

From:	郑宇浩	To:	投资部
Date:	2022/04/20	Class:	Week1 测试
Title:	AZYH001001	Version:	1.0.0

1. 摘要

黑色金属与煤炭下的 11 个商品期货品种之间存在有明显的产业链条，如锰硅和铁矿石均处于螺纹钢的上游。同一条产业链上的品种，价格可能存在相互影响，基于此现象构建出的跨品种交易策略被称为配对交易（Pair Trading）。价差套利（Spread Arbitrage）是配对交易中最主流的一种交易策略，通常通过寻找两品种之间的某一线性组合（协整检验）来构建平稳的价差序列，并基于该序列进行套利交易。但两品种之间的协整关系极易随市场的变化而变化，使得价差序列的波动率很大，均值回复的半衰期过长，从而导致构建出的交易策略非常不稳定。因此，降低价差序列的波动率、缩短均值回复的半衰期是提高策略有效性的关键。

通过简单地引入更多的交易品种可以使得投资组合的风险被合理地分散，所得到的价差序列也会较为稳定。[Box 和 Tiao^{\[1\]}](#)在 1977 年提供了一种典型分解(Canonical Decomposition)的方法可用于对多维时间序列的预测效果进行评估。当所选择的线性组合方式预测效果最差时，得到的残差序列最接近于白噪声，即可作为一条平稳价差序列供我们进行套利交易。

本投研报告将对黑色金属与煤炭下已有日频数据的 6 个商品期货品种进行每三个为一组的遍历搜索，结合所选品种在产业链上的供给关系，寻找在 Box-Tiao 典型分解下得到的平稳序列最适合进行价差套利的三个品种，构建配对交易组合。基于分解后得到的平稳序列，我们将依赖其均值回复的特性确定开仓点所在的上下轨道范围，同时尝试结合使用布林带突破信号以提升可套利空间。

最终确定的交易品种为螺纹钢、铁矿石和锰硅。当价差序列偏离均值 0.75 倍标准差后再次返回并触及轨道时进行开仓套利，待其回复均值附近时平仓。当价差序列发出布林带突破信号时，提前开仓以扩大可套利的价格区间。

2. 策略主要思想

2.1 策略主要思想，盈利来源，指标构建方式

处于同一条产业链上的商品期货品种可能具有较为相近的价格趋势，因此，某些品种之间极有可能存在某一线性组合，使得组合后的价差序列趋于平稳，具有均值回复的特性。本策略的主要思想就是寻找到黑色金属与煤炭产业链下存在该线性组合的配对交易品种，依赖其组合后的均值回复特性，在价差序列偏离均值时反向交易，等待其均值回复以获取利润。

策略的主要盈利空间取决于价差序列偏离均值时的开仓点，通常为两条设置在均值上下的阈值轨道。轨道与均值之间的距离决定了每次套利时预期可得的利润大小，但当轨道设置得越宽，策略开仓的条件就越严苛，交易频率也会随之降低。因此，轨道之间的宽度即可以作为策略的参数进行优化，也可以依据 [Puspaningrum^{\[2\]}](#)在 2009 年提出的优化方式进行预设。

本策略交易信号的发出依赖于两个指标，一为价差序列对阈值轨道的突破，二为价差序

列对布林带轨道的突破。当价差序列处于阈值轨道之外，重新返回并突破轨道时，策略将会开仓捕捉均值回复期间的收益。但为了提高可获利的价格区间，开仓点可提前至价差序列在高位反转时。因此，本策略额外地采用布林带轨道突破来提前捕捉价差序列的高位反转。详细的指标构建方式与指标计算过程中所用到的参数选择，将会在第三节中做进一步的解释。

