

# Study on Cross-sector Stock Pair Trading Strategies based on Cointegration

Jiamiao Chen

Shanghai University, 201800, China

†Email: 15958140355@163.com

## Abstract

In the context of the continuous development of the two financing businesses, this paper conducts a correlation analysis of the historical stock prices of 11 different industries between January 4, 2011 and December 31, 2020, selects two of the industries with the highest correlation and five representative stocks from each of them, tests the cointegration relationship between the two stocks with the highest correlation, and uses Python tools to conduct an empirical study of the matching trading. The results show that this strategy can help us to obtain significant positive returns within a short period of time, and is feasible and effective for implementation.

**Keywords:** Pair Trading; Cointegration Relationship; Python

## 基于协整的跨行业股票配对交易策略研究

陈佳苗

上海大学, 上海市 201800

**摘要:** 在两融业务不断发展的背景下, 本文通过对 2011 年 1 月 4 日至 2020 年 12 月 31 日间 11 个不同的行业的历史股价进行相关性分析, 选出其中两个相关性最高的行业并从中各选出五只具有代表性的股票, 检验其中相关性最高的两只股票的协整关系, 利用 Python 工具进行配对交易策略的实证研究, 结果发现该策略在短时间内可以帮助我们获得可观的正收益, 具有实施的可行性与有效性, 值得我们深入的学习与广泛的应用。

**关键词:** 配对交易; 协整关系; Python

## 引言

配对交易是一种建立在市场做多做空机制上的交易策略, 通过构建两种或两种以上的证券组合对于其价值的偏离进行多头或空头的交易来进行套利的一种交易模式。由于我国的金融市场起步晚, 配对交易策略在我国的研究与应用并不成熟, 但 2010 年 3 月 31 日沪深两地交易所正式启动了融资融券业务的试点工作后, 两融业务的出现结束了我国只能做多不能做空的单边交易市场制度, 为配对交易在我国的发展提供了更大的空间。融资融券业务也就是说投资者可以通过抵押一部分担保物向证券公司借入资金购买债券(融资业务)或是借入债券卖出(融券业务)。随着融资融券业务在我国金融市场的稳步发展、不断推进, 同时满足了个人投资者和机构投资者更加多样的投资需求, 也为我国金融市场的发展注入了更多的活力。常见的关于配对交易的研究大多基于期货市场的套利研究或是同一行业下的股票配对交易研究, 这大大限制了配对股票的选择范围, 因此本文将不同行业的股票纳入配对交易组合的选择范围利用 Python 工具进行研究, 通过实证研究分析跨行业具体股票统计套利的协整关系及收益效果。

# 1 理论背景

## 1.1 配对交易

配对交易的思想最早在 20 世纪 20 年代由华尔街交易员 Jesse Livermore 所提出，又可被称为价差交易或是统计套利交易，是一种风险小、收益又比较稳定的市场中性策略。具体的交易策略是首先在市场中寻找两只长期历史价格走势趋于一致的股票配成一对，以此来确保由这两只股票配对构成的价格差大致在一定的范围内波动。而后设定交易触发信号，在配对股票价格差发生正向偏离时，因预计价格差未来会恢复均值，从而做空价格走势强的股票同时做多价格走势弱的股票，当价差恢复到正常水平时平仓进行套利。当配对股票价格发生反向偏离时，反向建仓，同样在价差恢复后可平仓获利。

## 1.2 协整关系

Christian 和 Gianluigiorgioin 通过实证研究对比具有协整关系的配对股票与不具有协整关系的配对股票发现，具有协整关系的配对股票在进行套利时容易获取更高的收益，因此，在构建配对交易股票对时我们通常使用协整模型。协整模型最常用的方法就是通过 Engle 和 Granger 两步法对两个股票价格对数的时间序列进行协整关系的检验。具体操作是在两只股票价格对数都是一阶单整的前提下，对价格对数进行回归后得到的残差进行 ADF 单位根检验其序列是否平稳，若平稳则表示所选择的两个股票具有协整关系可以进行配对交易。

