏比：

$$\frac{\text{profit}}{\text{loss ratio}} = \frac{\text{total profit/number of profitable trades}}{\text{total loss/number of losing trades}}$$

上、下阈值的选择是通过计算机遍历不同的上、下分位数，选取预测胜率最高的阈值作为该因子最优阈值，并记录对应的胜率和盈亏比。

同一个因子，不同的窗口期构建出来的时间序列不同，在训练集上的 IC、胜率、盈亏比也不同。统一按照如下标准确定最优窗口期：

1. 当 IC 值变大、胜率变高时，按照胜率最高确定最优窗口期。
2. 当 IC 值变大，胜率变小时，盈亏比作辅助判断标准，可能还会结合一定的基本面经验，来确定最优窗口期。

最优窗口期确定后，基本面因子最终时间序列就确定下来，接下来介绍如何构建需求、库存、供给、价差这 4 个大类因子。

## 2.3 大类因子构造

### 2.3.1 基本面因子选取

结合因子在训练集上的 IC、胜率、盈亏比，以及基本面逻辑，人为挑选出了 41 个重要的基本面因子。用来构造最终的大类因子。下面对其中一些因子从基本面的角度出发给出解释。

供给类因子中，大部分单因子 IC 为正，意味着供应变化和标的价格的变化正相关，主要体现在月度供应数据和国产铅精矿价格上，违背常识。从基本面角度出发，因月度供应数据公布往往滞后 10-20 天左右。在数据公布之前，市场会根据冶炼厂检修动态预判供应边际变化，而滞后的产量数据显然无法反应市场预判行为。而铅精矿虽属供应端，但定价是期货价格减 TC，铅精矿 TC 往往长期稳定，铅精矿的价格天然的就和期货价格同向变动。窗口期方面，不含税再生铅利润和进口铅精矿价格的最优窗口期为 12 周，通常持续面临亏损才会加速炼厂减产落地。铅精矿加工费的最优窗口期为 51 周，

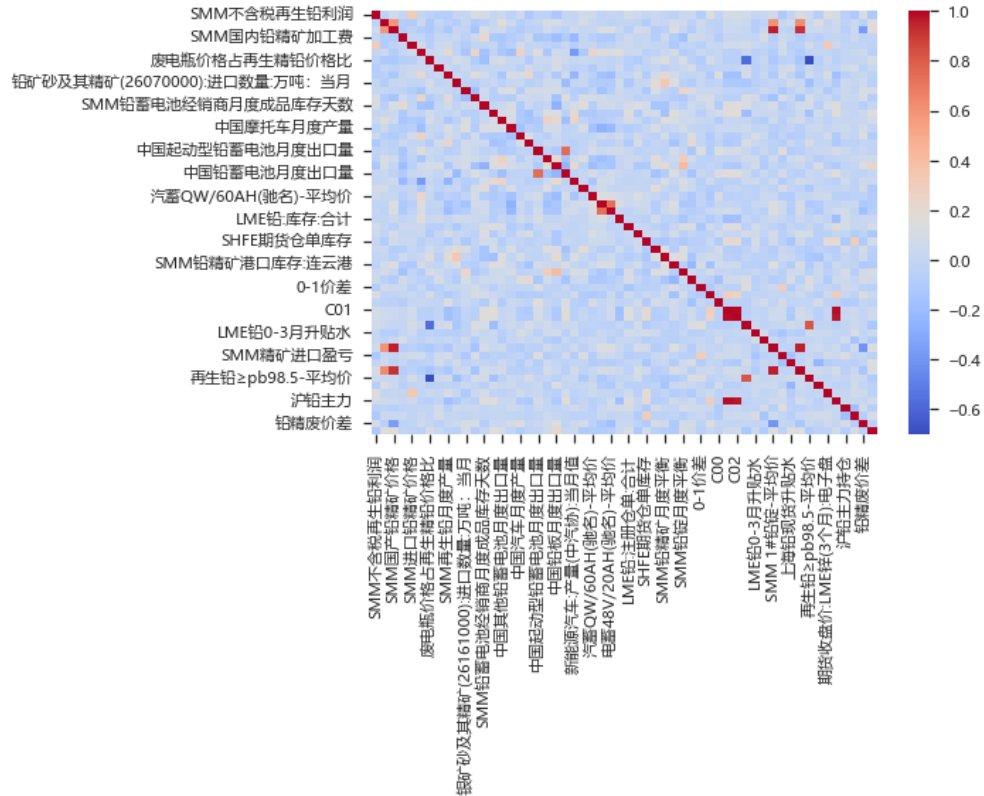
且 IC 值为负，铅精矿供应主要受季节性影响，TC 是矿山和冶炼厂博弈的结果，原生铅炼厂开工通常较平稳，TC 也通常长期不变，通常 TC 出现极端表现时影响才更明显。

需求类因子中，月度数据占比最高，数据变化缓慢，窗口期在 40 周附近，往往和铅价同向运动。周度数据为铅酸蓄电池价格，电池价格的高低反映终端消费的强弱，窗口期为 10 周时，IC 为负值。终端消费较差，经销商降价促销，会导致铅价的同步走低，而铅价走低之后，冶炼厂惜售情绪抬升，减产预期走强，反而可能刺激下一期的铅价走强。