2.2 交易配对选择逻辑

2.2.1 同一产业链且成交活跃品种

根据《主要产业链涵盖的品种》关系图，与螺纹钢处在同一条产业链上的主要是黑色金属和煤炭下属的 11 个品种：螺纹钢、铁矿石、热轧卷板、线材、硅铁、锰硅、不锈钢、焦煤、焦炭、动力煤。由于部分品种在 2017/03/19 回测开始前还未上市或才刚刚开始交易，活跃度不高，因此不被纳入考虑范围。最终满足在回测开始前一年 2016/03/19 就已经上市的且截止至 2022/03/18 仍在交易的期货品种共 6 个：螺纹钢、铁矿石、热轧卷板、锰硅、焦煤、焦炭，我们将对其逐一进行配对检验价差序列的平稳性。

2.2.2 Box-Tiao 典型分解

当选择了需要进行配对交易的品种之后，剩下的问题就是找到一组合适的线性组合方式，使得我们最终构建的价差序列平稳且均值回复的周期较为合适。这一过程相当于构建一个收益平稳的投资组合，配对交易的品种是组合内的标的物，而合适的线性组合方式即为各标的的权重。Box-Tiao 假设这样的投资组合 Π 满足如下的一阶向量自回归模型：

$$\Pi_t = \Pi_{t-1}A + Z_t$$

其中 $\Pi_{t-1}A$ 为在 $t-1$ 期对 t 期组合收益的线性回归预测值。二者的方差分别有如下关系：

$$\begin{aligned} Var(\Pi_t) &= \sigma_t^2 \mid Var(\Pi_{t-1}A) = \sigma_{t-1}^2 \mid Var(Z_t) = \Sigma \\ \sigma_t^2 &= \sigma_{t-1}^2 + \Sigma \\ 1 &= \frac{\sigma_{t-1}^2}{\sigma_t^2} + \frac{\Sigma}{\sigma_t^2} = v + \frac{\Sigma}{\sigma_t^2} \end{aligned}$$

由此可见，如果 $v = \frac{\sigma_{t-1}^2}{\sigma_t^2}$ 越小，由残差项 Z_t 构成的高斯白噪声就越占据主导地位，所构建的投资组合收益就越接近于是一个白噪声序列，平稳且带有较高的均值回复效果，是我们理想的价差套利序列。因此，求解线性组合的过程便转化为求解投资组合 Π_t 以实现 v 最小化。

令投资组合的权重向量为 x ，各品种的价格序列为 S_t ，其协方差矩阵为 Γ ，则有如下关系式：

$$\begin{aligned} \Pi_t &= S_t x \\ S_t x &= S_{t-1} A x + Z_t x \\ v &= \frac{x^T A^T \Gamma A x}{x^T \Gamma x} \end{aligned}$$

最小化 v 的过程可以转化为求解满足如下表达式的特征值 λ 的过程：

$$\det(\lambda\Gamma - A^T\Gamma A) = 0$$

可解得： $x = \Gamma^{-\frac{1}{2}}z$ ，其中， z 为如下矩阵的最小特征值对应的特征向量：

$$\Gamma^{-\frac{1}{2}}A^T\Gamma A\Gamma^{-\frac{1}{2}}$$

由于 A 为 S_t 对 S_{t-1} 线性回归后的参数， Γ 为 S_t 的协方差矩阵，所以，当对 S_t 进行去均值处理后，上式中的矩阵表达式可以改写为：

$$\begin{aligned} & \Gamma^{-\frac{1}{2}}A^T\Gamma A\Gamma^{-\frac{1}{2}} \\ &= (S_t^T S_t)^{-\frac{1}{2}}A^T(S_{t-1}^T S_{t-1})A(S_t^T S_t)^{-\frac{1}{2}} \\ &= (S_t^T S_t)^{-\frac{1}{2}}(S_{t-1}A)^T(S_{t-1}A)(S_t^T S_t)^{-\frac{1}{2}} \\ &= (S_t^T S_t)^{-\frac{1}{2}}(\hat{S}_t^T \hat{S}_t)(S_t^T S_t)^{-\frac{1}{2}} \end{aligned}$$

