

国联期货研究所

证监许可[2011]1773号

分析师:

黎伟

从业资格号: F0300172

投资咨询号: Z0011568

相关研究报告:

目 录

1、 引言	- 2 -
2、 研究框架	- 2 -
2.1 商品因子模型理论基础	- 2 -
2.2 常见数据处理方法	- 2 -
2.3 因子构建	- 3 -
2.3.1 动量因子	- 3 -
2.3.2 期限结构因子	- 4 -
2.3.3 库存类因子	- 5 -
2.3.4 套保压力因子	- 5 -
2.3.5 持仓类因子	- 5 -
2.3.6 波动率类因子	- 6 -
2.3.7 流动性因子	- 6 -
2.4 单因子检验	- 6 -
2.4.1 截面回归	- 7 -
2.4.2 分层回测	- 7 -
2.5 多因子检验	- 8 -
2.6 多因子模型多空策略构建方式	- 8 -
3、 小结	- 9 -

1、引言

多因子模型在股票市场投资中应用已十分广泛，近几年随着商品期货品种的不断丰富，越来越多的学者或机构将股票多因子模型应用到商品期货因子研究中。本文意在探讨商品多因子模型的研究框架，梳理现有主要商品因子构建方式，为后续因子改进、新因子挖掘、商品因子模型回测与优化等打下基础。

2、研究框架

2.1 商品因子模型理论基础

商品多因子模型其基本假设为商品期货收益率与因子间存在某种线性关系。一般可以用下面的公式进行刻画，即：

$$r_t = x_t f_t + u_t$$

假设有 N 个商品， K 个因子，则 r_t 为商品期货收益率向量 ($N \times 1$)， f_t 为该期因子收益率向量 ($K \times 1$)， x_t 表示因子暴露矩阵 ($N \times K$)， u_t 代表不能被因子所解释的部分。

在实务研究中，我们更多是需要探讨本期商品因子数据是否对下期商品期货收益率具有预测效应。因此我们更多是采用下期商品期货收益率和本期因子暴露做截面回归。

2.2 常见数据处理方法

任何模型建立前我们都需对底层数据进行相应处理。商品多因子模型其主要数据包含商品期货本身量价数据、库存仓单数据、会员持仓数据、基本面数据、宏观数据等，其数据可能会存在市场缺失、数据异常等特点，一般需要在测试前对数据进行相应清洗。

(1) 主力合约拼接及处理

无论投机类策略还是多空对冲类策略，最终都需要落实到具体回测标的。期货回测标的一般有两种可以选择模式：其一采用主力合约，其二采用加权指数合约。因部分品种换月合约价差较大，我们往往不能直接利用主力价格进行回测，为更贴近实际投资环境需要考虑展期时的处理规则。

主力合约的切换一般参考日持仓量或者成交量的变动作为基准，如：

(a) 若 T 日某晚月（按照日历月排序，更后即“晚”）月份合约持仓量高于目前主力合约持仓量，则在 $T+1$ 日将主力合约转换至该晚月份合约。同时为避免主力合约在短时期内出现来回切换的情形，可以规定主力合约仅向更晚月的合约月份进行单向切换。

(b) 主力合约的拼接可以采用收益率序列直接拼接形式，连续价格序列净值 $P_t = P_{t-1} * (1 + y)$ ，这

里 y 表示主力合约涨跌幅。需要注意的是采用该种收益序列在实际操作中仅能使用等面值移仓的方式进行，若采用等手数移仓的模式，则只能在程序回测时自行重新设置。

加权指数合约为各月份合约按成交量或持仓量加权所得，其曲线相对连续，但对于交易频率偏短期的交易策略其回测误差相对偏大。

(2) 缺失值的处理

原始数据尤其商品期货基本面数据很多时候存在数据缺失情形，对于缺失数据可采用如下两种方法：其一对缺失数据直接删除该时间段；其二采用前后临近数据取其均值进行填充。

(3) 数据去极值化

原始数据中可能存在部分异常值，这可能会给后期因子检验带来一定影响。实务中我们需要对数据进行去极值化处理，比较常见的处理方法有：MAD 去极值法、3 西格玛法、百分位法等。以 MAD 去极值法为例：

$$x_i = \begin{cases} x_{\text{median}} + n * \text{MAD}, & \text{if } x_i > x_{\text{median}} + n * \text{MAD} \\ x_{\text{median}} - n * \text{MAD}, & \text{if } x_i < x_{\text{median}} - n * \text{MAD} \\ x_i & \end{cases}$$

