

探寻股指期货跨品种的最优组合

另类交易策略系列之三十

报告摘要:

● 股指期货多品种的组合问题

随着上证 50 股指期货和中证 500 股指期货的推出,股指期货投资策略得到了很大的丰富。一方面,在不同品种期货上执行策略可以分散风险,有利于提高策略的多样化和资金容量;另一方面,不同品种股指期货的存在也使得指数的组合搭配变得方便可行。

本报告研究的是股指期货的隔夜策略,考虑了两种不同的股指期货多品种组合方式。第一种是给定组合为上证 50-中证 500 风格套利的组合;第二种是事先未给定组合,按照一定准则进行“最优组合”分析的情况。

● 最优组合的求解

在上证 50-中证 500 风格套利的前提下,我们可以建立多因子打分模型和多元线性回归模型来获取预测函数,根据预测函数确定多大盘空小盘或者是空大盘多小盘的组合来进行建仓。

在事先不给定组合的情况下,我们建立了一个目标函数:最大化预测函数和目标函数的相关系数。这个目标函数求解实际上是统计学习中的典型相关分析(CCA)模型,可以转换为一个特征值分解问题进行求解。通过求解该问题,我们可以同时获得组合的权重系数和预测函数的参数。

● 策略表现

从实证结果来看,本文提出的几个模型的表现都好于风格动量策略等单个预测因子的大小盘风格套利模型的表现。

在 2015 年 8 月份以后,单个预测因子的大小盘风格套利模型收益下滑,回撤变大,多预测因子构建的风格套利策略也经历了比较大的回撤,但是总体来看,收益回撤情况仍然好于单个预测因子的模型。其中,多因子打分模型在盈利把握不大的时候降低资金仓位,从而控制了策略的回撤,取得了明显好于单因子表现的收益回撤比和信息比。该模型从 2010 年至今,取得了 20.8% 的年化收益,历史最大回撤为 -10.5%。

在 2015 年 8 月份之后,市场的大小盘风格切换频繁,50-500 风格套利模型收益下滑,出现了明显的回撤。如果放松对大小盘风格套利的约束,参数不断更新的 CCA 模型的表现好于其他的几个策略。从 2010 年至今,取得了 17.2% 的年化收益,历史最大回撤为 -4.9%。

图 扩展窗口模式下 CCA 模型表现

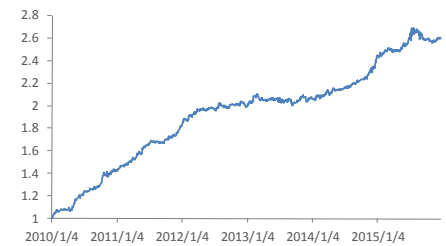


表 CCA 模型在 2010 年以来表现

年化收益率	17.2%
最大回撤率	-4.9%
胜率	58.5%
信息比	2.57

分析师: 张超 S0260514070002



020-87578291



zhangchao@gf.com.cn

相关研究:

另类交易策略系列之二十二: 2015-04-17
风格动量下的股指期货跨品种套利策略

另类交易策略系列之二十四: 2015-05-14
观日内趋势,察行业轮动,品种风格套利

联系人: 文巧钧



020-87578543



wenqiaojun@gf.com.cn

目录索引

一、交易策略概述.....	4
(一) 跨品种套利策略的再思考	4
(二) 策略的框架	4
二、交易模型介绍.....	5
(一) 给定组合的情况	5
(二) 未给定组合的情况	8
(三) 模型参数训练模式	9
(四) 因子筛选	10
三、实证分析.....	11
(一) 给定资产组合：单因子表现	11
(二) 给定资产组合：多因子打分模型	13
(三) 给定资产组合：多因子回归模型	14
(三) 未给定资产组合：CCA 模型	16
(四) 综合比较	19
四、总结与讨论.....	19

图表索引

图 1: 策略框架图	5
图 2: 因子打分模型示意图	6
图 3: 扩展窗口模式示意图	9
图 4: 涨跌幅因子表现	11
图 5: 最高价因子表现	12
图 6: 最低价因子表现	12
图 7: 资金强度因子表现	13
图 8: 多因子打分模型净值	14
图 9: 多因子回归模型净值	16
图 10: CCA 模型净值	17
图 11: 扩展窗口 CCA 模型的资金占比	18
图 12: 扩展窗口 CCA 模型的资金净多空头寸	18
表 1: 多因子组合打分	7
表 2: 因子的信息系数和信息比	10
表 3: 多因子打分模型不同年度的收益率	13
表 4: 多因子打分模型不同年度的最大回撤	14
表 5: 多因子打分模型性能指标	14
表 6: 多因子回归模型不同年度的表现	15
表 7: 多因子回归模型性能指标	15
表 8: CCA 模型不同年度的表现	17
表 9: CCA 模型性能指标	17
表 10: 不同模型性能指标综合比较	19

一、交易策略概述

（一）跨品种套利策略的再思考

随着上证 50 股指期货和中证 500 股指期货的推出，股指期货投资策略得到了很大的丰富。一方面，在不同品种期货上执行策略可以分散风险，有利于提高策略的多样化和资金容量；另一方面，不同品种股指期货的存在也使得指数的组合搭配变得方便可行。

此前，我们发布了《另类交易策略系列之二十二：风格动量下的股指期货跨品种套利策略》和《另类交易策略系列之二十四：观日内趋势，察行业轮动，品风格套利》两篇专题报告，提出了大小盘之间的跨品种套利策略，其中，上证 50-中证 500 指数之间的跨品种套利策略表现最好。该类策略通过观察上一个时间区间内上证 50 指数和中证 500 指数的走势强弱，对此后一个时间区间的上证 50 指数和中证 500 指数的走势相对强弱进行预测。具体来说，日频跨品种套利策略（另类交易策略系列之二十二）提出的 50-500 跨品种套利策略的做法是：

记上证 50 指数当日收盘价相对上一个交易日收盘价的涨幅为 x_{50} ，中证 500 指数当日收盘价相对上一个交易日收盘价的涨幅为 x_{500} ，则在每日现货收盘前进行判断：

（1）如果 $x_{50} > x_{500}$ ，则在收盘之后做多上证 50 股指期货（IH），用相等的资金做空中证 500 股指期货（IC），将多 IH 空 IC 的组合持有到下一个交易日收盘；

（2）如果 $x_{50} < x_{500}$ ，则在收盘之后做空 IH，用相等的资金做多 IC，将空 IH 多 IC 的组合持有到下一个交易日收盘。

在此基础上，有一些问题值得继续考虑和深入挖掘。

首先，上述策略中，通过指数收盘价当日涨跌幅的相对强弱来预测下一个交易日涨跌幅的相对强弱，除此之外，是否有其他“预测因子”可以进行这样的预测？如果有其他有效因子存在，不同的预测因子应该如何进行组合？

其次，50-500 风格套利组合是否是最优的一种指数组合配置？目前中金所推出了 IF、IH 和 IC 共 3 种股指期货，分别对应沪深 300 指数，上证 50 指数和中证 500 指数，这三种指数之间是否存在更加合理的搭配呢？

（二）策略的框架

为了解决上述问题，本篇专题报告首先提出了这样一个解决框架。

假设当前已经获得 $x_1, x_2, x_3, \dots, x_m$ 等有效的预测因子，包括价格和成交量等方面的因子。目的是预测上证 50 指数、沪深 300 指数和中证 500 指数等三个不同指数的组合的未来收益率。记 r_1, r_2, r_3 依次为上证 50 指数，沪深 300 指数和中证 500 指数的未来收益率， w_1, w_2, w_3 为投资三个不同指数的资金分配，则组合的收益率为 $y = \mathbf{w}\mathbf{r}^T = w_1r_1 + w_2r_2 + w_3r_3$ ，其中，收益率向量 $\mathbf{r} = [r_1, r_2, r_3]$ ，权重向量 $\mathbf{w} = [w_1, w_2, w_3]$ 。本文中，粗体表示向量或者矩阵，列向量 \mathbf{r}^T 表示行向量 \mathbf{r} 的转置。

