

索取号： 0213.9/3.114:180938019

密级： 公开

南京师范大学

应用统计硕士学位论文



基于协整理论的配对交易策略研究

研 究 生： 于梦航

指 导 教 师： 李启才 副教授

培 养 单 位： 数学科学学院

专业学位领域： 应用统计

完 成 时 间： 2020 年 3 月 20 日

答 辩 时 间： 2020 年 5 月 24 日

摘 要

配对交易策略由高盛投行投资家 Nunzio Tartaglia 领导的量化团队于上世纪 80 年代中期开发并成功实践，现已广泛应用于金融投资行业，并成为学术界的研究热点。配对交易是基于金融市场不均衡环境下实施的一种中性统计套利策略，配对过程涉及双边交易，做空股票的同时也要做多另一只股票，在对冲交易中频繁使用。2010 年 3 月 31 日我国正式开展融资融券业务，市场引入做空机制，配对交易受到了投资者的注意。随着两融业务的不断完善，配对交易有更大的发展空间。

配对交易的基本理念是选取两只价格高度相关的股票，它们之间的价差长期处于一个稳定的状态。价差出现偏差的时候，就做空价格偏高者做多价格偏低者，偏差消失时，反向操作，以此来套取利润。配对交易的运用，可以发现更合理的资产价格，提升市场定价效率，减少市场上的非理性行为。

本文基于协整理论进行配对交易分析，通过实证分析说明策略的有效性。考虑到相同行业的影响因素大部分相同，同行业资产之间会有长期均衡的关系等，最终选取了上海证券交易所融资融券标的股票中的 18 只银行股票作为研究对象。交易分为两步：选取股票对和制定交易策略。首先，对样本内数据进行相关性分析和协整分析，选择工商银行和建设银行作为一对。通过建立误差修正模型得到短期动态模型，并根据误差项系数得到了两只股票的对冲比例和价差序列。然后制定交易策略，分别使用固定标准差和时变标准差两种方法模拟价差的波动情况，确定交易信号。对于固定标准差，就是取样本内价差的标准差；对于时变标准差，根据 GARCH 模型导出的递推公式拟合标准差，分别用不同的标准差确定交易阈值，得到两种交易策略。最后，进行样本实证分析，得到不同策略下的交易情况。

实证结果表明，使用配对交易策略取得的收益率远高于无风险利率，说明配对交易在我国市场是有效的。同时，时变标准差下策略的效果优于固定标准差的效果，表明时变标准差更能反映出价差的变动，及时发现价差偏离正常水平，确定交易时机，取得更好的收益。

关键词：配对交易，融资融券，协整理论，GARCH 模型

Abstract

Developed and successfully practiced by a quantitative team led by Goldman Sachs Investment Bank investor Nunzio Tartaglia in the mid-1980s, paired trading strategy has been widely used in the financial investment industry and has become a research hotspot in academia. Paired trading is a neutral statistical arbitrage strategy implemented in a non-equilibrium financial market environment. The pairing process involves a bilateral trading, when building a short position, also having to build a long position. It is frequently used in hedge funds. On March 31, 2010, China officially launched the margin trading business, the short selling mechanism was introduced into the market and the pair trading had caught investors' attention. With the continuous improvement of the margin trading businesses, the pair trading has greater development space.

The basic idea of pair trading is to choose two stocks whose prices have a high correlation, and the spread of the two is in a stable state for a long time. When the spread diverges the historical standard deviations from equilibrium, the trader will short the high stock and buy the low stock. When the deviation disappears, the trader will move in the opposite operation to make a profit. The application of pair trading can find more reasonable prices, improve the efficiency of market pricing, and reduce the irrational behavior in the market.

In this paper, we study how to carry out the pair trading based on the cointegration theory and demonstrate the effectiveness of the strategy by empirical analysis. Considering that the influence factors of the same industry are mostly same, there will be a long-term equilibrium relationship during the same industry. We select 18 bank stocks of margin trading's underlying stocks from Shanghai Stock Exchange as research objects. Pair trading consists of two steps: selecting stock pairs and formulating trading strategies. Firstly, for the in-sample data, by conducting correlation analysis and cointegration analysis, ICBC and CCB are selected as a pair. The short-term dynamic model is obtained by establishing the error correction model, the hedging ratio and the spread sequence of two stocks are obtained according to the error term coefficient. Then, developing the trading strategy, we use different methods to fit the fluctuation of the spread, one is the fixed standard deviation, the other is the time-varying standard deviation, to determine the trading signal. For the fixed standard deviation, it's the standard deviation in sample of the spread; for the time-varying standard deviation, fitting the standard deviation with recurrence formula derived from the

GARCH model. Different trading thresholds are determined by using two standard deviations, two trading strategies are obtained. Finally, carrying out the empirical analysis of samples, we get the trading situation under different strategies.

The empirical results show that the rate of return obtained by the pair trading strategy is much higher than the risk-free rate, indicating that the pair trading is effective in the Chinese market. At the same time, the effect of time-varying standard deviation is better than the effect of fixed standard deviation, indicating that the former can better reflect the change of the spread, timely find the deviation of the spread from the normal level, determine the trading time, and obtain better returns.

Keywords: Paired Trading, Margin Trading, Cointegration Theory, GARCH Model

目录

第一章 绪论	1
1.1 研究背景及研究意义	1
1.1.1 研究背景.....	1
1.1.2 研究意义.....	2
1.2 文献综述	2
1.2.1 配对方法文献综述.....	3
1.2.2 基于协整理论的配对交易文献综述.....	4
1.3 研究内容	7
1.4 本文创新点	7
第二章 配对交易策略理论基础及具体操作方法	9
2.1 均值回归理论	9
2.2 相关性分析	9
2.3 时间序列预处理分析	10
2.3.1 平稳序列.....	10
2.3.2 纯随机序列.....	10
2.4 单位根检验	11
2.5 协整	12
2.5.1 协整概念.....	12
2.5.2 协整检验.....	13
2.6 误差修正模型	13
2.7 条件异方差模型	14
2.7.1 ARCH 检验	15
2.7.2 GARCH 模型	15
2.8 配对交易策略具体操作方法	17
2.8.1 选择股票对.....	17
2.8.2 确定交易策略.....	17
2.8.3 总收益率.....	18
第三章 实证分析	19
3.1 相关性分析	19
3.2 协整分析	21
3.3 误差修正模型	23

3.4 固定标准差	25
3.4.1 样本内实证.....	26
3.4.2 样本外实证.....	30
3.5 时变标准差	31
3.5.1 样本内实证.....	34
3.5.2 样本外实证.....	36
3.6 本章小结	37
第四章 结论与展望	38
4.1 结论	38
4.2 不足与展望	38
参考文献	40
致谢	43

第一章 绪论

1.1 研究背景及研究意义

1.1.1 研究背景

在二十世纪八十年代，配对交易策略由高盛投行投资家 Nunzio Tartaglia 成立的一个数量分析团队提出，于 1985 年投入实盘交易，带来了巨大的利润。这不仅让整个团队声名远扬，也让这种策略走进了人们的视野。配对交易是一种市场中性投资策略，可以降低证券市场的系统风险性，同时也是一种统计套利方法，基于统计学和计量经济学对一组证券进行研究分析，从而获得利润。其基本思想是选取两只历史相关性比较高的证券，它们的价格之间存在长期均衡关系。当一方价格变高另一方价格变低时，偏离了之前的均衡关系，这时就产生了套利机会，可以做空价格变高的一方，做多价格变低的一方，在价格关系回到正常时，再进行反向操作来获取利润。

2010 年 3 月 31 日起，融资融券交易正式在上海证券交易所、深圳证券交易所实施，两融交易开始在市场运行，这表明中国金融市场有了做空机制，双边交易机制终于在我国实现。融资融券的业务开展，为金融市场注入了活力和带来了挑战，能够有效地缓解资金压力，并且起到刺激、活跃市场的作用。而在此之前，我国金融市场的交易方式只有单侧做多，因为不能做空，这就很大程度地限制了配对交易策略的发展，不利于其应用。

两融交易由融资和融券构成，投资者要对具有业务资格的证券公司提供担保，融资表示投资者借入本金购入有价证券，之后在约定的时间内向证券公司偿还借入的资金；融券表示投资者借入证券并出售，之后在约定的时间内向证券公司偿还借入的证券。两融业务的实施，投资者能够借助融资服务做多证券，也可以通过融券做空证券，这就加快了人们对于双边交易策略的研究，其中配对交易这一经典的策略，自然得到了投资者的重视。

截至到 2019 年 12 月 18 日，融资融券余额即将突破万亿大关，我国沪深证券交易所两融业务标的证券数量由 950 只增加至 1600 只。其中，上交所的标的股票由原本的 525 只扩大到 800 只，深交所的标的股票由原本的 425 只扩大到 800 只。标的股票范围增加后，标的市值占市场总值的比例大于 80%。自从两融业务实施以来，标的覆盖范围逐步扩大，运作模式愈见成熟，证券公司风险控制水平逐渐提高，表明融资融券业务一直在稳步前进。标的股票的增加，丰富了投资者的交易方式，满足了投资者多样化的投资需求。

1.1.2 研究意义

配对交易在二十世纪八十年代已大放光彩,出色的应用让国外投资者对其青睐有加,在国外金融市场已被广泛地使用,在我国,随着融资融券业务不断的拓展,它在我国金融市场也越来越大显身手。对于股票配对交易来说,投资者选择两只具有长期均衡关系的股票,在价格出现短期的偏离时,可以通过融券做空价格偏高的股票,利用自有资金买入价格偏低的股票,当短期的偏离消失即它们的价格恢复均衡水平时,再进行反向平仓,以此来套取利润。配对交易既要做多股票同时也要做空股票,显然这是一种市场中性策略,因此无论市场是牛市还是熊市,投资者都能获得较稳定收入,特别是在熊市下,该策略会表现出很好的收益性,胜于其他类型的对冲基金。

配对的基本理念是均值回复,其产生是因为交易者的非理性行为,当市场中存在大量这样的非理性投资者时,股票价格的变动会产生惯性,可能在一段时间内都会上升,但在人们认清情况之后,发现并没有具体实际的基本因素支撑,这种非理性行为就会消失,从而股票的价格便会降回到原来的水平。同样,价格也可能会在非理性因素的驱使下跌落,然而当市场又回归理性时,价格便又会回复到之间的水平。因此,便有了实施配对交易的机会,在执行配对交易策略时,需要敏锐地发现股票价格偏离正常水平的时点,准确地判断交易时机、确定交易头寸以及实施决策技巧。

对于市场而言,配对交易的应用,可以促使发现更合理的价格,完善价格制定机制,提高市场定价能力,减少非理性行为,提升效率。对投资者而言,配对交易策略可以指导他们如何进行交易,克服非理性因素的影响,长期以往来获利。策略实行时没有太大的风险,但交易的时机很难得,要求投资者有比较高的分析能力。当然,其实质上是统计套利,利用统计学的方法在历史数据上进行研究,并不是无风险投资,需要投资者有一定的风险承受能力。长期来看,配对交易的收益率与股票指数收益率相当,但风险远低于股票,因此配对交易更受投资者的欢迎。

1.2 文献综述

配对交易属于典型的统计套利策略,利用统计学上的知识进行套利,自二十世纪八十年代中期被发现,已经引起了广大学者的注意,广泛地应用在了对冲基金上,对配对交易的研究也越来越多。自2010年来,我国证券交易所开通两融业务,证券市场具备了做空功能,刺激了配对交易的发展,我国学者也重点研究了配对交易策略。

研究配对交易的文章甚多,主要从两方面进行文献综述。首先,对于配对交易策略有不同的方法,介绍了比较著名的方法:协整方法、最小距离法、随机价差法等。然后,重点阐述怎样使用协整法进行配对交易。

1.2.1 配对方法文献综述

(1) 国外文献综述

Vidyamurthy (2004)^[1]观察到两个股票的对数价格可能存在协整关系,便提出了利用协整方法来进行股票的配对,对于符合条件的两只股票,将两个股票对数价格之间的偏离程度称为价差,在价差出现较大变化时,投资者可进入市场进行开仓,待价差回落,反向操作,投资者便可套利。

Gatev, Goetzmann, Rouwenhorst (2006)^[2]采用了最小距离法研究,选取了 1962 年至 2002 年的美国股票市场历史数据。对其标准化后,将序列两两配对作差,再对价差序列求平方和,选择平方和最小的一对股票。最终取得 11% 的年化收益率,说明最小距离法适用于美国的股票交易市场。Muslumov, Yuksel 等 (2009)^[3]使用距离法对伊斯坦布尔证券交易所进行配对交易测试,没有考虑任何限制条件,对前 20 个最佳组合的平均超额回报率为 5.4%,结果在统计意义上是显著的,但缺少对交易和卖空成本的考量,结果对自筹资金投资组合缺少经济意义。

Elliott, Van, Malcolm (2005)^[4]提出了用随机过程预测价差的方法进行配对交易,基于价差序列有均值回复的特点,在两个相似的股票价差出现波动偏离均衡水平时,运用高斯-马尔科夫链模型,可以预测出价差,并与后面的价差观测值相比较,发现合适的交易时机,表明随机价差法在金融市场上有很大的盈利潜力。Mudchanatongsuk, Primbs, Wong (2008)^[5]提出一个随机控制的方法解决配对交易问题,以股票的仿真数据进行说明,将股票价格之间的对数关系假设为一个 Ornstein-Uhlenbeck 过程,以此建立了一个投资组合优化问题,使用哈密顿-雅可比-贝尔曼方程可得到该控制问题的最优解。

Tourin, Yan (2013)^[6]提出了一个基于随机控制方法的模型来研究动态配对交易策略,该模型用来探索最优的投资组合,其中包含一个银行账户和两个相关配对的股票,目标是在固定的时间内实现收益最大化。选取了高盛集团和摩根大通在 2011 年 10 月 17 日精确到分的高频股票数据,研究结果表明该方法能够找到最优解,实现累计利润最大化,损失函数最小化,另外文章也关注了随机控制过程中的不确定性。Bogomolov (2013)^[7]提出了一种新的非参数配对交易方法,可以有效利用交易过程中的可变的统计信息,唯一的假设是价差过程波动率的统计特性保持合理的恒定,实证结果表明每月超额回报率从 1.4% 提高至 3.6%。

Goncu, Akyildirim (2016)^[8]提出了一个最佳配对交易的模型,假设带有噪声项的 OU 过程是一个具有广义双曲线边值的 Lévy 过程,模型显示价差的残差分布存在显著的均值回归、峰值、肥尾效应,使用商品期货数据进行实证,结果表明即使存在成本,模型也有很好的盈利能力。

