

证券研究报告—深度报告

金融工程

数量化投资

金融工程专题研究

2014年06月30日

专题报告

相关研究报告:

《国信量化 1 号——行业分析》——
2014-06-16
《检验量化基本面选股模型有效性的一类方法》——2014-06-11
《国信行业投资时钟系列: 基于长期趋势投资的行业成交量模型》——2014-04-04
《衍生品应用与产品设计系列之 VIX 介绍及 GSVX 编制》——2014-03-31
《期权投资策略系列之无风险套利策略》——
2014-03-31

证券分析师: 钱晶

电话: 021-60875163

E-MAIL: qianjing1@guosen.com.cn

证券投资咨询执业资格证书编码: S0980513070003

证券分析师: 林晓明

电话: 021-60875168

E-MAIL: linxiaom@guosen.com.cn

证券投资咨询执业资格证书编码: S0980512020001

基于协整方法与因子模型的配对交易策略

● 配对交易是统计套利的一种, 特点是不暴露市场方向风险

配对交易是利用股票价格之间存在的长期稳定的关系, 通过均值回归的特性去获取盈利的一种股票市场中性策略, 该策略的主要优势在于:

1. 通过同时在股票的多头和空头上下注, 规避掉市场及行业风险
2. 策略盈利主要来源于价差的均值回复, 与市场方向无关
3. 通过对大量的配对同时下注, 可以有效的控制风险和回撤

即使是在次贷危机, 股票市场中性策略也很好的控制住了回撤 (-6.27%), 当前采用该类策略的对冲基金代表有 CFM、AQR 等。

● 多种方法改进基于协整方法的配对策略, 效果明显

实证中我们发现, 交易成本会侵蚀掉交易的大部分利润, 这主要是由于持有头寸时间较长, 致使融券费用较高造成的。通过设置合适的止盈、止损点, 选择合适的最大持有期限等方法, 可以对策略进行改进。

1. 止盈点: 设置止盈点可以避免出现价差收窄至 0 附近后, 掉头又继续扩大的情况。另外, 通过这种方法也可以产生更多的交易机会, 并且可以显著的缩短持有时间, 降低融资成本。
2. 最大持有市场: 限定头寸的最长持有时间可以直接减少融券利息, 同时, 还能规避价差长期不回归甚至扩大的情况, 起到一定的止损作用。当然, 如果最长持有时间设置过短, 也会导致很多交易会未等到价差回归就被强行平仓, 使得单笔交易的平均利润下降。
3. 止损点: 止损点的设置, 并非为了提高收益, 而是控制风险, 因此有必要根据不同止损点的损益分布情况来确定合适的止损点。此外, 当配对数量较多, 单个头寸权重较小时, 可以放宽止损条件; 反之, 则应该设置较为严格的止损
4. 行业适用性: 实证表明, 2013 年, 配对交易在采掘、医药和银行等行业表现较好。

● 基于因子模型筛选配对, 可与协整方法互为补充

对股票建立因子模型, 根据共同趋势模型, 当股票的因子收益强相关, 且股票特殊收益的积分序列是平稳序列时, 即满足协整关系。该方法效率较高, 可对所有股票同时进行检验, 所筛选的配对也不限于同一行业。

但是因子模型筛选配对也有自身劣势。实证发现, 因子模型的配对数量要大大少于传统协整检验的方法, 且收益率也要略低于传统协整检验方法所筛选出的配对。因此我们认为在实践中, 可以同时参考这两种方法, 互为补充。

独立性声明:

作者保证报告所采用的数据均来自合规渠道, 分析逻辑基于本人的职业理解, 通过合理判断并得出结论, 力求客观、公正, 结论不受任何第三方的授意、影响, 特此声明。

内容目录

配对交易介绍.....	4
策略原理	4
策略表现	4
基于协整检验的配对交易策略	5
股票对筛选流程	5
开仓点	5
止盈点	6
最长持有时间	6
止损点	6
行业适用性	8
基于因子模型的配对交易策略	8
共同趋势模型	8
共同趋势模型与套利定价理论的联系	8
主成分分析法建立统计因子模型	9
利用因子模型确定协整关系	12
实证结果	13
国信证券投资评级.....	14
分析师承诺.....	14
风险提示.....	14
证券投资咨询业务的说明	14

图表目录

图 1: Barclay 股票市场中性策略指数 (1997 年 1 月~2014 年 3 月)	4
图 2: 不设止损的收益分布情况	7
图 3: 止损 $\Sigma = 7$ 的收益分布情况	7
图 4: 主成分分析法前 15 个因子对方差的解释度 (2013-1-1, 前 240 个交易日)	9
图 5: 第一特征组合 Vs 沪深 300 (2013-1-1~2013-12-31)	10

表 1: 不同开仓点测试结果 (2013-1-1~2013-12-31)	5
表 2: 不同止盈点测试结果 (2013-1-1~2013-12-31)	6
表 3: 不同最长持有时间测试结果 (2013-1-1~2013-12-31)	6
表 4: 不同止损点测试结果 (2013-1-1~2013-12-31)	7
表 5: 不同行业中的配对表现	8
表 6: 第二特征组合多头前 10 和空头前 10 行业分布情况	10
表 7: 第三特征组合多头前 10 和空头前 10 行业分布情况	11
表 8: 第四特征组合多头前 10 和空头前 10 行业分布情况	11
表 9: 第十二特征组合多头前 10 和空头前 10 行业分布情况	11
表 10: 协整方法筛选配对 VS 因子模型筛选配对	13