权重向量 \mathbf{w} 中的元素大于 0，则表示做多对应的指数，小于 0 则表示做空相应的指数。例如， $\mathbf{w} = [0.5, 0, -0.5]$ ，即 $w_1 = 0.5, w_2 = 0, w_3 = -0.5$ ，表示用

50%的资金做多上证 50 指数，50%的资金做空中证 500 指数，组合中不对沪深 300 指数进行交易。这个组合就是 50-500 风格套利策略中做多上证 50 指数，做空中证 500 指数的组合。

权重向量 w 中元素的绝对值之和 $\sum |w_i| = |w_1| + |w_2| + |w_3|$ 表示对组合进行建仓的资金占用，也就是仓位。例如， $|w_1| + |w_2| + |w_3| = 1$ ，即表示建仓占用的资金量为 1（满仓）； $|w_1| + |w_2| + |w_3| = 0.8$ ，即表示建仓占用的资金量为 80%。在本报告中，我们一般令 $|w_1| + |w_2| + |w_3| = 1$ ，也就是在不加杠杆情况下满仓操作。

权重向量 w 中元素之和 $\sum w_i = w_1 + w_2 + w_3$ 表示组合中没有完全对冲的单边头寸占比。例如， $w_1 + w_2 + w_3 = 0$ ，表示多头和空头的资金占比完全相等，没有单边头寸；当 $w_1 + w_2 + w_3 \neq 0$ 时，有单边头寸。

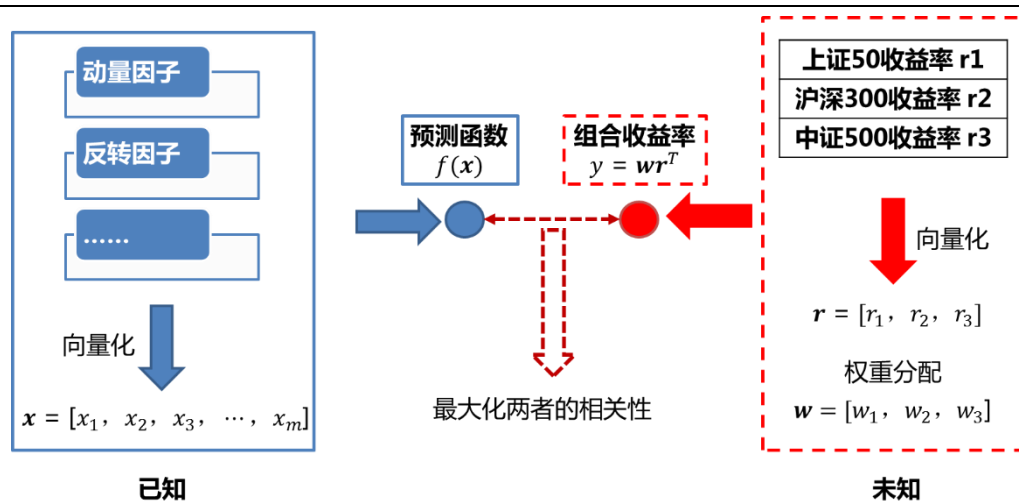
将预测因子向量化，记作 $x = [x_1, x_2, x_3, \dots, x_m]$ 。那么，现在的问题在于怎么样寻找合适的预测函数 $f(x)$ ，对组合收益率 $y = wr^T$ 进行预测。

通过多元统计方法，我们可以获得与 y 相关的预测函数 $f(x)$ 。

如果预测函数 $f(x)$ 和未来组合收益率 $y = wr^T$ 正相关，且 $f(x)$ 的均值为 0。则当 $f(x) > 0$ 时，应该做多该组合，预期收益率为 $PL = wr^T$ ；当 $f(x) < 0$ 时，应该做空该组合，预期收益率为 $PL = -wr^T$ 。实际交易中，收益率还应该扣掉交易成本。

策略框架如图 1 所示。

图1：策略框架图



数据来源：广发证券发展研究中心

二、交易模型介绍

（一）给定组合的情况

如何确定组合权重向量 $w = [w_1, w_2, w_3]$ 是该策略框架的核心问题之一。大体上有两种思路可以用来确定 w ：第一种是根据市场经验进行设定；第二种是通过统计模型进行计算。

首先，我们观察到 A 股市场存在显著的大小盘风格动量效应，而且上证 50 指数

和中证 500 指数的风格差异最大，因此，基于这样的经验知识，我们可以构建 50-500 的风格套利组合，即 $w = [0.5, 0, -0.5]$ 。做多该组合即用 50% 的资金做多上证 50 股指期货，另外 50% 的资金做空中证 500 股指期货；做空该组合即用 50% 的资金做空上证 50 股指期货，另外 50% 的资金做多中证 500 股指期货。

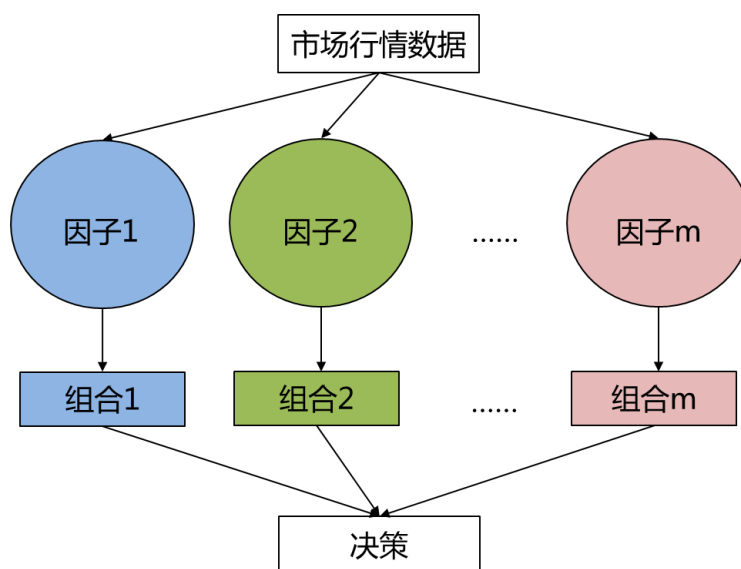
在这种情况下，不计交易成本时，做多组合的未来收益率 $y = w r^T = 0.5(r_1 - r_3)$ 。

假设我们当前一共有 m 个有效的因子， $x = [x_1, x_2, x_3, \dots, x_m]$ ，类似于股票多因子模型，我们有多种方式来根据多个因子确定最终的组合。

方法一：根据不同的因子分别进行打分，按照因子等权的方式来确定最终的组合。

这种策略相当于把资金等权划分，每一个因子对应一份资金，根据每个因子取值来对相应的资金进行建仓，然后综合起来。如图 2 所示。

图2：因子打分模型示意图



数据来源：广发证券发展研究中心

记 $\delta(x_i)$ 为根据第 i 个因子的打分 ($i = 1, 2, \dots, m$)， $\delta(x_i) = 1$ 表示根据因子 x_i 的打分为做多组合 $w = [0.5, 0, -0.5]$ ，即用一半的资金做多上证 50 指数，另外一半的资金做空中证 500 指数； $\delta(x_i) = -1$ 表示根据因子 x_i 的打分为做空组合 $w = [0.5, 0, -0.5]$ ，即用一半的资金做空上证 50 指数，另外一半的资金做多中证 500 指数。则最终的组合为不同因子的等权综合，即

$$w_{\text{多因子}} = \frac{[\delta(x_1)w + \delta(x_2)w + \dots + \delta(x_m)w]}{m} = \frac{\sum_{i=1}^m \delta(x_i)}{m} w \quad (1)$$

从资金占用 $\sum |w_i|$ 来看，单因子打分的资金占用为 100%，而多因子组合打分的资金占用为 $|\sum_{i=1}^m \delta(x_i) / m|$ ，可能小于 1。说明这种因子组合方式对资金仓位进行了调节。

