

基于最小距离法和协整理论的投资策略

黎明辉

(湛江幼儿师范专科学校 数学系, 广东 湛江 524084)

摘要:随着人民币国际化及 A 股市场的投资限制逐渐放宽, 国内市场的股价波动性进一步加剧, 增加了投资者的投资风险。在这种情况下, 配对交易作为一种较为稳健的投资组合策略逐渐为投资者所青睐。将上证 50 指数中的银行成分股作为研究对象, 以最小距离法对投资策略进行股票配对, 并分别采用最小距离交易策略和协整理论交易策略进行交易绩效实测, 对交易策略进行评估。研究发现, 配对交易在国内市场可以获得良好的收益。

关键词:配对交易; A 股市场; 最小距离法; 协整理论

中图分类号:F832 **文献标志码:**A **文章编号:**1671-1807(2023)11-0114-06

从 1994 年实行的盯住美元汇率制度到 2005 年汇改改为参考一篮子货币实行有浮动的汇率制度, 再到 2015 年汇改完善中间价报价机制, 中国金融市场对外开发的程度在进一步加深, 中国越来越成为经济全球化的重要组成部分。在这一大背景下, 中国国内金融市场的投资机会大大增加, 可供投资者选择的投资产品变得更加多元化, 投资者可以结合其自身风险偏好, 通过充分分散投资组合来获取其期望收益, 这无疑进一步推动了投资者的投资热情。与此同时, 中国对外开放的程度进一步加深也意味着国内经济与世界经济形势的联系更加密切, 诸如金融危机和信用危机等一些国际性的系统性风险对国内的金融市场的影响也会加大, 投资者面临的风险也增加了, 这也要求投资者在确定投资组合的时候要更加谨慎。因此, 投资者希望找到一种可以合理规避风险、获得稳定收益的投资策略。

配对交易由于构造了对冲组合, 投资风险较小, 是一种较好的中性投资策略, 也因此在国内的证券交易市场得到了广泛的应用。在国内证券市场开放部分股票做空制度下, 配对交易策略有望在实际市场交易中得到进一步的应用。

1 相关理论

配对交易作为一种新兴的投资交易策略, 主要包括配对期和交易期两个阶段。

进行配对交易首先要寻找合适的股票对进行

对冲组合的构建, 股票对中的股票应当具有相似的历史走势。由于 A 股市场的股票数量众多, 两两配对形成的股票对数量更是庞大, 故需要采用一些方法对股票池中的股票进行过滤。过滤方法包括行业内股票配对、产业链股票配对和财务管理股票配对等。陈晓芬和杨朝军^[1]通过对上证 180 指数进行行业内分类改进了股票配对方式。鲍鹏飞^[2]以沪深 300 指数成分股为股票池, 进行上下游股票配对, 分析其收益水平, 证实了通过产业链配对方法进行配对投资策略构建可以有效地提高策略的投资效益。

在挑选好配对的股票池以后, 对配对池众多的股票对, 可以进一步通过最小距离法和协整理论法进行定量研究, 从中选择相似性最高的股票对。郑青青和罗津^[3]通过对上证 50 指数成分股运用最小距离法测度了价格距离, 并进一步分析投资组合的收益。

挑选出合适的股票对后, 进入交易期。交易期的操作原理比较直观, 当股票对的价差偏离长期稳定状态时, 买入相对低估的股票, 做空相对高估的股票, 关键的问题在于价差偏离长期稳定状态具体达到什么程度时进行交易, 价差长期偏离稳定状态是否需要止损等。对于这些问题, 学者们也提出了相应的标准。Gatev 等^[4]学者提出: 利用 12 个月的股票数据进行计算, 将交易阈值设置成 2 倍的标准差, 当价差偏离设定的交易阈

收稿日期: 2023-02-03

基金项目: 湛江市非资助科技攻关计划项目(2021B01494)。

作者简介: 黎明辉(1996—), 男, 广东茂名人, 湛江幼儿师范专科学校, 助教, 硕士, 研究方向为量化投资与机器学习。

值时,触发配对交易,如果建仓后 6 个月价差未收敛,则强制平仓。国内学者通过应用不同的股票数据不断对交易阈值进行修正,做了一些改良。其中,丁涛^[5]将交易上限的阈值修正为 2.5 倍的标准差。