(2) 国内文献综述

方昊(2005)^[9]对于套利对象的选择,从经济意义上考虑,要有相同的影响因素,这样价格之间才会有趋同性,因此,选择了平均日换手率前十的封闭式基金作为研究对象。此外,在选取合适的股票对时,还要考虑到流动性、相关性和平稳性。最终选择的基金组合都取得了不错的收益,说明统计套利在基金市场上是有效的。仇中群,程希骏(2008)^[10]基于协整理论进行配对交易,利用沪深300股指期货的仿真数据对策略的有效性进行检验,取得了良好的结果,表明配对交易在商品期货中能取得有效的收益。张俊,李妍(2010)^[11]探讨了将交易分为两步,选择股票对和确定交易准则,对交易步骤进行了说明。

崔方达,吴亮(2011)^[12]采用最小距离法,选取了上证50指数的38只股票作为股票池,以400天为形成期,每25天重新配对,计算交易期内的超额收益率。并采用Bootstrap框架产生仿真数据,把随机进入市场的收益与配对交易的收益相比较。结果表明,配对交易策略可以取得收益且不受市场风险影响。胡论超、余乐安,汤铃(2016)^[13]在启动融资融券业务的背景下进行配对交易,将协整法与最小距离法结合起来,先是利用协整法挑选出候选股票,再根据最小距离原则选择最佳的股票对,实证结果显示使用该策略能取得超额收益。

蔡燕,王林,许莉莉(2012)^[14]基于协整理论使用了两种随机价差法,一种是假设价差服从O-U过程,另一种是基于Elliott框架下的价差,通过比较得出O-U过程的价差预测效果是最好的。王春峰,林碧波,朱琳(2013)^[15]根据累计偏差平方和最小的原则,选出股票对进行交易,结果表明取得的月化收益率在1%左右,说明配对交易在我国市场是可以获得稳定收益的。

刘永辉,张帝(2017)^[16]提出基于协整-OU过程的配对交易策略,使用浦发银行和华夏银行的2005年1月4日至2015年4月27日日收盘价数据实证,结果表明策略稳健,能够有较高收益。毕秀春,刘博,袁吕宁等(2019)^[17]假设股票价格服从几何布朗运动,为避免出现超量损失,在策略中引入止损条件,将最优阈值问题转变为最优控制问题,选取了2008年1月2日至2017年10月30日的北京银行和华夏银行历史数据作为样本数据,结果显示最大回撤率有所降低,表明加入止损条件后策略的有效性。

1.2.2 基于协整理论的配对交易文献综述

(1) 国外文献综述

Caldeira, Moura(2013)^[18]选择了圣保罗证券交易所2005年1月至2012年10月的数据进行估计,不仅使用协整方法配对,还根据样本内评估的盈利指标进行选择,结果表明该策略能够取得16.38%超额收益率。Miao(2014)^[19]使用美国股市2012年至2013年的高频股票数据,提出了动态配对交易策略,实证表明带来了超高额收益,

并且在股市表现不佳的时候,配对交易会取得更好的收益,表现该策略的市场中性特点。

Huck, Afawbo (2014)^[20]使用最小距离法和协整法分别对标准普尔 500 指数的成分股进行实证分析,研究不同的策略对交易性能的影响,结果表明在控制风险和交易成本后,距离发产生的收益并不显著,而协整法能够提供一个高、稳定的回报。Mikkelsen (2018)^[21]使用了 2005 年 1 月至 2014 年 12 月的高频每日数据,选取了奥斯陆证券交易所的 18 家海鲜公司股票进行配对交易研究,分别运用最小距离法和协整方法交易并比较了结果。本中还提到,在交易中还要考虑到公司的基本信息,这样才能提高模型的精度,文中两种策略在考虑到交易成本后都没有显著的利润,没有直接比较结果。

Rad, Low, Faff (2016)^[22]选取了 1962 年至 2014 年的整个美国股票市场的历史数据,分别使用距离法、协整法、连接法进行交易,结果显示所有策略都显示出积极的阿尔法值,其中,协整方法在动荡的市场条件下有更好的表现。Fallahpour, Hakimian, Taheri 等 (2016)^[23]将强化学习与协整方法相结合,基于协整理论的基础,使用强化学习研究目标函数,得到使其最大化的最优参数,选取了 2015 年 6 月至 2016 年 1 月的 25 只股票日数据研究,实证结果表明使用强化学习方法能够提高策略的执行效果,收益优于之前的方法。

Krauss (2017)^[24]研究了 100 多篇文献,将研究分为 5 组分别采用不同的方法,其中包含了协整方法,实证结果显示由于计量经济学对均衡关系的可靠性,基于协整方法的框架更加严谨和多样化,并指出多变量增强的策略更容易识别出虚假关系,因此多元统计套利更吸引注意。

Clegg, Krauss (2018)^[25]使用了 1990 年至 2015 年的标准普尔 500 指数里的成分股数据,使用局部协整方法得到合适的似然比检验结果,产生交易信号,结果证明使用部分协整方法扣除成本后取得的年化收益率超过了 12%,优于经典的基于距离或协整的收益。

(2) 国内文献综述

于玮婷 (2011)^[26]基于协整理论进行配对交易,选取融资融券的 90 只标的股票的 2007 年 1 月 1 日至 2009 年 12 月 31 日的股票数据研究,实证结果表明该策略下的收益要好于单纯持有到期的收益,同时风险也比较小,说明该策略的有效性。张河生, 闻岳春 (2013)^[27]运用协整理论研究配对交易策略,选取了 IF1209 与 IF1212 的 2012 年 5 月 4 日 10:06 至 2012 年 5 月 24 日 15:15 的分钟数据,构造交易函数,对里面的参数,在历史经验的基础上不断调优,选择获得最大收益率的参数。经过调参后,发现收益率大大提升,显示出参数调整的重要性。

丁涛 (2013)^[28]在两融业务的背景下,选取了中国银行和中信银行的日收盘数据,

基于经验策略的基础上,对策略改进为延后开仓,实证结果显示年收益率为6%,说明策略的可行性。麦永冠,王苏生(2014)^[29]提出了“王麦-折回首日”建仓策略,分别选择了上证50指数成分股、深证成分指数成分股以及香港恒生指数成分股的日股票数据,使用常见的GCR建仓与Herlem-ont方式分别与文章提出的新策略作比较,结果表明“王麦-折回首日”策略能够取得更好的收益。

邢恩泉,尹涛(2015)^[30]选取了2012年1月4日至2014年4月18日的大连商品交易所的8个商品期货,基于协整理论对配对策略进行改善,使用计算机快速计算的特点,寻找配对的最优组合和交易阈值,最终寻找到的最优组合是豆油和棕榈油,建仓信号为0.6倍的标准差。

欧阳红兵,李进(2015)^[31]选取了84对A+H股进行配对,以3年为样本内长度6个月为样本外长度,采用滑窗方法选取两个研究样本,以协整为基础,分别使用固定参数和时变参数研究策略的最佳阈值,将总利润最大化作为寻找最佳阈值的标准。实证结果显示,该方法是有用的,且固定参数模型的交易表现更好。

吴泽耕(2015)^[32]选取了2010年8月18日至2015年3月16日的银行股票,采用协整方法进行配对交易,对于价差并不仅仅观察其平稳性,还考虑到了对于时间的波动性,采用GARCH模型将这种波动性表示出来,计算出动态的标准差,经过实证表明都可取得较稳定的收益。张可欣(2017)^[33]选取了2011年1月4日至2016年9月30日的民生银行和浦发银行配对,基于协整理论,分别设定1.5倍和0.2倍的标准差作为开仓、平仓阈值,得到配对收益,并与单边股票收益进行对比,结果显示配对的收益好于单边交易。

吴添添(2018)^[34]研究了上证50指数的成分股,选取主要样本股构成一个投资组合模型,与上证50股指期货进行配对,运用协整分析判断上证指数与股指期货是否存在均衡关系,随后讨论了怎样制定交易规则,将协整得到的残差序列根据最大总收益率得到最佳建仓阈值、平仓阈值以及止损阈值,确定建仓时间和平仓时间,用线性回归计算出投资比例。

赵珊珊(2018)^[35]考虑了强化学习算法,采用相关系数法筛选出几对股票,再进行协整关系检验,通过检验后,运用强化学习算法将模型参数改进为动态参数,得出新模型,并与固定参数模型进行比较,得出新模型的表现优于传统的固定参数模型。

冯玉茹(2019)^[36]选择了A股上948只股票作为股票池,选择宁波银行与平安银行进行配对。在制定交易策略时,设置了两种判断方法,一种是传统模型下的交易信号;另一种是GARCH模型下的交易信号。最后进行实证分析,计算出两种方法下的收益率,结果显示两种策略下的收益率都不错,同时也将两种结果对比,发现基于GARCH模型下的策略收益率更好。

1.3 研究内容

随着融资融券业务的不断发展，标的股票也越来越多，本文主要研究配对交易在证券市场的应用，利用统计分析的知识使用该策略进行套利。交易主要分为两步，第一步是选择合适的股票对，它们的价格要存在长期均衡关系，一般来说，这两只股票最好属于同一或相近行业，这表明它们的价格影响因素大部分是相同的，同时两只股票价格之间要有很高的相关性，它们的价格变动也是相似的；第二步是制定详细的交易规则，要发现合适的开仓平仓信号以及对冲比例。

本文主要分为四个章节，每个章节具体内容如下：

第一章绪论，该部分首先介绍了配对交易的含义和我国融资融券业务的发展及现状，投资者可通过融券业务借助配对交易策略在市场中获利；然后分别介绍了国内外配对交易相关研究，了解到配对交易有三种操作方法，最小距离法、随机价差模型法、协整理论方法，而本文研究的是基于协整理论进行配对交易。

第二章配对交易策略理论基础及具体操作方法，先是介绍了配对交易的理论基础，包括经济和数学上的相关知识，介绍完理论后，对交易的具体操作方法进行说明，为实证做准备。

第三章实证分析，经过详细讨论，最终从上交所规定的 800 只融资融券标的股票中选取其中的银行股作为股票池，根据相关性和协整分析，从中选定工商银行和建设银行作为股票对，分别采用固定标准差和时变标准差研究交易信号点，得出两种情况下的收益率并相互比较。

第四章结论与展望，该部分对本文的工作进行总结，并提出不足与展望。

1.4 本文创新点

相比于以往的研究，本文的创新点有：

(1) 2019 年交易所增加了一些标的股票，涉及各个行业，扩充了股票池和配对的选择范围，本文使用最新的标的股票进行实证分析。

(2) 为计算价差的时变标准差，本文运用 GARCH 模型预测条件方差时，在本期结束后，及时更新当期的残差项，进行下一期条件方差的预测，在下一期结束后，重复上述操作，又可进行下下期的预测。在此过程中，没有超过 1 步的预测值，以确保算出的时变标准差能准确地反映价差的波动情况。

(3) 在以往的论文中，寻找交易开仓平仓点时，通常使用静态阈值，以 1 倍标准差作为交易阈值，三倍标准差作为止损阈值。本文并未直接给定交易阈值，而是动态地寻找最佳阈值，在一定范围内，以一定值作为步长，循环得出不同阈值下的收益率，通过比较，将收益率最大值对应的阈值作为交易阈值。对于止损阈值，也不是简

单地使用三倍标准差，而是使用止损比例来控制损失风险，也采取了循环的方法，选取合适的比例。本文考虑动态阈值，其优势是可以灵活改变阈值大小，选择最佳值，但计算复杂度也随之增加，需要的运算时间变长。

第二章 配对交易策略理论基础及具体操作方法

2.1 均值回归理论

均值回归是以正态分布假设为基础,认为事物在长期的变化过程中,总有向均衡位置靠拢的倾向,这种现象是 Francis 在大量的甜豌豆种子实验中发现的,实验发现子豌豆直径的分布比母豌豆直径的分布更为紧凑。均值回归原理在自然领域、社会现象中都有验证,也刺激了各种风险承担和预测理论的产生,在股票投资中,均值回归也有着重要的应用,是金融学上的一个重要理论,股票价格无论是高于或低于价值中枢都会以很大概率回归到中枢附近。也就是说,价格不管是上涨还是下跌都不会一直持续下去,最终一定会回归到价格均值附近,价格上升太快时,就会向均值移动回落;价格下降太快时,就会向均值移动回升。因此,从长期上看,均值回归理论有很强的指导意义。

根据这一原理,在一段时间内,股价如果持续上升或下降出现价格偏离平均值,在未来股价很可能会向均值回归,那么对于投资者来讲,出现价格偏离的那段时间就是最好的交易时机,但在实际中,交易的暴露时间是比较短的,因为市场上大量的投资者都会注意到这个价格偏差,从而导致价格很快会被修正,因此要依靠统计上的技术分析及时获得交易信号点。

对于配对交易来说,选取的两只股票具有长期均衡关系,一般是指它们的价差处于长期稳定的状态。但当其中一只股票受某种因素影响,其价格发生较大的变动,那么两者之间的价差会脱离正常值,平衡会被打破,一方价格相对来讲会高,而另一方价格则偏低。在此时买入价低者卖出价高者,依据均值回归原理,变动方的价格变动状态不会一直持续下去,待其价格回复正常时,即两者的价差又回归正常时,进行反向操作。这样就完整地一次配对交易,以此获取利润。

2.2 相关性分析

进行配对交易,首先就是选择一对合适的股票,两者的价格之间要存在很高的相关性,这样才可能会存在长期的均衡关系。1890 年英国数学家皮尔森设计了相关系数指标,可以衡量两个变量之间的相关程度,将两个股票序列分别记为 x_t, y_t ,它们之间的相关系数记为 γ ,计算公式如下:

$$\gamma = \frac{\sum_1^n (x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_1^n (x_t - \bar{x})^2} \sqrt{\sum_1^n (y_t - \bar{y})^2}}. \quad (2.1)$$

γ 的取值在-1 至 1 之间,其绝对值越接近 1,表示 x_t 和 y_t 有很强的相关性。若 γ 接

近 1, 表示两者有很强的正相关性; 若 γ 接近 -1, 表示两者有很强的负相关性; 若 γ 接近 0, 则表示两者不相关。在本文中, 为了选择合适的股票对, 将股票池里的股票两两配对, 选择相关系数高的进行研究。

2.3 时间序列预处理分析

2.3.1 平稳序列

在时间序列分析里, 平稳性是基本的一个概念, 平稳有两种定义, 按照约束条件的不同要求, 能够分成严平稳和宽平稳。

1、严平稳

严平稳的要求比较严格, 它认为只有当所有序列的统计性质不因时间的平移而改变, 即具有时间平移不变性, 该序列才能被认为是平稳的, 具体定义如下, 设 $\{X_t\}$ 为一时间序列, 对任意正整数 m , 任取 $t_1, t_2, \dots, t_m \in T$, 对任意整数 τ , 有:

$$F_{t_1, t_2, \dots, t_m}(x_1, x_2, \dots, x_m) = F_{t_1, t_2, \dots, t_m}(x_{1+\tau}, x_{2+\tau}, \dots, x_{m+\tau}).$$

