



商品期货配对交易系列（二）配对交易策略设计

华泰期货研究所 量化组

罗 剑

量化组组长

✉ luojian@htfc.com

从业资格号：F3029622

报告摘要：

投资咨询号：Z0012563

本报告为商品期货配对交易系列的第二篇，将在篇一商品期货配对交易系列（一）《配对交易的理论》延伸讨论，从配对交易策略的资产选择、配对交易策略的可交易性测试和配对交易策略的设计，为投资者介绍配对交易构建、测试与的策略设计的基础框架和方法。

一、期货配对交易策略的分类

参考 Ganapathy Vidyamurthy (2004) 研究，再结合配对交易的三大要素，将配对交易基础策略分为以下两类：**基于统计套利的配对交易**和**基于基本面套利的配对交易**。

基于统计套利的配对交易

对于期货的统计套利配对交易基于预估价格均衡水平，倾向于买入低估值合约，卖出高估值的合约，维持市场中性的头寸，具体步骤包括，**配对标的选择、配对可交易性测试和配对交易设计**。

(一) 配对标的选择

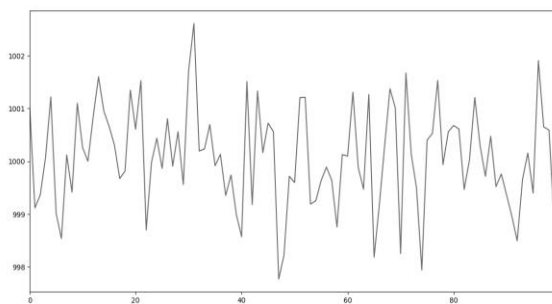
配对标的选择的方法目的为确定配对交易存在平稳关系或具有类似性质、影响因素的相对关系，具体方法可以分为**协整法、因子暴露法和距离法**。

协整法

协整关系需要对配对交易的资产通过平稳性检验和协整性检验来确定。

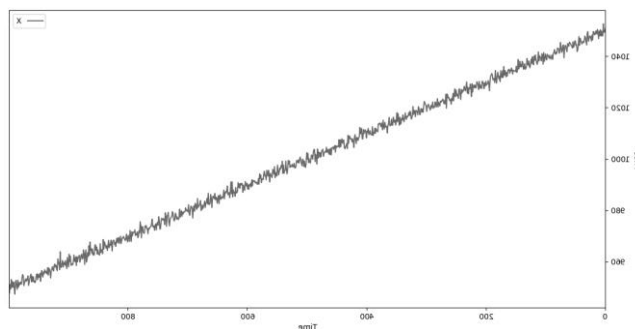
首先平稳序列与非平稳序列参考图 1、2。图 1 平稳序列即为一个以 1000 为均值的噪声序列，且可以看出它的期望与方差协方差均不随时间而改变。图 2 为一个趋势向上的时间序列，且期望随着时间向上变化，即为非平稳序列。

图 1: 平稳序列



数据来源：华泰期货研究所

图 2: 非平稳序列



数据来源：华泰期货研究院

平稳性检验——协整性检验步骤如下：

- (1) 检验配对资产 a,b 时间序列是否平稳
- (2) 若非平稳序列，进行差分序列，直到检验序列服从同阶单整
- (3) 进行协整性检验，判断是否存在协整关系

因子暴露法

Ganapathy Vidyamurthy (2004) 指出统计套利中的协整法，同时将买入低估的 A 股票资产与卖出 B 股票资产的收益率分解为非平稳性部分和平稳性部分，公式如下：

$$\log(\text{Price}_t) = n_t + \varepsilon_t \quad (3.1)$$

$$\log(\text{Price}_{a,t}) - \log(\text{Price}_{b,t}) = (n_{a,t} + \varepsilon_{a,t}) - (n_{b,t} + \varepsilon_{b,t}) \quad (3.2)$$

$$= (n_{a,t} - n_{b,t}) + (\varepsilon_{a,t} - \varepsilon_{b,t}) \quad (3.3)$$

$$= r_t^c + r_t^s \quad (3.4)$$

公式 (3.1) 首先将资产收益分解为非平稳部分 n_t 和平稳部分 ε_t ，公式 (3.4) 即将套利交易收益分解为非平稳的趋势收益 r_t^c 和平稳的残差收益 r_t^s 。