2.2.3 半衰期与平稳性检验

经过 Box-Tiao 典型分解后可以得到每一组配对品种的线性组合 $x = \Gamma^{-\frac{1}{2}}z$ ，再对组合后的价差序列进行均值回复的半衰期求解与平稳性检验，即可对比得出最适合做价差套利的配对组合。该组合首先需要满足的要求是所生成的价差序列必须平稳，其次是均值回复的半衰期越短越好。

在对 $C_6^3 = 20$ 对配对品种进行 Box-Tiao 典型分解后产生的价差序列进行 ADF 检验之后，有 11 组在 95%置信度下通过了平稳性检验，但仅有 5 组半衰期在 21 个交易日以内：

【表 2.2.3-1】黑色金属与煤炭产业链价差套利组合半衰期

品种名称	品种代码	半衰期
螺纹钢+铁矿石+锰硅	rb+i+SM	20
螺纹钢+焦炭+焦煤	rb+j+jm	19
铁矿石+锰硅+焦煤	i+SM+jm	17
铁矿石+焦炭+焦煤	i+j+jm	20
热轧卷板+焦炭+焦煤	hc+j+jm	21

其中，第一组螺纹钢、铁矿石和锰硅同属黑色金属，因此，从产业链的角度上分析，其价格的趋同效应应该要强于其他四组跨类别的组合，故选此作为最终的价差套利组合。

值得注意的是，选用多少商品期货品种去构建价差套利组合，以及选择多长时间的半衰期进行筛选，均是可以调整的参数，但往往所选品种越多，波动率越小，均值回复越快。

3. 指标分析

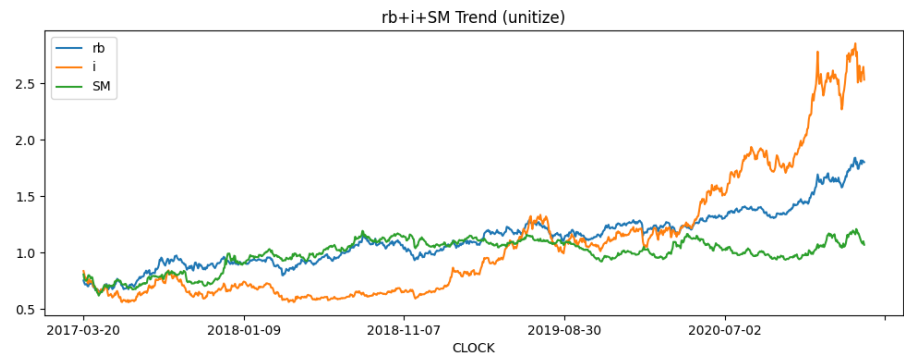
3.1 价差构造方式

在 2.2.2 节中，基于多品种投资组合时间序列服从一阶向量自回归模型的假设，我们已经详细介绍了如何通过最小化模型的可预测性（ $v = \frac{\sigma_{\epsilon}^2 - 1}{\sigma_{\epsilon}^2}$ ）来确定能产生均值回复效果最优的价差序列的组合权重，故在此便不再多进行赘述。本策略样本内回测区间为 2017/03/19 至 2021/03/19，利用此区间内的数据计算得到的组合权重如下表：

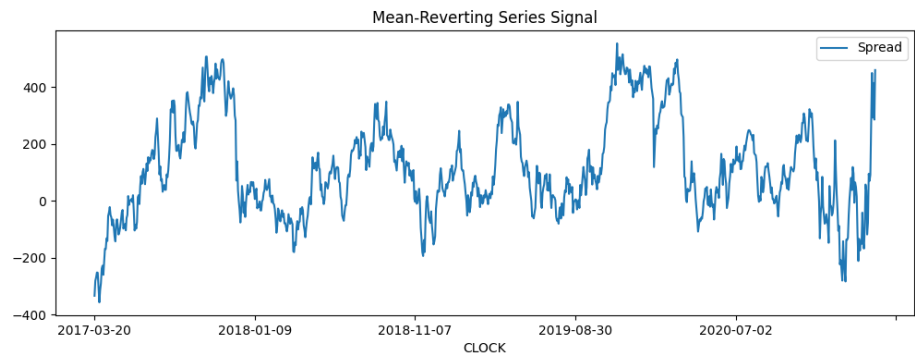
【表 3.1-1】价差套利组合权重

品种名称	品种代码	权重	合约乘数	合约份数
螺纹钢	rb	1	10	100
铁矿石	i	-0.75038229	100	8
锰硅	SM	-0.12836106	5	26