当且仅当所有的因子预测结论一致时， $\sum_{i=1}^m \delta(x_i) / m = 1$ 或者 -1 ，此时从预测模型来看，预测因子结论一致，盈利的可能性比较大，应该满仓投资；而当有因子预测结论与其他因子不一致时，说明盈利的可能性比较低，此时应该降低资金仓位，不采取满仓投资的策略。

如表 1 所示，假如当前一共有 $m=4$ 个有效的因子，根据第 1 个因子， x_1 ，确定的组合为 $\delta(x_1)=1$ ，即做多 50 指数，做空 500 指数；根据第 2 个因子和第 3 个因子， x_2 和 x_3 ，也获得了同样的结论， $\delta(x_2)=\delta(x_3)=1$ ；但是，根据第 4 个因子获得的结论为 $\delta(x_4)=-1$ ，则综合得分为

$$w_{\text{多因子}} = \frac{1+1+1-1}{4} w = 0.5w = [0.25, 0, -0.25]$$

此时，资金占用 $\sum |w_i| = 0.5$ ，表示多因子组合打分的资金占用为 50%。因为 4 个因子预测的结果不完全一致，有 3 个因子认为应该做多组合，有 1 个因子认为应该做空组合，“看多”的因子比“看空”的因子净多了 2 个（“净多头”的因子数目占总预测因子数目的 50%），因此采取 50%的仓位进行投资。

表 1: 多因子组合打分

因子	打分	组合中 50 权重	组合中 300 权重	组合中 500 权重	资金仓位
x_1	1	0.5	0	-0.5	100%
x_2	1	0.5	0	-0.5	100%
x_3	1	0.5	0	-0.5	100%
x_4	-1	-0.5	0	0.5	100%
综合	0.5	0.25	0	-0.25	50%

数据来源：广发证券发展研究中心

方法二：对不同的因子分别赋予一定的权重，按照该权重对因子进行加权组合。

这里有两种不同的组合方式：1、根据不同的因子分别计算出预测组合之后，对组合进行加权，相当于在前文方法一的基础上，对因子打分的组合不是等权进行，而是加权组合；2、直接在因子预测阶段进行加权处理。本报告中仅讨论第 2 种组合方式，实际上可以通过多个预测因子的多元线性回归模型来实现。

假设当前有 m 个有效的预测因子 $x = [x_1, x_2, x_3, \dots, x_m]$ ，在上证 50 指数和中证 500 指数上，第 i 个预测因子都有因子取值，分别为 $x_{i,50}$ 和 $x_{i,500}$ ，则我们先进行因子标准化处理，令

$$x_{i,50-500} = x_{i,50} - x_{i,500}, i = 1, 2, \dots, m$$

即预处理之后的因子值为该因子在 50 指数和 500 指数的取值之差。标准化之后所有因子的向量表示可以一并记为 $x_{50-500} = [x_{1,50-500}, x_{2,50-500}, x_{3,50-500}, \dots, x_{m,50-500}]$ 。

在预测因子的基础上，构建预测函数 $f(x_{50-500})$ ，对组合收益率 $y = \mathbf{w}^T = 0.5(r_1 - r_3)$ 进行预测。令预测函数为

$$f(x_{50-500}) = \mathbf{v}^T x_{50-500} = \sum_{i=1}^m v_i x_{i,50-500} \quad (2)$$

其中， $\mathbf{v} = [v_1, v_2, \dots, v_m]$ 为权重系数，表示不同因子在预测模型中的权重。回归模型可写成

$$y = f(x_{50-500}) + e \quad (3)$$

其中， e 为噪声项，这是一个线性回归问题，权重向量 \mathbf{v} 可以通过最小二乘方法 (OLS) 解得。

(二) 未给定组合的情况

上一节中，给定资产组合 $w = [0.5, 0, -0.5]$ ，将组合限定在上证 50 指数和中证 500 指数上，而且限定为多空对冲的情况。这是因为我们根据“历史经验”获得了“A 股市场存在显著的大小盘风格动量效应”这一结论。组合中所采用的上证 50 指数和中证 500 指数之间的风格差异是最大的。

这里有两点问题需要指出。首先，如果 A 股市场风格动量效应不是那么显著，这种组合搭配是否能够成立？例如，在海外市场，我们能否构建类似的组合？其次，这样构建的组合是不是最优的一种组合方式呢？换一句话来说，如果不事先给定资产组合的形式，我们能否根据历史数据，获得“最优”的资产组合呢？

这里，我们同时考虑上证 50，沪深 300 和中证 500 这 3 种指数。将 m 个有效的因子 $x = [x_1, x_2, x_3, \dots, x_m]$ 进行因子标准化处理，获得预测因子向量

$x_{50, 300, 500} = [x_{1,50-mean}, x_{2,50-mean}, \dots, x_{m,50-mean}, x_{1,300-mean}, x_{2,300-mean}, \dots, x_{m,300-mean}, x_{1,500-mean}, x_{2,500-mean}, \dots, x_{m,500-mean}]$ ，一共有 $3m$ 个预测因子。其中， $x_{1,50-mean} = x_{1,50} - (x_{1,50} + x_{1,300} + x_{1,500})/3$ 表示因子值 $x_{1,50}$ 标准化之后的值；实际上，在这 $3m$ 个因子中，有 m 个因子是冗余因子，可以通过其他 $2m$ 个因子计算出来（例如， $x_{1,50-mean} + x_{1,300-mean} + x_{1,500-mean} = 0$ ），因此可以剔除出预测模型。因子的线性预测函数为

$$f(x) = vx_{50,300,500}^T \quad (4)$$

需要预测的组合收益率为 $y = wr^T$ ，与上一节不同的是，此时组合权重 w 是未知的。但是在以下条件下，我们可以获得“最优”的组合权重：

在“最优”的预测函数和资产组合下，预测函数值与组合收益率具有最大的相关性，即最优参数对 $\{w, v\}$ 为使得以下目标函数最大的参数

$$\max_{w,v} \text{corr}(f(x), y) = \max_{w,v} \text{corr}(vx_{50,300,500}^T, wr^T) \quad (5)$$

注意到，上一节的多元线性回归模型可以认为是在事先给定 w 的前提下，公式 (5) 的一种特殊情况。

在 $f(x)$ 线性的情况下，模型的参数 w 和 v 可以同时估计出来。这个可以通过典型相关分析 (Canonical Correlation Analysis, CCA) 获得。下边我们对求解过程进行简单推导。

简单起见，记 $s = x_{50, 300, 500}$ ，则我们的目标函数是最大化 vs^T 和 wr^T 的相关系数

$$\max_{w,v} \text{corr}(vs^T, wr^T) = \frac{v\Sigma_{12}w^T}{\sqrt{v\Sigma_{11}v^T}\sqrt{w\Sigma_{22}w^T}} \quad (6)$$

其中， $\Sigma_{12} = \text{cov}(s, r)$ ， $\Sigma_{11} = \text{cov}(s, s)$ ， $\Sigma_{22} = \text{cov}(r, r)$ 分别为向量 s 和 r 的自协方差矩阵和互协方差矩阵，且有

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

因此， Σ_{11} ， Σ_{12} ，和 Σ_{22} 为增广向量 $[s, r]$ 的协方差矩阵的矩阵分块。

考虑到 w 和 v 经过伸缩之后，相关系数不变，即设 $v_{new} = av$ ， $w_{new} = bw$ ，有 $\text{corr}(v_{new}s^T, w_{new}r^T) = \text{corr}(avs^T, bwr^T) = \text{corr}(vs^T, wr^T)$ ，其中 a 和 b 为任意

大于 0 的数（标量），因此可以将(6)式中的分母值进行固定，将优化问题转换为

$$\begin{aligned} \max_{\mathbf{w}, \mathbf{v}} \quad & \mathbf{v} \Sigma_{12} \mathbf{w}^T \\ \text{s.t.} \quad & \mathbf{v} \Sigma_{11} \mathbf{v}^T = 1, \mathbf{w} \Sigma_{22} \mathbf{w}^T = 1 \end{aligned} \quad (7)$$