则称时间序列 $\{X_t\}$ 为严平稳序列。

严平稳的要求比较高, 不容易直接实现, 在实际研究中经常假定的是平稳性较弱的形式, 即宽平稳性。

2、宽平稳

宽平稳没有那么严格, 对于一个序列, 约束其均值和协方差具有时间平移不变性, 具体定义如下, 如果时间序列 $\{x_t\}$ 满足如下三个条件^[38]:

- (1) 任取 $t \in T$, 有 $Ex_t^2 < \infty$;
- (2) 任取 $t \in T$, 有 $Ex_t = \mu$, μ 为常数;
- (3) 任取 $t, s, k \in T$, 且 $k + s - t \in T$, 有 $\gamma(t, s) = \gamma(k, k + s - t)$ 。

则称时间序列 $\{x_t\}$ 为宽平稳序列, 从上面的约束条件可看出, 宽平稳序列 x_t 的一阶矩、二阶矩都是小于无穷的。从严平稳的定义可知, 对于低阶距存在的严平稳序列 x_t , 那么 x_t 也是宽平稳的。反之, 是不一定成立的, 在实际应用中, 一般考虑宽平稳。

2.3.2 纯随机序列

并不是所有的平稳序列都值得建模, 只有序列之间具有密切的相关关系, 才值得研究, 假使序列值之间不存在任意联系即没有相关性, 这说明该序列不含任何记忆, 过去的行为对将来的发展没有丝毫影响, 含有上述性质的序列就是纯随机序列, 对于理论分析而言, 该序列没有研究价值。

纯随机序列的定义如下, 如果时间序列 $\{x_t\}$ 满足以下性质:

- (1) 任取 $t \in T$, 有 $Ex_t = \mu$;

(2) 任取 $t, s \in T$, 有

$$\gamma(t, s) = \begin{cases} \sigma^2, & t = s \\ 0, & t \neq s \end{cases}.$$

称时间序列 $\{x_t\}$ 为纯随机序列, 也称白噪声序列。

2.4 单位根检验

为了检验序列的平稳性, 学者们发展了多种统计检验方法, 其中用的较多的是单位根检验。通过检验特征根是在单位圆内还是单位圆外, 来判断序列的平稳性, 常用的单位根检验法有 DF 检验、ADF 检验。Dickey 和 Fuller 发明了 DF 检验, 该方法假设时间序列是由 1 阶自回归过程产生, ADF 检验则是将 1 阶扩展至 p 阶, 故又称增广 DF 检验, 文章使用 ADF 方法检验序列平稳性。

对任一 p 阶自回归时间序列:

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t, \quad (2.2)$$

它的特征方程为:

$$\lambda^p - \phi_1 \lambda^{p-1} - \cdots - \phi_p = 0.$$

对于式 (2.2), 若方程全部特征根均位于单位圆内, 即:

$$|\lambda_i| < 1, i = 1, 2, \dots, p,$$

那么序列 x_t 平稳。

如果其中一个特征根落在单位圆上, 不妨设:

$$\lambda_1 = 1,$$

那么序列存在单位根, x_t 非平稳, 此时式 (2.2) 的系数之和为 1:

$$\lambda^p - \phi_1 \lambda^{p-1} - \cdots - \phi_p = 0 \xrightarrow{\lambda=1} 1 - \phi_1 - \cdots - \phi_p = 0 \Rightarrow \phi_1 + \cdots + \phi_p = 1.$$

通过上面的分析, 要想检验 p 阶自回归序列的平稳性, 可以考察自回归系数的和是否为 1, 若为 1, 表示不平稳。

为便于检验, 对式 (2.2) 进行等价变换, 得

$$\nabla x_t = (\phi_1 + \cdots + \phi_p - 1)x_{t-1} - (\phi_2 + \cdots + \phi_p)\nabla x_{t-1} - \cdots - \phi_p \nabla x_{t-p+1} + \varepsilon_t,$$

简记为:

$$\nabla x_t = \rho x_{t-1} - \beta_1 \nabla x_{t-1} - \cdots - \beta_{p-1} \nabla x_{t-p+1} + \varepsilon_t, \quad (2.3)$$

式 (2.3) 中:

$$\rho = \phi_1 + \cdots + \phi_p - 1,$$

$$\beta_j = -\phi_{j+1} - \cdots - \phi_p, j = 1, 2, \dots, p-1.$$

若序列平稳, 有:

$$\phi_1 + \cdots + \phi_p < 1,$$

等价于 $\rho < 0$ 。

若序列不平稳，必有单位根，由上面的分析知：

$$\phi_1 + \cdots + \phi_p = 1.$$

等价于 $\rho = 0$ 。

原假设、备择假设可以写为：

$$H_0: \rho = 0 \leftrightarrow H_1: \rho < 0.$$

构建如下检验统计量：

$$\tau = \frac{\hat{\rho}}{S_{(\hat{\rho})}}. \quad (2.4)$$

其中， $S_{(\hat{\rho})}$ 为参数 ρ 的样本标准差，通过蒙特卡洛方法，可以得到 τ 统计量临界值表。

由于序列非平稳的原因有很多，判断序列是否平稳时，对于不同构成成分的非平稳序列来说，通常将其分为三大类型。

第一种是无常数、无趋势的 $AR(p)$ 过程：

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t,$$

第二种是有常数、无趋势的 $AR(p)$ 过程：

$$x_t = \mu + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t,$$

第三种是既有常数又有线性趋势的 $AR(p)$ 过程：

$$x_t = \mu + \beta t + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t.$$

通常，对序列用 ADF 检验法分析时，会将上面三种类型的检验结果求出来，只要有一个结果是拒绝原假设的，就能判断没有单位根，序列平稳；反之，只有全部检验结果均没有拒绝原假设，序列才被认为是非平稳的。

2.5 协整

2.5.1 协整概念

在实际应用中，金融时间序列通常是不平稳的，虽然有些序列与序列之间有着密切的长期均衡关系，但为了防止出现虚假回归问题，不能轻易使用经典线性回归模型，虽可以对序列差分得到平稳的序列，但在差分过程中可能会丢失原本的信息，模型估计效果会受到影响。为了研究非平稳序列之间的有效信息，Engle 和 Granger 提出了协整概念，记自变量序列为 $\{x_t\}$ ，响应变量序列为 $\{y_t\}$ ，构造回归模型：

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t.$$

若回归模型的残差序列 ε_t 是平稳的，可认为 y_t 与 x_t 存在协整关系。其中有一个前

提, 对于序列 x_t 与 y_t , 虽然不要求它们是平稳的, 但是要求它们都是 d 阶单整, 即序列里至少经过 d 阶差分才能平稳, 记为 $x_t \sim I(d), y_t \sim I(d)$ 。

2.5.2 协整检验

要想对非平稳序列建立线性方程, 变量之间必须存在协整关系, 否则得出的回归模型可能是虚假的, 因此要进行协整检验, 记为 EG 检验, 防止出现虚假模型。

一般情况下, 我们想得到的结果是非平稳序列间具有协整关系, 故 EG 检验的原假设、备择假设分别为:

H_0 : 非平稳序列之间不存在协整关系

H_1 : 非平稳序列之间存在协整关系

从回归残差的平稳性可看出序列间是否存在协整关系, 所以上述假设等价于:

H_0 : 回归残差序列 ε_t 非平稳

H_1 : 回归残差序列 ε_t 平稳

进行检验前, 先要检验原序列的单整性, 必须确定他们都是同阶单整才能进行下面的步骤。EG 检验依照如下两个步骤进行, 步骤一, 建立响应变量与输入变量的回归模型:

$$y_t = \hat{\beta}_0 + \hat{\beta}_1 x_t + \varepsilon_t. \quad (2.5)$$

式(2.5)中, $\hat{\beta}_0, \hat{\beta}_1$ 是最小二乘估计值;

步骤二, 检验残差序列 ε_t 平稳性, 若残差检验平稳, 则序列间存在协整关系, 若残差检验不平稳, 则序列间不存在协整关系。

若通过了 EG 检验, 说明序列间存在协整关系, 具有长期均衡关系, 此时可以建立经典回归模型来描述均衡关系。

2.6 误差修正模型

协整衡量的是序列间长期均衡状态, 但在现实情况中, 往往存在短期波动, 便产生了误差修正模型, 主要形式于 1978 年由 Davidson 和 Hendry 等提出, 常常作为补充协整模型, 描述短期的非均衡状态。

模型假定变量 y 和 x 具有协整关系, 但 y 和 x 在短期往往不均衡, 可以采用自回归分布滞后模型来描述, 设分布形式为(1, 1)阶:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \mu y_{t-1} + \delta_t,$$

上面的模型表明当期的 y 值, 不仅与当期的 x 值有关, 和上一期的 y 、 x 值均有关系, 对上式变形可得:

$$\nabla y_t = \beta_1 \nabla x_t - \lambda(y_{t-1} - \alpha_0 - \alpha_1 x_{t-1}) + \delta_t, \quad (2.6)$$

其中:

$$\lambda = 1 - \mu, \quad \alpha_0 = \beta_0 / (1 - \mu), \quad \alpha_1 = (\beta_1 + \beta_2) / (1 - \mu).$$

其中 $(y_{t-1} - \alpha_0 - \alpha_1 x_{t-1})$ 项就是上一期的非均衡误差, 用 ecm 表示误差项, 式(2.6)可以简记为:

$$\nabla y_t = \beta_1 \nabla x_t - \lambda ecm_{t-1} + \delta_t.$$

这说明本期波动 ∇y_t 主要受三方面短期波动的影响: 输入序列的当期波动 ∇x_t ; 上一期的误差 ecm_{t-1} ; 当期纯随机误差 δ_t 。 $-\lambda$ 为误差修正项系数, 一般情况下 $|\mu| < 1$, 由 $\lambda = 1 - \mu$ 可以确定 $\lambda > 0$, 表明 ecm_{t-1} 对本期波动 ∇y_t 修正水平, 当 $ecm_{t-1} > 0$, 即上期真实值比估计值大时, $-\lambda ecm_{t-1}$ 会使 ∇y_t 减少, 将 y_t 拉向均衡水平; 当 $ecm_{t-1} < 0$, 即上期真实值比估计值小时, $-\lambda ecm_{t-1}$ 会使 ∇y_t 增加, 将 y_t 拉向均衡水平。

2.7 条件异方差模型

根据误差修正模型可得到协整向量, 依照其中的参数可求出价差, 对于价差的研究本文主要使用两种方法, 第一种是计算其样本标准差, 第二种是计算其时变方差, 分别用这两种标准差来研究怎样发现交易信号点。对于样本标准差, 就是利用样本内的数据求出标准差, 以此作为整体标准差, 这其实是假设价差序列的方差是不变的, 但是在金融时间序列中, 序列的标准差往往会呈现一定的波动性, 有一定的集群效应, 在某些时段的序列波动持续偏大, 在某些时段的序列波动持续偏小, 并不满足方差齐性, 因此提出条件异方差模型用于解决异方差问题。

对于要研究的时间序列 r_t , 可以用简单时序模型表示, 假设其服从平稳 ARMA(p, q)过程, 可写为:

$$r_t = \mu_t + \varepsilon_t, \quad (2.7)$$

其中, $\mu_t = \phi_0 + \phi_1 r_{t-1} + \cdots + \phi_p r_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$ 。

考虑 t 时刻已获取历史信息 F_{t-1} , 特别地 F_{t-1} 包含过去的一切线性函数, 那么 r_t 的条件均值 μ_t 和条件方差 σ_t^2 可表示为:

$$\begin{aligned} \mu_t &= E(r_t | F_{t-1}), \\ \sigma_t^2 &= \text{Var}(r_t | F_{t-1}) = E[(r_t - \mu_t)^2 | F_{t-1}], \end{aligned} \quad (2.8)$$

结合公式(2.7)和(2.8), 有:

$$\sigma_t^2 = \text{Var}(r_t | F_{t-1}) = \text{Var}(\varepsilon_t | F_{t-1}). \quad (2.9)$$

由式(2.9)可知原序列的条件方差和残差的条件方差是相等的, 对原序列条件方差的研究可以转为对残差序列条件方差的研究。 σ_t^2 的模型称为 r_t 的波动率方程, 条件异方差模型就是描述 σ_t^2 的演变。

2.7.1 ARCH 检验

条件异方差模型可分为两种：第一种是用确定的函数来描述方差的演变，如 ARCH 模型、GARCH 模型；第二种是用随机方程来描述，如随机波动率模型。

本文使用第一种方法描述方差的演变，在使用具体模型前，先要进行 ARCH 效应检验，只有检验结果显著时，才可以拟合 ARCH 模型。这是一种特殊的异方差检验，不仅要求序列具有异方差性，而且是由某种自相关性造成的异方差，本文利用 LM 方法进行 ARCH 检验。

LM 检验是由 Engle 在 1982 年提出，其构造思想是：对有着集群效应的残差序列来讲，通常残差平方序列存在自相关，那么就能使用自回归模型拟合残差平方：

$$\varepsilon_t^2 = \omega + \sum_{j=1}^q \alpha_j \varepsilon_{t-j}^2 + e_t.$$

这样，可通过检验上述方程的显著性来考察残差序列的方差齐性。

如果方程成立，必有参数 $\alpha_j \neq 0$ ，意味着残差平方序列具有自相关性，可得到自相关信息；反之不成立，则所有参数 $\alpha_j (j = 1, \dots, q)$ 均为零，表示残差平方序列没有自相关性。所以，LM 检验实际上是对残差平方序列 ε_t^2 自回归方程的显著性检验，假设条件为：

$$H_0: \text{残差平方序列纯随机} \leftrightarrow H_1: \text{残差平方序列具有自相关性}$$

对残差平方序列构造 q 阶自回归方程，假设条件等价于：

$$H_0: \lambda_1 = \lambda_2 = \dots = \lambda_q = 0 \leftrightarrow H_1: \lambda_1, \lambda_2, \dots, \lambda_q \text{ 不全为零}$$

记总误差平方和为 $SST = \sum_{t=q+1}^T \varepsilon_t^2$ ，自由度为 $T-q-1$ ，回归平方和为 $SSR = SST - SSE$ ，自由度为 q ， SSE 为回归方程残差平方和， $SSE = \sum_{t=q+1}^T e_t^2$ ，自由度为 $T-2q-1$ ，则 LM 检验统计量为：

$$LM(q) = \frac{(SST - SSE)/q}{SSE/(T - 2q - 1)}.$$

原假设成立时， $LM(q)$ 统计量近似服从自由度为 $q-1$ 的 χ^2 分布。当统计量的 P 值小于给定的显著性水平，认为该残差序列方差非齐，残差平方序列可用自回归模型表示。

