

清华大学本科毕业生论文开题报告
学院：交叉信息研究院
导师：唐平中

姓名：陈炜艺
学号：2010011352
班级：计科00

论文题目：配对交易——相对价值套利的算法实现与实验分析

研究背景

1980年代，华尔街量化分析师Nunzio Tartaglia建立了一个组织，该组织囊括了各名校的物理、数学和计算机科学研究员，目的在于挖掘资本市场中的套利机会。Tartaglia的组织采用了当时学术导向用的复杂统计模型发展高科技交易项目，这个项目在当时就能通过自动交易系统运行。自动交易系统在当时是首创，一度震惊华尔街，原因在于它某种程度上替代了交易员（Trader）的职能，最关键的是它量化了人为策略，有计划并且恒久快捷的交易都让人认同它的巨大潜力。

除此之外，Tartaglia也在自动交易系统的应用中发现了某些证券之间的价格同步性，即倾向于向同一形势发展。它们通过买入卖空在1987年就创造了极大的财富——一年在该策略给它们公司摩根斯坦利（Morgan Stanley）带来的收益达到5千万美元。虽然原摩根斯坦利在之后两年业绩不佳，在1989年分拆成了大摩（JP Morgan）和现摩根斯坦利（Morgan Stanley），配对交易（pairs trading）从此不再是该公司的私有策略，并开始在“市场中性”（Market Neutral）投资策略领域风靡，许多个人投资者和对冲基金公司也不断改进和进行该策略的开发与运营。

众所周知，任何套利策略的风靡和知名都会导致它的回报不断减小。在纽约时报（New York Times）的一篇报道中，过去Tartaglia手下的一个软件开发师David Shaw，现已经是一间相当成功的量化组织D.E. Shaw的总裁提及，量化套利策略的利润现已经缩水，他的公司之所以能成功，主要因素还是在于较早使用该技术进入市场。Tartaglia更以一种哲学性的口吻描述他的配对策略，他说人们一般喜欢在股票涨了以后才买入，而不愿意去买跌了的股票。换言之就是发现了策略好用后才加以利用，但为时已晚。那到底现在配对交易者是否还能从满是过激的无规律个人投资者中有规律地盈利？这是本论文的目的所在。

研究目的

综上所述，华尔街出现了量化套利（quantitative arbitrage）并长期成为风靡金融数学世界的交易方法。其中一种短期的套利策略被称之为配对交易（pairs trading）。该策略在华尔街风靡至少20年，并且现在仍是对冲基金和投资银行最主流的统计套利（statistical arbitrage）工具之一。配对交易实质概念比较简单，简而言之，如果能找到两只历史价格走势相似的股票，当它们之间的价差（spread）足够宽时，卖空走势偏高的股票，同时买入走势偏低的股票。如果未来重复它们价格走势相似的历史特性，相对价格将会收敛，之前的套利策略就能盈利。该策略仅仅是基于历史价格动态和简单的逆势原理（contrarian principle），难以让人相信它真的可以赚钱。如果美国资本市场在任意时刻都是充分有效的，配对交易在风险调整后是难以有回报的，当然真正的市场无法达到任意时刻充分有效，因此套利策略的优越在于实验模型的精益求精，该论文的目的正是在引入计算机算法

后进行不断实验构造一个最高历史收益性的配对交易策略。同时也将回答背景中提出的问题“配对交易者是否还能从满是过激的无规律个人投资者中有规律地盈利？”，证明现在的逆势回报有一部分实质来自于人们对公司新闻信息的过激反应，而非价格回归模型中假设的长期相对均衡。

研究内容

该论文中，我将对配对交易的风险和回报特性进行研究，数据为2004至2013年的美国股市日数据。由计算机算法、代码实现细节、金融应用三个层面进行介绍——

计算机算法辅助模型构造

首先会根据著名书籍《量化方法与分析》（Quantitative method and analysis）介绍的经典配对交易策略模拟构架，建造三步模型：

- 一、数据挖掘（Data Mining）算法进行高协整（Co-integration）配对的搜索
- 二、机器学习（Machine Learning）算法进行高盈利性（Profitability）配对的筛选
- 三、后台测试（Back test）策略，生成金融交易报告

初期采用简单的算法进行配对的发现、选择和交易，研究几种直接、自融资交易法则的盈利效果。根据之前的简单实验，我发现对于好的配对股票组合，年收益大致在11%。虽然配对策略只是挖掘股票信息的临时成分，但可以理论证明利润并不只是像文献所述的仅有均值回归，对此也将该论文算法层面陈述。

代码实现

该论文的实验实现以Python为主，代码将兼顾风险因子，以便合理研究结果的稳定性，即不只包含广泛使用的因子如价格数据，同时也兼顾低频的机构因素如破产风险，均加以量化。另外，在后台模拟测试（back test）代码中也会考虑金融微观因素如买入与卖出滑移（slippage），卖空利率（short-selling interest）和交易成本（transaction cost），更趋近真实交易。

实验分析与金融应用

考虑现实因素会降低超额收益的大小，但在初步实验看来，配对交易依然我的样本数据和改进模型中依然保持着盈利性，精确地说，在近十年的后台测试中有八年都维持着正回报。

在基于以上算法的模型构造和代码实现后，我将进入实验分析阶段进行各模型的搭配模拟测试效果，包括改造回归模型，多种数据挖掘算法对配对进行搜索选择，多种机器学习算法对高利润型配对进行筛选，以及对量化因素可靠性的时间序列分析，最终以金融分析中的夏普比率（Sharpe ratio）进行策略优劣的比较。完成该部分试验后，将在论文中介绍收益性效果最好（即夏普比率最高）的算法搭配组合和策略模型，展现金融分析报告。