2 配对期实证研究

2.1 最小距离法

最小距离法是由 Gatev 等^[4]提出的,是配对交易中一种较为常见的股票相似性计算方法。 P_t^i 为股票 i 在第 t 天的价格。可以计算得到单期收益率,公式为

$$r_t^i = \frac{P_t^i - P_{t-1}^i}{P_{t-1}^i}, \quad t = 1, 2, \dots, T \tag{1}$$

用 $\hat{p}_t^i$ 表示股票 i 在第 t 天的经过标准化处理后的去量标价格,一般 $\hat{p}_t^i$ 通过当天回溯 t 天的时间段累积收益率表示为

$$\hat{p}_t^i = \prod_{\tau=1}^t (1 + r_\tau^i) \tag{2}$$

由式(2)可以计算股票 X 和股票 Y 的标准去量标价格离差平方和 $SSD_{X,Y}$,公式为

$$SSD_{X,Y} = \sum_{t=1}^T (\hat{p}_t^X - \hat{p}_t^Y)^2 \tag{3}$$

采用同行业股票池构建方法,采取将同板块(上证 50 指数)中的成分股中的同行业(银行业)股票作为股票池寻找配对股票对。具体的成分股票如表 1 所示。

将配对期阶段的时间段定为 2018 年 1 月 1 日至 2018 年 12 月 31 日,并将上证 50 指数的 10 只银行成分股放入股票池,通过国泰君安数据库获取该 10 只股票在配对期阶段的每日收盘价数据作为股票价格的代表数据,通过将股票池的股票价格两两组合计算标准化价格偏差平方和,可以取得标准化价格偏差平方和最小的一组,将其作为配对股票对。计算结果如表 2 所示。

表 1 配对交易股票池的成分股票

序号	股票代码	股票名称
1	600000	浦发银行
2	600016	民生银行
3	600036	招商银行
4	601166	兴业银行
5	601288	农业银行
6	601398	工商银行
7	601328	交通银行
8	601818	光大银行
9	601939	建设银行
10	601988	中国银行

数据来源:国泰君安数据库。

由表 2 的结果可知,SSD 计算结果最小的股票对为建设银行(601939)和工商银行(601398),通过 python 程序可以画出建设银行(601939)和工商银行(601398)在配对期的股价走势图(图 1)。

由图 1 可知,建设银行(601939)和工商银行(601398)在配对期的股价整体走势大致相同,可以粗略认为可以进行配对交易策略^[6],其具体的走势在多大程度上相同,是否能够构造完善的交易策略需要进一步观察,以下通过协整检验进一步验证。

2.2 协整检验法

协整方法是一种非平稳时间序列分析方法,其原理为两个或两个以上的非平稳时间序列虽然单个来看是非平稳的时间序列,但是互相之间具有长期稳定的关系,图像上看表现为随着时间推移,时间序列数据有着相似的走势,可以通过协整理论对这种关系进行描述^[7]。应用协整理论进行分析,要求两个及两个以上的非平稳时间序列构造的线性模型序列为平稳时间序列,若满足该条件,则说明存在协整关系,反之则说明不存在。

应用 E-G 两步法进行协整关系检验,需要先检查两只股票的对数价格序列是否为同阶单整序列。如果两只股票的对数价格序列为一阶单整序列,那

表 2 股票池配对股票所得标准化价格偏差平方和 SSD

SSD	600000	600016	600036	601166	601288	601398	601328	601818	601939	601988
600000		1.60	3.77	1.31	3.75	1.96	2.48	3.03	1.78	2.35
600016			7.64	4.21	7.49	4.24	5.51	6.32	3.83	5.12
600036				0.76	0.20	0.91	0.44	0.25	1.07	0.54
601166					0.78	0.43	0.31	0.48	0.50	0.31
601288						0.64	0.40	0.17	0.83	0.35
601398							0.47	0.50	0.05	0.19
601328								0.16	0.63	0.10
601818									0.64	0.18
601939										0.33
601988										