2.7.2 GARCH 模型

通过 ARCH 检验后，可以构造 ARCH(q) 模型，模型的全称是自回归条件异方差模型，可表达为：

$$\varepsilon_t = \sigma_t a_t,$$

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2.$$

其中 $a_t \sim^{i.i.d} N(0,1)$ ， $\omega > 0$ ， $\alpha_j \geq 0 (j = 1, 2, \dots, q)$ ，本质是用移动平均模型拟合方

差, 由于 MA 模型自相关系数 q 阶截尾, 故 ARCH 模型适用短期自相关。

在有些问题中, 残差序列的方差是长期自相关的, 如果使用 ARCH 模型拟合, 会估计出较高的 q 值, 这将增加模型计算难度还可能影响精度。因此, 提出了 GARCH 模型, 在 ARCH 模型的基础上, 增加考虑了异方差函数的 p 阶自相关性而形成的, 模型可表示为:

$$\varepsilon_t = \sigma_t a_t,$$

$$\sigma_t^2 = \omega + \sum_{j=1}^q \alpha_j \varepsilon_{t-j}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2.$$

其中, $\omega > 0$, $\alpha_j, \beta_i \geq 0 (j = 1, 2, \dots, q, i = 1, 2, \dots, p)$, $\sum_{j=1}^q \alpha_j + \sum_{i=1}^p \beta_i < 1$, 它可以有效地拟合具有长期记忆性的异方差函数, 显然 ARCH 模型就是 GARCH 模型的一个特例, ARCH(q)模型实际上就是 $p=0$ 的 GARCH(p, q)模型。

在实际应用中, 只用到低阶的 GARCH 模型, 如 GARCH(1,1)、GARCH(1,2)、GARCH(2,1)模型等, 在本文中, 使用最常见且最多使用的 GARCH(1,1)模型:

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \quad (2.10)$$

其中, $0 \leq \alpha_1, \beta_1 < 1, (\alpha_1 + \beta_1) < 1$, 对此模型的前一步预测, 假定 h 为预测原点, 有:

$$\sigma_{h+1}^2 = \omega + \alpha_1 \varepsilon_h^2 + \beta_1 \sigma_h^2,$$

其中, ε_h 和 σ_h^2 在时间指标为 h 时是已知的, 因此向前一步预测为:

$$\sigma_h^2(1) = E(\sigma_{h+1}^2 | F_h) = \omega + \alpha_1 \varepsilon_h^2 + \beta_1 \sigma_h^2, \quad (2.11)$$

对向前多步预测, 用 $\varepsilon_t = \sigma_t a_t$ 把 (2.10) 改写成:

$$\sigma_{t+1}^2 = \omega + (\alpha_1 + \beta_1) \sigma_t^2 + \alpha_1 \sigma_t^2 (a_t^2 - 1),$$

当 $t=h+1$ 时, 此方程变为:

$$\sigma_{h+2}^2 = \omega + (\alpha_1 + \beta_1) \sigma_{h+1}^2 + \alpha_1 \sigma_{h+1}^2 (a_{h+1}^2 - 1),$$

因为 $E(a_{h+1}^2 - 1 | F_h) = 0$, 故以 h 为预测原点的波动率向前 2 步预测满足:

$$\sigma_h^2(2) = \omega + (\alpha_1 + \beta_1) \sigma_h^2(1),$$

一般地, 我们有

$$\sigma_h^2(l) = \omega + (\alpha_1 + \beta_1) \sigma_h^2(l-1), l > 1,$$

对上式重复迭代, 可得到前行 l 步预测能写成:

$$\sigma_h^2(l) = \omega \frac{1 - (\alpha_1 + \beta_1)^{l-1}}{1 - (\alpha_1 + \beta_1)} + (\alpha_1 + \beta_1)^{l-1} \sigma_h^2(1).$$

考虑 GARCH(1,1)模型, 若 σ_1^2 是已知的, 则 GARCH(1,1)模型的 σ_t^2 可以递推得到, 在一些应用中, 将 ε_t 的样本方差作为 σ_t^2 的初始值比较好。根据上面的公式, 就可以将条件标准差递推出来。

本文里,对于价差序列要想研究其波动性,先根据其自身的信息建立一个 ARMA 模型,消除线性依赖,接着检验残差序列的异方差性,若 ARCH 检验在统计上显著,则用波动率模型表示条件方差随时间的变化。

对价差序列建立一个波动率模型需要以下 4 个步骤:

- (1) 分析序列的各种性质,构建计量经济模型,本文选择 ARMA(p,q)模型;
- (2) 利用 ARCH 检验考察模型残差的异方差性;
- (3) 若残差存在异方差性,需指定一个波动率模型,对 ARMA(p,q)模型和波动率模型进行联合估计;
- (4) 对得到的模型进行检验。

2.8 配对交易策略具体操作方法

上面介绍了配对交易涉及的理论知识,下面介绍实证时配对交易的具体操作步骤,主要分为两步,第一步是选择合适的股票对,第二步是确定交易信号。

2.8.1 选择股票对

对股票池里的股票两两配对并且计算它们的相关系数,相关系数越接近于 1 表明股票之间有很高的相关性,考虑相关系数较高的股票对为研究对象,选择相关系数最高的一组股票拟作为交易对象。

仅有相关性是不够的,下面还要运用统计学知识判断它们的价格是否具有协整关系,只有具有协整关系才能确定两者存在长期均衡关系,适合配对交易。所以,两个股票价格序列只有通过协整检验,才能进行配对交易。

2.8.2 确定交易策略

通过协整建立的模型是静态的,我们可以使用误差修正模型弥补长期静态模型的不足。通过误差修正模型,可以建立短期的动态模型,判断变量在短期波动中偏离其长期均衡关系的程度,且模型中误差项的系数也反映了短期波动偏离长期均衡时的调整力度。

根据误差项表达式 $ecm_{t-1} = y_{t-1} - \alpha_0 - \alpha_1 x_{t-1}$,可知协整向量为 $(1, -\alpha_1)$,在配对交易中,两个股票的对冲比例正是误差项里的系数 α_1 ,即当买入 1 头寸的某种股票时,就要卖出与之配对的 α_1 头寸股票,反之亦然。由此,得到了两只股票之间的价格关系,将两只股票的价差记为 $spread_t$,计算公式为: $spread_t = y_t - \alpha_1 x_t$,接下来计算价差的标准差。

由上可知,两股价间存在长期的均衡关系,价差序列应该是平稳的,对价差进行单位根检验证明其是平稳序列。为方便分析,可以求出价差序列的均值,将价差序列

去中心化得到序列 $mspread_t$ ，使去中心化后的序列均值为零，计算公式为 $mspread_t = spread_t - \text{mean}(spread_t)$ 。主要采用两种方法计算价差序列的标准差，一种是计算样本内的样本标准差，另一种是建立波动率模型计算时变标准差，可看出第一种是对固定标准差进行研究，第二种是对随时间变化的时变标准差进行研究。

得到标准差后，就可以用来发现信号交易点了。一般来说，使用标准差的 w 倍作为交易阈值，当去中心化后的价差绝对值大于交易阈值时，就出现了交易机会。记去中心化后的价差记为 $mspread_t$ ，标准差记为 σ ，交易情况可分为两种，具体规则如下：

(1) 当 $mspread_t > w * \sigma$ 时，进行开仓操作卖出价格高的买入价格低的，当 $mspread_t < 0$ 时，进行反向平仓；(2) 当 $mspread_t < -w * \sigma$ 时，进行开仓操作卖出价格高的买入价格低的，当 $mspread_t > 0$ 时，进行反向平仓。在对样本内数据的研究中，带入不同的 w ，得到不同阈值下的收益率，选取收益最高的对应的阈值。

根据分析，理想的情况是当去中心化后的价差突破设定的阈值时，能够等到其回落到零附近，但在实际的交易市场中，充满各种未知性，可能价差会越来越远离阈值，在这段期间内，投资者可能承受的损失会很大，即使未来差价会回落，但有的投资者会因承受不了巨大的损失而提前平仓。因此，为了控制交易的亏损风险，需再设定一个止损点，一旦亏损大于设定值时，就要及时平仓。本文使用的是最简单的止损方法：定额止损法，将亏损额设置成一个固定的比例，在开仓后的等待期间，一旦损失大于固定比例就要强制平仓。

2.8.3 总收益率

在交易期间会进行多次配对交易，每一次交易完，即可计算出本次配对的收益率 r_i ，暂时不考虑交易成本， r_i 的计算公式为：

$$r_i = \frac{M_1 - M_2 + M_3 - M_4}{M},$$

其中， M_1 、 M_3 分别代表开仓、平仓时做空的金额， M_2 、 M_4 分别代表开仓、平仓时做多的金额， M 表示每次配对交易前的总金额，每完成一次配对交易，都要及时更新总金额 M 。有了单次的收益率，可以计算总体的收益率，计算公式为：

$$r_{total} = (1 + r_1)(1 + r_2) \cdots (1 + r_k) - 1. \quad (2.12)$$

其中， r_{total} 表示交易期间的总收益率， k 表示配对的次数， $r_i (i = 1, 2, \dots, k)$ 表示每次配对后的收益率。总收益率可以用来衡量交易策略的收益情况，其值自然是越大越好，在交易阈值的选取上，就是根据总收益率最高的原则来选择的。

第三章 实证分析

在进行配对交易时，一般通过融券做空其中一只股票，因此选择要选择融资融券标的的股票进行研究。上交所侧重国有中大型企业，而深交所侧重创投和中小企业，显然上交所的优质股更多，所以选择上交所的标的股票。将同行业的股票进行配对，不仅股票之间的相关性高，所取得的收益也高。上交所融资融券业务的 800 只标的股票中，主要的行业有银行业、房地产、证券期货业、医药制造业、有色金属等。历史经验表明，银行业股票之间的相关性相比于其它行业较高，且配对的效果也好，因此选择了银行业作为研究对象。

从标的中选择了 24 只银行业股票作为股票池，研究选取的期间为 2017 年至 2019 年。其中南京银行、北京银行分别在 2017 年 7 月 31 日、2017 年 4 月 14 日停牌，西安银行、长沙银行、成都银行、紫金银行分别于 2019 年 3 月 1 日、2018 年 9 月 26 日、2018 年 1 月 31 日、2019 年 1 月 3 日上市，这六只股票数据部分缺失，将它们从股票池中去掉。此时，股票池中还有 18 只银行股，选择 2017 年 1 月 3 日至 2019 年 12 月 13 日的历史收盘价作为研究对象。其中，2017 年 1 月 3 日至 2018 年 12 月 28 日作为样本内数据，用来获取交易参数，2019 年 1 月 2 日至 2019 年 12 月 13 日作为样本外数据，用于实施得到的交易策略。

本文使用前复权收盘价，前复权是对股票价格进行修正，将因除权除息造成的价格突变消除掉，以现有价位为基准将以前的价格复权，可反映历史价格在现阶段的真实成本。在金融时间序列分析中，对价格取自然对数可以一定程度上减少模型残差的异方差性和波动性，提高预测精度，故在下面的研究中，使用的均为取对数后的股票前复权收盘价数据。实证数据来源于 WIND 数据库，使用的统计软件是 R 语言。

3.1 相关性分析

选择好的股票池里有 18 只股票，分别是：浦发银行（600000.SH）、华夏银行（600015.SH）、民生银行（600016.SH）、招商银行（600036.SH）、江苏银行（600919.SH）、杭州银行（600926.SH）、常熟银行（601128.SH）、兴业银行（601166.SH）、上海银行（601229.SH）、农业银行（601288.SH）、交通银行（601328.SH）、工商银行（601398.SH）、光大银行（601818.SH）、建设银行（601939.SH）、中国银行（601988.SH）、贵阳银行（601997.SH）、中信银行（601998.SH）、苏农银行（603323.SH）。

由前面的分析可知，配对交易的第一步就是选择一对股票，首先计算它们的相关系数矩阵，绘制出相关系数矩阵的热力图：

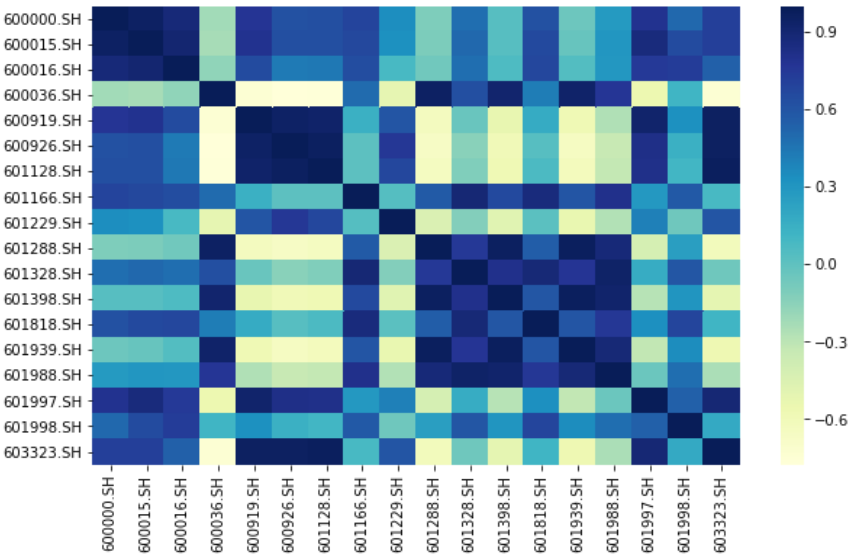


图 3.1 相关系数矩阵的热力图

其中，颜色越深表示相关系数越接近 1，颜色越浅表示相关系数越接近-1，从图 3.1 可看出除了对角线还有其它颜色很深的方格，这表示有一部分股票两两之间存在很高的正相关性，下表列出相关系数最高的前十组：

表 3.1 相关系数最高的前十组

银行名称	银行名称	相关系数
工商银行	建设银行	0.9775
农业银行	建设银行	0.9757
常熟银行	苏农银行	0.9728
杭州银行	常熟银行	0.9667
农业银行	工商银行	0.9665
江苏银行	苏农银行	0.9627
杭州银行	苏农银行	0.9589
浦发银行	华夏银行	0.9564
招商银行	农业银行	0.9521
江苏银行	杭州银行	0.9446

从表 3.1 可看出这十组彼此之间有很高的相关性，相关系数均在 0.9 以上，前三组之间的相关系数更是都在 0.97 以上。选择相关系数最高的一组：工商银行和建设银行，将它们配成一对，记工商银行对数价格序列为 $\{lny_t\}$ ，建设银行对数价格为

$\{lnx_t\}$ ，画出两个对数价格序列在 2017 年 1 月 3 日至 2018 年 12 月 28 日的时间序列图：



图 3.2 工商银行和建设银行对数价格时间序列图

其中，虚线表示工商银行对数价格，实线表示建设银行对数价格。图 3.2 表明在样本内期间，两只股票的对数价格的波动趋势非常相似，选择它们进行配对应该是不错的决定。下面研究对数价格之间的协整关系，判断两者是否存在长期均衡关系。