配对交易介绍

策略原理

配对交易是海外对冲基金最早使用的一种统计套利策略。其基本思想是：如果股票 P 和股票 Q 所代表的公司有着及其相似的属性，那么它们的股票价格之间应该有长期的稳定关系。如果用 P_t 和 Q_t 分别代表相应的股票价格序列，则可表示为：

$$\ln\left(\frac{P_t}{P_0}\right) = \alpha t + \beta \ln\left(\frac{Q_t}{Q_0}\right) + X_t$$

其中， X_t 是一个平稳序列，具有均值回复的特性。 α 是一个小量，在 t 较小的情况下， αt 可以忽略不计。这意味着，按照 β 构建股票 P 和 Q 的多空组合，该组合会围绕着一均衡值上下波动。

依据这一特性，可以构建这样一种策略：当 X_t 较小时，做多 1 元钱的 P，做空 β 元钱的 Q；反之，当 X_t 较大时，做空 1 元钱的 P，做多 β 元钱的 Q。如果均值回复发生，便可以从中获利。

该策略的主要优势在于：

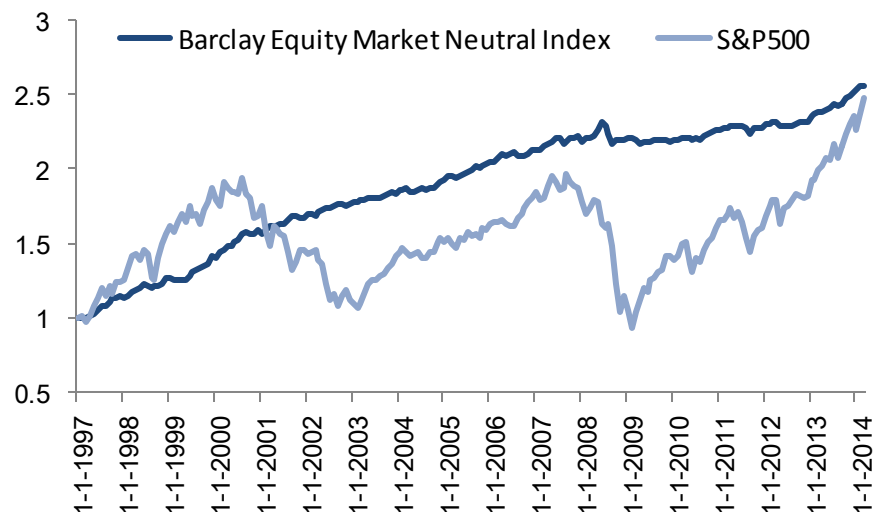
1. 通过同时在股票的多头和空头上下注，规避掉市场及行业风险
2. 策略盈利主要来源于价差的均值回复，与市场方向无关
3. 通过对大量的配对同时下注，可以有效的控制风险和回撤

由于这些优势，基于统计套利的股票多空策略在海外对冲基金中也占有一席之地，比较著名的有 CFM、AQR 等。

策略表现

虽然没有专门的配对交易策略指数，但是从 Barclay 编制的股票市场中性策略指数中我们可以一窥该类策略的风险收益特征。从图中可以发现，即使是在次贷危机，股票市场中性策略也很好的控制住了回撤（-6.27%）。这也是其不对市场暴露方向性风险的优势所在。

图 1：Barclay 股票市场中性策略指数（1997 年 1 月~2014 年 3 月）



资料来源：wind；国信证券经济研究所整理

基于协整检验的配对交易策略

配对交易的核心在于选取股票对，我们首先介绍最常用的筛选方法——协整检验。

股票对筛选流程

1. 对同一行业内的股票进行两两之间的协整检验。之所以在同行业内选，主要是因为这样选出来的股票之间的协整关系有一定的逻辑基础，其协整关系更可靠，另一方面也大大减少了计算量。
2. 计算收益率相关系数，剔除掉小于阈值的配对。实证中我们发现有些股票配对是负相关关系，同时也满足协整关系。尽管我们也有可能通过他们的协整关系去获利，但是这类交易往往需要从基本面去做更深入的分析，因此我们将这类配对剔除。
3. 对价差进行标准化处理，计算零穿越次数和零穿越所用时间的方差和均值，剔除达不到阈值的配对。在进行配对交易时，我们希望价差序列具有很强的均值回复特性，我们通过零穿越（zero-crossing）去量化均值回复的程度。一般而言，穿越的次数越多，穿越间隔的均值越小，方差越小，越符合配对交易的要求。

开仓点

以沪深 300 指数的成分股为研究对象，从 2013 年 1 月 1 日起，每个月月初按照股票对筛选流程选出股票对。对当月内发出建仓信号的股票对进行交易，当价差回归至平仓条件时平仓，每笔交易的最大持有期为 40 个交易日，若 40 个交易日价差仍未回复则强行平仓。

表 1：不同开仓点测试结果（2013-1-1~2013-12-31）

开仓 sigma	收益率均值 (未扣成本)	收益率均值 (扣除成本)	交易次数	胜率	持有时间均值
4.0	2.04%	0.59%	180	57.22%	34.97
3.5	2.26%	0.82%	228	61.40%	34.62
3.0	1.98%	0.54%	307	63.84%	34.83
2.5	1.56%	0.16%	422	60.66%	33.45
2.0	1.63%	0.34%	571	60.77%	30.62
1.5	1.64%	0.42%	774	61.31%	28.49
1.0	1.23%	0.13%	1099	59.70%	25.22