对问题(7)，可以构建 Lagrangian 等式进行求解，

$$L = \mathbf{v} \Sigma_{12} \mathbf{w}^T + \frac{\lambda_1}{2} (\mathbf{v} \Sigma_{11} \mathbf{v}^T - 1) + \frac{\lambda_2}{2} (\mathbf{w} \Sigma_{22} \mathbf{w}^T - 1) \quad (8)$$

对向量 \mathbf{v}^T 和 \mathbf{w}^T 分别进行求导，可以获得

$$\begin{aligned} \Sigma_{12} \mathbf{w}^T &= \lambda_1 \Sigma_{11} \mathbf{v}^T \\ \Sigma_{21} \mathbf{v}^T &= \lambda_2 \Sigma_{22} \mathbf{w}^T \end{aligned} \quad (9)$$

因此有

$$\begin{aligned} \mathbf{v} \Sigma_{12} \mathbf{w}^T &= \lambda_1 \mathbf{v} \Sigma_{11} \mathbf{v}^T = \lambda_1 \\ \mathbf{w} \Sigma_{21} \mathbf{v}^T &= \lambda_2 \mathbf{w} \Sigma_{22} \mathbf{w}^T = \lambda_2 \end{aligned} \quad (10)$$

由于 $\mathbf{v} \Sigma_{12} \mathbf{w}^T = \mathbf{w} \Sigma_{21} \mathbf{v}^T$ ，所以最优解的情况下目标函数满足 $\lambda_1 = \lambda_2 = \lambda = \text{corr}(\mathbf{v} \mathbf{s}^T, \mathbf{w} \mathbf{r}^T)$ ，如果协方差矩阵 Σ_{22} 可逆，则由(9)式有

$$\mathbf{w}^T = \frac{1}{\lambda} \Sigma_{22}^{-1} \Sigma_{21} \mathbf{v}^T$$

且

$$\Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \mathbf{v}^T = \lambda^2 \Sigma_{11} \mathbf{v}^T$$

这是一个广义特征值问题，如果协方差矩阵 Σ_{11} 也可逆，则有

$$\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \mathbf{v}^T = \lambda^2 \mathbf{v}^T$$

这是以 λ^2 为特征值， \mathbf{v}^T 为特征向量的特征值问题。类似的，向量 \mathbf{w} 也可以获得。

因此，预测函数 $f(\mathbf{x}) = \mathbf{v} \mathbf{x}_{50,300,500}^T$ 和组合收益率 $y = \mathbf{w} \mathbf{r}^T$ 的参数可以通过 CCA 模型同时进行求解。

CCA 模型求得的资金权重向量 \mathbf{w} 不一定满足 $\sum |w_i| = 1$ ，但是将 \mathbf{w} 经过伸缩 $\mathbf{w}_{new} = b \mathbf{w}$ 之后，对 CCA 求解获得的向量 \mathbf{v} 和最优相关系数 λ 没有影响，因此在求得的资金权重向量 \mathbf{w} 的基础上，可以通过 $\mathbf{w}_{new} = \mathbf{w} / \sum |w_i|$ 获得实际的资金权重分配。

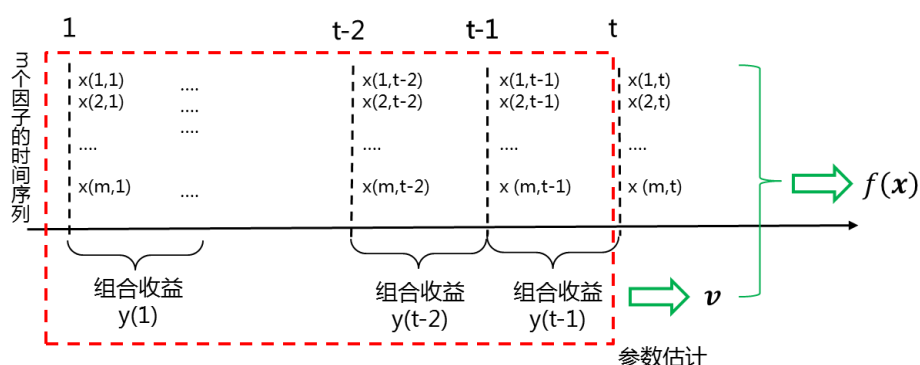
（三）模型参数训练模式

由前文可以看到，多因子回归模型和 CCA 模型中都涉及到参数优化的问题。需要通过一段样本内数据（历史数据）进行处理，训练出模型的参数，用于样本外的交易。

本篇报告考虑两种不同的参数训练模式。第一种为固定窗口的训练样本挑选，事先选取一段样本内的数据，训练好模型参数之后，所有样本外数据都用该参数来处理。在这一种模式下，我们只训练一次模型。

第二种是扩展窗口的训练样本挑选模式，在每一次预测前，都把所有的历史样本选出来，重新训练模型，将新训练出来的模型参数用于预测。在这种模式下，我们每次预测都需要重新训练模型。例如，为了做 t 时刻的预测，我们将时刻 1 以来直到时刻 $t - 1$ 的所有数据都拿过来，用于训练模型，如图 3 所示。

图3：扩展窗口模式示意图



数据来源：广发证券发展研究中心

（四）因子筛选

因子的挑选是多因子策略中非常重要的一步，在本报告的预测模型中，也是同样重要的环节。

由于本报告关注的是日间的交易，因此，我们在此主要从短期的价量数据中寻找有效的因子。借鉴多因子的思路，因子的有效性可以通过信息系数（IC）和信息比（IR）来衡量。其中，每一个时间段的IC值可以通过截面数据求得：

$$IC_t = \text{corr}(x_t, r_{t+1})$$

其中， x 是不同指数的因子取值， r 是不同指数的收盘价（相对前收盘价）涨跌幅。由于不同时间段的IC值不一致，这里，我们考察IC的平均值。IR的值为IC的平均值与标准差之比。

$$IR = \text{mean}(IC_t) / \text{std}(IC_t)$$

因子信息系数IC和因子信息比IR的绝对值越大，则认为该预测因子越有效。根据2007年至2009年的数据，我们挑选了以下4个有效因子，如表2所示。

表2：因子的信息系数和信息比

因子	IC 均值	信息系数 IR	方向
日收盘价涨跌幅	0.159	0.186	正向
当日最高价	-0.134	-0.161	负向
当日最低价	-0.136	-0.163	负向
资金强度	0.144	0.181	正向

数据来源：广发证券发展研究中心，天软科技

其中，日收盘价涨跌幅指的是股价今日收盘价相对于上一交易日收盘价的涨跌幅，由于涨跌停的限制，该因子取值范围为[0.9, 1.1]。由表2可以看到，该因子的信息比最高，为0.186。

当日最高价为股价在日内的最高价，为了使得不同指数的最高价具有可比性，本报告中通过除以股价的当日收盘价来进行标准化，即为：最高价因子=最高价/收盘价。该因子取值不小于1。

当日最低价为股价在日内的最低价，在这里通过除以股价的当日收盘价来进行标准化，即为：最低价因子=最低价/收盘价。该因子的取值不大于1。

资金强度为指数在日内的资金流入流出强度，根据高频数据计算出来，假如指数在第 T 分钟的成交金额为 $M(T)$ ，整个交易日的成交金额为 $Amount$ 。则令资金强度 MS 的初始值为 0，从当日开盘之后开始，在第 T 分钟

1、如果指数上涨，则 $MS := MS + M(T)$ ；

2、如果指数下跌，则 $MS := MS - M(T)$ 。

直到收盘时刻。最后，标准化之后的资金强度为：资金强度因子 = $MS/Amount$ 。该因子的取值范围为 $[-1, 1]$ 。

三、实证分析

数据选取：上文中因子的信息系数和信息比是通过考察 2007 年至 2009 年的数据获得的。下文选取 2010 年 1 月 1 日至 2015 年 12 月 25 日以来的数据进行回测。