基于上述两个理论前提，本文将通过实证研究的方法利用 Python 工具对不同行业的股票进行配对交易，利用对冲的手段进行套利分析。

# 2 实证分析

## 2.1 行业选择

### 2.1.1 数据来源

为了对跨行业的股票进行配对交易分析，本文根据 wind 一级行业分类，将行业划分为以下 11 类：电信服务、能源、公共事业、房地产、材料、金融、信息技术、工业、可选择消费、医疗保健以及日常消费。数据上选取了 2011 年 1 月 4 日至 2019 年 12 月 31 日的日数据作为形成期的样本数据，与 2020 年 1 月 2 日至 2020 年 12 月 31 日的日数据作为样本外的回测数据进行实证分析与研究。

### 2.1.2 最优配对行业选择

首先，为了从 11 个行业中选择最优的两个配对行业，对 11 个行业各自的相关性进行检验结果如表 1 所示：

表 1 行业相关性分析

	A	B	C	D	E	F	G	H	I	J	K
A	1.0000	0.1684	0.9434	0.8774	0.8723	0.7522	0.9005	0.9563	0.9245	0.8247	0.5471
B	0.1684	1.0000	0.0605	-0.0983	0.3108	-0.2348	-0.1333	0.2025	-0.0373	-0.2648	-0.2937
C	0.9434	0.0605	1.0000	0.9448	0.8657	0.8239	0.9498	0.9687	0.9612	0.8957	0.6068
D	0.8774	-0.0983	0.9448	1.0000	0.8717	0.9134	0.9715	0.9151	0.9742	0.9531	0.7408
E	0.8723	0.3108	0.8657	0.8717	1.0000	0.7460	0.8484	0.9280	0.8986	0.7894	0.6070
F	0.7522	-0.2348	0.8239	0.9134	0.7460	1.0000	0.8763	0.7406	0.8417	0.9211	0.9091
G	0.9005	-0.1333	0.9498	0.9715	0.8484	0.8763	1.0000	0.9286	0.9833	0.9661	0.7113
H	0.9563	0.2025	0.9687	0.9151	0.9280	0.7406	0.9286	1.0000	0.9601	0.8393	0.5217

I	0.9245	-0.0373	0.9612	0.9742	0.8986	0.8417	0.9833	0.9601	1.0000	0.9429	0.6508
J	0.8247	-0.2648	0.8957	0.9531	0.7894	0.9211	0.9661	0.8393	0.9429	1.0000	0.8044
K	0.5471	-0.2937	0.6068	0.7408	0.6070	0.9091	0.7113	0.5217	0.6508	0.8044	1.0000

(注:表中字母 A 为电信服务、B 为能源、C 为公共事业、D 为房地产、E 为材料、F 为金融、G 为信息技术、H 为工业、I 为可选择消费、J 为医疗保健、K 为日常消费)

根据表 1 结果分析可知,11 个行业中信息技术行业与可选消费行业的相关性最高为 0.9833,因此本文将选择信息技术与可选择消费两个行业作为进一步分析的样本,选择不同行业中具体股票进行配对交易的研究。

## 3.2 股票选择

### 3.2.1 数据选取

本文分别选取了信息技术与可选择消费行业中最具有代表性的 5 只股票作为样本进行下一步分析,这十只股票分别为以下十只:

(1) 信息技术行业:工业富联(601138.SH)、上海钢联(300226.SZ)、京东方 A(000725.SZ)、中兴通讯(000063.SZ)、TCL 科技(000100.SZ)

(2) 可选择消费行业:上汽集团(600104.SH)、美的集团(000333.SZ)、苏宁易购(002024.SZ)、海尔智家(600690.SH)、格力电器(000651.SZ)