资料来源：wind；国信证券经济研究所整理

我们首先测试了在不同开仓 Sigma 下（止盈 Sigma = 0，不设止损），该策略每笔交易 PnL 的情况，可以发现：

1. Sigma 越大，交易次数越少，单笔交易的收益均值越高
2. Sigma 越大，回归需要的时间越长，因此单笔交易的持续时间均值越长
3. Sigma 越大，胜率不一定越高，主要原因是越高的价差越有可能是公司基本面出现了变化所导致的，这种情况下价差很有可能会继续扩大，而我们在该测试中并未设止损，因此 Sigma 越大，胜率不一定高。
4. 权衡收益率均值（扣除成本）和交易次数，Sigma = 3 或 3.5 作为开仓点较为合适
5. 交易成本会侵蚀掉交易的大部分利润，这主要是由于持有头寸时间较长，致使融券费用较高造成的（相当一部分交易价差未回归而持有到期）。当然，我们可以通过设置合适的止盈、止损点，选择合适的最大持有期限等方法对策略进行改进，我们将一一进行说明。

止盈点

在前文的策略中,我们是等到标准化后的价差回归至 0 时进行平仓,在实证中,我们发现有很多交易的价差虽然也发生均值回归,但在距离 0 一定距离处长期停滞,甚至出现掉头又继续扩大的情况。

为了避免这种情况,我们采用到达止盈点提前平仓的方式。另外,通过这种方法也可以产生更多的交易机会,并且可以显著的缩短持有时间,降低融资成本。

实证表明,当开仓 $\text{Sigma} = 3$ 时,平仓 Sigma 设为 0.5 至 2.5 都能带来更高的扣除成本后的平均收益,其中平仓 Sigma 为 1.5 或者 2.0 时,表现最好。

表 2: 不同止盈点测试结果 (2013-1-1~2013-12-31)

平仓 sigma	收益率均值 (未扣成本)	收益率均值 (扣除成本)	交易次数	胜率	持有时间均值
0.0	1.98%	0.53%	307	63.84%	34.83
0.5	2.03%	0.65%	307	64.50%	32.96
1.0	2.09%	0.83%	315	66.03%	29.66
1.5	2.32%	1.19%	319	71.47%	25.99
2.0	2.14%	1.18%	330	75.76%	21.18
2.5	1.73%	0.95%	365	80.27%	16.19

资料来源: wind; 国信证券经济研究所整理

最长持有时间

由于融券成本较高,要减少融券利息,最直接的方法就是限定头寸的最长持有时间。同时,设定合适的最长持有时间,还能规避价差长期不回归甚至扩大的情况,起到一定的止损作用。

当然,如果最长持有时间设置过短,也有其劣势。很多交易会未等到价差回归就被强行平仓,导致单笔交易的平均利润下降。

实证表明,当开仓 $\text{Sigma} = 3$ 、止盈 $\text{Sigma} = 1.5$ 时,最大持有时间设为 40 个交易日时,扣除成本后的单笔交易的收益率均值最高。

表 3: 不同最长持有时间测试结果 (2013-1-1~2013-12-31)

最大持有时间	收益率均值 (未扣成本)	收益率均值 (扣除成本)	交易次数	胜率	持有时间均值
5	1.05%	0.68%	305	59.34%	4.82
10	1.39%	0.87%	314	64.01%	8.98
20	1.47%	0.70%	319	62.38%	15.91
30	1.82%	0.85%	314	64.33%	21.45
40	2.32%	1.19%	319	71.47%	25.99
50	2.38%	1.13%	321	69.78%	29.41
60	1.97%	0.60%	297	69.70%	32.68

资料来源: wind; 国信证券经济研究所整理

止损点

设置止损的目的主要是为了避免由于公司基本面出现变化而导致的价差持续扩大的情况。从实证结果来看,设置止损点并不能提高单笔收益率均值,并且止损点设置的越低,单笔收益率均值越低,主要原因是很多原本可以赚钱的交易,由于触发了止损而认赔出局。

我们设定开仓 $\text{Sigma} = 3$ 、止盈 $\text{Sigma} = 1.5$,最大持有时间为 40 个交易日,对不同的止损 Sigma 进行实证测试,结果如下表所示:

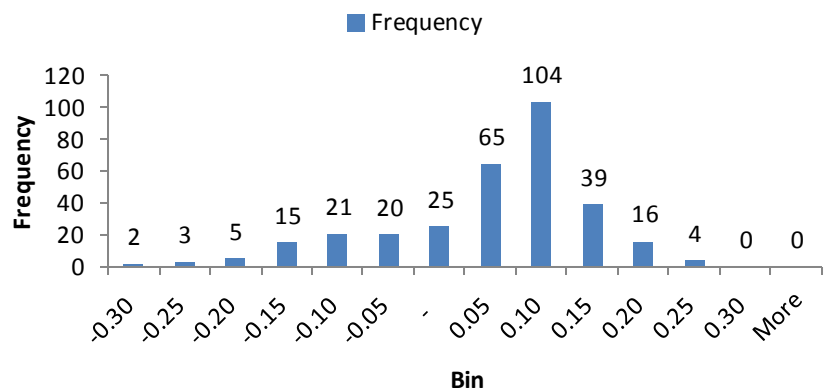
表 4: 不同止损点测试结果 (2013-1-1~2013-12-31)