因子暴露法基于普通协整法原则，但是利用因子模型分解资产收益组成，并且利用线性模型寻找配对交易资产 a,b 的线性关系 β 。

如果资产 a,b 存在协整关系，具有线性关系并且资产 a,b 的收益率可以由共同因子解释，资产 a,b 的趋势收益将趋同。假设资产 a 和资产 b 的因子暴露矩阵分别为 x 和 γx ，因子收益率为 $b=(b_1 + b_2 + b_3 + \dots + b_n)$ 。

表 1: 风险因子模型

	资产 a	资产 b
因子风险因素	$x = (x_1 + x_2 + x_3 + \dots + x_n)$	$\gamma x = (\gamma x_1 + \gamma x_2 + \gamma x_3 + \dots + \gamma x_n)$
收益率因子模型	$r_a = (b_1 x_1 + b_2 x_2 + b_3 x_3 + \dots + b_n x_n) + \varepsilon_{a,t}$	$r_b = \gamma(b_1 x_1 + b_2 x_2 + b_3 x_3 + \dots + b_n x_n) + \varepsilon_{b,t}$

数据来源：Ganapathy Vidyamurthy (2004) 华泰期货研究所

由表 1 中 $r_a - \gamma r_b$ 结合公式 (3.4) 可得 $\log(\text{Price}_{a,t}) - \gamma \log(\text{Price}_{b,t}) = [(b_1 x_1 + b_2 x_2 + b_3 x_3 + \dots + b_n x_n) - \gamma(b_1 x_1 + b_2 x_2 + b_3 x_3 + \dots + b_n x_n)] + (\varepsilon_{a,t} - \varepsilon_{b,t})$ ，假设资产 a 与资产 b 满足具有线性关系条件，即可以以检验 $(\varepsilon_{a,t} - \varepsilon_{b,t})$ 是否为平稳序列为标准选择配对。

距离法

Gatev, E., Goetzmann, W. N., and Rouwenhorst, K. G. (2006) 研究中提到可以回溯历史时间窗口，从 n 个资产 n(n-1)/2 个两两配对的欧几里得距离 SSD，利用欧几里得距离 SSD 最小的组合选择为配对组合，计算欧几里得距离 SSD 公式如下：

$$\overline{SSD}_{i,j,t} = \frac{1}{T} \sum_{t=1}^T (P_{i,t} - P_{j,t})$$

(二) 配对可交易性测试

经过上述方法对配对交易目标的选择完成后，还需要对选择的配对进行可交易性分析，包括配对的线性关系检验和配对交易目标的残差稳定性进行检验。

线性关系检验

通过回归方程 (3.5) 的右边等式得到均衡水平 μ 和期望为零的 ε_t ，为了确保时间序列呈均值回复特征，配对价差将在均衡水平 μ 附近来回摆动，线性关系的确定保证了他们的均衡水平 μ 稳定。

$$\log(\text{Price}_{a,t}) - \beta \log(\text{Price}_{b,t}) = \mu + \varepsilon_t \quad (3.5)$$

检验方法可以通过**配对时间序列的线性回归**和多因子线性回归的 R^2 拟合度和回归系数 β 的 P-Value 值判断。

残差稳定性进行检验

根据之前所设想的理想配对交易环境，包括一个稳定的均衡水平和一个平稳的残差序列，即时间序列具有明显的均值回复性。均值回复，即在价格偏离均衡水平一定程度后，会向均值水平收敛的过程，所以除了检验时间序列的平稳性，还可以通过检验均值回复的**击穿均衡水平率**作为检验指标。在单位时间内击穿均衡水平的次数越多，说明均值回复速度越快，即均值回复特性越强。

同时，可以采用 bootstrap 检验方法比较随机进入市场与配对交易收益率比较，利用蒙特卡罗仿真技术匹配配对交易信号的利润是否具有偶然性。最简单的方法是模拟采用随机方法进入市场，计算配对交易信号优于随机进入市场的比例，通常判断配对交易为优秀的标准以获得收益在 80% 的比例上优于随机模拟收益。

(三) 配对交易设计

在配对交易的设计中，参考配对交易的技术分析要素，需要重点考虑两个问题，首先要确定**如何确定均衡水平**，再者需要考虑**配对交易价差的趋势**。以下本报告列举常用的解决方法仅供参考。