库存类因子中，数据频率分布较均匀，最优窗口期多在 30–40 周，社会库存数据由 SMM 发布，对未来行情的判断有一定的指引性，通常库存越高意味着市场供大于求的形势更严峻，铅下行压力越大，该类因子 IC 为负，符合基本面逻辑。

价差类因子都是日度数据，包含现货价、近远月价差和现货升贴水、精废价差等，铅的振幅不及其他金属，价差数据通常较小，标准化处理以后，虽然抵消了部分量纲和量级的变化，也抹除了部分独特的因子信息。跨期价差的 IC 值较为不稳定，随着窗口期的变动正负切换较为频繁。而精废价差在不同窗口期的 IC 值均为负，这也符合基本面逻辑，当铅价走低，再生铅利润被压缩，惜售情绪抬升，对原生铅贴水收窄，甚至出现精废价格倒挂，下游倾向采购原生铅，铅可交割品牌库存减少，带动下一期铅价低位反弹。窗口期方面，除 LME3–15 升贴水的窗口期低于 10 周之外，其余窗口期基本在 40–50 周至间。

图：IC 相关性



资料来源: SMM, Wind, 国投安信期货

### 2.3.2 合成大类因子及大类因子在训练集表现

筛选出重要的基本面因子后，下面通过因子在训练集上的胜率加权得到四个大类的因子：

$$Class_{j,t} = \sum Fator_{j,i,t} * Win\_ratio_i$$

其中，

1.  $j$ 代表供给、需求、库存和价差中的一个大类，
2.  $Class_{j,t}$ 代表构造的最终大类因子在  $t$  期的值，由该大类中所有基本面因子的  $t$  期值通过胜率加权求和得到。
3.  $Fator_{j,i,t}$ 是大类  $j$  中的因子  $i$  在  $t$  期的因子值
4.  $Win\_ratio_i$ 代表因子  $i$  在训练集上的胜率

由此得到四个大类因子的时间序列。在训练集上，通过迭代，分别找到四个大类因子对应的最优开仓阈值，形成交易信号。统计发现，供给、需求、库存以及价差大类单因子的胜率分别为 0.68，0.71，0.68 以及 0.73。



分析单个基本面因子对其所在大类因子的贡献，发现：

对供给大类因子而言，最优阈值往往靠近历史序列的最大最小值，这意味着因子值出现极端变化时开仓的胜率会有提升。供给大类中每个单因子的胜率基本维持在 0.61 附近，而废电瓶价格占再生精铅价格比这个基本面因子的胜率最高为 0.65，由于再生铅产能过剩，废电瓶持续供不应求，废电瓶往往比铅更扛跌。作为再生铅的生产原料，废电瓶价格占再生精铅价格的比越高，再生铅利润越薄，炼厂挺价意愿越强。SMM 再生铅月度产量的盈亏比最高为 5.66，在相似胜率下，该因子的收益更好。再生铅占比已经超过了 53%，再生企业利润薄，检修成本低，可以灵活安排检修；而原生铅检修成本高，开工整体平稳，市场往往更加关注再生铅炼厂生产动态，来调整对未来供应边际变化的预期。

需求大类因子的阈值以及胜率情况均和供给大类因子相似，胜率最高的是 SMM 铅蓄电池企业月度成品库存天数，约为 0.67，基本面上看，铅蓄电池企业月度成品库存天数越高，说明电池销售越差，需求越弱，对铅价的影响比较直观。而中国启动型铅蓄电池月度出口量这个基本面因子的盈亏比最高，为 6.16，暂时无法从基本面逻辑中给出更多合理解释。

价差大类的基本面因子胜率均值是 0.59，0-2 价差和 C01(即次月合约收盘价)胜率最高，同时 C01 的盈亏比相对亮眼为 4.01。次月合约和主力合约收盘价在最近的两次换月中间是重合的，由于对因子做了滞后一周处理，C01 因子的胜率和盈亏比高，某种程度也体现了沪铅价格的动量效应。

库存大类中的基本面因子胜率整体略高于其他大类中的因子，多数位于 0.62 附近，其中胜率最高的是 LME 铅注销仓单，而 LME 铅注册仓单的盈亏比最优。库存是供求平衡的最终体现，注册、注销仓单是供需动态调整的结果，所以预测胜率较好也合理。

### 3 交易策略

每周五，分别比较四个大类因子的最新值和其对应的最优阈值，生成四个大类因子的开仓信号。

为了减少开仓次数，通过计算机在训练集上迭代出最优阈值，只有当因子值超过一定阈值，才会出现开仓信号；因子未突破阈值的情况下，对应平仓或不开仓。



一个大类因子的开仓信号按照如下规则生产：

如果  $IC > 0$ ：

$$Class_j \text{ 的开仓信号} \begin{cases} 1 & Class_j > upper \\ 0 & upper_a > Class_j > lower \\ -1 & Class_j < lower \end{cases}$$