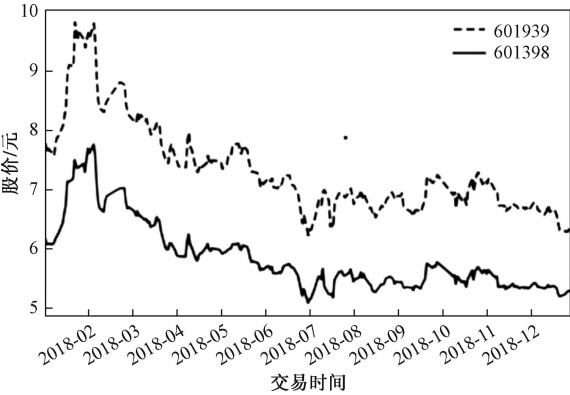


图 1 建设银行(601939)和工商银行(601398)在配对期的股价走势

么进一步应用最小二乘法构造回归方程为

$$\ln P_t^Y = \alpha + \beta \ln P_t^X + \epsilon_t \tag{4}$$

从而得到回归系数 $\hat{\alpha}$ 和 $\hat{\beta}$, 构造残差估计值

$$\epsilon_t = \ln P_t^Y - (\hat{\alpha} + \hat{\beta} \ln P_t^X) \tag{5}$$

检验序列 $\{\epsilon_t\}$ 的平稳性。如果序列 $\{\epsilon_t\}$ 平稳, 则说明 $\{\ln P_t^X\} (t = 0, 1, 2, \cdots, T)$ 序列和 $\{\ln P_t^Y\} (t = 0, 1, 2, \cdots, T)$ 序列具有协整关系。

应用协整理论来验证在运用最小距离法计算处理的最优配对股票对建设银行(601939)和工商银行(601398)在配对期时间阶段(2018 年 1 月 1 日至 2018 年 12 月 31 日)是否存在长期稳定的协整关系。

首先通过 python 程序分别对建设银行(601939)和工商银行(601398)的对数价格序列进行单位根检验, 判断其是否为一阶单整序列, 检验结果如表 3 所示。

由表 3 可知, 在 1% 的显著性水平下, 建设银行(601939)和工商银行(601398)的对数价格序列均不平稳且对数价格序列的一阶差分序列均平稳, 即两者的对数价格序列均为一阶单整序列。故应用最小二乘法进行回归分析, 以建设银行(601939)的对数价格序列为因变量, 工商银行(601398)的对数价格序列为自变量, 结果如表 4 所示。

表 3 股票对序列单位根检验结果

序列	单位根检验结果	一阶差分处理后 单位根检验结果
建设银行(601939) 对数价格序列	-1.283 (0.637)	-7.724*** (0.000)
工商银行(601398) 对数价格序列	-1.351 (0.606)	-5.188*** (0.000)

注: 括号内为 P 值; * 为 $P < 0.1$, ** 为 $P < 0.5$, *** 为 $P < 0.01$ 。

由表 4 可知, 回归方程各项系数均在 1% 的显著性水平上显著不为 0, 且回归方程在 1% 的显著性水平上显著, R^2 接近 1, 说明方程的拟合效果较好。进一步地, 可以得到回归模型的残差序列 Spread, 对残差序列进行单位根检验结果如表 5 所示。

由表 5 可知, 在 1% 的显著性水平下, 可以认为回归模型的残差序列 Spread 为平稳时间序列。由于在建设银行(601939)和工商银行(601398)的对数价格序列均不平稳且对数价格序列的一阶差分序列均平稳的同时, 建设银行(601939)的对数价格序列对工商银行(601398)的对数价格序列进行回归得到的残差序列为平稳序列, 故可以认为建设银行(601939)和工商银行(601398)的对数价格序列存在长期稳定的协整关系。

2.3 设置交易信号

根据配对交易的交易策略, 在配对股票期选择出配对股票, 并在交易期通过设置交易信号进行股票交易。参考较为常用的配对交易策略, 将 12 个月的时间长度作为股票配对的形成期, 6 个月的时间长度作为股票配对的交易期。具体来说, 针对不同的交易策略, 设定的交易信号也不同。