这里 x_{median} 表示序列中位数，MAD 表示各因子与中位数绝对偏差后取其中位数。

(4) 数据标准化

由于不同因子量纲的不同，无法直接对其进行截面回归，需要对相关数据进行标准化。数据标准化常见方法有极差标准化法、Z-Score 标准化法等。以 Z-Score 标准化法为例：

$$x_i = \frac{x_i - \mu}{\sigma}$$

标准化后的序列将变为以 0 为均值，1 为标准差的新序列。

2.3 因子构建

2.3.1 动量因子

动量效应顾名思义就是过去表现较好的品种在未来一段时间亦可能有较好表现，目前市面上已有大量衡量动量或趋势类的因子被挖掘。如：

表 1：部分动量因子计算方法

因子名称	计算公式	说明
简单截面动量	$\frac{S_T - S_{T-K+1}}{S_{T-K+1}}$	过去 K 天累计收益率
稳健动量	Mean(C_Rank _t)，C_Rank _t 为 t 时刻根据截面排名赋分标准化后数值	过去 K 天截面收益率排序后的标准化得分均值

趋势系数	$\frac{S_T - S_{T-K+1}}{\sum abs(S_t - S_{t-1})}$	简单动量基础上充分考虑路程的不同
日内动量	$\frac{1}{K} \sum_{t=T-K+1}^T y_t$	过去 K 天日内涨幅均值
隔夜动量	$\frac{1}{K} \sum_{t=T-K+1}^T t \text{ 日跳空比例}$	过去 K 天隔夜跳空均值
RSI 动量因子	A / (A+B), A=过去 K 天收盘涨幅之和; B=过去 K 天收盘跌幅之和	过去 K 天累计涨幅与累计绝对涨跌幅比值
乖离率	$\frac{\text{收盘价} - \text{过去 K 天收盘价均值}}{\text{过去 K 天收盘价均值}}$	过去 K 天乖离率

数据来源: Wind 国联期货研究所

2.3.2 期限结构因子

期限结构主要用于刻画同一品种不同期限月份合约之间的价差关系,该因子亦是商品期货市场历史最为悠久的一类因子。一般而言,当近月相对远月价格更强势,或现货相对期货价格更高时,此时做多期货往往具有相对优势。因其与动量类因子在构造逻辑上有明显区分,其亦能起到较好的互补效果。常见的衡量期限结构因子的方法有:

表 2: 部分期限结构因子计算方法

因子名称	计算公式	说明
展期收益率	$\frac{1}{K} \sum_{t=T-K+1}^T [\ln(S_t^{\text{近月}}) - \ln(S_t^{\text{远月}})] * 365 / \text{间隔时间}$	近月较远月年化升水率过去 K 天均值,用对数收益率替代
基差动量	$\prod_{t=T-K+1}^T (1 + \text{ret}_t^{\text{近月}}) - \prod_{t=T-K+1}^T (1 + \text{ret}_t^{\text{远月}})$	过去 K 天持续做多近月合约、做空远月合约收益率
主次价差	$\frac{1}{K} \sum_{t=T-K+1}^T [\ln(S_t^{\text{主力}}) - \ln(S_t^{\text{次主力}})] * 365 / \text{间隔时间}$	主力较次主力年化升水率过去 K 天均值,用对数收益率替代
期现价差	$\frac{1}{K} \sum_{t=T-K+1}^T [\ln(\text{现货}) - \ln(S_t^{\text{主力}})] * 365 / \text{间隔时间}$	现货较主力合约年化升水率过去 K 天均值,用对数收益率替代

数据来源: Wind 国联期货研究所

当然,亦有部分投资者不采用均值的形式,仅用最新值进行计算。

2.3.3 库存类因子

库存是商品供需基本面因子的核心指标之一，其一般可以分为交易所库存和社会库存。交易所库存一般可以用交易所注册仓单进行替代；社会库存的数据来源则相对分散，一般由第三方数据机构调研获得。该类因子的计算一般采用同比或环比变化率，同时为了避免数据毛刺太大，可以采用均值的方式进行平滑。如下：