交易模式：在每个交易日的收盘前建仓，持有至下一个交易日收盘。

交易费用：这里是模拟对股指期货进行交易，因此交易成本按照双边万分之二来计算。

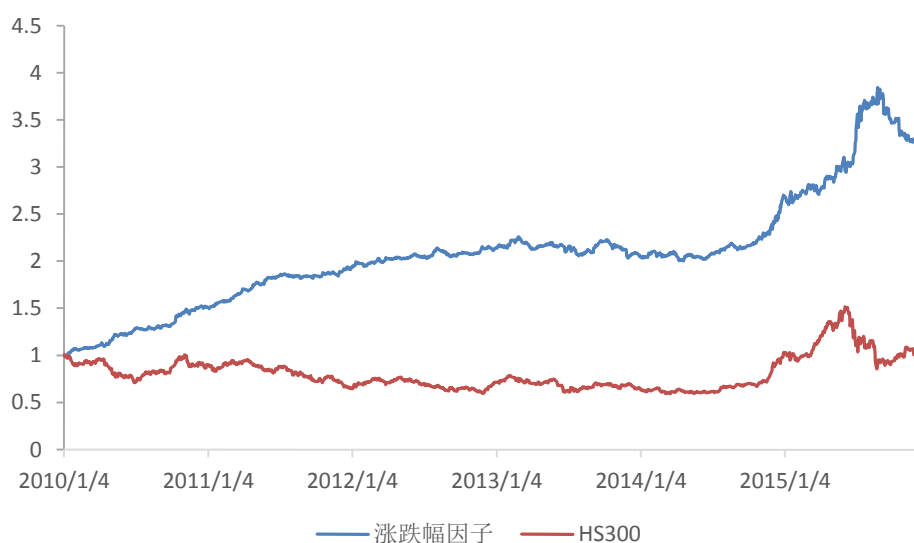
首先看一下单因子的表现。

（一）给定资产组合：单因子表现

我们首先看一下在给定组合为 50-500 风格套利组合时的单个预测因子表现。

涨跌幅因子用于 50-500 风格套利的表现如图 4 所示，从 2010 年 1 月以来，年化收益率 21.3%，最大回撤-17.5%，年化收益/回撤比为 1.22。策略在 2015 年 8 月份以来出现了比较大的回撤，说明这一段时间风格切换速度明显加快。

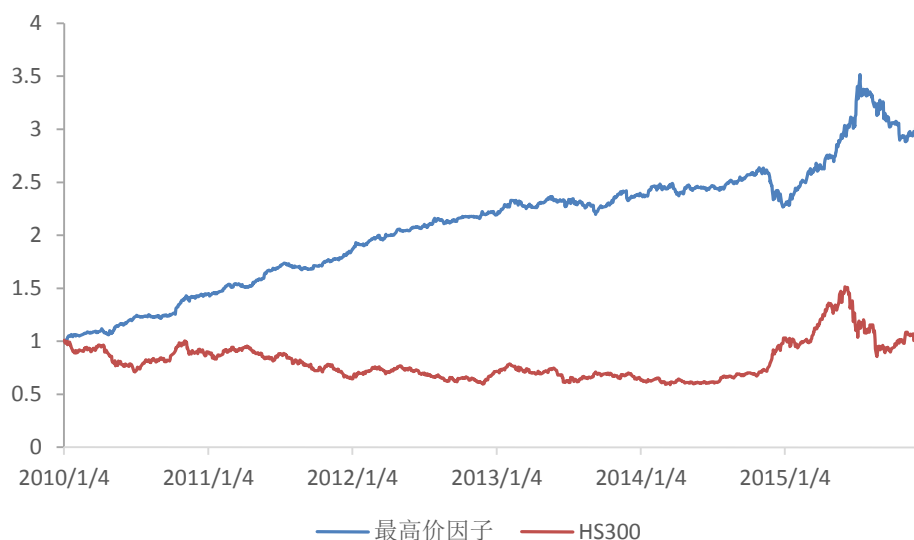
图4：涨跌幅因子表现



数据来源：广发证券发展研究中心，天软科技

最高价因子用于 50-500 风格套利的表现如图 5 所示，从 2010 年 1 月以来，年化收益率 20.6%，最大回撤-18.1%，年化收益/回撤比为 1.14。

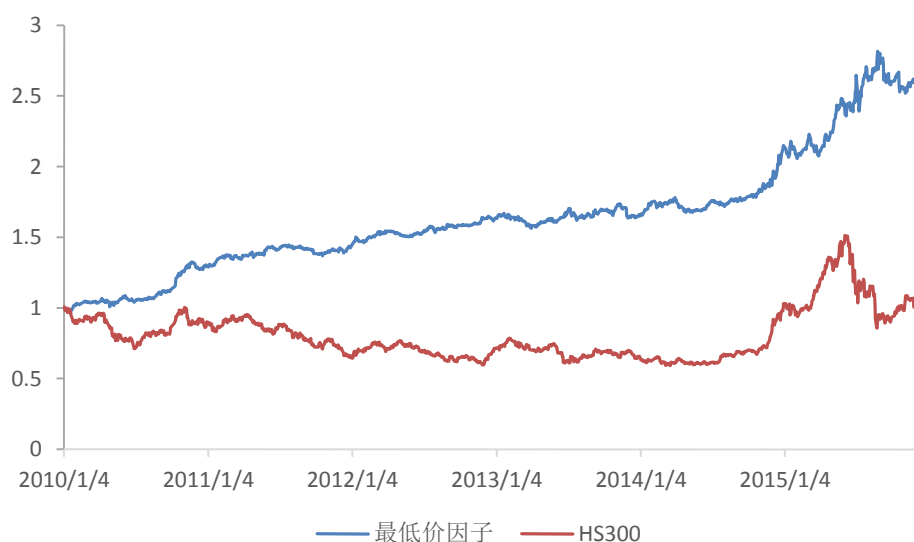
图5：最高价因子表现



数据来源：广发证券发展研究中心，天软科技

最低价因子用于 50-500 风格套利的表现如图 6 所示，从 2010 年 1 月以来，年化收益率 17.6%，最大回撤-10.6%，年化收益/回撤比为 1.66。

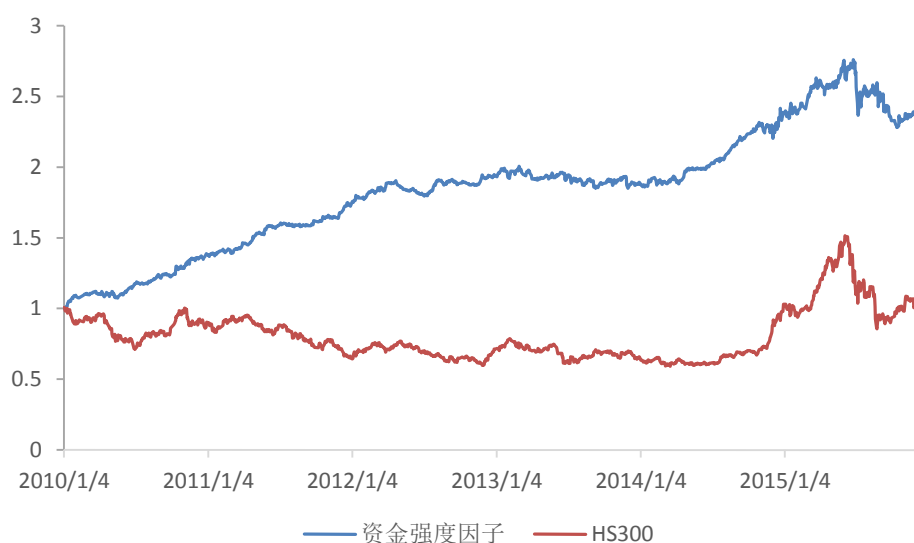
图6：最低价因子表现



数据来源：广发证券发展研究中心，天软科技

资金强度因子用于 50-500 风格套利的表现如图 7 所示，从 2010 年 1 月以来，年化收益率 15.5%，最大回撤-17.4%，年化收益/回撤比为 0.89。

图7：资金强度因子表现



数据来源：广发证券发展研究中心，天软科技

（二）给定资产组合：多因子打分模型

按照公式（1）的等权打分模型，可以获得多因子打分模型。这里一共涉及4个不同的预测因子，因此组合打分和仓位的计算方式类似于表1。当4个因子预测结论一致时，按照预测方向进行满仓交易，即 $\sum |w_i| = 1$ 。当因子预测方向不一致时，仓位会降低。