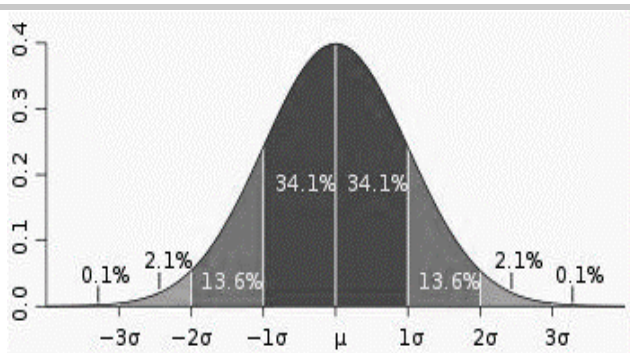
标准差

配对交易者利用配对交易的价差时间序列的标准差判断偏离程度作为入场依据，标准差可以定义的数学计算公式如下：

$$\sigma(r) = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (3.6)$$

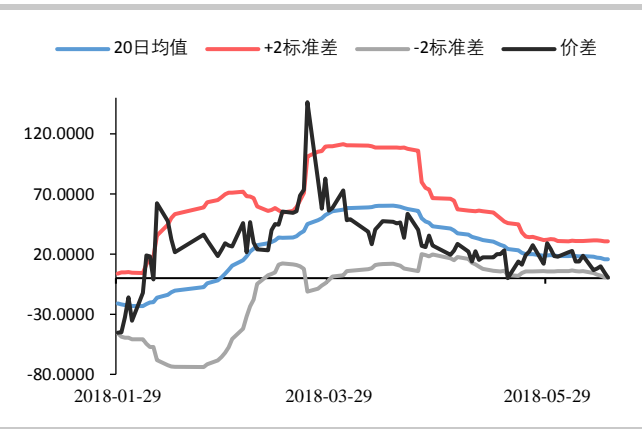
数据的时间序列所有数值减去其平均值的平方和，所得结果除以该组数之个数，将结果所得值开根号，所得之数就是这组数据的标准差，反映了数据的离散程度，假设数据的时间序列呈正态分布，距平均值小于一个标准差之内的数值范围占比率为全部数值之68%。两个标准差之内的比率合起来为95%。正负三个标准差之内的比率合起来为99%。

图 3: 正态分布标准差概率图



数据来源：网络图片 华泰期货研究所

图 4: 沪深 300 指数与 IF1806 价差的标准差系统



数据来源：Wind 华泰期货研究院

图 4 构造了最简单的两倍标准差入场系统，即传统的类布林带指标，以沪深 300 指数与 IF1806 指数期货合约的 20 日收盘价差为均值，上下两倍标准差为判断价差偏离异常的标准，当价差大于或小于±2倍标准差时入场，预估价差将向均值收敛，并在均值水平出场。

z-score

z-score 亦称为标准分数，是一个数与平均数的差再除以标准差的过程，可以表示一个分数距离平均数的相对标准距离，具体公式如下：

$$z = \frac{x - \mu}{\sigma} \quad (3.7)$$

其中 μ 为总体均值， $x - \mu$ 为离均差， σ 表示总体标准偏差。

图 5: z-score 标准分数图

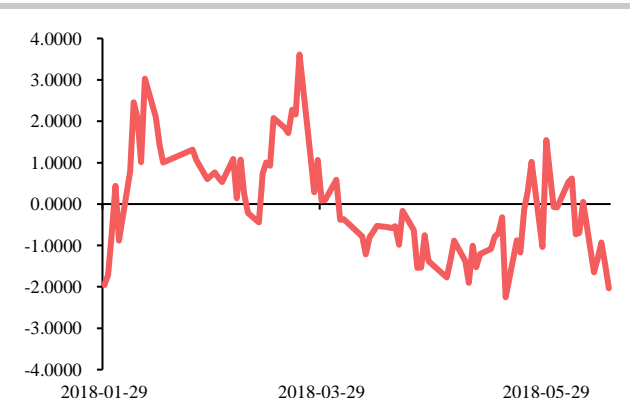
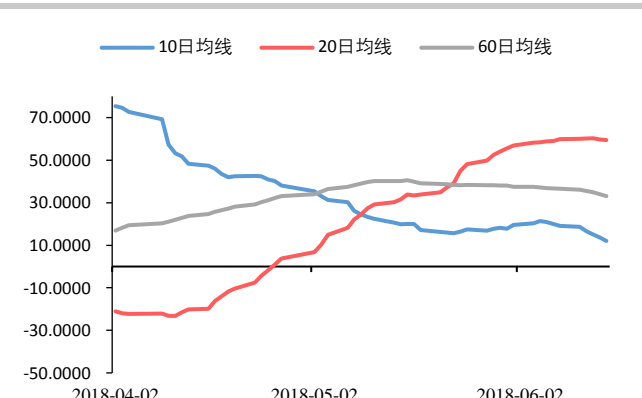