【图 3.1-1】价差套利组合复权除均值后趋势图



【图 3.1-2】价差序列



3.2 均值回复轨道突破指标构建及表现

价差序列的均值回复轨道由均值加减某个特定倍数的标准差得到，具体的倍数作为策略的参数可供后续进行优化。理论上，价差序列超出轨道的天数占比应当接近于对应的正态分

布概率：

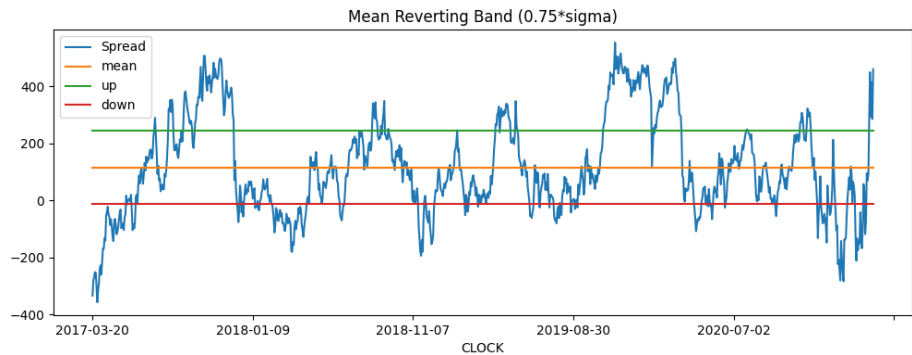
$$\frac{N \cdot I(\mu - n\sigma < P_t < \mu + n\sigma)}{N} \approx \Phi_{\mu,\sigma}(\mu + n\sigma) - \Phi_{\mu,\sigma}(\mu - n\sigma)$$

同时，我们希望有接近一半的交易天数都处于轨道之外，以保证足够的交易频率。通过遍历标准差的倍数，最终决定将其预设为 0.75。

【表 3.2-1】均值回复轨道宽度

均值回复轨道	轨道内理论交易天数	轨道内实际交易天数
$\mu \pm 0.75 \times \sigma$	0.5467	0.561

【图 3.2-1】均值回复轨道



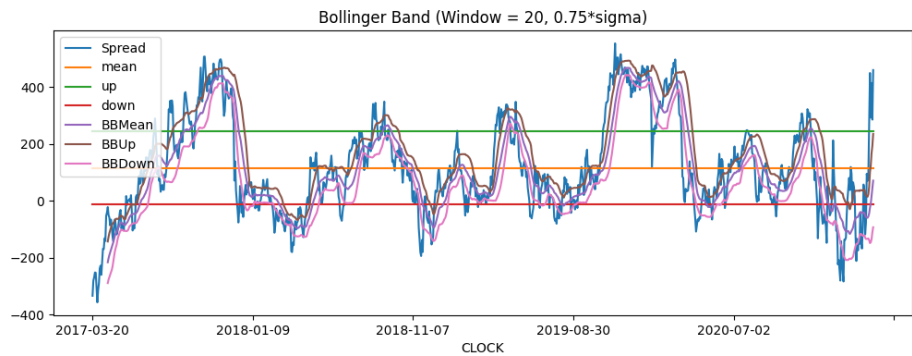
该信号发出的交易频率较低且盈利空间有限，故引入布林带突破指标加以优化。详细的交易方式以及信号、仓位模拟图将会在下一节中给出。