3.2 协整分析

在第二章的理论部分提到协整关系的前提是两个序列都是同阶单整，因此在对工商银行和建设银行之间进行协整研究前，要分别检查这两只股票的平稳性。采用 ADF 检验方法，分别采用三种类型的单位根检验，检验结果在下表列出，其中，‘NC’表示无常数均值，无截距类型，‘C’表示有常数均值，无截距类型，‘CT’表示有常数均值，有截距类型，且下列检验中延迟阶数均为 1：

表 3.2 lny 和 lnx 的 ADF 检验结果

变量	检验类型	ADF 值	1%临界值	5%临界值	10%临界值	P 值	平稳性
	型		值	值	值		
lny	NC	0.6360	-2.5800	-1.9500	-1.6200	0.8191	不平稳
	C	-2.1684	-3.4400	-2.8700	-2.5700	0.2501	不平稳
	CT	-1.6268	-3.9800	-3.4200	-3.1300	0.7361	不平稳

续表 3.2

	NC	0.4979	-2.5800	-1.9500	-1.6200	0.7750	不平稳
lnx	C	-2.1478	-3.4400	-2.8700	-2.5700	0.2578	不平稳
	CT	-1.6576	-3.9800	-3.4200	-3.1300	0.7230	不平稳

由表 3.2 可知, 对于变量 lny , 三种检验结果的 ADF 值均大于 1%、5%、10%水平下的临界值, 在显著性水平取 5%时, 相应的 P 值都大于 5%, 所以可以认为序列 lny 不平稳, 即工商银行对数价格序列不平稳。对于变量 lnx , 同上分析, 可知建设银行对数价格序列不平稳。分别对序列 lny 、 lnx 作差分, 记差分后的序列为 dif_lny 、 dif_lnx , 再分别用三种类型对差分后的序列做单位根检验, 结果如下表:

表 3.3 lny 和 lnx 差分后序列的 ADF 检验结果

变量	检验类型	ADF 值	1%临界值	5%临界值	10%临界值	P 值	平稳性
	NC	-14.2309	-2.5800	-1.9500	-1.6200	0.0100	平稳
dif_lny	C	-14.2457	-3.4400	-2.8700	-2.5700	0.0100	平稳
	CT	-14.3265	-3.9800	-3.4200	-3.1300	0.0100	平稳
	NC	-14.727	-2.5800	-1.9500	-1.6200	0.0100	平稳
dif_lnx	C	-14.7319	-3.4400	-2.8700	-2.5700	0.0100	平稳
	CT	-14.8061	-3.9800	-3.4200	-3.1300	0.0100	平稳

由表 3.3 的检验结果知, 对于变量 dif_lny , 三种检验结果的 ADF 值均小于 1%、5%、10%水平下的临界值, 且相应的 P 值也都是小于显著性水平 5%, 所以拒绝原假设, dif_lny 序列不存在单位根, 是平稳的, 即工商银行对数价格差分序列是平稳的, 所以原序列是一阶单整, 有 $lny_t \sim I(1)$ 。对于变量 dif_lnx , 同上分析可知建设银行对数价格序列也是一阶单整, 有 $lnx_t \sim I(1)$ 。由上面的分析可知序列 lny 、 lnx 都是一阶单整的, 满足协整检验的前提, 接下来进行 EG 两步检验法。

首先拟合线性回归方程, 模型参数估计结果如下:

表 3.4 线性回归模型参数估计结果

	估计值	T 值	P 值
截距项	-0.1447	-8.3110	9.6e-16
系数项	0.9659	101.9770	<2e-16

根据表 3.4 的回归结果, 截距项和系数项对应的 P 值均小于 1%显著性水平, 因此这两个估计值都显著不为零, 线性回归方程可表示为:

$$\ln y_t = -0.1447 + 0.9659 \ln x_t + e_t.$$

其次, 用估计出的回归方程计算出残差序列记为 e , 对残差序列进行平稳性检验, 结果如下:

表 3.5 残差序列 e 的 ADF 检验结果

变量	检验类型	ADF 值	1%临界值	5%临界值	10%临界值	P 值	平稳性
			值	值	值		
	NC	-2.2429	-2.5800	-1.9500	-1.6200	0.0245	平稳
e	C	-2.2401	-3.4400	-2.8700	-2.5700	0.2233	不平稳
	CT	-2.2708	-3.9800	-3.4200	-3.1300	0.4636	不平稳

由表 3.5 知, 在运用无常数均值、无趋势类型模型对残差序列进行 ADF 检验时, 其 ADF 值大于 1%水平下的临界值, 小于 5%、10%水平下的临界值, 故在 5%的显著性水平下拒绝原假设, 残差序列不存在单位根, 所以序列是平稳的。

由上面两步可得到工商银行和建设银行对数价格是存在协整关系的, 即两对数价格之间存在长期的均衡关系, 可以用来配对交易。

3.3 误差修正模型

从上面的协整分析知, 工商银行对数价格和建设银行对数价格之间存在长期的均衡关系, 然而通过协整建立的模型是静态的, 使用误差修正模型则可以建立短期的动态模型, 考虑到两个对数价格序列均是一阶单整的, 建立 (1, 1) 阶分布滞后模型, 模型可以表示为:

$$\ln y_t = \beta_0 + \beta_1 \ln x_t + \beta_2 \ln x_{t-1} + \mu \ln y_{t-1} + \varepsilon_t,$$

对上式进行差分处理, 得到下面变形:

$$\nabla \ln y_t = \beta_0 + \beta_1 \nabla \ln x_t + (\beta_1 + \beta_2) \ln x_{t-1} + (\mu - 1) \ln y_{t-1} + \varepsilon_t,$$

对上式进行线性回归, 估计结果中截距项对应的 P 值为 0.9783, 不拒绝为零的原假设, 故将上面式子中截距项去掉, 变为:

$$\nabla \ln y_t = \beta_1 \nabla \ln x_t + (\beta_1 + \beta_2) \ln x_{t-1} + (\mu - 1) \ln y_{t-1} + \varepsilon_t,$$

对模型进行参数估计, 结果如下:

表 3.6 误差修正模型参数估计结果

	估计值	T 值	P 值
β_1	0.8052	47.8700	$<2e-16$
$\beta_1 + \beta_2$	0.0188	2.3570	0.0188
$\mu - 1$	-0.0211	-2.3480	0.0193

由表 3.6 可知, 几个系数估计结果对应的 P 值都是小于 5% 的, 在 5% 的显著性水平下显著不为零, 故模型可表示为:

$$\nabla \ln y_t = 0.8052 \nabla \ln x_t + 0.0188 \ln x_{t-1} - 0.0211 \ln y_{t-1} + \varepsilon_t,$$

可变形为:

$$\nabla \ln y_t = 0.8052 \nabla \ln x_t - 0.0211 (\ln y_{t-1} - 0.8910 \ln x_{t-1}) + \varepsilon_t.$$

其中, 误差修正项的系数为 -0.0211, 表明当实际情况与均衡状态发生偏离时, 短期偏离会以 2.11% 幅度反向调整, 使其向长期均衡状态回复。可以得到误差修正项表达式 $ecm_{t-1} = \ln y_{t-1} - 0.8910 \ln x_{t-1}$, 协整向量为 (1, -0.8910), 由此可以确定两只股票开仓的对冲比例, 当买入 1 头寸的工商银行股票时, 就要卖出 0.8910 头寸的建设银行, 反之亦然。

工商银行和建设银行对数价格之间的均衡关系可表示为:

$$spread_t = \ln y_t - 0.8910 \ln x_t.$$

由上面的分析知两个对数股价是有着长期均衡关系的, 误差修正模型是对协整模型的补充, 因此, 价差序列应该是平稳的, 对其作单位根检验, 结果如下:

表 3.7 价差序列 $spread$ 的 ADF 检验结果

变量	检验类型	ADF 值	1%临界值	5%临界值	10%临界值	P 值	平稳性
	NC	-2.2066	-2.5800	-1.9500	-1.6200	0.0271	平稳
$spread$	C	-2.1499	-3.4400	-2.8700	-2.5700	0.2570	不平稳
	CT	-2.1413	-3.9800	-3.4200	-3.1300	0.5184	不平稳

由表 3.7 知, 在运用无常数均值、无趋势类型模型对价差序列进行 ADF 检验时, 其 ADF 值小于 5%、10% 水平下的临界值, 故在 5% 的显著性水平下拒绝原假设, 序列不存在单位根, 所以价差序列是平稳的。

为了便于研究, 将价差序列去中心化, 即减去其均值 $mean_{spread}$, 得到序列 $mspread_t$, 其中:

$$mspread_t = spread_t - mean_{spread},$$

$$mean_{spread} = mean(spread_t).$$

注意，这里的均值 $mean_{spread}$ 是样本内价差的均值，为-0.0074，下标 t 的范围在2017年1月3日至2018年12月28日。

本文使用两种方法计算标准差，使用不同的标准差研究交易阈值、止损比例及收益情况。第一种是计算固定标准差，第二种是计算时变标准差，比较两种方法下的总收益率大小。

3.4 固定标准差

对去中心化后的价差序列求样本标准差 $\sigma_0 = 0.0286$ ，由前面得出两只股票的对冲比例为0.8910，将开仓阈值记为 $w * \sigma_0$ ，制定交易规则如下：（1）当 $mspread_t > w * \sigma_0$ 时开仓，卖出1个头寸的工商银行股票，买入0.8910个头寸的建设银行股票，在 $mspread_t < 0$ 时反向平仓；（2）当 $mspread_t < -w * \sigma_0$ 时开仓，买入1个头寸的工商银行股票，卖出0.8910个头寸的建设银行股票，在 $mspread_t > 0$ 时反向平仓。

考虑到不同的阈值会选出不同的开仓点，这也会影响收益情况，将 w 的值设定在 $(0, 15)$ 范围内，每次步长为0.0001，根据公式(2.12)，计算出不同阈值下的总收益率，将开仓阈值与对应的总收益率作图，如下：

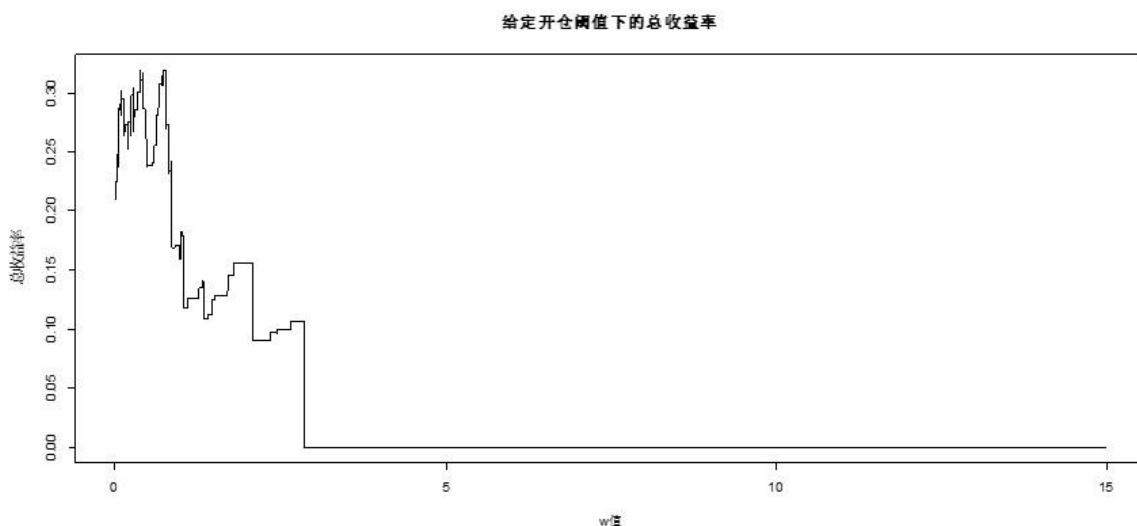


图 3.3 给定开仓阈值下的总收益率

其中，横坐标表示样本标准差的 w 倍，纵坐标表示特定阈值对应的总收益率，从图 3.3 可看出，在开仓阈值大于 $5\sigma_0$ 后，收益率恒为 0，这说明超出一定范围时，就不存在交易机会，收益率自然为 0。因此，就在 $(0, 15\sigma_0)$ 的范围内寻找使总收益率

最大的交易值，求解出最优阈值为 $0.7327\sigma_0$ ，其对应的总收益率为 31.95%。

3.4.1 样本内实证

将开仓阈值设置为 $0.7327\sigma_0$ ，在样本内期间利用上面的交易策略进行交易，先找出交易点，随后再计算收益，画出了开仓点平仓点位置，其中，实线表示价差序列，虚线表示开仓阈值即 $\pm 0.7327\sigma_0$ ，空心点表示开仓点，红色实点表示平仓点，交易情况如图 3.4:

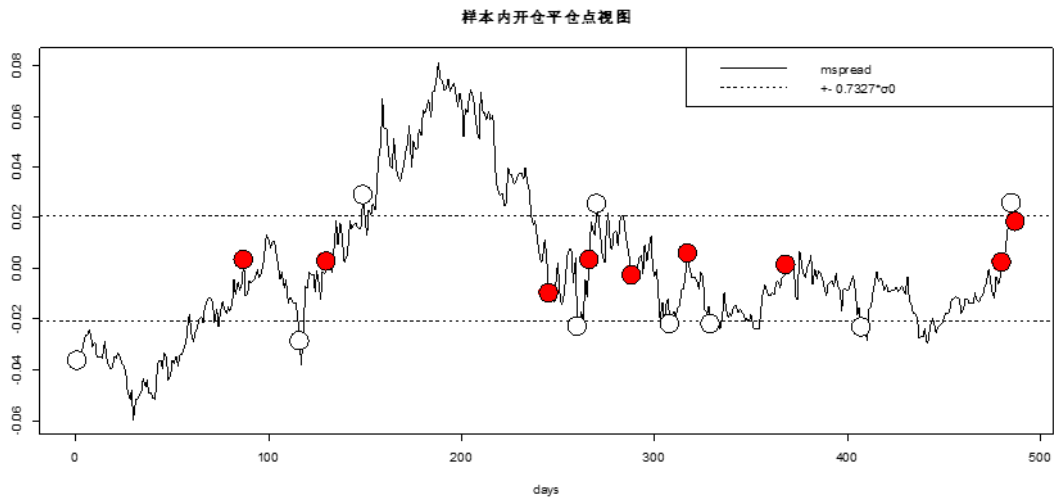


图 3.4 开仓阈值为 $0.7327\sigma_0$ 时样本内开仓平仓点视图

从图 3.4 可看出共进行 9 次交易，交易详情如下：

表 3.8 开仓阈值为 $0.7327\sigma_0$ 时样本内交易详情

交易序号	操作	时间	交易方式	收益率
1	开仓	2017 年 1 月 3 日	买工行卖建行	1.84%
	平仓	2017 年 5 月 15 日	卖工行买建行	
2	开仓	2017 年 6 月 27 日	买工行卖建行	1.74%
	平仓	2017 年 7 月 17 日	卖工行买建行	
3	开仓	2017 年 8 月 11 日	卖工行买建行	7.15%
	平仓	2018 年 1 月 2 日	买工行卖建行	
4	开仓	2018 年 1 月 23 日	买工行卖建行	3.44%
	平仓	2018 年 1 月 31 日	卖工行买建行	
5	开仓	2018 年 2 月 6 日	卖工行买建行	-0.05%
	平仓	2018 年 3 月 9 日	买工行卖建行	
6	开仓	2018 年 4 月 10 日	买工行卖建行	3.90%
	平仓	2018 年 4 月 23 日	卖工行买建行	
7	开仓	2018 年 5 月 11 日	买工行卖建行	5.94%
	平仓	2018 年 7 月 6 日	卖工行买建行	