图 6: 沪深 300 指数与 IF1806 价差均线水平



数据来源：华泰期货研究所

数据来源：Wind 华泰期货研究院

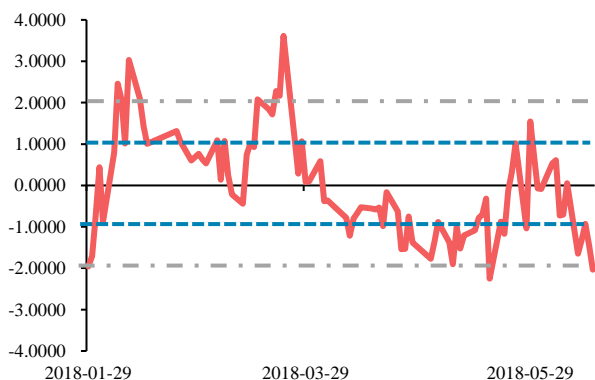
配对交易在应用 z-score 方法是主要为了方便对比不同配对组合的偏离程度，z-score 以标准差倍数的形式表示，方便选择多组配对组合中偏离程度最大或收敛可能性最高的组合。

图 5 将图 4 的沪深 300 指数与 IF1806 价差进行 z-score 标准分数化，图 4 的价差走势和图 5 的 z-score 的走势一致，但相对图 4，图 5 的 z-score 标准分数更清楚表达价差处于多少倍的标准差偏离水平，更容易衡量偏离程度。

标准差作为配对交易判断偏离程度的标准并不完美，由于标准差计算依赖历史数据窗口选择，而历史数据窗口的错误选择容易引起**均衡水平的预估**和**价差收敛趋势**的预估两个问题。

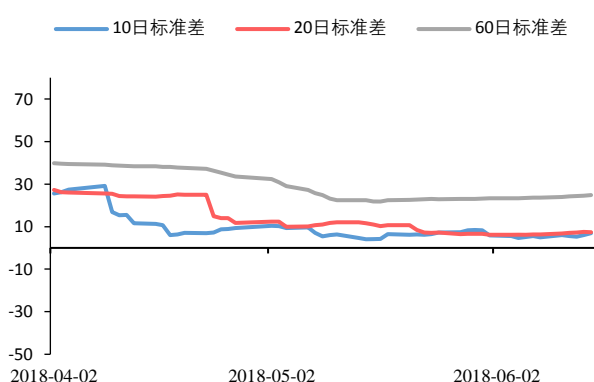
简单用收盘价均值预估均衡水平，对历史窗口参数极为敏感，图 6 展示以不同周期的均线作为均衡水平相差较远，容易真实均衡水平的错误预估。

图 7: z-score 标准分数图



数据来源：华泰期货研究所

图 8: 沪深 300 指数与 IF1806 价差不同周期标准差



数据来源：Wind 华泰期货研究院

同样，不同的周期选择会引起标准差水平的差异，导致利用标准差判断价差收敛趋势的错误预估，如图 8 沪深 300 指数与 IF1806 价差不同周期标准差，时间窗口参数的敏感性较强差异较大。

多重均衡水平设置

考虑均衡水平对于配对交易策略的收益起决定性作用，可将设置多重均衡水平应对真实均衡水平的漂移。如图 7 z-score 标准分数图，在原本只设定 0 为均衡水平，入场择时为 z-score 超过 ± 2 水平，收敛至 0 附近出场，从观测期 2018 年 1 月至 2018 年 6 月只有 3 次机会。若考虑同时将 z-score 值 ± 1 也同时设为均值水平，在 z-score 值 ± 1 均衡水平上平移 ± 1 倍标准差作为入场时点，大大增加了观测期间的获利机会。

多重窗口设置

鉴于时间窗口对于简单利用历史均值确定均衡水平的参数设置比较敏感，可以利用多重窗口平滑均衡水平预测精度，赋予远近价格数据不同的权重，包括不同的加权算法。

(1) 加权移动平均

加权移动平均 (Weighted moving average, WMA) 指计算平均时个别数据乘以不同数值，n 日 WMA 的当期一个数值乘以 n、上一期的乘以 n-1，如此类推，一直到 0，具体公式如下：

$$WMA_n = \frac{price_n * n + price_{n-1} * (n-1) + price_{n-2} * (n-2)}{n + (n-1) + (n-2)}$$