分年度的收益和回撤情况如表3和表4所示。从收益率来看，多因子打分模型获得了20.8%的年化收益率，与单因子收益第二高的最高价因子的年化收益率接近，略逊于涨跌幅因子21.3%的表现；而且每年都获得了正收益，其中2013年的表现比较差，只有2.2%的累计收益，其他年份的收益都超过了16%。从最大回撤来看，4个单因子模型的最大回撤都超过-10%，甚至接近-18%，而多因子模型的最大回撤为接近-10%，回撤明显小于不同的单因子模型。

由于多因子打分模型的最大回撤小于单因子模型，因此，策略的信息比和最大回撤明显优于单因子模型。一方面，可以归功于多策略的风险分散和仓位控制；另一方面，我们注意到，多因子打分模型的胜率比单因子模型至少高了1.4个百分点，说明多因子打分模型在预测性能上也有所提高。

表3：多因子打分模型不同年度的收益率

因子	涨跌幅因子	最高价因子	最低价因子	资金强度因子	多因子打分
2010	51.6%	44.1%	29.7%	38.5%	42.8%
2011	26.1%	27.3%	9.9%	24.5%	23.7%
2012	12.0%	19.5%	14.8%	12.3%	16.5%
2013	-4.3%	9.0%	0.7%	-2.9%	2.2%
2014	31.9%	-4.8%	30.4%	26.3%	22.1%
2015	17.7%	35.3%	23.4%	-0.1%	21.2%
年化收益率	21.3%	20.6%	17.6%	15.5%	20.8%

数据来源：广发证券发展研究中心，天软科技

表 4: 多因子打分模型不同年度的最大回撤

因子	涨跌幅因子	最高价因子	最低价因子	资金强度因子	多因子打分
2010	-3.6%	-5.1%	-5.7%	-4.3%	-3.2%
2011	-2.6%	-3.6%	-5.3%	-2.2%	-2.0%
2012	-4.4%	-2.3%	-2.9%	-5.7%	-2.1%
2013	-10.0%	-7.2%	-6.4%	-7.7%	-4.5%
2014	-5.0%	-14.0%	-5.8%	-4.9%	-3.2%
2015	-17.5%	-18.1%	-10.6%	-17.4%	-10.5%
历史最大回撤	-17.5%	-18.1%	-10.6%	-17.4%	-10.5%

数据来源: 广发证券发展研究中心, 天软科技

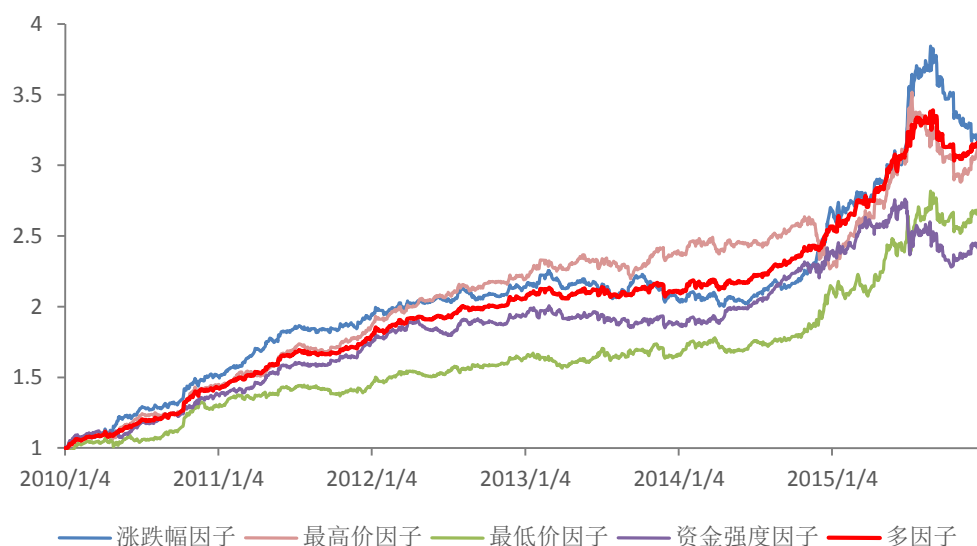
表 5: 多因子打分模型性能指标

因子	涨跌幅因子	最高价因子	最低价因子	资金强度因子	多因子打分
信息比	1.85	1.80	1.56	1.39	2.38
收益回撤比	1.22	1.14	1.66	0.89	1.98
胜率	55.1%	57.6%	56.0%	56.6%	59.0%

数据来源: 广发证券发展研究中心, 天软科技

不同单因子和多因子打分模型的净值表现如图 8 所示。可以看到, 多因子打分模型的收益曲线比较平滑而且回撤比较小。

图 8: 多因子打分模型净值



数据来源: 广发证券发展研究中心, 天软科技

(三) 给定资产组合: 多因子回归模型

在给定资产组合为 50-500 风格套利的情况下, 多因子回归模型的 4 因子预测函数为:

$$f(x) = v_1 \times (50 \text{ 涨跌幅} - 500 \text{ 涨跌幅}) + v_2 \times (50 \text{ 最高价} - 500 \text{ 最高价}) + v_3 \times (50 \text{ 最低价} - 500 \text{ 最低价}) + v_4 \times (50 \text{ 资金流} - 500 \text{ 资金流})$$

资产组合 $w = [0.5, 0, -0.5]$ 的收益率为:

$$y = 0.5 \times (50 \text{ 次日收益率} - 500 \text{ 次日收益率})$$

根据 $f(x)$ 对 y 进行预测。

需要确定的参数为因子的权重系数 $v = [v_1, v_2, v_3, v_4]$ 。

在固定窗口模式下,选取 2007 年至 2009 年的数据作为样本内数据,训练好模型参数后,2010 年以来,一直沿用该参数。

在扩展窗口模式下,每个交易日建仓前选取 2007 年至上一交易日的的数据作为样本内数据,训练模型参数。

不同年度的收益与回撤如表 6 所示,综合性能表现如表 7 所示。可以看到。两种参数训练样本窗口选取模式下的结果差不多,均获得了 23% 左右的年化收益率,但是在 2015 年 8 月份以来,仍然出现了比较大的回撤,最大回撤为 -17.9% 左右,信息比和收益回撤比均优于单因子模型(表 6)。

从图 9 的净值曲线上可以看到,两种训练窗口下的收益曲线基本一致。说明这种多因子回归策略模型受建模样本选取的影响比较小。

表 6: 多因子回归模型不同年度的表现

年份	累计收益率		最大回撤	
	固定窗口	扩展窗口	固定窗口	扩展窗口
2010	52.3%	59.6%	-3.6%	-3.6%
2011	29.9%	35.6%	-2.2%	-2.2%
2012	18.5%	17.4%	-2.2%	-2.2%
2013	13.3%	4.4%	-4.8%	-7.0%
2014	18.2%	26.8%	-4.4%	-4.4%
2015	6.9%	10.4%	-17.9%	-17.7%