续表 3.8

8	开仓	2018 年 8 月 30 日	买工行卖建行	3.49%
	平仓	2018 年 12 月 19 日	卖工行买建行	
9	开仓	2018 年 12 月 26 日	卖工行买建行	0.91%
	平仓	2018 年 12 月 28 日	买工行卖建行	

从表中,可看出在样本内期间,一共进行了 9 次交易,其中只有第 5 次的收益是负值,亏损了 0.05%,其余 8 次收益均为正,最大的一次收益为 7.15%,样本内期间总收益率达到了 31.95%。

观察表 3.8,第 1 次交易从 2017 年 1 月 3 日开始至 2017 年 5 月 15 日结束,历时 86 个交易日。根据交易策略,在 2017 年 1 月 3 日,去中心化后的价差小于设定的负的开仓阈值,由价差公式可知工商银行的价格偏低,因此买入工商银行股票卖出建设银行股票,进行开仓操作,接下来会经历一段等待期,等价差回落到零附近,再进行平仓。下表列出在等待期间的收益率:

表 3.9 等待期间交易日收益率

等待天数	收益率	等待天数	收益率	等待天数	收益率	等待天数	收益率
1	0.00%	23	-1.02%	45	-1.41%	67	0.98%
2	0.23%	24	-1.35%	46	-1.63%	68	1.29%
3	0.83%	25	-1.75%	47	-2.56%	69	1.57%
4	1.08%	26	-3.01%	48	-2.18%	70	1.34%
5	1.08%	27	-3.41%	49	-1.53%	71	-0.08%
6	1.30%	28	-2.71%	50	-1.73%	72	0.55%
7	0.90%	29	-4.44%	51	-1.05%	73	-0.28%
8	0.35%	30	-3.41%	52	-1.62%	74	0.07%
9	0.29%	31	-3.56%	53	-1.17%	75	0.70%
10	-0.36%	32	-3.34%	54	-1.43%	76	0.30%
11	-0.36%	33	-3.16%	55	-1.23%	77	0.35%
12	-0.36%	34	-2.38%	56	-0.95%	78	0.73%
13	-0.38%	35	-2.78%	57	-0.28%	79	0.53%
14	0.45%	36	-2.41%	58	0.20%	80	0.65%
15	-0.51%	37	-3.03%	59	-0.98%	81	1.76%
16	-0.89%	38	-2.88%	60	-1.23%	82	1.08%
17	-1.40%	39	-3.11%	61	-0.96%	83	1.55%
18	-1.09%	40	-3.41%	62	-0.43%	84	1.02%
19	-0.49%	41	-2.18%	63	0.10%	85	0.74%
20	-0.51%	42	-1.55%	64	0.02%	86	1.84%
21	-0.29%	43	-1.53%	65	-0.08%		
22	-0.66%	44	-2.08%	66	0.73%		

从上表可看出,等待期间有的交易点收益是负值,这表示表示亏损,其中在第 29

个交易点即 2017 年 2 月 20 日亏损最大，亏损了 4.44%。因为这是样本内数据分析，计算出的样本标准差可以描述数据的离散性，在使用交易策略时可以等到价差回落，对于样本外数据就不一定了。并且在实际交易中市场瞬息万变，在漫长的等待期里，投资者不能预知价差是否会回落，如果价差变得越来越大，那么损失就越大，当损失超过一定程度时，投资者可能就会放弃等待强制平仓了。所以，为了控制风险减少损失，会设置一个止损点，本文采用的是定额止损法，将亏损额设置成一个固定的比例，当损失大于该比例时，就强制平仓。

基于上面讨论，在原来交易规则的基础上，对其补充完善，在开仓后的等待期间，在每个交易日都计算收益率，当亏损大于给定的比例时，为了防止未来出现更大的损失，就及时平仓。一般情况下，强制平仓时，价差往往会更偏离开仓阈值，根据均值回归原理，价格不会一直远离平均值，会逐渐向均值回复，因此价差在后面也会向零回归。此时寻找下一个开仓点时，不能以价差大于阈值为信号点，因为在强制平仓后的一段时期内，价差都是处在偏离状态，若是仓促开仓可能还会出现为了止损而强制平仓的情形，因此要等价差开始回落时再考虑开仓事宜。

本文中，设定强制平仓后，若价差比较大，则等其绝对值回落到开仓阈值内后，从那之后再依照规则寻找开仓点。开仓阈值为 $0.7327\sigma_0$ ，分别带入 1%至 5%，每次步长为 0.5%的止损点计算收益率，得到每个止损点对应的收益率、交易次数和总收益率，如下：

表 3.10 固定标准差、不同止损点下样本内交易情况

止损比例	交易次数	总收益率
1%	12	17.03%
1.5%	12	18.46%
2%	11	13.26%
2.5%	10	17.12%
3%	10	16.52%
3.5%	10	14.23%
4%	10	13.48%
4.5%	9	31.95%
5%	9	31.95%

从表 3.10 可看出，随着止损比例的增加，交易次数变少，总收益率的变化则有增有减，当止损比例大于等于 4.5%时，交易次数和总收益率就不再变化，不难发现，

这和未设置止损点的交易情形一致，说明样本内等待期间，投资者承受的最大损失未超过 4.5%，所以没有触发止损条件。

在实际交易中，投资者能接受的损失一般在 3% 左右，表 3.10 表明当止损点设为 3% 时，交易次数为 10 次，总收益率为 16.52%，虽然没有 4.5% 止损点得到的总收益高，但有效地控制了可能的损失。对于风险承受能力比较小的投资者来说，3% 止损点既能有效地降低风险，又能取得不错的收益率，比较适用。因此不单一地看总收益率，适当地考虑风险，将止损比例设为 3%，按照上面完善的规则，当损失超过 3% 时，就强制平仓提前结束交易，开仓阈值为 $0.7327\sigma_0$ ，在样本内区间依照此策略交易，画出开仓平仓点图：

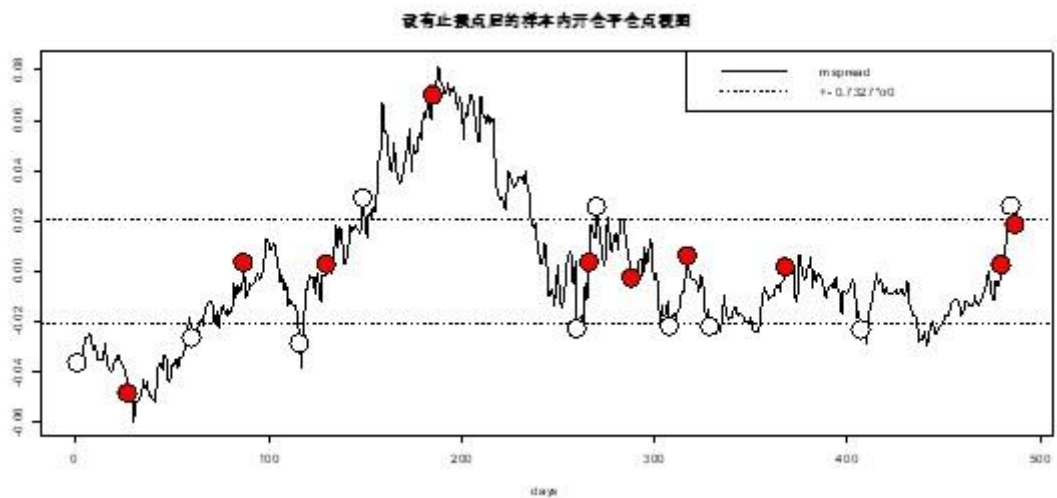


图 3.5 设有止损点后的样本内开仓平仓点视图

从图 3.5 可看出，对于第一组交易点，在开仓后 27 天后触发到止损点就强制平仓了，随后在价差又回落到开仓阈值以内后再继续执行策略，交易详情如下：

表 3.11 增加止损点后的样本内交易详情

交易序号	操作	时间	交易方式	收益率
1	开仓	2017 年 1 月 3 日	买工行卖建行	-3.01%
	平仓	2017 年 2 月 15 日	卖工行买建行	
2	开仓	2017 年 4 月 5 日	买工行卖建行	2.58%
	平仓	2017 年 5 月 15 日	卖工行买建行	
3	开仓	2017 年 6 月 27 日	买工行卖建行	1.74%
	平仓	2017 年 7 月 17 日	卖工行买建行	
4	开仓	2017 年 8 月 11 日	卖工行买建行	-3.14%
	平仓	2017 年 10 月 9 日	买工行卖建行	
5	开仓	2018 年 1 月 23 日	买工行卖建行	3.44%
	平仓	2018 年 1 月 31 日	卖工行买建行	

续表 3.11

6	开仓	2018 年 2 月 6 日	卖工行买建行	-0.05 %
	平仓	2018 年 3 月 9 日	买工行卖建行	
7	开仓	2018 年 4 月 10 日	买工行卖建行	3.90%
	平仓	2018 年 4 月 23 日	卖工行买建行	
8	开仓	2018 年 5 月 11 日	买工行卖建行	5.94%
	平仓	2018 年 7 月 6 日	卖工行买建行	
9	开仓	2018 年 8 月 30 日	买工行卖建行	3.49%
	平仓	2018 年 12 月 19 日	卖工行买建行	
10	开仓	2018 年 12 月 26 日	卖工行买建行	0.91%
	平仓	2018 年 12 月 28 日	买工行卖建行	

从表 3.11 可看出,使用修改后的交易策略,共进行了 10 次交易,7 次收益为正,3 次收益为负,总收益率达到了 16.52%。其中有 2 次是触碰到了止损点强制平仓,例如第一次交易中,在 2017 年 2 月 15 日亏损达到了 3.01%超出了设定的 3%,为了防止后面出现更大的亏损,因此直接平仓结束交易。对比于表 3.8,我们知道若是一直等下去在 2018 年 5 月 15 日,价差会回落到零附近收益会达到 1.84%,但在真实交易中不能预知未来,投资者不一定都能等来符合平仓条件的点,因此要强制平仓后。虽然总收益率低于之前的,但增加止损点能有效地控制交易风险,防止出现更大的损失。

3.4.2 样本外实证

利用完善后的规则,对样本外数据进行分析,首先求出价差,利用公式:

$$spread_t = \ln y_t - 0.8910 \ln x_t,$$

其中下标 t 的范围在 2019 年 1 月 2 日至 2019 年 12 月 13 日,再利用公式:

$$mspread_t = spread_t - mean_{spread}.$$

其中 $mean_{spread} = -0.0074$,这样就把样本外的价差序列求出来了。

使用样本内的样本标准差,使用完善后的交易规则,得到样本外开仓平仓点如下:

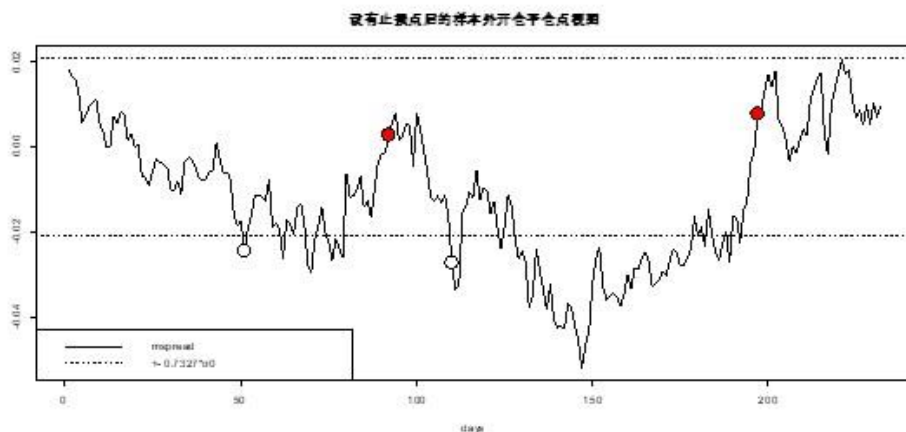


图 3.6 设有止损点后的样本外开仓平仓点视图

从图 3.6 可看出在样本外共进行两次交易，具体交易情况如下：

表 3.12 样本外交易详情

交易序号	操作	时间	交易方式	收益率
1	开仓	2019 年 3 月 20 日	买工行卖建行	3.25%
	平仓	2019 年 5 月 22 日	卖工行买建行	
2	开仓	2019 年 6 月 18 日	买工行卖建行	2.32%
	平仓	2019 年 10 月 25 日	卖工行买建行	

从表 3.12 可知，在样本外区间时，进行了两次交易，均是正收益，总收益率为 5.65%。

下面使用时变标准差进行配对交易。

3.5 时变标准差

由表 3.7 可知价差序列 $spread_t$ 是平稳的，画出序列的自相关、偏自相关图，如下：

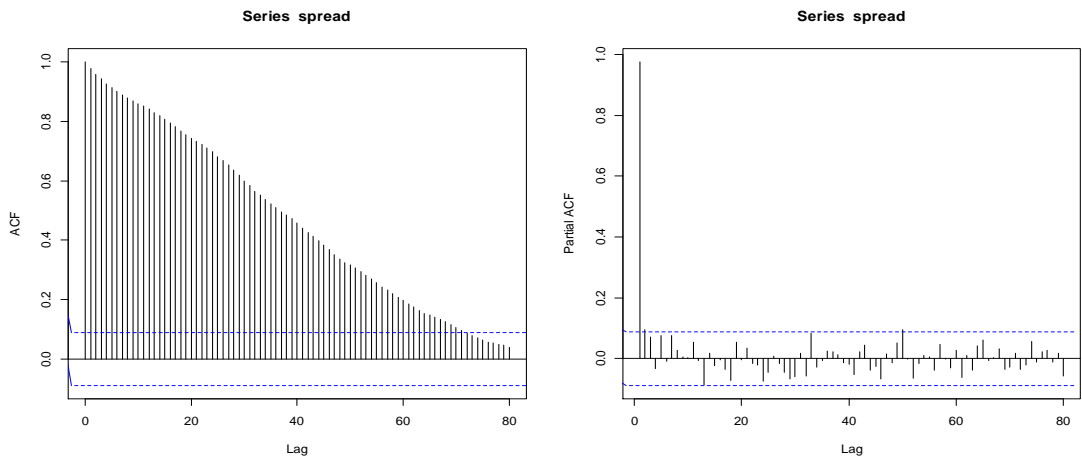


图 3.7 价差序列自相关、偏自相关图

如上图结果显示，价差序列自相关图呈直线下降趋势，偏自相关系数 1 阶后呈拖尾状态，对价差序列拟合 $ARMA(p,q)$ 模型， p 、 q 的取值均值 1 至 3 之间，根据 AIC 值越小模型的效果越好的准则，选取 $ARMA(3,1)$ 作为拟合价差序列的模型，模型估计结果如下：