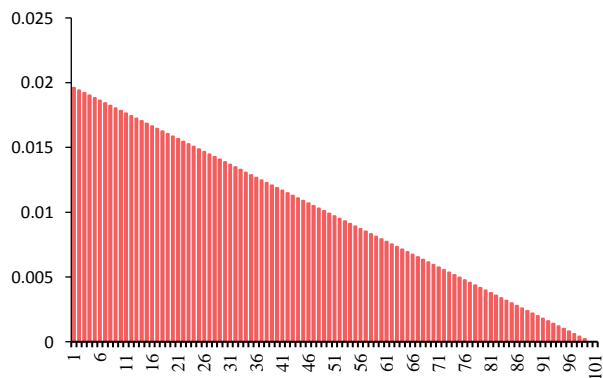
(2) 指数加权移动平均

Exponential Moving Average, EMA 或 EWMA 称作指数加权移动平均，是以指数式递减加权的移动平均。各数值的加权按时间久远而指数式递减，越近的数据权重越高，较久远的数据权重递减，但仍维持也给予一定的权重。

递减速度以常数 α 决定， α 数值介乎 0 至 1。具体公式如下：

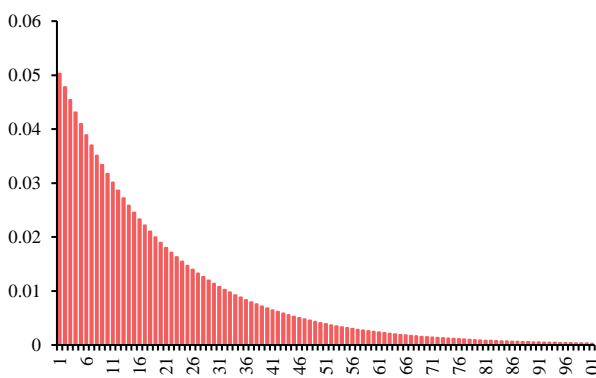
$$EWMA = \frac{price_n(1-a)^n + price_{n-1}(1-a)^{n-1} + \dots + price_2 * (1-a) + price_1}{(1-a)^n + (1-a)^{n-1} + \dots + (1-a) + 1}$$

图 9: WMA 加权移动平均



数据来源：华泰期货研究所

图 10: EWMA 指数加权移动平均



数据来源：华泰期货研究院

基于基本面套利的配对交易

基本面套利的配对交易是一种依据基本面逻辑的策略，收益源区别于统计套利，其由基本面因素驱动，根据基本面因素的驱动因子逻辑和收敛的类型，可分为风险套利和无风险套利。

（一） 风险套利

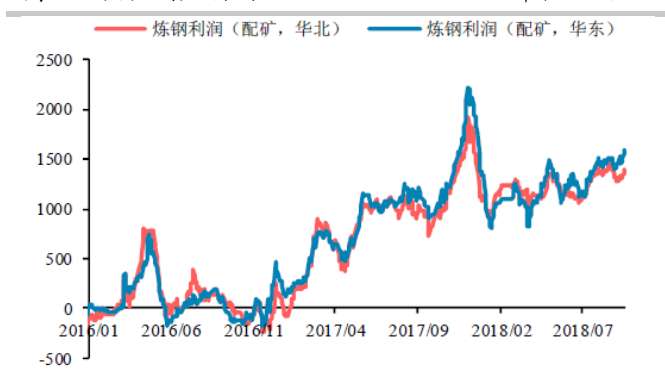
风险套利具有理论收敛的特征，但却不是绝对收敛，因为风险套利的收敛性受风险因子影响，而风险因子可能导致理论收敛不成立。期货的产业套利就属于风险套利，产业利润水平、生产工艺配方和上下游产业情况等均可能成为产业套利不收敛的风险因子。