止损 sigma	收益率均值 (未扣成本)	收益率均值 (扣除成本)	交易次数	胜率	持有时间均值
不设止损	2.32%	1.19%	319	71.47%	25.99
8.0	2.22%	1.15%	318	70.75%	24.24
7.5	1.90%	0.87%	318	69.50%	23.10
7.0	1.97%	0.97%	318	69.18%	22.21
6.5	2.02%	1.05%	317	68.14%	21.43
6.0	1.19%	0.30%	316	62.97%	19.13
5.5	0.98%	0.17%	315	58.41%	17.15
5.0	0.86%	0.13%	313	80.27%	14.89

资料来源: wind; 国信证券经济研究所整理

由于止损点的设置, 并非为了提高收益, 而是控制风险, 因此我们有必要根据不同止损点条件下, 策略每笔交易的收益率分布情况来确定合适的止损点。

图 2: 不设止损的收益分布情况

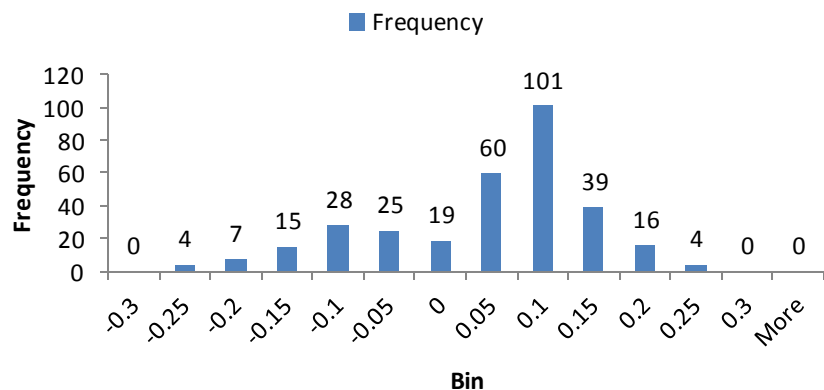


资料来源: wind; 国信证券经济研究所整理

例如, 比较不设止损和止损 $\text{Sigma} = 7$ 两种情况下, 每笔交易收益率的分布图。可以发现, 设置止损能够更好的控制单笔交易可能出现的最大亏损。

此外, 止损点的设置与单个配对头寸的权重也有关。当配对数量较多, 单个头寸权重较小时, 可以放宽止损条件; 反之, 则应该设置较为严格的止损。

图 3: 止损 $\text{Sigma} = 7$ 的收益分布情况



资料来源: wind; 国信证券经济研究所整理

行业适用性

配对交易在不同行业所出现的交易机会次数、胜率和收益率也有很大区别。2013 年，采掘、有色、银行、医药生物和机械设备交易机会较多，其中采掘、医药和银行的收益率情况表现较好。而农林牧渔、建材、传媒、商贸、食品饮料、计算机等，虽然收益率不错，但是交易次数较少，并不具有很强的代表性。

表 5：不同行业中的配对表现（2013-1-1~2013-12-31）

行业	收益率均值	收益率	交易次数	胜率	持有时间均值
采掘	4.88%	3.95%	86	79.07%	20.36
有色金属	1.20%	0.00%	49	67.35%	28.08
银行	2.21%	0.98%	33	69.70%	28.82
医药生物	4.52%	3.39%	24	79.17%	25.88
机械设备	-2.01%	-3.20%	21	57.14%	27.76
非银金融	-0.63%	-2.01%	16	68.75%	32.81
化工	2.56%	1.29%	13	69.23%	29.77
国防军工	3.51%	2.14%	12	66.67%	32.58
汽车	-1.84%	-3.02%	11	63.64%	27.27
房地产	-1.21%	-2.59%	9	55.56%	33.00
钢铁	2.39%	1.33%	8	75.00%	24.13
电气设备	-0.33%	-1.37%	7	57.14%	23.29
综合	-2.44%	-3.70%	5	60.00%	29.80
交通运输	-5.32%	-6.48%	5	60.00%	26.60
计算机	8.22%	7.16%	4	100.00%	24.00
食品饮料	6.69%	5.37%	4	75.00%	31.25
商业贸易	5.36%	4.22%	4	75.00%	26.25
传媒	6.57%	6.11%	3	100.00%	7.33
建筑装饰	5.99%	5.03%	2	100.00%	21.00
公用事业	-6.08%	-7.58%	2	50.00%	36.50
农林牧渔	9.25%	8.69%	1	100.00%	10.00
Grand Total	2.32%	1.19%	319	71.47%	26.00

资料来源：wind；国信证券经济研究所整理

基于因子模型的配对交易策略

除了通过协整方法去筛选配对外，我们也可以基于因子模型去筛选配对。基于因子模型筛选，比通过协整方法去筛选，更为有效率。下面，我们首先梳理一下相关内容：

共同趋势模型

共同趋势模型的基本思想是：一个时间序列可以分解为平稳成分和非平稳成分两部分。如果两个时间序列存在协整关系，那么这两个时序的线性组合应当可以将非平稳成分消去，仅留下平稳成分。