表 3.13 $ARMA(3,1)$ 模型参数估计结果

	ar1	ar2	ar3	ma1	intercept
估计值	0.1378	0.6660	0.1660	0.7285	-0.0088
标准误差	0.1602	0.1552	0.0448	0.1585	0.0132

得到模型为:

$$\text{spread}_t = -0.0088 + 0.1378\text{spread}_{t-1} + 0.6660\text{spread}_{t-2} + 0.1660\text{spread}_{t-3} + \delta_t \\ + 0.7285\delta_{t-1} + \varepsilon_t.$$

接下来进行模型显著性的检验,对模型的残差序列进行之后 1 阶白噪声检验,得到 P 值为 0.9727,不拒绝原假设,即残差序列是白噪声序列,说明对价差序列的线性信息已全部提取完。下面观察残差的波动性,画出 ARMA(3,1)模型残差序列图:

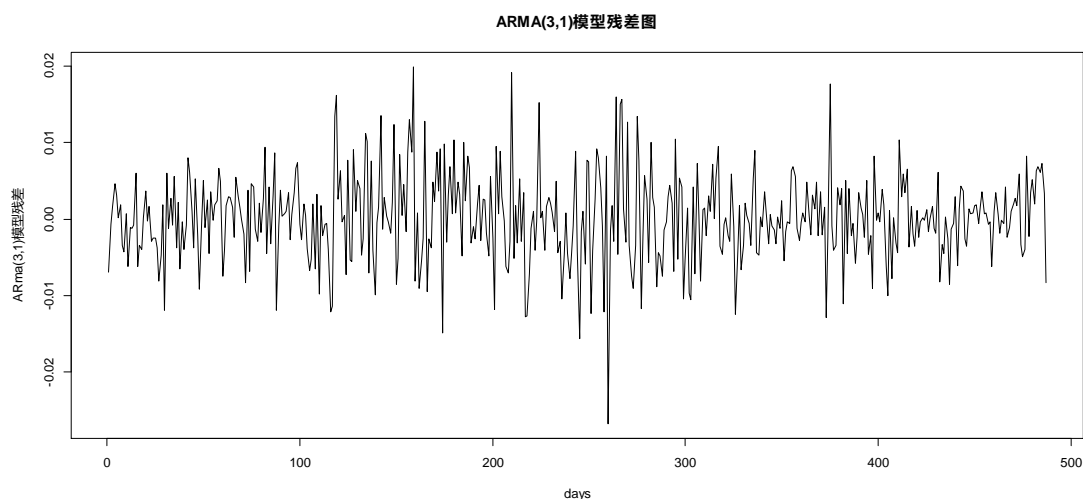


图 3.8 ARMA(3,1)模型残差序列图

从图 3.8 可看出,残差的集群波动比较显著,在中间部分波动性比较大,两端相比波动较小,说明其很大可能是异方差,进行滞后 15 期的 ARCH-LM 检验,检验结果如下:

表 3.14 滞后 15 期的 ARCH 检验结果

滞后期	Chi-squared	P 值
15	67.746	1.122e-08

得到的 P 值均小于显著性水平 5%,拒绝原假设,故残差序列存在 ARCH 效应,它的方差非齐性且具有长期相关性,下面构造 GARCH(1,1)模型,得到估计结果如下:

表 3.15 GARCH(1,1)模型估计结果

	估计值	T 值	P 值
mu	3.947e-05	0.1740	0.8621
omega	1.203e-06	1.5510	0.1208

续表 3.15

alpha1	0.1085	2.5880	0.0097
beta1	0.8598	15.5620	<2e-16

由表 3.15 可得估计结果为:

$$\sigma_t^2 = 0.000001203 + 0.1085\varepsilon_{t-1}^2 + 0.8598\sigma_{t-1}^2.$$

根据向前一步预测公式: $\sigma_h^2(1) = \omega + \alpha_1\varepsilon_h^2 + \beta_1\sigma_h^2$, 将 GARCH 模型估计参数带入, 可得 $\sigma_h^2(1) = 0.000001203 + 0.1085\varepsilon_h^2 + 0.8598\sigma_h^2$, 其中, ε_h^2 为 ARMA 模型的第 h 期的残差平方, 残差可由真实值减去估计值得出, σ_h^2 为第 h 期的条件方差, 只要知道初始值 σ_1^2 即可将 σ_t^2 递推求出, 通常将 ARMA(3,1)模型的残差序列 ε_t 的样本方差作为 σ_1^2 , $\sigma_1^2 = \text{var}(\varepsilon_t) = 3.3145 \times 10^{-5}$. 迭代求出 σ_t^2 后, 将其开方便可得到时变标准差。

接下来, 利用求出的时变标准差, 找到合适的开仓阈值, 先不考虑止损点的存在, 交易规则如下: (1) 当 $\text{mspread}_t > w * \sigma_t$ 时开仓, 卖出 1 个头寸的工商银行股票, 买入 0.8910 个头寸的建设银行股票, 在 $\text{mspread}_t < 0$ 时反向平仓; (2) 当 $\text{mspread}_t < -w * \sigma_t$ 时开仓, 买入 1 个头寸的工商银行股票, 卖出 0.8910 个头寸的建设银行股票, 在 $\text{mspread}_t > 0$ 时反向平仓。

将 w 值设定在 $(0, 15)$ 中, 每次步长为 0.0001, 计算出不同阈值下的收益率, 将开仓阈值与对应的收益率作图, 如下:

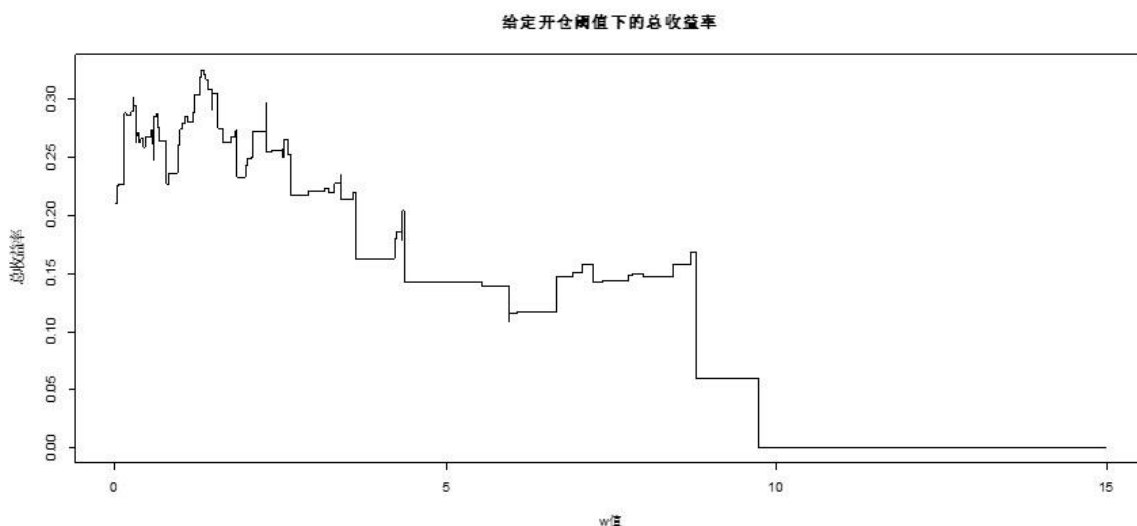


图 3.9 给定开仓阈值下的总收益率

从图 3.9 可看出开仓阈值在大于 $10\sigma_t$ 后, 收益率就为 0, 可认为在阈值较大后, 就不存在交易机会了。因此, 就在 $(0, 15\sigma_t)$ 的范围内寻找使总收益率最大的值, 求解出最优阈值为 $1.3099\sigma_t$, 其总收益率最大为 32.54%。

3.5.1 样本内实证

将开仓阈值设置为 $1.3099\sigma_t$ ，在样本内期间利用上面的交易策略进行交易，画出了开仓点平仓点位置，其中，实线表示价差序列，虚线表示开仓阈值即 $\pm 1.3099\sigma_t$ ，空心点表示开仓点，红色实点表示平仓点，如图 3.10：

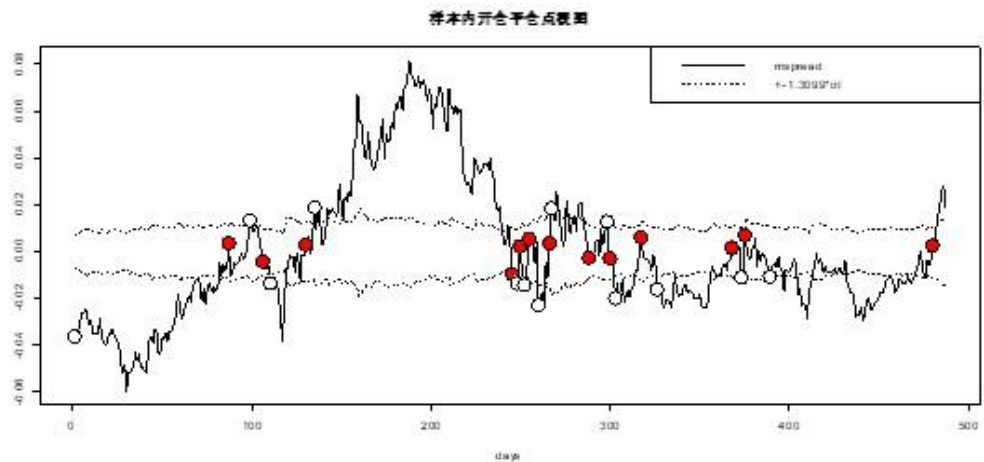


图 3.10 时变标准差下样本内开仓平仓点视图

交易详情见下表：

表 3.16 时变标准差下样本内交易详情

交易序号	操作	时间	交易方式	收益率
1	开仓	2017 年 1 月 3 日	买工行卖建行	1.84%
	平仓	2017 年 5 月 15 日	卖工行买建行	
2	开仓	2017 年 6 月 2 日	卖工行买建行	1.39%
	平仓	2017 年 6 月 13 日	买工行卖建行	
3	开仓	2017 年 6 月 19 日	买工行卖建行	-0.52%
	平仓	2017 年 月 17 日	卖工行买建行	
4	开仓	2017 年 7 月 24 日	卖工行买建行	6.70%
	平仓	2018 年 1 月 2 日	买工行卖建行	
5	开仓	2018 年 1 月 5 日	买工行卖建行	1.92%
	平仓	2018 年 1 月 9 日	卖工行买建行	
6	开仓	2018 年 1 月 11 日	买工行卖建行	0.95%
	平仓	2018 年 1 月 16 日	卖工行买建行	
7	开仓	2018 年 1 月 23 日	买工行卖建行	3.44%
	平仓	2018 年 1 月 31 日	卖工行买建行	
8	开仓	2018 年 2 月 1 日	卖工行买建行	-1.16%
	平仓	2018 年 3 月 9 日	买工行卖建行	
9	开仓	2018 年 3 月 23 日	卖工行买建行	0.63%
	平仓	2018 年 3 月 27 日	买工行卖建行	

续表 3.16

10	开仓	2018 年 3 月 30 日	买工行卖建行	3.53%
	平仓	2018 年 4 月 23 日	卖工行买建行	
11	开仓	2018 年 5 月 8 日	买工行卖建行	4.85%
	平仓	2018 年 7 月 6 日	卖工行买建行	
12	开仓	2018 年 7 月 13 日	买工行卖建行	2.77%
	平仓	2018 年 7 月 17 日	卖工行买建行	
13	开仓	2018 年 8 月 6 日	买工行卖建行	2.40%
	平仓	2018 年 12 月 19 日	卖工行买建行	

表 3.16 显示共进行 13 次交易,其中 2 次交易收益为负,其余 11 次均为正收益,经过计算总收益达 32.54%。相比于前面的固定标准差,交易次数增加了,说明时变标准差更擅于发现交易机会。

接下来增加止损点的限制条件,分别带入 1%至 5%,每次步长为 0.5%的止损点计算收益率,得到每个止损点对应的收益率、交易次数和总收益率,如下:

表 3.17 时变标准差、不同止损点下样本内交易情况

止损比例	交易次数	总收益率
1%	18	14.05%
1.5%	17	14.46%
2%	16	10.81%
2.5%	16	8.20%
3%	15	7.83%
3.5%	14	15.30%
4%	14	14.41%
4.5%	14	21.12%
5%	13	32.54%

从表 3.17 可看出,随着止损比例的增加,交易次数变少,总收益率的变化有增有减,当止损比例大于等于 5%时,交易次数和总收益率就不再变化,和未设置止损点的交易情形一致,说明在样本内等待期间,投资者承受损失的比例未超过 5%,没有触发止损条件。

在对固定标准差止损点的研究中,考虑到风险因素和总收益率,将比例设置为 3%,但对于时变标准差来说,当止损比例为 3%时,总收益率反而是里面最低的。考虑到时变标准差对时间的函数,相比于固定的样本标准差,能够及时反映价差波动的大小,当价差变大脱离均衡水平时,可能存在的损失会变大,故可将止损比例适当变大,再

结合表 3.17, 当止损比例逐渐接近 5% 时, 总收益率呈现慢慢变大的趋势, 不妨将止损比例设为 5%。

当止损比例为 5% 时, 样本内交易未触发止损点, 交易详情见图 3.10、表 3.16。

3.5.2 样本外实证

根据上面对固定标准差样本外数据分析, 已将样本外价差序列 $mspread_t$ 求出, 其中 t 的范围在 2019 年 1 月 2 日至 2019 年 12 月 13 日。对于时变标准差, 根据递推公式可将其求出, 使用完善后的交易规则, 得到样本外开仓平仓点如下:

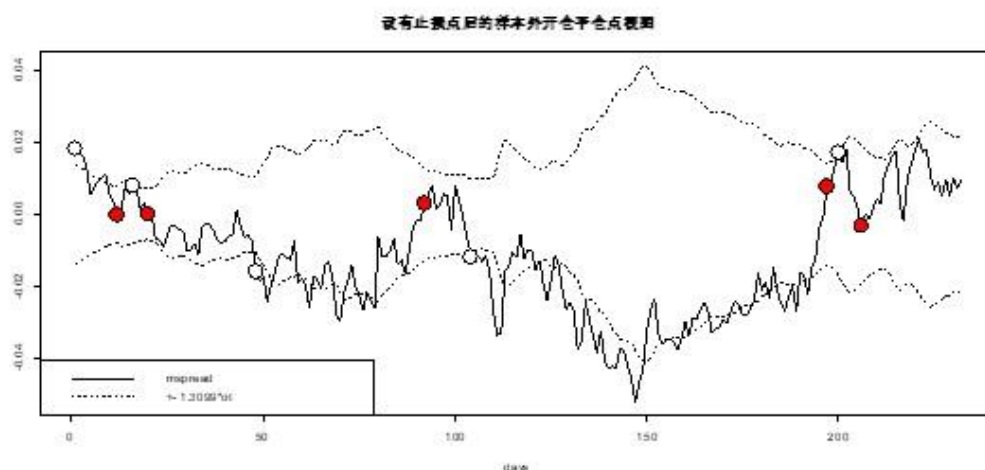


图 3.11 样本外开仓平仓点视图

从图 3.11 可看出在样本外共进行 5 次交易, 具体交易情况如下:

表 3.18 样本外交易详情

交易序号	操作	时间	交易方式	收益率
1	开仓	2019 年 1 月 2 日	买工行卖建行	2.30%
	平仓	2019 年 1 月 17 日	卖工行买建行	
2	开仓	2019 年 1 月 23 日	买工行卖建行	1.32%
	平仓	2019 年 1 月 29 日	卖工行买建行	
3	开仓	2019 年 3 月 15 日	买工行卖建行	2.05%
	平仓	2019 年 5 月 22 日	卖工行买建行	
4	开仓	2019 年 6 月 10 日	买工行卖建行	0.53%
	平仓	2019 年 10 月 25 日	卖工行买建行	
5	开仓	2019 年 10 月 30 日	买工行卖建行	2.32%
	平仓	2019 年 11 月 7 日	卖工行买建行	

从表 3.18 可知, 在样本外区间时, 进行了 5 次交易, 均是正收益, 总收益率为 8.81%。

3.6 本章小结

本章先是根据相关性和协整分析,从股票池中选定工商银行和建设银行作为股票对,进行配对交易。随后利用误差修正模型得到两只股票的对冲比例,并以此算出价差。接着分别使用固定标准差和时变标准差,得到两种情况下最佳开仓阈值、止损比例以及交易情况,下面将两种实证结果汇总比较。

在样本内期间,分别将两种结果汇总在表格中,如下:

表 3.19 样本内固定标准差、时变标准差交易概况

标准差	交易阈值	止损比例	交易次数	总收益率
固定标准差	$0.7327\sigma_0$	3%	10	16.52%
时变标准差	$1.3099\sigma_t$	5%	13	32.54%

从表 3.19 可看出在样本内期间,对于交易次数,固定标准差下交易次数少于时变标准差下的次数,这表明时变标准差更能及时反映两个股票的价差变动,擅于抓住交易机会。对于总收益率,时变标准差下的收益几乎是固定标准差下的两倍,表明时变标准差下交易策略优于固定标准差下。

总的来讲,两种标准差下配对交易策略都能取得不错的收益率,即使是固定标准差下也能取得 16.52% 的总收益率。一般国债利率被称为无风险利率,2017 年 1 月 3 日的 1 年期国债利率为 2.75%、2017 年 1 月 3 日的 1 年期国债利率为 3.68%,配对交易策略下的收益相比于无风险利率收益高出很多。

在样本外期间,两种结果汇总如下:

表 3.20 样本外固定标准差、时变标准差交易概况

标准差	交易阈值	止损比例	交易次数	总收益率
固定标准差	$0.7327\sigma_0$	3%	2	5.65%
时变标准差	$1.3099\sigma_t$	5%	5	8.81%

表 3.20 显示在样本外期间,使用固定标准差进行配对交易,共进行 2 次交易,可获得 5.65% 的总收益率,使用时变标准差,共进行 5 次交易获得 8.81% 的总收益。可看出时变标准差相比于固定标准差,交易次数多,取得的收益也高。显然,时变标准差较于固定标准差,配对交易的效果更好。

2019 年 1 月 2 日的 1 年期国债利率为 2.50%,从总收益率方面看,无论使用哪种标准差,配对交易策略带来的收益都高于无风险利率,说明使用配对交易策略能取得不错的收益。

第四章 结论与展望

4.1 结论

配对交易属于统计套利,本质是一种市场中性策略,自从在二十世纪八十年代中期被提出,在市场上取得了不俗的表现,应用十分广泛。交易所经过融资融券试点后,在2010年3月31日正式实施两融业务,市场引进了做空机制,这意味着我国的金融市场更加完善,也为配对交易的发展提供了条件。

本文通过融券业务进行配对交易,基于协整理论采用两种方法进行交易,都取得了正收益,说明使用配对策略是有效的,适用于我国市场。考虑到同一行业的影响因素大部分相同,价格变动相像,会有长期均衡的关系。选取了上交所融资融券标的股票中的18只银行股票对数价格历史数据作为股票池,样本区间为2017年1月3日至2019年12月13日的股票日数据,其中,2017年和2018年的数据为样本内数据,其余数据为样本外数据。

首先,对于样本内数据进行相关分析,计算出18只股票的相关系数矩阵,选择相关系数最高的工商银行和建设银行进行配对,同时还要对这一组进行协整检验,结果显示通过检验,证明这两只股票对数价格存在长期均衡的关系,可以配对。接下来,使用误差修正模型得到短期动态模型,并根据误差项系数得到了两只股票的对冲比例和价差序列。然后,研究价差的标准差,以此确定交易信号,分别使用固定标准差和时变标准差,在设置了止损比例后,得到两种方法下的样本内最佳收益率,固定标准差下的总收益率为16.52%,而时变标准差下的总收益率达到32.54%。最后,对样本外数据分别使用两种方法,固定标准差下的样本外总收益率为5.65%,时变标准差下样本外总收益率为8.81%。

从实证结果可看出,一方面,使用配对交易策略取得的收益率远高于无风险利率,说明配对交易能够很好应用于我国的股票市场;另一方面,无论是样本内还是样本外,使用时变标准差的效果都优于固定标准差的效果,表明时变标准差更能反映出价差的变动,能取得更好的收益。

4.2 不足与展望

本文有一定的局限性,有不足之处可以改进:

(1) 交易所两融业务涉及到的标的股票含盖很多行业,在实证时只选取了代表性的银行业股票进行研究,忽略了其它行业之间的配对,在未来的研究中,还可以选

择地产经营、钢铁、汽车等行业进行分析，考察配对交易策略在各个行业中的效果如何。并且，不同行业之间也能进行交叉配对，如将一只银行股票和一只房地产股票配对，研究跨行业配对交易策略的有效性。

（2）本文计算交易收益时，没有考虑到交易成本，在真实的交易环境中，进行一次交易会涉及到各种交易费用，在后续的研究中要模拟真实市场的交易，就要考虑到各种费用。

（3）本文研究的是以天为单位的低频股票数据的配对，对于当日价格变化较大的股票来说，容易错失套利机会，未来可以研究高频数据，这样能更加及时把握交易机会，且配对的研究范围也可以扩展到期货、基金之间，提高投资的多样化。

（4）本文采用的是固定取样研究，选取固定的样本区间，在接下来的研究中可以考虑滚动取样研究，即每隔一段时间对序列进行更新，再对新的序列进行研究，观察能否进行交易，若可以交易，则更新参数，比较不同时期的差异，如对冲比例、交易阈值等，观察不同时期参数的大小变化。

参考文献

- [1] Ganapathy Vidyamurthy. Pairs trading : quantitative methods and analysis[M]. John Wiley & Sons, Inc, 2004
- [2]Evan Gatev, William N. Goetzmann, K. Geert Rouwenhorst. Pairs Trading: Performance of a Relative-Value Arbitrage Rule[J]. The Review of Financial Studies, 2006, Vol.19(3): 797-827
- [3] Alovsat Muslumov, Asli Yuksel, Aydin Yuksel. The Profitability of Pairs Trading in an Emerging Market Setting: Evidence from the Istanbul Stock Exchange[J]. Empirical Economics Letters, 2009, Vol.8(5): 495-500
- [4]Robert J. Elliott, John Van Der Hoek, William P. Malcolm. Pairs trading[J]. Quantitative Finance, 2005, Vol.5(3): 271-276
- [5]Supakorn Mudchanatongsuk, James A. Primbs, Wilfred Wong. Optimal pairs trading: A stochastic control approach[A]. 2008
- [6]Agnès Tourin, Raphael Yan. Dynamic pairs trading using the stochastic control approach[J]. Journal of Economic Dynamics and Control, 2013, Vol.37(10): 1972-1981
- [7] Timofei Bogomolov. Pairs trading based on statistical variability of the spread process[J]. Quantitative Finance, 2013, Vol.13(9): 1411-1430
- [8]Ahmet Goncu, Erdinc Akyildirim. A stochastic model for commodity pairs trading[J]. Quantitative Finance, 2016, Vol.16(12): 1843-1857
- [9]方昊.统计套利的理论模式及应用分析——基于中国封闭式基金市场的检验[J].统计与决策,2005,12: 14-16
- [10]仇中群, 程希骏.基于协整的股指期货跨期套利策略模型[J].系统工程,2008,12: 26-29
- [11]张俊, 李妍.配对交易策略:一个文献评述[J].中国证券期货,2010,12: 28-29
- [12]崔方达, 吴亮.配对交易的投资策略[J].统计与决策,2011,23: 156-159
- [13]胡伦超, 余乐安, 汤铃.融资融券背景下证券配对交易策略研究——基于协整和距离的两阶段方法[J].中国管理科学,2016,24(4): 1-9
- [14]蔡燕, 王林, 许莉莉.基于随机价差法的配对交易研究[J].金融理论与实践,2012,8: 30-35
- [15]王春峰, 林碧波, 朱琳.基于股票价格差异的配对交易策略[J].北京理工大学学报(社会科学版),2013,1: 71-75
- [16]刘永辉, 张帝.基于协整-OU 过程的配对交易策略研究[J].管理评论,2017,29(9): 28-

36

- [17]毕秀春, 刘博, 袁吕宁, 张曙光.带止损条件的配对交易最优阈值[J].系统科学与数学,2019,39(7): 1117-1141
- [18]João Frois Caldeira, Guilherme Valle Moura. Selection of a Portfolio of Pairs Based on Cointegration: A Statistical Arbitrage Strategy[J]. Revista Brasileira de Finanças, 2013, Vol.11(1): 49-80
- [19]George J. Miao. High Frequency and Dynamic Pairs Trading Based on Statistical Arbitrage Using a Two-Stage Correlation and Cointegration Approach[J]. International Journal of Economics and Finance, 2014, Vol.6(3): 96
- [20]Nicolas Huck, Komivi Afawubo. Pairs trading and selection methods: Is cointegration superior?[J]. Applied Economics, 2014, Vol.47(6): 599-613
- [21]Andreas Mikkelsen. Pairs trading: the case of Norwegian seafood companies[J]. Applied Economics, 2018, Vol.50: 303-318
- [22]Hossein Rad, Rand Kwong Yew Low, Robert Faff. The profitability of pairs trading strategies: distance, cointegration and copula methods[J]. Quantitative Finance, 2016, Vol.16(10): 1541-1558
- [23]Saeid Fallahpour, Hasan Hakimian, Khalil Taheri, Ehsan Ramezanifar. Pairs trading strategy optimization using the reinforcement learning method: a cointegration approach[J]. Soft Computing, 2016, Vol.20(12): 5051-5066
- [24]Christopher Krauss. Statistical arbitrage pairs trading strategies: review and outlook [J]. Journal of Economic Surveys, 2017, Vol.31(2): 513-545
- [25]Matthew Clegg, Christopher Krauss. Pairs trading with partial cointegration[J]. Quantitative Finance, 2018, Vol.18(1): 121-138
- [26]于玮婷.基于协整方法的统计套利策略的实证分析[J].科学决策,2011,3: 70-85
- [27]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
- [28]丁涛.配对交易策略在 A 股市场的应用与改进[J].中国商论,2013,: 24-26
- [29]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,2014,1: 30-40
- [30]邢恩泉, 尹涛.协整模型的配对交易策略优化[J].经济数学,2015,32(1): 65-69
- [31]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
- [32]吴泽耕.基于 GARCH 模型银行股配对交易应用研究[D].兰州: 兰州财经大学,2015
- [33]张可欣.配对交易策略统计方法及 R 语言实现[D].北京: 北京工业大学,2017

- [34]吴添添.基于上证 50 指数成分股与股指期货的配对交易方法与实现[D].昆明: 云南财经大学,2018
- [35]赵珊珊.基于强化学习算法的配对交易策略研究[D].合肥: 安徽大学,2018
- [36]冯玉茹.配对交易策略在我国 A 股市场研究——基于传统模型和 GARCH 模型[D].济南: 山东大学,2019
- [37] William K. Bertram. Analytic solutions for optimal statistical arbitrage trading[J]. Physica A, 2010, Vol.389(11): 2234-2243
- [38]王燕编著.基于 R 应用的统计学丛书 时间序列分析 基于 R[M].北京: 中国人民大学出版社.2015

致谢

时光荏苒，两年的研究生生活即将结束，在南京师范大学求学的两年时间里，我的导师、同学和家人给予了很大的帮助，在毕业之际，向曾经帮助、支持过我的人们表示最真挚的感谢。

首先，由衷地感谢我的指导老师李启才副教授，感谢李老师在这两年在学习和生活中对我的关心、帮助和指导。在本文的撰写过程中，得到了李老师的悉心指导，对于论文的选题和思路，李老师提供了很多宝贵、中肯的建议，并且在文章初步完成后，多次审阅全文，提出修改意见，让我改进文章。李老师严谨的工作态度、一丝不苟的科研精神深深地影响着我，是我学习的榜样。

其次，我要感谢院里老师对我的教导和培养，感谢这些老师的教诲，让我掌握了很多知识，为以后的工作做准备。同时我还要感谢每一位给予我帮助的同学，在学习上，会耐心地替我解疑答惑；在生活上，会全力帮我解决问题。

然后，我要感谢我的家人和朋友对我的支持和鼓励，他们是我坚强的后盾，特别是我的家人，更是让我感受到了陪伴的力量。

最后，衷心地感谢在百忙中抽出时间对本文进行审阅的专家老师，由于本人的知识和写作水平有限，文章会有一些不足，恳请各位老师指教。