剔除上海钢联与美的集团两只时间跨度较短的股票,对剩下八只股票进行相关性检验,结果如表 2 所示:

表 2 个股相关性分析

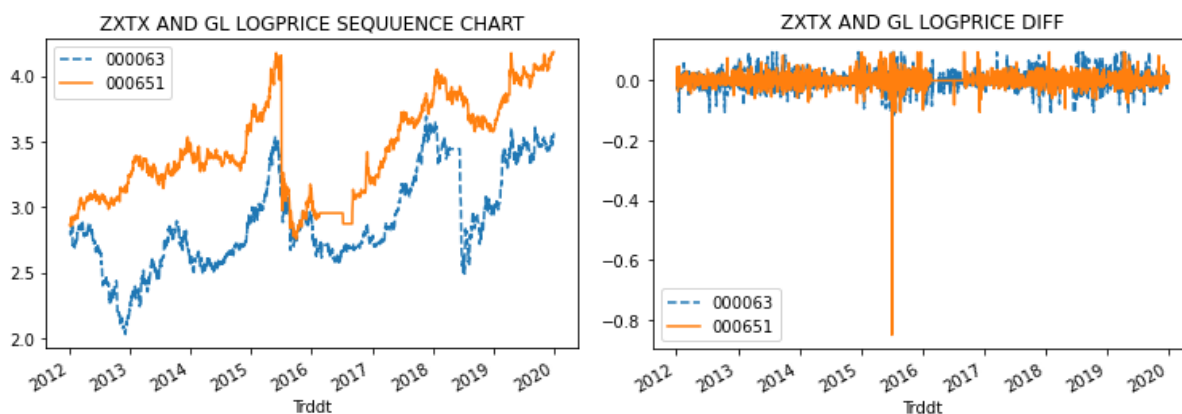
	000063	000100	000651	600104	600690	002024	000725	300226
000063	1.0000	0.5482	0.8324	0.7035	0.5428	0.5842	0.8243	0.5729
000100	0.5482	1.0000	0.4081	0.4930	0.4951	0.7370	0.6613	0.7349
000651	0.8324	0.4081	1.0000	0.6503	0.7376	0.3843	0.7242	0.6081
600104	0.7035	0.4930	0.6503	1.0000	0.3936	0.6340	0.8123	0.4400
600690	0.5428	0.4951	0.7376	0.3936	1.0000	0.3736	0.5646	0.6481
002024	0.5842	0.7370	0.3843	0.6340	0.3736	1.0000	0.5975	0.6248
000725	0.8243	0.6613	0.7242	0.8123	0.5646	0.5975	1.0000	0.4901
300226	0.5729	0.7349	0.6081	0.4400	0.6481	0.6248	0.4901	1.0000

根据表 2 结果分析可知,8 只股票中中兴通讯(000063.SZ)与格力电器(000651.SZ)的相关性最高为 0.8324,且它们分别属于信息技术行业与可选择消费行业,满足了本文跨行业股票配对交易的研究主题,因此本文将运用中兴通讯与格力电器股票的交易数据设计配对交易策略,将 2012 年 1 月 4 日至 2019 年 12 月 31 日作为配对的形成期,以及将 2020 年 1 月 4 日至 2020 年 12 月 31 日作为交易期,用 Python 实现配对交易策略。

### 3.2.2 协整关系检验

根据 Engle 和 Granger 两步法,在构建协整模型前,首先对中兴通讯(000063.SZ)与格力电器(000651.SZ)两只股票的对数价格进行 ADF 单位根的平稳性检验,结果显示如表 3 所示,两者的对数价格单位根检验 P 值均>0.05,不能拒绝原假设,也就是说两个股票价格对数的时间序列并不平稳,因此对两者的价格对数取一阶差分再次进行单位根检验,这次结果显示 P 值远小于 0.05,也就是表明此时序列平稳,因此两个股票价格对数均满足了协整模型一节单整的前提,可进行下一步分析。