因此，要满足协整，两个时间序列中的非平稳成分的一阶差分序列必然完全相关。而协整系数，也可以通过一阶差分序列进行回归计算得到。

共同趋势模型与套利定价理论的联系

套利定价理论（APT）将股票收益分解为因子收益和特殊收益。该理论认为，如果两个证券对相同的风险因子有相同的暴露，那么他们的期望收益也应当是相同的。

可以发现，APT 中的因子收益与共同趋势模型中非平稳成分的一阶差分序列正好对应；而特殊收益，正好与平稳成分的一阶差分序列相对应。因此，如果我们首先建立一个因子模型，那么只要满足：

1. 股票的因子收益强相关
2. 股票特殊收益的积分序列是平稳序列

那么我们就可以认为两只股票具有协整关系。下文，我们介绍如何用主成分分析的方法建立统计因子模型，并基于该模型去寻找具有稳定协整关系的股票对。

主成分分析法建立统计因子模型

主成分分析（PCA）是一种常用的构建统计因子模型的方法，其目的是找到一些互不相关的因子，并使得观察到的股票收益率可以很好的用因子收益的线性组合来解释。PCA 的困难之处在于如何解读所提取的因子，这点我们稍后说明。

我们用 PCA 方法对沪深 300 成分股构建因子模型，过程如下：

1. 提取沪深 300 成分股过去 240 个交易日的交易价格（前复权），计算收益率序列并做标准化处理。

$$Y_{it} = \frac{R_{it} - \bar{R}_{it}}{\bar{\sigma}_{it}}$$

2. 计算沪深 300 成分股的相关性系数矩阵，即：

$$C_{ij} = \frac{1}{T} \sum_{t=1}^T Y_{it} Y_{jt}$$

3. 计算矩阵 C_{ij} 的特征根和特征向量，将特征根按降序排列：

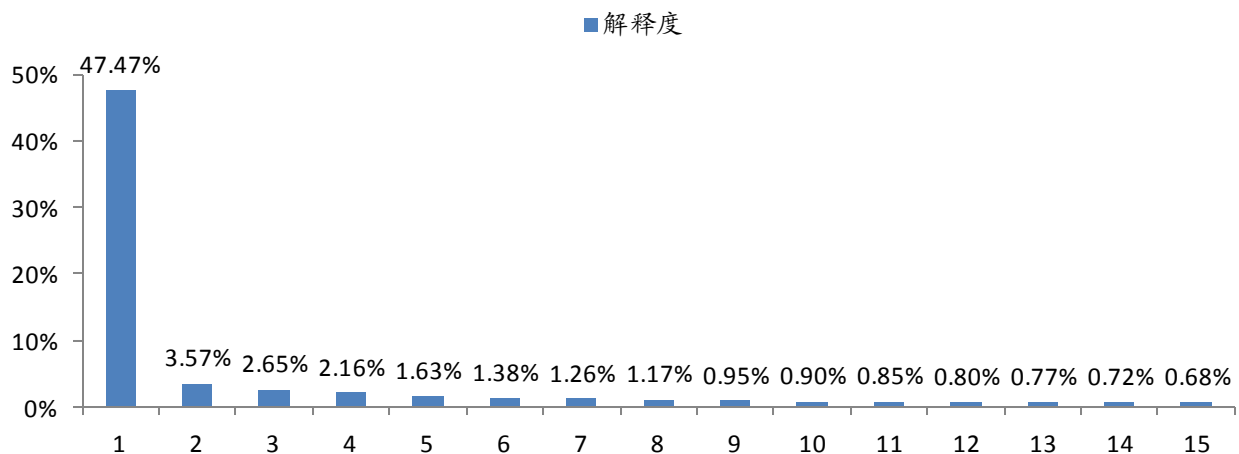
$$\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_N$$

特征根 λ_i 对应的特征向量记为：

$$v^i = (v_1^i, v_2^i \dots v_N^i)$$

其中，第 i 个特征根 λ_i 代表了第 i 个因子所解释的方差，前 15 个因子对收益率方差的解释度如图所示：

图 4：主成分分析法前 15 个因子对方差的解释度（2013-1-1，前 240 个交易日）



资料来源：wind；国信证券经济研究所整理

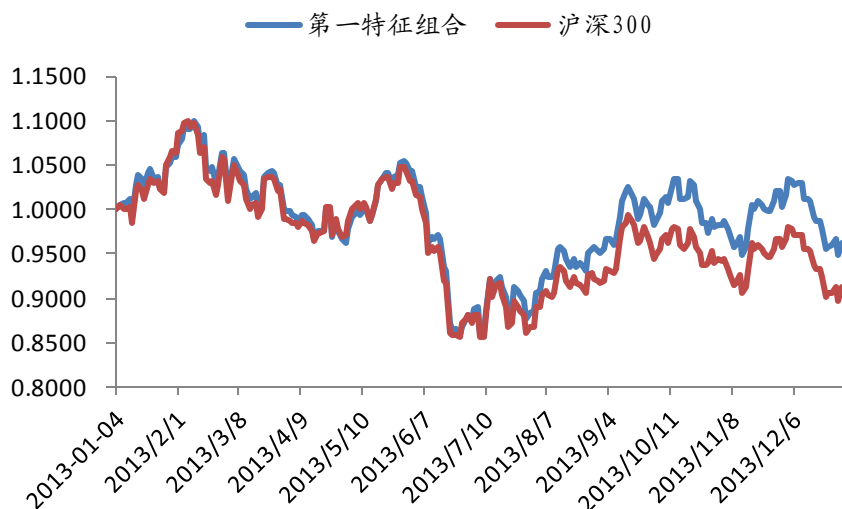
在 PCA 所提取的 N 个因子中，大部分是噪音因子，对于方差的解释度很小，需要将他们剔除，同时，我们也希望所采用的因子有一定的直观意义，而不仅仅是数学上处理的结果。