3.3 布林带突破指标构建及表现

【表 3.3-1】指标计算公式

指标名称	指标代码	计算公式
布林带中轨	BBMean	Moving Average(Spread _t , 20)
布林带上轨	BBUp	BBMean+0.75*Moving Std(Spread _t , 20)
布林带下轨	BBDown	BBMean-0.75*Moving Std(Spread _t , 20)

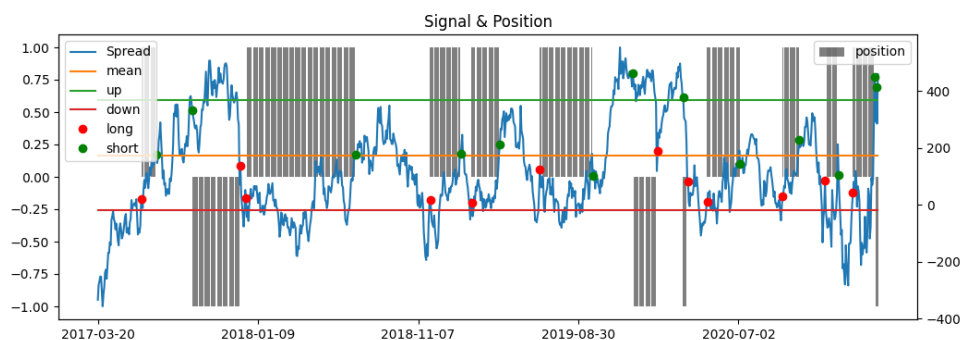
【图 3.3-1】布林带轨道



【表 3.3-2】交易方式描述清单

交易要素	说明
交易标的	rb 螺纹钢；i 铁矿石；SM 锰硅
交易合约	连续合约
交易频率	日频
交易成交时间	次日盘内成交
信号计算方法	详见【表 3.2-1】与【表 3.3-1】
信号计算使用的价格类型	后复权收盘价
信号交易规则	<p>当价差序列处于均值回复轨道上方时，若出现下突破布林带下轨，或出现下突破均值回复上轨道，则在下一个交易日盘内卖出 100 份 rb 螺纹钢连续合约，买入 8 份 i 铁矿石连续合约和 26 份 SM 锰硅连续合约（交易份数指定规则详见【表 3.1-1】）；</p> <p>当价差序列处于均值回复轨道下方时，若出现上突破布林带上轨，或出现上突破均值回复下轨道，则在下一个交易日盘内反向交易上述连续合约；</p> <p>在开仓状态下，当价差序列穿越均值回复中轨（均值）时，平仓所有正在交易的合约。</p>
模拟成交使用的价格类型	开盘价与收盘价均值
交易成本估计	参照不同商品期货交易所的规定，rb 螺纹钢与 i 铁矿石连续合约均收取成交额 的 0.03% 作为手续费，SM 锰硅连续合约每份合约收取 3 元手续费
是否考虑了换月	否
收益计算类型	绝对收益
收益计算使用价格类型	开盘价与收盘价均值

【图 3.3-2】信号与仓位模拟图



对比【图 3.2-1】均值回复轨道可以发现，引入了布林带突破指标后，部分位于均值回复轨道上方的开仓点都要相对地有所提前，即做空的价位更高，从而能够获得更多的收益。值得注意的是，布林带生成过程中采用了 20 个交易日作为移动平均的窗口。该数值可作为参数在后续中进行优化，此处使用了价差序列的均值回复半衰期作为预设值。半衰期反映的是价差序列在偏离均值后，完成回复过程的一半所花费的时间。同样，布林带上下轨之间的距离也采用了和均值回复轨道宽度相同的标准差倍数，0.75 进行预设，也可在后续的策略回测中再进行优化。