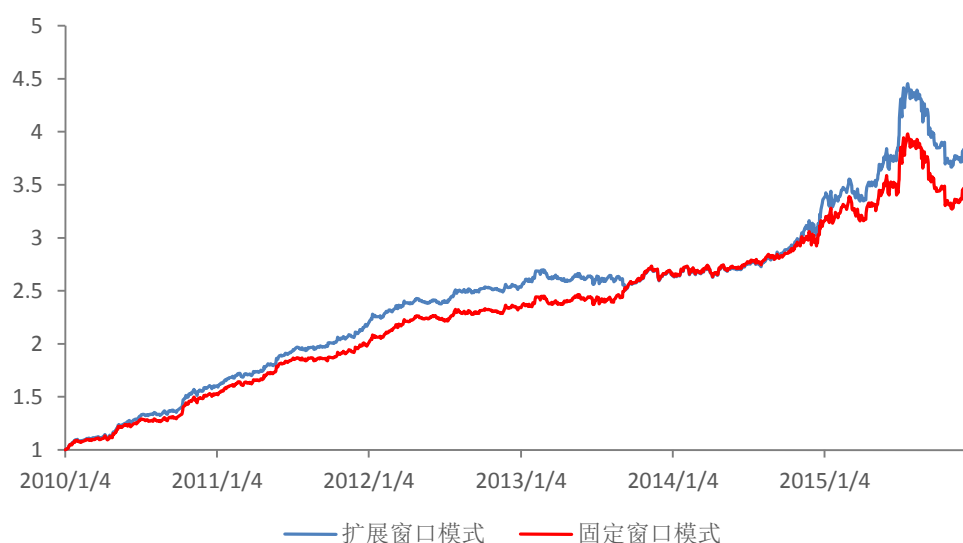
数据来源: 广发证券发展研究中心, 天软科技

表 7: 多因子回归模型性能指标

	年化收益率	最大回撤	信息比	收益回撤比	胜率
固定窗口模式	22.4%	-17.9%	1.93	1.25	57.4%
扩展窗口模式	24.4%	-17.7%	2.09	1.38	58.0%

数据来源: 广发证券发展研究中心, 天软科技

图9：多因子回归模型净值



数据来源：广发证券发展研究中心，天软科技

（三）未给定资产组合：CCA 模型

在未给定资产组合的情况下，我们不对资金组合的上证 50 指数、沪深 30 指数和中证 500 指数进行限制，即资金组合权重 w 未知。在对输入因素进行标准化之后，输入函数可以写成：

$$f(x) = v_1 \times (50 \text{ 涨跌幅} - \text{平均涨跌幅}) + v_2 \times (50 \text{ 最高价} - \text{平均最高价}) + v_3 \times (50 \text{ 最低价} - \text{平均最低价}) + v_4 \times (50 \text{ 资金流} - \text{平均资金流}) + v_5 \times (300 \text{ 涨跌幅} - \text{平均涨跌幅}) + v_6 \times (300 \text{ 最高价} - \text{平均最高价}) + v_7 \times (300 \text{ 最低价} - \text{平均最低价}) + v_8 \times (300 \text{ 资金流} - \text{平均资金流})$$

其中，平均涨跌幅 = (50 涨跌幅 + 300 涨跌幅 + 500 涨跌幅) / 3。因此 (500 涨跌幅 - 平均涨跌幅) 可以拆分成 (50 涨跌幅 - 平均涨跌幅) 和 (300 涨跌幅 - 平均涨跌幅) 的线性组合，为了使得协方差矩阵可逆，这一项可以省去。

需要确定的参数为因子的权重系数 $v = [v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8]$ 和资金组合权重 w 。参数可以由 CCA 模型解出。

类似的，CCA 模型的参数可以在固定窗口模式下进行求取，也可以在扩展窗口模式下进行求取。

不同年度的收益与回撤如表 8 所示，综合性能表现如表 9 所示。总体而言，固定窗口模式下，CCA 模型的年化收益率为 20.9%，最大回撤为 -16.2%；在扩展窗口下，该模型的年化收益率为 17.2%，最大回撤为 -4.9%。

可以看到，两种参数训练样本窗口选取模式下的结果在 2014 年之后有较大差异。

特别是在 2015 年，如果沿用此前的参数，该模型在 2015 年下半年有-16.2%的最大回撤，而如果参数不断更新，则最大回撤仅有-4.9%。

从图 10 的净值曲线上可以明显看到两种训练样本选取窗口下的收益曲线在 2014 年以来的差异。

表 8: CCA 模型不同年度的表现

年份	累计收益率		最大回撤	
	固定窗口	扩展窗口	固定窗口	扩展窗口
2010	39.9%	42.2%	-3.2%	-3.2%
2011	28.4%	27.7%	-1.5%	-1.4%
2012	14.3%	10.1%	-3.1%	-2.8%
2013	5.9%	3.1%	-4.3%	-4.8%
2014	27.1%	18.6%	-3.5%	-2.0%
2015	13.0%	6.2%	-16.2%	-4.9%

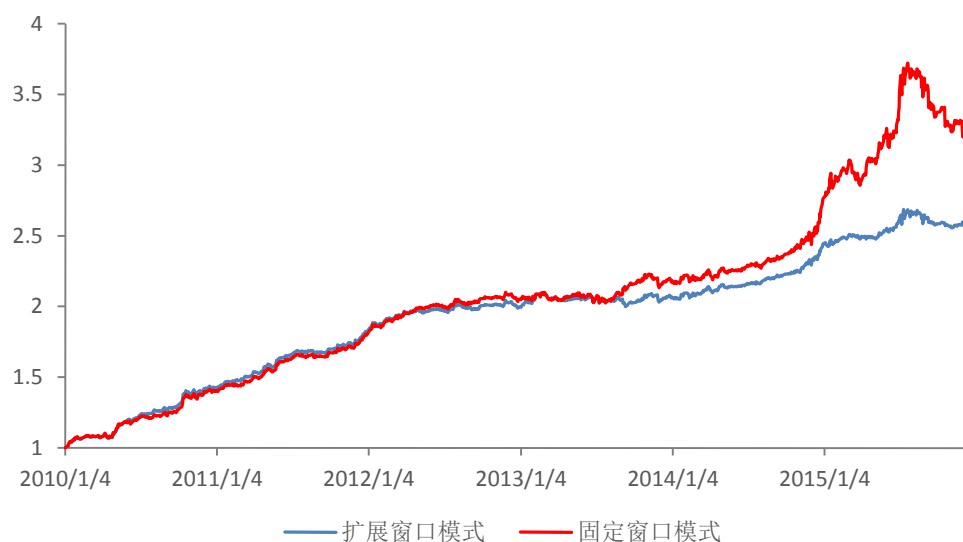
数据来源：广发证券发展研究中心，天软科技

表 9: CCA 模型性能指标

	年化收益率	最大回撤	信息比	收益回撤比	胜率
固定窗口模式	20.9%	-16.2%	2.15	1.29	58.0%
扩展窗口模式	17.2%	-4.9%	2.57	3.53	58.5%

数据来源：广发证券发展研究中心，天软科技

图10: CCA模型净值



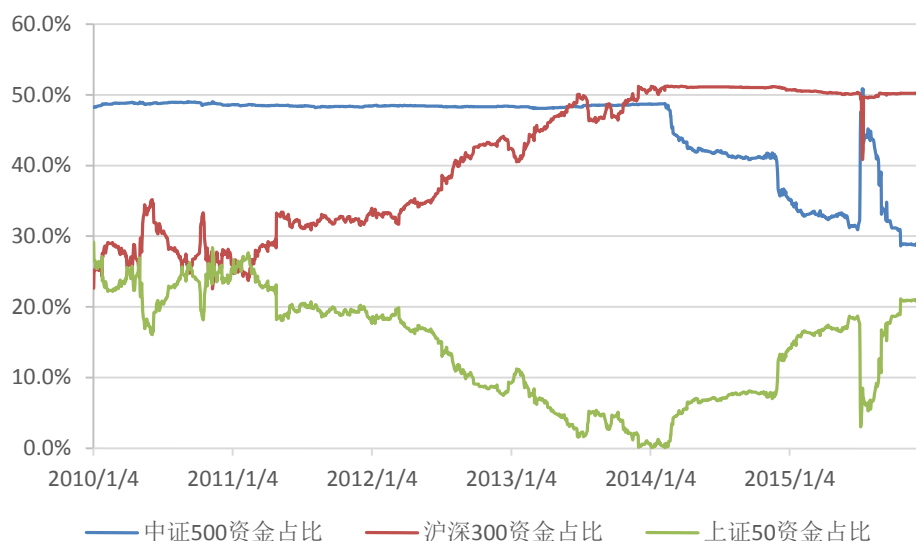
数据来源：广发证券发展研究中心，天软科技

图11表示扩展窗口模式下，每日资金分配中上证50指数、沪深300指数和中证500指数的权重占比。图12表示扩展窗口模式下，每日持仓的资金净多空头寸，即 $w_1 + w_2 + w_3$ 。由图12可见，每日的资金净头寸基本上在0附近，说明CCA模型选出的组合是一个风格套利的组合。需要指出的是，固定窗口模式下的每日资金权重就是扩展