由于第 i 个因子实际上是一个特征组合，该组合在第 j 只股票上的权重可以通过特征向量获得，即 $w_j^i = v_j^i / \sigma_j$ ，该特征组合的收益率序列可以表示为：

$$F_t^i = \sum_{j=1}^N \frac{v_j^i}{\sigma_j} R_{jt}$$

实证可以发现，第 1 个因子在所有股票上的权重都是正数，我们可以近似将它看成是市场组合。从图中也可以看到，通过样本内数据提取的第一特征组合，在样本外期间与市场的走势相当一致。

图 5：第一特征组合 Vs 沪深 300（2013-1-1~2013-12-31）



资料来源：wind；国信证券经济研究所整理

除第 1 个特征组合之外，其余特征组合中的股票权重有正有负，通过对权重排序我们发现，一些特征组合中的股票有很强的行业聚集现象。因此，这些特征组合可以赋予直观意义上的解释，即行业多空组合。例如，从表中可以发现，第 2 特征组合可以认为是一个多医药生物，空银行的组合。

表 6：第二特征组合多头前 10 和空头前 10 行业分布情况

多头前 10	权重	行业	空头前 10	权重	行业
云南白药	1.21%	医药生物	民生银行	-1.21%	银行
东阿阿胶	1.21%	医药生物	浦发银行	-1.24%	银行
华润三九	1.12%	医药生物	中信银行	-1.30%	银行
恒瑞医药	1.12%	医药生物	中国石油	-1.39%	采掘
天士力	1.11%	医药生物	光大银行	-1.40%	银行
同仁堂	1.10%	医药生物	交通银行	-1.68%	银行
双鹭药业	1.03%	医药生物	中国银行	-1.89%	银行
华兰生物	1.03%	医药生物	建设银行	-2.08%	银行
上海家化	1.00%	化工	工商银行	-2.18%	银行
康美药业	0.95%	医药生物	农业银行	-2.19%	银行

资料来源：wind；国信证券经济研究所整理

第 3 特征组合中，多头部分有一些杂乱，但基本可以认为是一个多银行，空有色金属的组合。

表 7：第三特征组合多头前 10 和空头前 10 行业分布情况

多头前 10	权重	行业	空头前 10	权重	行业
农业银行	1.73%	银行	江西铜业	-0.59%	有色金属
中国银行	1.48%	银行	五矿发展	-0.59%	商业贸易
天士力	1.25%	医药生物	中金黄金	-0.60%	有色金属
建设银行	1.25%	银行	山东黄金	-0.60%	有色金属
万科 A	1.23%	房地产	辰州矿业	-0.60%	有色金属
工商银行	1.23%	银行	中色股份	-0.61%	有色金属
长江电力	1.22%	公用事业	中国铝业	-0.68%	有色金属
光大银行	1.13%	银行	驰宏锌锗	-0.68%	有色金属
中国建筑	1.13%	建筑装饰	锡业股份	-0.74%	有色金属
交通银行	1.08%	银行	紫金矿业	-0.76%	有色金属

资料来源：wind；国信证券经济研究所整理

第 4 特征组合中，多头和空头部分的行业聚集现象仍然非常明显，可以认为是一个多银行，空房地产的组合。

表 8：第四特征组合多头前 10 和空头前 10 行业分布情况

多头前 10	权重	行业	空头前 10	权重	行业
工商银行	3.03%	银行	首开股份	-0.88%	房地产
农业银行	2.89%	银行	北京城建	-0.90%	房地产
中国银行	2.64%	银行	华侨城 A	-0.94%	房地产
交通银行	2.37%	银行	中南建设	-0.95%	房地产
光大银行	2.19%	银行	金融街	-0.98%	房地产
建设银行	1.97%	银行	荣盛发展	-1.06%	房地产
中信银行	1.65%	银行	保利地产	-1.09%	房地产
招商银行	1.51%	银行	金地集团	-1.17%	房地产
浦发银行	1.48%	银行	招商地产	-1.20%	房地产
兴业银行	1.48%	银行	万科 A	-1.21%	房地产

资料来源：wind；国信证券经济研究所整理

以此类推，我们挨个查看每个特征组合。一直到第 12 特征组合，我们仍然可以发现较明显的行业聚集现象。第 12 特征组合可以大致认为是一个多公用事业，空国防军工、汽车的组合。

表 9：第十二特征组合多头前 10 和空头前 10 行业分布情况

多头前 10	权重	行业	空头前 10	权重	行业
大唐发电	1.77%	公用事业	洪都航空	-1.04%	国防军工
国电电力	1.67%	公用事业	中国卫星	-1.05%	国防军工
山东钢铁	1.26%	钢铁	宇通客车	-1.10%	汽车
华能国际	1.18%	公用事业	上海建工	-1.10%	建筑装饰
广深铁路	1.15%	交通运输	中国船舶	-1.16%	国防军工
内蒙华电	0.99%	公用事业	长安汽车	-1.17%	汽车
中能股份	0.95%	公用事业	航空动力	-1.32%	国防军工
大北农	0.88%	农林牧渔	上汽集团	-1.36%	汽车
中南传媒	0.84%	传媒	长城汽车	-1.37%	汽车
梅花集团	0.84%	食品饮料	华域汽车	-1.46%	汽车

