北京工商大學

硕 士 学 位 论 文

配对交易策略中交易信号动态化研究

学科专业 ： 数量经济学

研究方向 ： 金融投资计量

作者姓名 ： 由林青

指导教师 ： 王琴英 副教授

所在学院 ： 经济学院

二〇一八年五月

Research on Dynamics of Trading Signals in Trading Strategy

Dissertation Submitted to

Beijing Technology and Business University

in Partial Fulfillment of the Requirement

for the Degree of

Master of Economics

by

You Linqing

（Econometrics）

Dissertation Supervisor：Wang Qinying

May 2018

摘 要

配对交易策略是一种市场中性策略，被广泛应用在金融市场中，其基本思路是通过统计历史数据，寻找不同资产之间的价格变化规律，来建立资产的回归模型。与传统方法不同，配对交易策略关注的是不同资产之间的相对价差，而不是绝对价差，因此，具有较低的风险。

配对交易主要分为配对资产构建和交易信号构建两个部分。配对资产构建的主要目的寻找价格变动相似的资产，来建立配对池，而交易信号的构建关注的是开仓和平仓的时机。

本文重点研究的是交易信号，在传统配对交易信号的基础上，通过引入遗传算法实现交易信号的动态化，具体来说，将传统交易信号模型产生的结果，作为动态化寻优的初始范围，采用遗传算法进行动态寻优。同时针对传统遗传算法，收敛速度慢、过早收敛和实时性不足的缺点，本文引入自学习遗传算法、基于Metropolis 准则的遗传算法和两级递阶的算法对传统遗传算法进行优化。最后，结合这几种改进遗传算法的特点，采用 Stacking 算法融合，提出一种优化后的交易信号动态化模型，同时，总结了程序化交易在应用配对交易策略中的一般流程。

本文的实证部分，以数字货币市场中的比特币和比特币现金、商品期货市场中的焦煤主力合约和焦炭主力合约为例，通过分析累计利润、累计成交量、夏普比率来验证本文改进后优化算法的实用性。

关键字：统计套利；配对交易；交易信号动态化；遗传算法

Abstract

Pairing trading strategy is a market-neutral strategy, It is widely used in financial markets. The basic idea is to build a regression model of assets which depends on statistical historical data in order to search for the law of price changes between different assets. Unlike the traditional methods, Pairing trading strategy is concerned with the relative spread between different assets, rather than the absolute spread, therefore, this method is with a lower risk.

Pairing transactions are mainly divided into two parts, one are paired assets and the other are trading signals. The main purpose of paired assets to find the relatively similar price changes in assets so as to establish a pooling. The construction of trading signals focused on the timing of opening and closing positions.

The focus of this paper are trading signals. Based on the traditional trading signals, the introduction of genetic algorithm is to achieve the trading signal dynamics, specifically, the results of the traditional trading signal model is regarded as the initial range of dynamic optimization, According to the characteristics of the optimization function, the genetic algorithm is used to search for the dynamic optimization. At the same time, Because of the shortcomings of traditional genetic algorithm, such as slow convergence, premature convergence and lack of real-time performance, we adapt genetic algorithm based on Metropolis criterion and two Hierarchical algorithm to optimize the traditional genetic algorithm. Finally, By combining the characteristics of these improved genetic algorithms with the Stacking algorithm, this paper proposes an optimized algorithm of trading signal dynamics. At the same time, combined with our own experience, this paper puts forward the application of programmatic trading in the application of the matching trading strategy General process.

The empirical part of this paper is based on the case of bitcoin and bitcoin cash in the digital money market, the coking coal and the coke in the commodity futures market. The paper analyzes the accumulated profits, the accumulated turnover and the Sharpe ratio to verify the practicality of the algorithm.

Key words: statistical arbitrage, pairing trading, trading signal dynamics; genetic algorithm

目　录

第 1 章 绪论

1.1 研究背景及意义

1.1.1 研究背景

著名经济学家法玛在 1965 年提出有效市场假说。但是金融市场中的很多现象是无法用有效市场假说进行解释的。比如当投资者对一些消息过度反应时，会引起金融资产的价格出现较大的波动，严重偏离均衡价格。然而，恰恰因此这一点，吸引大量投机者进入市场。我国期货市场起步相对较晚，商品期货交易量相对其它金融产品不是特别的活跃，很多投机者以亏损离场。随着现代统计学和计算机的发展，部分投资者开始运用统计方法对资产的价格进行建模，通过计算机发出交易指令，同时建立空头和多头来降低系统性风险，实现稳定的收益，由于商品期货市场本身存在做空机制，使得该方法能够在市场进行操作，投资者能够利用价格偏离的机会获利，使得价格在均值附近振荡，相对稳定。