图 a、b 分别为中兴通讯(000063.SZ)与格力电器(000651.SZ)对数价格以及一节差分后价格的时间序列图,如下所示:



(a) 中兴通讯与格力股票价格对数时序图

(b) 中兴通讯与格力股票价格对数一阶差分时序图

图 1 时间序列图

第二步，对两者的价格对数进行 OLS 回归得到一个残差序列，同样对残差序列进行单位根检验，结果显示如表 3 所示 P 值 $<0.05$ ，拒绝原假设，残差序列平稳，也就证明了两只股票具有协整关系。为了更加便于观察残差序列的变化，本文对残差序列进行去中心化的处理（残差-残差的平均数），结果显示如图 2 所示：

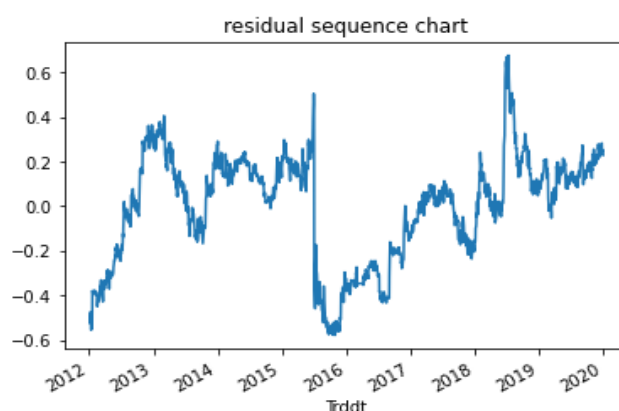


图 2 残差去中心化时序图

上述 5 次 ADF 检验结果具体如下所示：

表 3 ADF 检验结果

合约	T 值	P 值	延迟
000063	-1.496	0.535	3
000651	-1.587	0.494	16
000063 (差分)	-23.701	0.000	2
000651 (差分)	-10.343	0.000	15
残差	-2.844	0.004	11

### 3.3 配对交易策略构建

#### 3.3.1 阈值设定

根据交易期配对交易的残差特征，本文在交易期内分别设定  $\mu \pm 1.2\sigma$  和  $\mu \pm \sigma$  为开仓与平仓的阈值，并将  $\mu \pm 2.2\sigma$  设置为因协整关系可能破裂而强制平仓的阈值，具体交易规则如下：

- (1) 当价差上穿  $\mu + 1.2\sigma$  时，做空配对股票，反向建仓（卖出格力电器，同时买入中兴通讯，中兴通讯与格力电器股票的资金比值为对两者价格对数进行 OLS 回归时得到的  $\beta$  值）；
- (2) 当价差下穿  $\mu + \sigma$  时，做多配对股票，进行反向平仓套利；
- (3) 当价差下穿  $\mu - 1.2\sigma$  时，做多配对股票，正向建仓（买入格力电器，同时卖出中兴通讯，中兴通讯与格力电器股票的资金比值为对两者价格对数进行 OLS 回归时得到的  $\beta$  值）；
- (4) 当价差再次回复到  $\mu - \sigma$  的上方时，做空配对股票，进行正向平仓套利；
- (5) 当价差突破  $\mu \pm 2.5\sigma$  时，为了止损及时进行强制平仓操作。

### 3.3.2 模拟交易结果

本文的回测时间选择的是 2020 年 1 月 2 日至 2020 年 12 月 31 日两只股票的日交易数据，通过 Python 构造交易信号函数自动进行模拟交易，交易期配对交易残差序列如图 3 所示，可以看出整个交易期间共进行了一次建仓与一次平仓：