2.3.1 以最小距离法为基础的交易信号设定

首先利用前 12 个月的股票交易数据, 计算建设银行(601939)与工商银行(601398)的股票标准化价格序列差 SSD. pair 序列, 并通过 python 程序计算出 SSD. pair 序列的数学期望 μ 和标准差 σ 。在交易期内, 设定 $\mu \pm 1.2\sigma$ 为开仓和平仓的阈值。

交易期的具体交易规则如下:

- 1) 当标准化价格序列差 SSD. pair 上穿 $\mu + 1.2\sigma$ 时, 做空配对股票, 反向建仓(做空工商银行股票, 同时做多建设银行股票)。
- 2) 当标准化价格序列差 SSD. pair 下穿 μ 时, 做多配对股票, 反向平仓(做多工商银行股票, 同时做空建设银行股票)。
- 3) 当标准化价格序列差 SSD.

表 4 建设银行的对数价格序列对工商银行的对数价格序列回归结果

变量	回归系数	标准误	t	P
c	-0.090 7	0.015	-5.906	0.000
工商银行 601399	0.932 8	0.008	120.607	0.000
$R^2 = 0.984, P(F\text{-statistic}) = 0.000$				

表 5 回归模型的残差序列 Spread 的单位根检验结果

序列	单位根检验结果
回归模型的残差序列 Spread	-3.360*** (0.012)

注: 括号内为 P 值; * 为 $P < 0.1$, ** 为 $P < 0.5$, *** 为 $P < 0.01$ 。

pair 下穿 $\mu - 1.2\sigma$ 时,做多配对股票,正向建仓(做多工商银行股票,同时做空建设银行股票)。

4)当标准化价格序列差 SSD. pair 上穿 μ 时,做空配对股票,正向平仓(做空工商银行股票,同时做多建设银行股票)。

2.3.2 以协整理论法为基础的交易信号设定

首先利用前 12 个月的股票交易数据,将建设银行(601939)对数价格序列对工商银行(601398)的对数价格序列运用最小二乘法进行回归,得到回归方程后提出斜率 $\hat{\beta}$,并应用 Python 程序求出回归方程中的残差序列 Spread 的数学期望 μ 和标准差 σ 。在交易期内,设定 $\mu \pm 1.5\sigma$ 和 $\mu \pm 0.2\sigma$ 为开仓和平仓的阈值,将 $\mu \pm 2.5\sigma$ 视为协整关系可能破裂强制平仓的阈值。

交易期的具体交易规则^[8]如下:

1)当价差 Spread 上穿 $\mu + 1.5\sigma$ 时,做空配对股票,反向建仓(做空工商银行股票,同时做多建设银行股票,建设银行与工商银行股票的资金比值为 $\hat{\beta}$)。

2)当价差 Spread 下穿 $\mu + 0.2\sigma$ 时,做多配对股票,反向平仓(做多工商银行股票,同时做空建设银行股票)。

3)当价差 Spread 下穿 $\mu - 1.5\sigma$ 时,做多配对股票,正向建仓(做多工商银行股票,同时做空建设银行股票,建设银行与工商银行股票的资金比值为 $\hat{\beta}$)。

4)当价差 Spread 又回复到 $\mu - 0.2\sigma$ 上方时,做空配对股票,正向平仓(做空工商银行股票,同时做多建设银行股票)。

5)当价差 Spread 突破 $\mu \pm 2.5\sigma$ 时,及时平仓。

2.4 交易期配对交易实证研究及投资绩效分析

在选定配对交易的股票对后,分别运用最小距离法投资策略和协整理论法交易策略进行交易,最后选取最优的投资策略并进行评价。

为了更为直观地阐述实证分析过程,需要做一些假设。假设内容为:交易初期,账户的资产总价值为 2 000 元现金;假设开仓交易信号触发时做多被低估的股票,做空被高估的股票;不考虑交易成本。^[9]

首先通过最小距离法确立最优配对交易股票对,由表 2 的结果可知建设银行(601939)和工商银行(601398)是最优的配对交易股票对。样本时间范围为 2018 年 1 月 1 日至 2019 年 6 月 28 日,其中 2018 年 1 月 1 日至 2018 年 12 月 31 日为股票配对期,2019 年 1 月 2 日至 2019 年 6 月 28 日为股票交