表 3：库存因子计算方法

因子名称	计算公式	说明
仓单因子	$\ln(T \text{ 期仓单移动均值}) - \ln(T-K \text{ 期仓单移动均值})$	仓单移动平均后间隔 K 天的环比变化率，用对数收益率替代
社会库存因子	$\frac{1}{K} \sum_{t=T-K+1}^T [\ln(\text{库存}_t) - \ln(\text{库存}_{t-1})]$	过去 K 期库存的环比变化率均值，用对数收益率替代变化率

数据来源：Wind 国联期货研究所

2.3.4 套保压力因子

对冲压力假说认为净多头（空头）的投机者能够从净空头（多头）的套期保值者那里获得相应的风险补偿。由于实际操作过程中我们无法有效区分交易者的性质，有部分机构可以采用该时间段内持仓量变动与累计成交量的比值进行近似，该值越高可以认为套期保值参与者比例相对更高，做多该类合约能够获得更多的风险补偿。因子定义如下：

$$\text{套保压力因子} = \frac{\text{abs}(OI_T - OI_{T-K+1})}{\sum_{t=T-K+1}^T VOL_t}$$

其中 OI 代表当日持仓量，VOL 代表当日成交量。

2.3.5 持仓类因子

期货持仓量公开数据有两类，一类是各个品种合约的具体持仓数量；另一类是交易所公布的前二十会员多空净持仓数量。一般有如下几种计算方式：

表 4：部分持仓类因子计算方法

因子名称	计算公式	说明
总持仓变化率	$\ln(OI_T) - \ln(OI_{T-K+1})$	过去 K 天持仓量环比变化率，用对数收益率替代
会员持仓因子 1	$\frac{T \text{ 期多空净头寸} - (T-K+1) \text{ 期多空净头寸}}{\text{abs}((T-K+1) \text{ 期多空净头寸})}$	会员多空净头寸环比变化率

会员持仓因子 2	$\frac{\text{多空净头寸} - \text{过去}K\text{期多空净头寸均值}}{\text{abs}(\text{过去}K\text{期多空净头寸均值})}$	会员持仓多空净头寸相比过去K天净头寸均值变化率
----------	--	-------------------------

数据来源: Wind 国联期货研究所

2.3.6 波动率类因子

波动率因子亦一类重要的风险因子, 计算波动率的方法有很多, 最常用的为日度对数收益率的年化标准差, 除此之外亦有其他一些波动率计算方式, 如:

表 5: 各波动率计算公式

方法	公式
Parkinson	$\sigma = \sqrt{\frac{1}{4m \ln 2} \sum_{t=1}^T (\ln \frac{H_t}{L_t})^2}$
Garman-Klass	$\sigma = \sqrt{\frac{1}{m} \sum_{t=1}^m \frac{1}{2} (\ln \frac{H_t}{L_t})^2 - \frac{1}{m} \sum_{t=1}^m (2 \ln 2 - 1) (\ln \frac{C_t}{C_{t-1}})^2}$
Rogers-Satchell	$\sigma = \sqrt{\frac{1}{m} \sum_{t=1}^m ((\ln \frac{H_t}{C_t})(\ln \frac{H_t}{O_t}) + (\ln \frac{L_t}{C_t})(\ln \frac{L_t}{O_t}))}$
Garman-Klass Yang-Zhang	$\sigma = \sqrt{\frac{1}{m} \sum_{t=1}^m \left(\frac{1}{2} (\ln \frac{h_t}{l_t})^2 - (2 \ln 2 - 1) (\ln \frac{c_t}{o_t})^2 + \ln \frac{L_t}{C_{t-1}} \right)}$
Yang-Zhang	$\sigma = \sqrt{\sigma_{\text{隔夜}}^2 + k \sigma_{\text{当天}}^2 + (1 - k) \sigma_{\text{RS}}^2}$ <p>其中: $\sigma_{\text{隔夜}}^2 = \frac{1}{m-1} \sum_{t=1}^m \left[\ln \left(\frac{o_t}{c_{t-1}} \right) - \overline{\ln \left(\frac{o_t}{c_{t-1}} \right)} \right]^2$</p> $\sigma_{\text{当天}}^2 = \frac{1}{m-1} \sum_{t=1}^m \left[\ln \left(\frac{c_t}{o_t} \right) - \overline{\ln \left(\frac{c_t}{o_t} \right)} \right]^2, k = \frac{0.34}{1 + \frac{m+1}{m-1}}$ $\sigma_{\text{RS}}^2 = \sqrt{\frac{1}{m} \sum_{t=1}^m ((\ln \frac{h_t}{c_t})(\ln \frac{h_t}{o_t}) + (\ln \frac{l_t}{c_t})(\ln \frac{l_t}{o_t}))}$