1、 钢厂利润

钢厂利润为黑色系收市场最为关注的产业套利模式，一吨螺纹钢根据生产工艺可近似表达为：

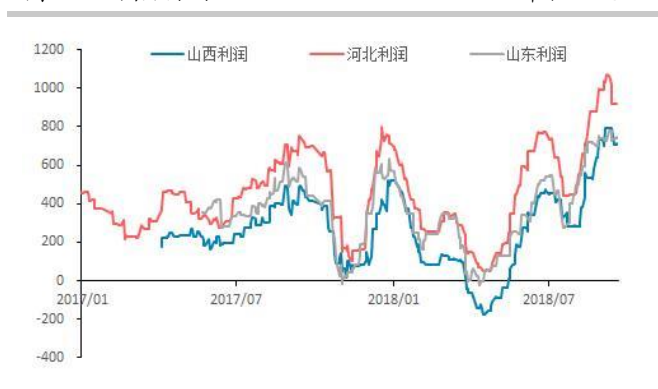
$$1 \text{ 吨螺纹钢} = 1.6 \text{ 吨铁矿} + 0.5 \text{ 吨焦炭} + \text{杂费（常数）}$$

图 11: 模拟钢厂利润 单位：元/吨



数据来源：华泰期货研究所

图 12: 炼焦利润 单位：元/吨



数据来源：Wind 华泰期货研究院

从图 11 的模拟钢厂利润走势可以发现，2016 年与 2018 的钢厂利润平衡点由于政策性原因，均衡水平有所偏移，且近两年整体黑色系波动性较高，受盘面资金和限产政策，钢厂利润走势也波动较大。

在进行产业生产利润配比的套利时，除了需要考虑从基本面和政策面考虑钢厂利润的均衡水平，还要了解由于盘面期货合约手数限制，并未能考虑实际钢厂利润中涉及的相关其他成本，有一定的风险敞口。同时也需要从统计套利的角度预估钢厂利润均衡水平。

2、 炼焦利润

焦煤和焦炭同属于黑色产业链，且焦煤为焦炭的生产原料，从传统炼焦工艺显示，1 吨焦炭的生产需要大概 1.37 吨的焦煤。

$$1 \text{ 吨焦炭} = 1.37 \text{ 吨焦煤} / 1.17 - 185(\text{回收物}) * 1.17 + 100$$

3、大豆压榨利润

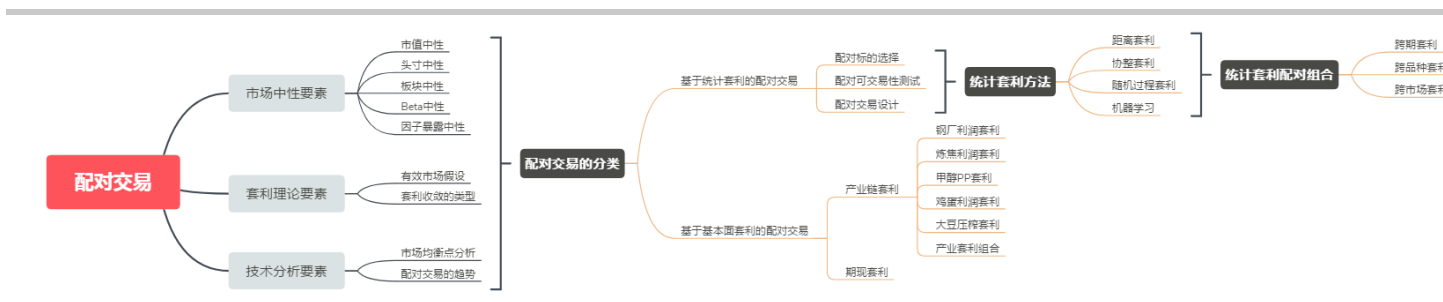
豆油主要利用进口大豆压榨，进口大豆的出油率达到 20% 高于国产大豆 16% 出油率，因此大豆压榨利润主要关注进口大豆成本，豆油价格和豆粕价格，具体公式可近似于：

$$100\% \text{进口大豆} = 19.2\% \text{豆油} + 78.5\% \text{豆粕} + 2.1\% \text{损耗}$$

(二) 期现套利

期货市场存在实物交割制度，随着期货合约到期日临近，临近的期货合约价格将向可交割的期货合约标的物价格收敛，但在期货合约到期日前，期货价格与现货价格的影响因子相似但不完全相同，如从现货价格受便利收益因子影响，期货价格受交割成本因子，包括仓储成本、运输成本和制作可交割仓单相关费用影响，同时也受投资者未来对价格的预期和资金流因子影响，所以在交割日来临前可发生期现价格偏离，出现套利机会。

图 13: 配对交易基础与分类



数据来源：华泰期货研究所

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2018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、29层04单元

电话：400-6280-888

网址：www.htfc.com