易期,使用历史收盘价数据进行分析。

2.4.1 最小距离法投资策略

根据配对交易中的最小距离法投资策略,首先计算配对期建设银行(601939)与工商银行(601398)的股票标准化价格序列差 SSD. pair 序列,利用 python 程序进行绘图如图 2 所示。

从图 2 可知,在配对期,由建设银行(601939)与工商银行(601398)构成的股票对均存在开仓平仓的获利交易机会。因此,可以进一步看交易期的交易情况(图 3)。

由图 3 可知,交易期建设银行(601939)与工商银行(601398)的股票标准化价格序列差大多数时候都位于 $\mu + 1.2\sigma$ 以上,平仓机会较少,获利机会基本没有,难以形成有效的投资交易策略,故可以认为,以建设银行(601939)与工商银行(601398)这一配对股票对而言,最小距离法为基础的投资策略是不成功的。

2.4.2 协整理论法投资策略

根据配对交易中的协整理论法投资策略,首先利用前 12 个月的股票交易数据,将建设银行

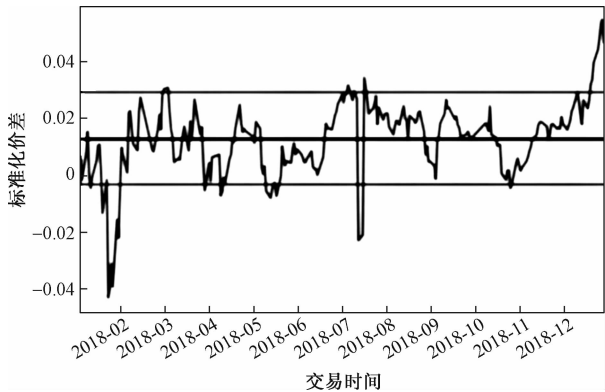


图 2 配对期建设银行(601939)与工商银行(601398)的股票标准化价格序列差

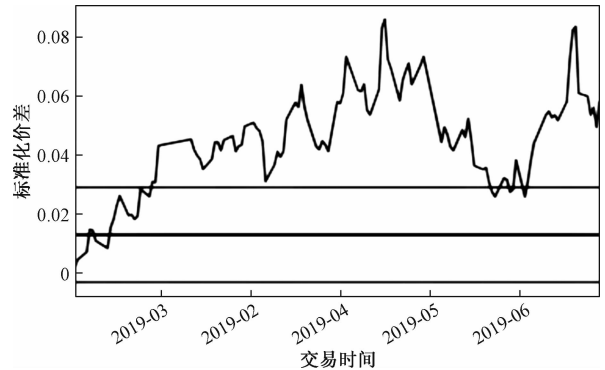


图 3 交易期建设银行(601939)与工商银行(601398)的股票标准化价格序列差

(601939)对数价格序列对工商银行(601398)的对数价格序列运用最小二乘法进行回归,得到回归方程后提出斜率 $\hat{\beta}$ 和截距 $\hat{\alpha}$,并应用 Python 程序求出回归方程中的残差序列 Spread 的数学期望 μ 和标准差 σ 。其中,计算得出的 $\hat{\beta}$ 为 0.932 8, $\hat{\alpha} = -0.090 7$ 。故可知,建设银行 $= -0.090 7 \hat{\alpha} + 0.932 8 \hat{\beta} + \epsilon_t$ 。

在交易期,一旦触发了交易信号,便根据建设银行(601939)和工商银行(601398)的股价相对高估情况进行交易^[10]。如当建设银行(601939)被相对高估,而工商银行(601398)被相对低估时,便做空建设银行(601939)股票,同时做多工商银行(601398)股票,且保证做多工商银行(601398)的股票分数与建设银行(601939)股票的分数的比值等于 $\hat{\beta} = 0.932 8$ 。

利用 python 程序可以将残差序列 Spread 可视化,并求出其数学期望 μ 和标准差 σ 。其中,计算得出的 $\mu = -3.751 700 \times 10^{-15}$, $\sigma = 0.012 157$,残差序列 Spread 可视化结果如图 4 所示。