数据来源: 国联期货研究所

除此外, 亦有投资者用日收益率标准差的平方与日收益率均值比值来描述期货波动特征。

2.3.7 流动性因子

市场流动性往往用于度量期货品种的交易活跃水平, 其计算方式可以用平均买卖价差、换手率、成交持仓比、成交量与日收益率绝对值比值等来表达。如:

$$\text{流动性因子} = \frac{1}{K} \sum_{t=T-K+1}^T \frac{\text{日成交量}}{\text{abs}(\text{日收益率})}$$

2.4 单因子检验

各因子对未来标的期货收益率是否具有预测效应, 因子检验是必不可少的过程。截面回归和分层

回测是两种常见的检验方法。前者主要检验因子在截面上是否具有预测性能，其更注重商品期货收益率与因子间的线性程度；后者则重点关注因子排序后两端品种的多空对冲综合收益情况。

2.4.1 截面回归

(1) 简单线性回归

将因子经过去缺失值填充、MAD 去极值、标准化处理后，带入截面回归方程，其中因变量为各期货品种的 T+1 期收益率向量，自变量为具体因子打分，可以采用标准化后的因子值进行替代。具体如下：

$$R_{t+1} = \alpha_t + \beta * X_t + \varepsilon_t$$

这里 R_{t+1} 表示商品期货 T+1 期收益率向量， β 为待求解的因子收益率， X_t 表示 T 期因子暴露向量，可以用标准化后的因子值替代。实际操作中可以采用滚动截面回归的方式（如每周），观察累计因子收益率情况。

(5) 因子有效性检验

因子是否有效常见的检验方式为通过检验信息系数 IC 和信息比率 IR 进行判断。IC 表示因子暴露向量与期货下期收益率的截面相关系数，而 IR 近似计算为 IC 序列均值与标准差的比率。实际操作中为防止各因子暴露值差距过大，常使用 RankIC 与 RankIC_IR 来进行替代。其具体公式如下：

$$\text{RankIC}_t = \frac{\text{cov}(\text{sort}(X_t), \text{sort}(R_{t+1}))}{\sqrt{\text{var}(\text{sort}(X_t))\text{var}(\text{sort}(R_{t+1}))}}, \quad \text{RankIC_IR}_t = \frac{E(\text{RankIC}_t)}{\sqrt{\text{var}(\text{RankIC}_t)}}$$

其中 $\text{sort}(X_t)$ 表示 t 期因子暴露值排序， $\text{sort}(R_{t+1})$ 表示 t+1 期商品期货收益率排序。一般认为 RankIC_t 绝对值大于 0.03，该因子与下期资产收益率相关性较强，且 RankIC_IR_t 大于 0.5 时因子稳定获取超额收益的能力较强。

2.4.2 分层回测

因子有效性检验的另外一种有效直接的方法是根据因子大小排序构建多空组合，该方法也称为分层回测法，该方法的直接好处是测试组合具有直接可投资性，但无法确保其对其他因子的暴露中性，即测试结果可能亦受其他因子的共同影响。具体步骤如下：

(1) 确定回看期 K 和调仓周期 L（如 L=5，代表每周）、测试期货截面品种；

(2) 计算回看期因子值并经过填充、去极值、标准化处理；

(3) 根据因子值打分排序，分为 N 组（如 5 组），做多打分前 20% 品种，同时做空后 20% 品种。

上述过程中另一个重要参数是组内期货品种之间的仓位配比，常见的配置方式有：等面值加权、波动率倒数加权等。

2.5 多因子检验

多因子检验首先需要解决的问题便是因子间的多重共线性问题，共线性会导致某些因子在投资组合上重复暴露，从而影响组合长期表现。剔除共线性可以从因子构建逻辑和因子间的相关性进行综合考虑，如果因子逻辑类似或相关性较高，则可留下单因子检验中表现更好的因子，舍弃不好的因子。数学上可以采用主成分分析（PCA）进行降维，亦可考虑因子正交化的方式来剔除因子间的共线性。