如果  $IC < 0$ ：

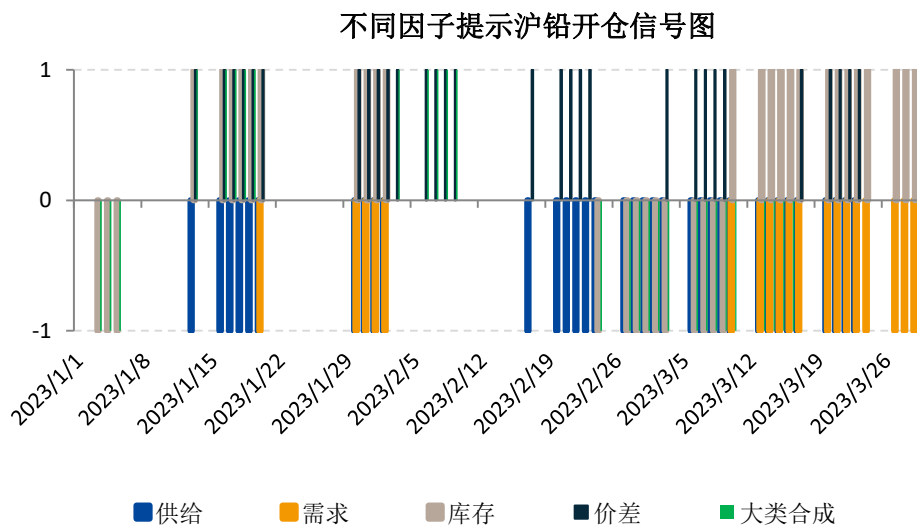
$$Class_j \text{ 的开仓信号} \begin{cases} -1 & Class_j > upper \\ 0 & upper_a > Class_j > lower \\ 1 & Class_j < lower \end{cases}$$

然后将四个大类的开仓信号等权相加形成最终的大类合成信号。具体示意图见表。

开仓信号	Class_供应	Class_需求	Class_库存	Class_价差	大类合成因子
第 N-1 周	1	0	-1	1	1
第 N 周	-1	1	1	1	2
第 N+1 周	0	0	0	0	0
第 N+2 周	1	-1	1	-1	0

下图展示了 2023 年 1 月-3 月这三个月的时间内，通过以上交易策略，分别得到的供给、需求、库存、价差大类因子提供的开仓信号。并由四个大类因子的开仓信号等权相加得到的大类合成因子的开仓信号。

图：大类因子开仓信号分布



## 4 实证结果

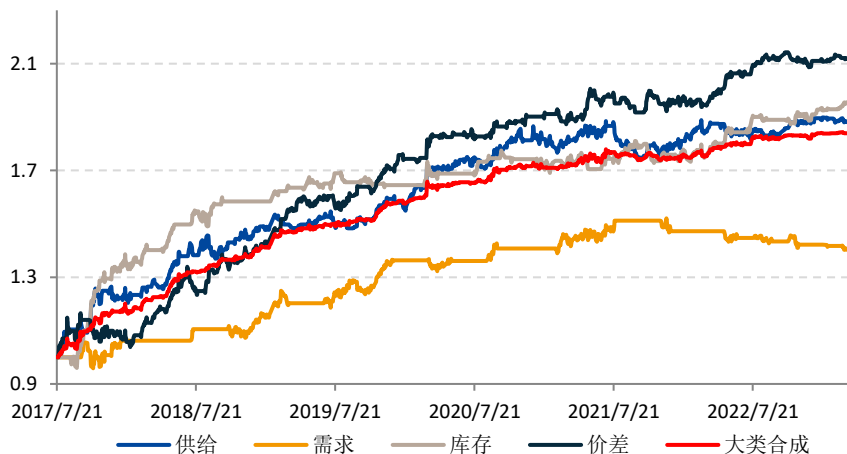
根据我们构建的大类因子和大类合成因子，利用阈值开仓法在训练集和测试集进行实证检验。在 2021 年以后的测试集上，价差因子、库存因子表现出了比较好的盈利性，供给大类因子表现平平，而需求大类因子在测试集上表现不好。

从基本上理解，价差和库存因子频率较高，信息保留较全面，能更好把握周度的波动和趋势，所以表现较好。根据我们铅基本面研究经验，市场在交易现实和预期之间频繁切换，再生炼厂通过灵活调节检修计划来改变市场对未来供应的预期，单纯根据供给端的数据作为开仓依据，缺乏预期指引，供给大类因子在测试集上很难延续训练集中的表现。需求端大类因子中，月度数据占比非常高，数据发布滞后，信息含量不足，很难捕捉沪铅在周度下的市场反应。

合成大类因子信号由四个大类因子信号等权求和得来，因加入了需求因子，策略累计收益下降，不过得益于综合了基本面的每个大类因子，收的波动降低，夏普比提高。总体而言，合成大类因子在测试集上依然取得了年化收益 6%，回撤 2%，夏普比 3% 的不错效果。

图：基本面因子净值曲线

铅大类因子净值图



资料来源: Wind, 国投安信期货