由图 4 可知,残差序列 Spread 的走势大致平稳,且已经验证过残差序列 Spread 的平稳性,检验结果是在 1%的显著性水平下,残差序列 Spread 是显著平稳的。因此,可以进一步分析在交易期的交易情况。

同样的,利用 Python 程序将残差序列 Spread 在交易期内的数据情况可视化,如图 5 所示。

由图 5 可知,交易期的大部分时间中,残差序列 trading Spread 均在 $\mu + 1.5\sigma$ 和 $\mu - 1.5\sigma$ 之间围绕均值上下波动,2019 年 1 月至 2019 年 6 月底这 6 个月的时间内,配对交易信号触发不多(共 6 次),向上突破 $\mu + 1.5\sigma$ 一次,向下突破 $\mu - 1.5\sigma$ 五次,存

在开仓和平仓机会,且残差序列 trading Spread 整体波动情况较为稳定,故认为交易期存在交易获利机会,进一步建立 Python 交易程序模拟交易情况。由假设条件,交易初期账户只拥有 2 000 元现金作为总资产,不持有任何股份,将交易过程中持有的工商银行(601398)股份数记为 ShareY(当 ShareY 小于 0 时意味着卖空工商银行股票),现金持有额记为 Cash,账户总资产记为 Asset,可得交易期模拟交易情况如图 6 所示。

通过 python 程序,可以进一步调出账户最后 5 天的交易情况如表 6 所示。

由图 6 和表 6 可知,整个交易期期间,账户总资产始终保持在 2 000 元以上,整体上保持一个稳定的增长,观察现金曲线图,由于开仓可能需要现金,现金曲线有升有降,在整个交易期现金曲线呈现缓慢上升的趋势,且保持在 2 000 元以上。观察总资产曲线,截至 2019 年 6 月 28 日,账户最终资产总额为 2 547.94 元,同时完成到期平仓,账户的期间总收益为 547.94 元,转化计算年化收益率约为 54.8%,从收益金额和收益率上来看,投资策略的绩效较为可观。