窗口模式下在2010年第一个交易日的资金分配权重。

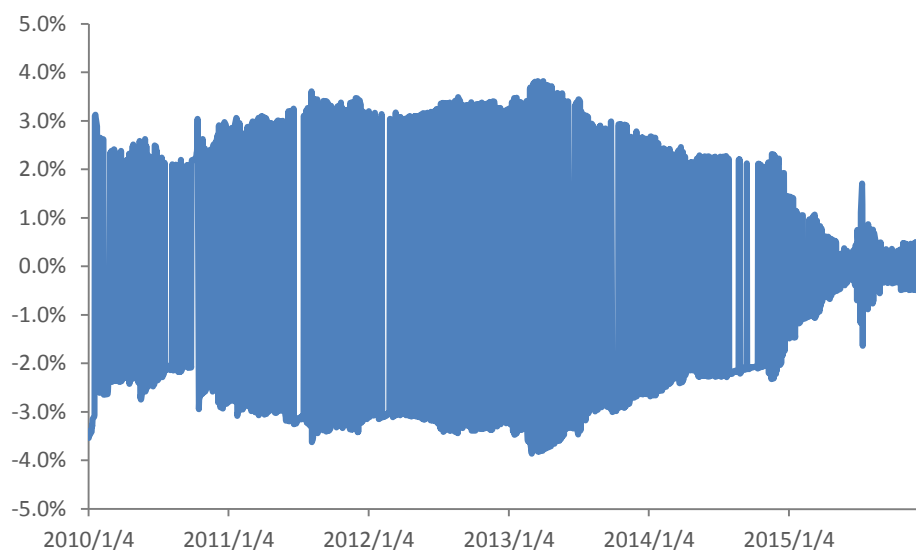
从资金占比上来看，固定窗口模式在2014年之前大约一半的资金用于持仓中证500指数，另外一半的资金用于持仓上证50和沪深300指数，且头寸方向相反。由图11和图12所示，在2014年之前，扩展窗口模式下获得的模型中与固定窗口模式的结论一致，这表示这段时间大小盘的风格动量显著；但是在2014年之后，中证500指数的资金占用明显偏离50%，这表示日频的大小盘风格动量现象有所减弱。

图11：扩展窗口CCA模型的资金占比



数据来源：广发证券发展研究中心，天软科技

图12：扩展窗口CCA模型的资金净多空头寸



数据来源：广发证券发展研究中心，天软科技

（四）综合比较

以上几种不同模型性能的综合比较如表 10 所示，与表 5 相比，可以看到，不同的几种多预测因子综合考虑的模型性能都要优于单因子模型（信息比和收益回撤比更高）。其中信息比和收益回撤比最高的两个模型是扩展窗口建模的 CCA 模型和多因子打分的 50-500 风格套利模型。扩展窗口建模的 CCA 模型放松了大小盘风格套利的约束，因此在风格切换频繁的 2015 年下半年依然能够控制风险；多因子打分模型由于考虑到在不同因子预测结论不一致时降低资金仓位，也取得了比较好的收益和较小的回撤。

表10：不同模型性能指标综合比较

	50-500 线性回归模型		CCA 模型		50-500 多因子打分模型
	固定窗口建模	扩展窗口建模	固定窗口建模	扩展窗口建模	
年化收益率	22.4%	24.4%	20.9%	17.2%	20.8%
最大回撤	-17.9%	-17.7%	-16.2%	-4.9%	-10.5%
信息比	1.93	2.09	2.15	2.57	2.38
年化收益回撤比	1.25	1.38	1.29	3.53	1.98
胜率	57.4%	58.0%	58.0%	58.5%	59.0%

数据来源：广发证券发展研究中心，天软科技

四、总结与讨论

本篇报告对不同品种股指期货的组合进行了比较深入的研究。首先，本篇报告提出了一个多个预测因子的研究框架，用不同的预测因子来建立预测模型，从而建立合适的股指期货跨品种组合。

从预测因子的选取上来看，可以通过类似多因子的思路，通过因子信息系数（IC）和因子信息比（IR）等指标来衡量因子的有效性。

对于股指期货跨品种组合，有两种研究思路。首先是基于 A 股市场的大小盘风格轮动来进行考虑，我们挑选了风格差异最大的上证 50 指数和中证 500 指数，通过多空对冲的方式进行获利。在多个预测因子组合的问题上，可以按照多因子打分的形式和回归模型的形式分别建立起 50-500 的预测模型。

第二种研究思路是基于市场数据来进行的，在这种思路下，我们不加大小盘风格套利的约束，通过预测因子来选取最合适的品种组合。在这里，预测函数和品种组合通过解决多元统计分析中的最优化问题来获得。

从实证结果来看，本文提出的几个模型的表现都好于风格动量策略等单个预测因子的模型。即使是在 2015 年 8 月份之后，市场的大小盘风格切换频繁，本文提出的几个模型都获得了比较好的表现。其中信息比和收益回撤比最高的两个模型是扩展窗口建模的 CCA 模型和多因子打分的 50-500 风格套利模型。扩展窗口建模的 CCA 模型放松了对大小盘风格套利的约束，因此在风格切换频繁的 2015 年下半年依然能够控制风险；多因子打分模型在盈利把握不大的时候降低了资金仓位，也取得了比较好的收益和较小的回撤。

在后续的研究中，一方面可以加入新的预测因子进行测算，另一方面也可以考虑其他不同周期的风格轮动效应和相应的预测因子选取。

风险提示

策略模型并非百分百有效，市场结构及交易行为的改变以及类似交易参与者的增多有可能使得策略失效。

广发证券—行业投资评级说明

- 买入： 预期未来 12 个月内，股价表现强于大盘 10%以上。
- 持有： 预期未来 12 个月内，股价相对大盘的变动幅度介于-10%~+10%。
- 卖出： 预期未来 12 个月内，股价表现弱于大盘 10%以上。

广发证券—公司投资评级说明

- 买入： 预期未来 12 个月内，股价表现强于大盘 15%以上。
- 谨慎增持： 预期未来 12 个月内，股价表现强于大盘 5%-15%。
- 持有： 预期未来 12 个月内，股价相对大盘的变动幅度介于-5%~+5%。
- 卖出： 预期未来 12 个月内，股价表现弱于大盘 5%以上。

联系我们

	广州市	深圳市	北京市	上海市
地址	广州市天河北路 183 号 大都会广场 5 楼	深圳市福田区金田路 4018 号安联大厦 15 楼 A 座 03-04	北京市西城区月坛北街 2 号 月坛大厦 18 层	上海市浦东新区富城路 99 号 震旦大厦 18 楼
邮政编码	510075	518026	100045	200120
客服邮箱	gfyf@gf.com.cn			
服务热线	020-87555888-8612			

免责声明

广发证券股份有限公司具备证券投资咨询业务资格。本报告只发送给广发证券重点客户，不对外公开发布。

本报告所载资料的来源及观点的出处皆被广发证券股份有限公司认为可靠，但广发证券不对其准确性或完整性做出任何保证。报告内容仅供参考，报告中的信息或所表达观点不构成所涉证券买卖的出价或询价。广发证券不对因使用本报告的内容而引致的损失承担任何责任，除非法律法规有明确规定。客户不应以本报告取代其独立判断或仅根据本报告做出决策。

广发证券可发出其它与本报告所载信息不一致及有不同结论的报告。本报告反映研究人员的不同观点、见解及分析方法，并不代表广发证券或其附属机构的立场。报告所载资料、意见及推测仅反映研究人员于发出本报告当日的判断，可随时更改且不予通告。

本报告旨在发送给广发证券的特定客户及其它专业人士。未经广发证券事先书面许可，任何机构或个人不得以任何形式翻版、复制、刊登、转载和引用，否则由此造成的一切不良后果及法律责任由私自翻版、复制、刊登、转载和引用者承担。