接着将筛选后的因子进行多因子回归建模，以当期因子为自变量，下期期货收益率为因变量进行截面多元线性回归。如下：

$$\begin{bmatrix} R_{1,t+1} \\ R_{2,t+1} \\ R_{3,t+1} \\ \vdots \\ R_{N,t+1} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \alpha_t + \begin{bmatrix} \beta_{1,t}^1, \beta_{1,t}^2, \beta_{1,t}^3, \dots, \beta_{1,t}^p \\ \vdots \\ \beta_{N,t}^1, \beta_{N,t}^2, \beta_{N,t}^3, \dots, \beta_{N,t}^p \end{bmatrix} \begin{bmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \\ \vdots \\ X_t^p \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \\ \vdots \\ \varepsilon_{N,t} \end{bmatrix}$$

这里 X 为待求解的因子收益率向量， β 为因子暴露矩阵， α_t 为回归模型中的截距项，类似为股票多因子模型中的国家因子， ε 为残差项。

通过对因子累计收益率的分析，便可直观观测到各因子对期货收益率的贡献大小和稳定程度。同时，通过分析截面回归 R^2 水平高低，可以分析各因子对商品期货收益率的解释力。

2.6 多因子模型多空策略构建方式

商品多因子模型基础应用便是因子多空策略的构建，其构建方式有多种可以参考。如完全借鉴单因子分层测试思路，对因子进行合成加权打分，做多综合打分前 20% 的品种，做空综合打分后 20% 的品种。因子合成加权打分方式亦有多种，如因子标准化后直接等权相加形成综合打分，或者使用各因子排序后进行赋分等。

下面我们介绍另一种完全依据多因子模型预测期货收益率后进行多空对冲方式。具体步骤如下：

- (1) 依据上节中多因子模型采用 OLS 法估算出 α_t 和 X_t ；
- (2) 依据求解出来的当期因子收益率估算下期因子收益率，如采用简单移动平均法、加权移动平均法等；
- (3) 利用新的因子收益率带入多元回归模型中估算再下一期商品期货收益率；
- (4) 做多预测收益率排名前 20% 的品种，同时做空预测收益率排名后 20% 的品种，品种仓位分配可采用等面值或波动率倒数加权等方式。

3、小结

本文主要就商品单因子和多因子模型的研究框架做了简单的梳理,给出了目前常见的一些商品因子构建方式,为后续因子改进、新因子挖掘、商品因子模型回测与组合优化等打下基础。当然整个框架仍有精进空间,如缺少品种仓位的动态调整机制、风险因子暴露的约束、板块总仓位的约束等。后期我们将对单因子及多因子模型进行具体回测及优化,挖掘出新的有效因子,为投资者做一定参考。

免责声明

本报告中信息均来源于公开资料，我公司对这些信息的准确性和完整性不作任何保证。

报告中的内容和意见仅供参考，并不构成对所述期货操作的依据。由于报告在撰写时融入了研究员个人的观点和见解以及分析方法，如与国联期货发布的其他信息有不一致及有不同的结论，未免发生疑问，本报告所载的观点并不代表国联期货公司的立场，所以请谨慎参考。我公司及其研究员对使用本报告及其内容所引发的任何直接或间接损失概不负责。

本报告所提供资料、分析及预测只是反映国联期货公司在本报告所载明日期的判断，可随时修改，毋需提前通知。

本报告版权归国联期货所有。未经书面许可，任何机构和个人不得进行任何形式的复制和发布。如遵循原文本意的引用，需注明引自“国联期货公司”，并保留我公司的一切权利。

期市有风险 投资需谨慎

联系方式

国联期货研究所无锡总部

地址：无锡市金融一街8号国联金融大厦6楼（214121）

电话：0510-82758631

传真：0510-82758630

国联期货研究所上海总部

地址：上海市浦东新区滨江大道999号高维大厦9楼（200135）

电话：021-60201600

传真：021-60201600