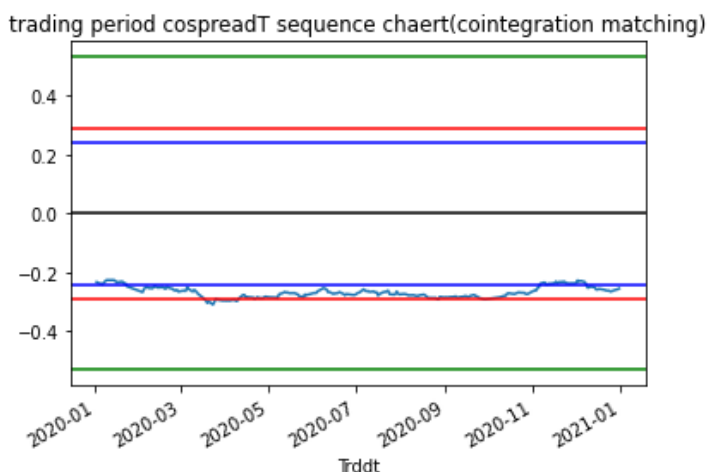


图 3 交易期残差序列图（配对交易）

本文假设初始投资资金为两万元，经过一年的模拟交易时间，最终资产规模涨到了 39015.94 万元（如图 4 所示），年收益率高达 95.08%，远高于同期的基准收益率，样本回测结果显示此次配对交易带来了显著的正收益，这表示具有协整关系的股票配对交易的套利策略是有效的，配对交易策略值得得到更加深入的研究以及更加广泛的应用。

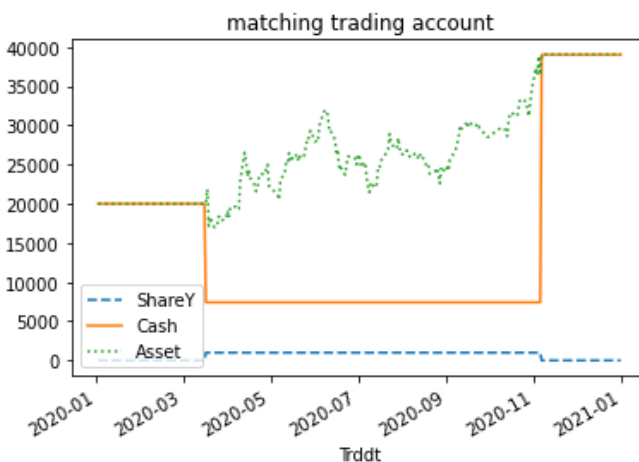


图 4 配对交易结果

### 3 结论

在两融业务不断发展的背景下，本文通过对于 2011 年 1 月 4 日至 2020 年 12 月 31 日不同行业间的具有协整关系的股票构建配对交易，通过实证分析发现统计套利的策略在一年的时间内可以帮助我们获得可观的正收益，具有可实际操作的可行性与有效性。但同时本文在关于阈值的设定方面的方法较为简单，并没有通过复杂的算法进行更深一步优化设计，而是根据两只股票配对后的特征进行简单的阈值确定，因此后续还可以通过机器学习的方法对实证过程进行进一步的改善。

### REFERENCES

- [1] Du WJ. A study on cross-period arbitrage portfolio of soybean meal futures based on cointegration[J]. Business Development Economics,2021(03):57-60.
- [2] Zhong Mingxue. An empirical study of stock cointegration relationship among different industries. Time Finance. 2020:39-41
- [3] Ma Lixia. Research on statistical arbitrage based on cointegration model and empirical analysis. Reform and Openness.2016:25-26
- [4] Tang, Luwei; Xu, Qingjuan; Luo, Wenting. A study of stock matching based on cointegration. Times Finance.2019:71-73
- [5] Feng Yuru. A study on pairwise trading strategy in China's A-share market. Shandong University. 2019
- [6] Zhu Lirong; Su Xin; Zhou Yong. A study on statistical arbitrage based on China's futures market. Mathematical Statistics and Management.2015:164-174

#### 【作者简介】



陈佳苗（1997.12.18），女，汉族，硕士，学生，现就读于上海大学金融专业。  
Email: 15958140355@163.com