3.4 仓位管理方法

标的：多标的（rb 螺纹钢连续合约、i 铁矿石连续合约、SM 锰硅连续合约）

调仓周期：按日调仓

信号描述：在收盘之后用当日的后复权收盘价乘以相应的组合权重得到当日价差 S_t

如果 S_{t-1} 在均值回复轨道上方且 S_t 下穿布林带下轨或下穿均值回复上轨，做空

如果 S_{t-1} 在均值回复轨道下方且 S_t 上穿布林带上轨或上穿均值回复下轨，做多

如果存在仓位且 S_t 穿过均值回复中轨（均值），清仓

仓位方法描述：

每日交易品种：固定为由螺纹钢、铁矿石和锰硅构成的配对交易组合

每日交易资金分配：每次开仓固定交易 rb 螺纹钢连续合约 100 份，按照合约乘数及组合权重对应地交易相应份数的 i 铁矿石与 SM 锰硅连续合约，详见【表 3.1-1】

收益计算方法描述：

$$Profit_t = \sum_{i \in [rb, i, SM]} (Signal_i \times (Price_{i,t} - Price_{i,t-1}) \times Multiplier_i \times Volume_i)$$

$$Return_t = \frac{Profit_t}{Allocation_0}$$

$$Total Profit_t = \sum_{i=0}^t Profit_i$$

$$Total Return_t = \frac{Total Profit_t}{Allocation_0}$$

交易示例描述：

2017 年 6 月 11 日收盘计算价差，满足做多信号

2017 年 6 月 12 日执行交易算法，最终以开盘价与收盘价的均值在收盘前完成 100 份 rb 螺纹钢连续合约的多头，8 份 i 铁矿石连续合约空头和 26 份锰硅连续合约空头

2017 年 6 月 13 日收盘计算价差，并未发出平仓信号

2017 年 7 月 5 日收盘计算价差，满足平仓信号

2017 年 7 月 6 日执行交易算法，最终以开盘价与收盘价的均值在收盘前完成对所有交易品种的平仓，同时，当日并未发出开仓信号

2017 年 7 月 7 日空仓，收盘计算价差，并未发出开仓信号

4. 策略表现

4.1 策略表现

起始资金：1000000

交易份数：rb: 100, i: 8, SM: 26

交易手续费：rb: 0.01%, i: 0.01%, SM: 3RMB/份

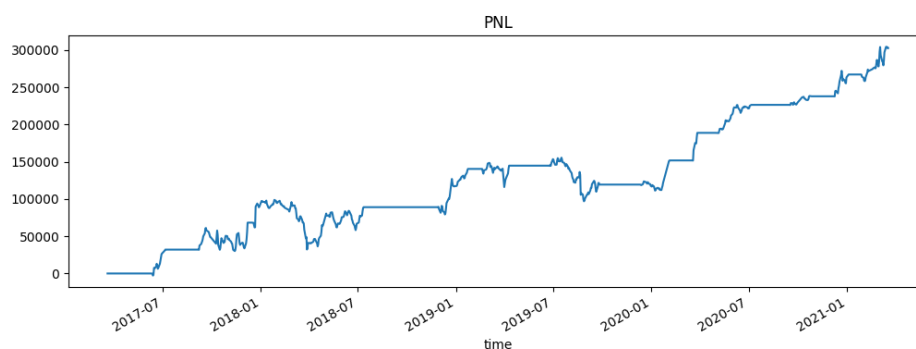
样本内回测周期：2017/03/19-2021/03/19

样本外回测周期：2021/03/19-2022/03/18

【表 4.1-1】样本内策略性能指标

指标	数值
年化收益率	7.5612%
最大回撤	66616
夏普比率	1.8118
卡玛比例	1.1350×10^{-6}
胜率	84.6154
盈亏比	2.2122
平均持仓周期	59.3077