资料来源：wind；国信证券经济研究所整理

从第 13 特征组合开始，我们发现行业聚集现象基本消失，因此我们选用前 12 个特征根所对应的因子，这些因子对收益率方差的解释度合计达 64.80%。

利用因子模型确定协整关系

1. 在提取出因子收益后，利用股票的历史收益率和前 12 个因子的历史收益率回归就可以计算得到该股票在每个因子上的暴露 (loading factor)，从而建立以下模型：

$$R_s = \sigma_s \sum_{k=1}^m \beta_{sk} F_k + \sigma_s \left(\sqrt{1 - \sum_{k=1}^m \beta_{sk}^2} \right) G_s$$

其中， $m = 12$ ， G_s 则满足 $E(G_s) = 0$ ， $\text{Var}(G_s) = 1$ 。这样我们就建立了股票的统计因子模型。

2. 前文中，我们已经提到，要满足协整关系，股票的共同趋势项必须强相关，在因子模型下，也就是 2 只股票可以用因子解释的收益部分是强相关的。

股票 A、B 共同趋势项的方差分别为：

$$\text{var}(A^{cf}) = \sigma_A^2 \sum_{k=1}^m \beta_{Ak}^2, \quad \text{var}(B^{cf}) = \sigma_B^2 \sum_{k=1}^m \beta_{Bk}^2$$

股票 A、B 共同趋势项的协方差为：

$$\text{cov}(A^{cf}, B^{cf}) = \sigma_A \sigma_B \sum_{k=1}^m \beta_{Ak} \beta_{Bk}$$

因此，股票 A、B 共同趋势项的相关系数为：

$$\text{corr}(A^{cf}, B^{cf}) = \frac{\sum_{k=1}^m \beta_{Ak} \beta_{Bk}}{\sqrt{\sum_{k=1}^m \beta_{Ak}^2} \sqrt{\sum_{k=1}^m \beta_{Bk}^2}}$$

只有当相关系数大于一定阈值时，我们才认为股票 A、B 有可能满足协整关系。

3. 要满足协整关系，除了共同趋势项的相关系数要大于一定阈值外，还有一个条件，就是残余项的积分序列必须是平稳序列。为此，我们可以先求出协整系数：

$$\gamma_{AB} = \frac{\text{cov}(A^{cf}, B^{cf})}{\text{var}(B^{cf})} = \frac{\sigma_A \sum_{k=1}^m \beta_{Ak} \beta_{Bk}}{\sigma_B \sqrt{\sum_{k=1}^m \beta_{Bk}^2}}$$

再利用 ADF 法检验 $\ln(P_A) - \beta \ln(P_B)$ 是否为平稳序列，如果满足平稳条件，我们就认为股票 A、B 是满足协整关系的，并纳入我们的配对池中。

实证结果

跟传统协整检验相比，利用因子模型筛选配对的一个主要特点就是不限于同一行业。我们知道，在用传统的协整检验法筛选配对时，我们是对同一行业内的股票进行两两之间的协整检验来筛选配对的，这有 2 方面的原因：

1. 协整检验是纯数学的方法，因此需要通过一些基本面的条件去加以限制，以保证股票之间的协整关系有一定的逻辑基础。
2. 对所有股票进行两两之间的协整检验计算量极大，仅针对同一行业的股票进行协整检验可以大大减少计算量。

而因子模型则比较好地解决了这 2 个问题，这使得我们可以去筛选跨行业的满足配对交易条件的股票配对。

当然，在实证中，利用因子模型筛选配对也有其自身的劣势。比如我们发现，在选用相同的相关性系数阈值的情况下，因子模型的配对数量要大大少于传统协整检验的方法。在相同参数设置下，因子模型所筛选出的配对的收益率，也要低于传统协整检验方法所筛选出的配对。

在实践中，我们认为可以同时参考这两种方法，互为补充。并且，除了统计因子模型之外，也可以考虑其他的因子模型，例如一些基本面因子模型等。

表 10：协整方法筛选配对 VS 因子模型筛选配对

配对筛选方法	收益率均值	收益率均值 (扣除成本)	交易次数	胜率	持有时间均值	不同行业配对
协整方法	2.32%	1.19%	319	71.47%	25.99	0
因子模型	1.92%	0.95%	82	63.41%	21.40	57

资料来源：wind；国信证券经济研究所整理

国信证券投资评级

类别	级别	定义
股票 投资评级	推荐	预计 6 个月内，股价表现优于市场指数 20%以上
	谨慎推荐	预计 6 个月内，股价表现优于市场指数 10%-20%之间
	中性	预计 6 个月内，股价表现介于市场指数 $\pm 10\%$ 之间
	回避	预计 6 个月内，股价表现弱于市场指数 10%以上
行业 投资评级	推荐	预计 6 个月内，行业指数表现优于市场指数 10%以上
	谨慎推荐	预计 6 个月内，行业指数表现优于市场指数 5%-10%之间
	中性	预计 6 个月内，行业指数表现介于市场指数 $\pm 5\%$ 之间
	回避	预计 6 个月内，行业指数表现弱于市场指数 5%以上