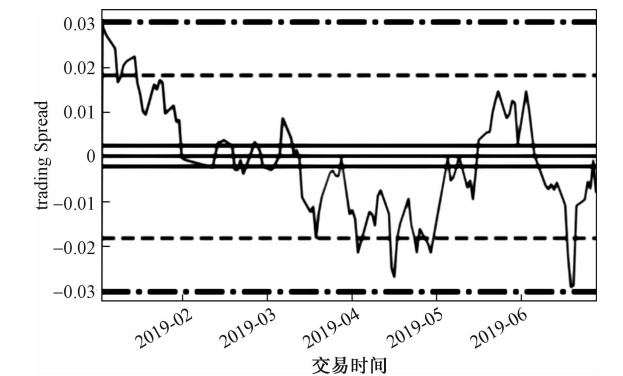


图 5 交易期残差序列 trading Spread

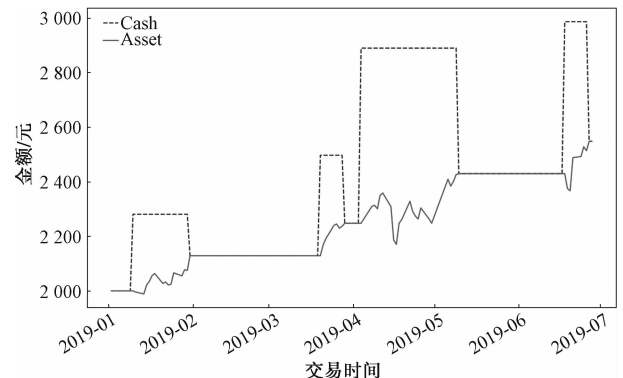


图 6 交易期交易情况

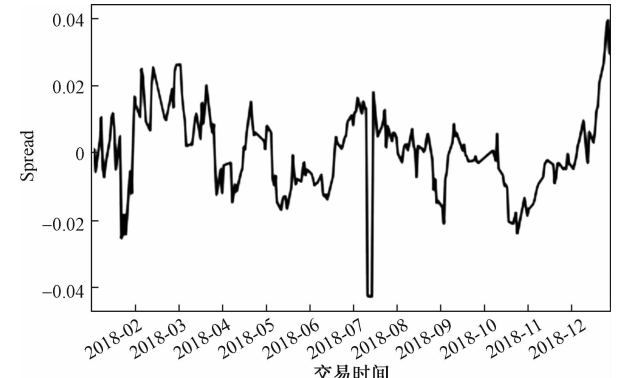


图 4 配对期残差序列 Spread

表 6 交易期最后 5 d 账户交易情况

交易日期	仓位	ShareY/股	ShareX/股	Cash/元	Asset/元
2019/6/24	1	1 000	—856.683 815 2	2 985.971 883	2 492.244 298
2019/6/25	1	1 000	—856.683 815 2	2 985.971 883	2 527.912 68
2019/6/26	1	1 000	—856.683 815 2	2 985.971 883	2 513.645 327
2019/6/27	0	0	0	2 547.944 813	2 547.944 813
2019/6/28	0	0	0	2 547.944 813	2 547.944 813

3 结论

通过最小距离理论和协整理论进行股票配对交易回测和实证分析。首先通过最小距离法完成了配对期股票对的选择,并通过最小距离法交易策略和协整理论法交易策略对配对股票对在交易期的交易情况进行分析,通过比较两者的交易情况和记录投资绩效进行了研究策略的绩效总结。研究发现,通过协整理论法投资策略进行统计套利,可以获得较好的投资收益。

参考文献

[1] 陈晓芬,杨朝军. 配对交易中配对与交易策略双重改进下的行业实证检验[J]. 上海金融, 2017(2): 67-74.

[2] 鲍鹏飞. 配对交易在我国股票市场上的实证研究[D]. 成都: 电子科技大学, 2017.

[3] 郑青青,罗津. “两融”背景下证券配对交易投资组合的策

略研究[J]. 价格理论与实践, 2017(11): 118-121.

[4] Gatev E, Goetzmann W N, Rouwenhorst K G. Pairs trading: performance of a relative value arbitrage rule[J]. Social Science Electronic Publishing, 2006, 19 (3): 797-827.

[5] 丁涛. 配对交易策略在 A 股市场的应用与改进[J]. 中国商贸, 2013(5): 24-26, 28.

[6] 张河生, 闻岳春. 基于参数调整的协整配对交易策略: 理论模型及应用[J]. 西部金融, 2013(1): 11-16.

[7] 胡文伟, 胡建强, 李湛, 等. 基于强化学习算法的自适应配对交易模型[J]. 管理科学, 2017, 30(2): 148-160.

[8] 欧阳红兵, 李进. 基于协整技术配对交易策略的最优阈值研究[J]. 投资研究, 2015, 34(11): 79-90.

[9] 麦永冠, 王苏生. WM-FTBD 配对交易建仓改进策略及沪深港实证检验[J]. 管理评论, 2014, 26(1): 30-40.

[10] 沈银芳, 郑学东, 徐信喆. 基于混合 Copula 的 ETF 配对交易策略[J]. 浙江大学学报(理学版), 2016, 43(3): 271-278.

Investment Strategy Based on Minimum Distance Method and Cointegration Theory

LI Minghui

(Department of Mathematics, Zhanjiang Preschool Education College, Zhanjiang 524084, Guangdong, China)

Abstract: With the internationalization of the RMB and the gradual relaxation of investment restrictions in the A-share market, the volatility of stock prices in domestic market has further intensified, which increases investors’ investment risks. Under such circumstances, pair trading, as a relatively stable investment portfolio strategy, is gradually favored by investors. The bank constituent stocks in the Shanghai Stock Exchange 50 Index were used as the research object, and the minimum distance method was used to carry out stock matching for the investment strategy, and the minimum distance trading strategy and the cointegration theory trading strategy were used to measure the trading performance respectively, and the trading strategy was evaluated. It is found that pairing transactions can get good benefits in the domestic market.

Keywords: pair trading; A-share market; minimum distance method; cointegration theory