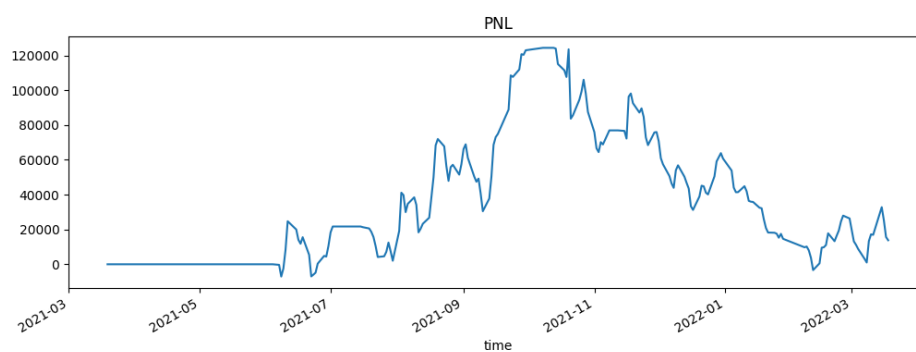
【图 4.1-1】样本内 PNL 曲线



【表 4.1-2】样本外策略性能指标

指标	数值
年化收益率	1.3788%
最大回撤	127679
夏普比率	0.1643
卡玛比例	1.0799×10^{-7}
胜率	50%
盈亏比	1.1247
平均持仓周期	66.25

【图 4.1-2】样本外 PNL 曲线

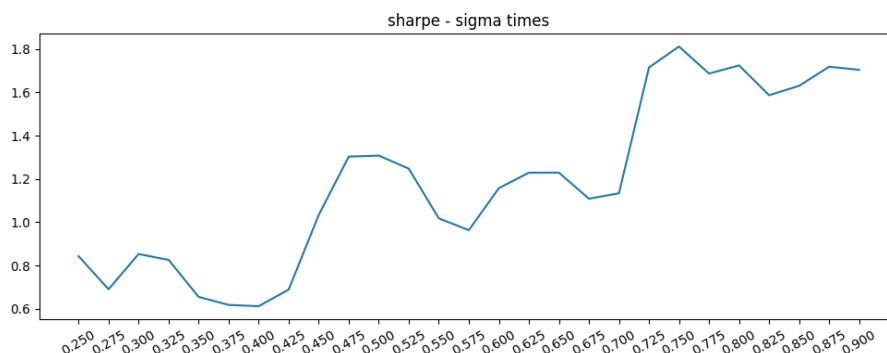


该策略的样本外表现一般，最可能的原因就是过去一年里，用于配对交易的三个品种之间的一阶向量自回归关系发生了变化，导致线性组合不能在样本外产生平稳的价差序列。

4.2 参数敏感性分析

均值回复轨道的宽度为 $2k$ 倍标准差， k 的大小决定了可套利的价格区间，也决定了策略的交易频率（详见【表 3.2-1】）。

【图 4.2-1】标准差倍数 k 与策略样本内夏普比关系图



该策略标准差倍数参数在 0.75 左右表现较为平稳，0.7 倍标准差以下，由于轨道过窄可能会发出更多的噪声信号从而降低夏普比率。

5. 总结

本策略利用 Box-Tiao 典型分解寻找到螺纹钢、铁矿石和锰硅之间的有效线性组合，使其产生的价差序列足够平稳，能够用于构建套利策略。最终确定的组合权重为：rb: 1, i: -0.75, SM: -0.13。同时，由于产生的价差序列均值回复的半衰期为 20 个交易日，将其预设为布林带移动平均值计算时的滑动窗口也较为合理。轨道距离均值之间的宽度， k 倍标准差是根据轨道能够涵盖接近 50% 的交易天数来确定的。根据参数敏感性测试的结果来看，当 $k=0.75$ 时，该参数较为稳定。

该策略的样本外表现不佳，最主要的原因是多品种之间的价格趋同效应发生了变化，因此，仅依赖于一次固定周期的样本内测试来永久确定多品种之间的线性组合是极其不具拓展性的。故未来最主要的改进方向是将多品种之间的线性组合关系求解由静态改为滚动进行，或至少是按周期对线性组合进行重新求解，以更新最优的价差序列。