分析师承诺

作者保证报告所采用的数据均来自合规渠道，分析逻辑基于本人的职业理解，通过合理判断并得出结论，力求客观、公正，结论不受任何第三方的授意、影响，特此声明。

风险提示

本报告版权归国信证券股份有限公司（以下简称“我公司”）所有，仅供我公司客户使用。未经书面许可任何机构和个人不得以任何形式使用、复制或传播。任何有关本报告的摘要或节选都不代表本报告正式完整的观点，一切须以我公司向客户发布的本报告完整版本为准。本报告基于已公开的资料或信息撰写，但我公司不保证该资料及信息的完整性、准确性。本报告所载的信息、资料、建议及推测仅反映我公司于本报告公开发布当日的判断，在不同时期，我公司可能撰写并发布与本报告所载资料、建议及推测不一致的报告。我公司或关联机构可能会持有本报告中所提到的公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取提供投资银行业务服务。我公司不保证本报告所含信息及资料处于最新状态；我公司将随时补充、更新和修订有关信息及资料，但不保证及时公开发布。

证券投资咨询业务的说明

证券投资咨询业务是指取得监管部门颁发的相关资格的机构及其咨询人员为证券投资者或客户提供证券投资的相关信息、分析、预测或建议，并直接或间接收取服务费用的活动。

证券研究报告是证券投资咨询业务的一种基本形式，指证券公司、证券投资咨询机构对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向客户发布的行为。

国信证券经济研究所团队成员

宏观			策略			技术分析		
董德志	021-60933158		郇 彬	021-6093 3155		闫 莉	010-88005316	
钟正生	010-88005308		马 轲	021-60933157				
林 虎	010-88005302		孔令超	021-60933159				
固定收益			大宗商品研究			互联网		
董德志	021-60933158		马 轲	021-60933157		王学恒		
赵 婧	021-60875174		郇 彬	021-6093 3155		郑 剑		
刘 鹏			郑 东	010-66025270		李树国		
魏玉敏								
医药生物			社会服务(酒店、餐饮和休闲)			家电		
贺平鸽	0755-82133396		曾 光	0755-82150809		王念春	0755-82130407	
杜佐远	0755-82130473		钟 潇	0755-82132098		曾 婵	0755-82130646	
林小伟	0755-22940022							
邓周宇	0755-82133263							
李少思	021-60933152							
通信服务			电子			环保与公共事业		
程 成	0755-22940300		刘 翔	021-60875160		陈青青	0755-22940855	
李亚军	0755-22940077		陈 平	021-60933151		徐 强		
			卢文汉	021-60933164				
军工			机械			非金属及建材		
朱海涛	0755-22940097		朱海涛	0755-22940097		黄道立	0755-82130685	
			陈 玲	021-60875162		刘 宏	0755-22940109	
			成尚汶	010-88005315				
房地产			食品饮料			汽车及零配件		
区瑞明	0755-82130678		刘 鹏	021-60933167		丁云波		
			龙 飞	0755-82133920				
电力设备			新能源			建筑工程		
杨敬梅	021-60933160		张 弢	010-88005311		邱 波	0755-82133390	
						刘 萍	0755-22940678	
传媒与文化			零售、纺织服装及快销品			基础化工		
陈财茂	010-88005322		朱 元	021-60933162		李云鑫	021-60933142	
农林牧渔			轻工造纸			计算机		
杨天明	021-60875165		邵 达	0755-82130706		孙艺峻	010-88005323	
赵 钦	021-60933163							
银行			金融工程					
李关政			林晓明	021-60875168				
			吴子昱	0755-22940607				
			周 琦	0755-82133568				
			钱 晶	021-60875163				

国信证券机构销售团队

华北区（机构销售一部）			华东区（机构销售二部）			华南区（机构销售三部）		
王晓健	010-66026342 13701099132 wangxj@guosen.com.cn		叶琳菲	021-60875178 13817758288 yelf@guosen.com.cn		魏 宁	0755-82133492 13823515980 weining@guosen.com.cn	
李文英	010-88005334 13910793700 liwying@guosen.com.cn		崔鸿杰	021-60933166 13817738250 cuihj@guosen.com.cn		邵燕芳	0755-82133148 13480668226 shaoyf@guosen.com.cn	
赵海英	010-66025249 13810917275 zhaohy@guosen.com.cn		李 佩	021-60875173 13651693363 lipei@guosen.com.cn		段莉娟	0755-82130509 18675575010 duanlj@guosen.com.cn	
原 玮	010-88005332 15910551936 yuanyi@guosen.com.cn		汤静文	021-60875164 13636399097 tangjingwen@guosen.com.cn		郑 灿	0755-82133043 13421837630 zhengcan@guosen.com.cn	
甄 艺	010-66020272 18611847166		梁轶聪	021-60873149 18601679992 liangyc@guosen.com.cn		徐 冉	0755-82130655 13923458266 xuran1@guosen.com.cn	
杨 柳	18601241651 yangliu@guosen.com.cn		唐泓翼	13818243512		颜小燕	0755-82133147 13590436977 yanxy@guosen.com.cn	
王耀宇	18601123617		吴 国	15800476582		赵晓曦	0755-82134356 15999667170 zhaoxxi@guosen.com.cn	
陈孜譞	18901140709		储贻波	18930809296		刘紫薇	13828854899	
						许樱之	18688989863	