配对交易作为统计套利中的一种重要的交易策略，被广泛应用在高频交易中，其基本思路是通过统计方法和历史数据，寻找不同资产之间的价格变化规律，来建立资产的回归模型，与传统方法不同，配对交易策略更关注的是不同资产的相对价差，而不是绝对价差，因此，这种策略更加稳定。

然而，目前对于配对交易的研究更多是在配对资产的选择、组合及交易的过程，对交易信号的研究比较少。传统的做法是将过去一段时间内均值加减 0.75 个标准差作为开仓和平仓的信号，但这种方法存在很多问题：

（1） 获取的信号有偏差：

如果采用静态阈值的方法，需要我们根据一段时间内的历史数据计算均值和标准差，但这个时间窗口是很难确定的，因为时间窗口内，资产如果出现较大的经济结构变化或者政策因素，那么获取的信号是有偏差的，这势必会导致盈利的降低，甚至亏损。

（2） 忽略了不同资产的基本面：

采用长期的静态阈值，存在一个重要的假设，即配对的资产的基本面是一样的，例如黄金和白银，当一种价格上涨时，另外一种价格也会上涨，但有的时候，由于基本面的影响，两种商品的价格可能出现相反的方向，而静态阈值恰恰忽略了这一点。

（3） 反应滞后

静态阈值是根据历史推算出来的，如果一个市场比较有效，我们很难单纯从历史的均值信息中，获得较好的回报，考虑到静态阈值的滞后性，很难根据当时的市场情况，快速反应，因此，会失去很多机会。

1.1.2 研究意义

目前，中国强调要“走出去”，更多的和世界接轨，那么就需要中国适当放松对资本市场的管制，让国际投资者进入中国的金融市场，因此，进行风险对冲是一个重要的投资理念，本文主要通过对配对交易中交易信号动态化的研究帮助个人投资者和机构投资者更好的进行理性投资和资产配置。

（1）理论意义：

目前，对于配对交易中交易信号动态化的研究比较少，更多的还是采用以 0.75 倍的标准差作为交易信号，本文通过引入改进的遗传算法，将配对交易方法和非线性动态最优相结合，对该领域的发展具有促进作用。

（2）现实意义：

本文结合遗传算法，实现配对交易中交易信号动态化的研究，能有效发现焦煤和焦炭之间价值的偏离，帮助投资者发现市场的机会，同时促进商品期货市场能够更加有效的存在。另外，通过在数字货币市场的实践，进一步验证了方法的有效性，为投资者选择数字货币资产提供了一种投资方法，其中获得的一些启示和结论，能够为广大投资者提供参考，同时，对监管层也具有一定的借鉴意义。

1.2 国内外相关领域研究综述

统计套利是一种基于模型的套利策略，通过从资产的历史交易数据找寻规律，发现两个或者两个以上的资产之间存在的套利机会，然后通过模型拟合资产变化规律，设定交易阈值，用计算机程序根据市场的实时信息自动发出交易信号而进行套利。统计套利具有较高的收益同时具有较低的风险，因为两种资产的价差被认为是与市场无关的，因此这种策略也被称为市场中性策略。统计套利策略主要分为四类，分别是距离策略、时间序列策略、随机控制策略和配对交易策略，其中，配对交易策略是其最常用的方法。配对交易策略主要利用两个相关资产的价差具有均值回复性来进行对冲，当两个资产的价差超过一定的阈值时，交易者买入一种资产同时卖出另一种资产，当价差收敛时，交易者做相反的操作来平仓，进而获取收益。

但是，在应用配对交易策略时存在两个主要的问题：第一是确定可以用来进行配对交易的资产。第二是确定两种配对资产的进场信号、出场信号和止损信号。

针对问题一，经济统计的集大成者 Alexander[1]，Alexander 和 Dimitriu[2][3][4]

经过多年的研究提出了完整的协整框架，为确定进行配对交易的最优资产奠定了重要的理论基础，但并未用现实的金融资产做统计回测模拟。Girma and Paulson[5]，在考虑交易成本后，应用协整模型对原油、汽油、石油期货产品进行套利，年收益达到 15%。这一模型的优势来源于其所选标的自身具有生产关系上的强相关性，另外，大豆及其制成品[6]、天然气和电价期货[7]，这两个配对交易也能获取超额收益，但金银间无套利机会[8]。Elliot[9]在美国股票市场上，以标普 500 中的股票为实证研究对象，进一步印证了协整理论在选择配对股票对的合理性，发现通过协整理论组合的配对资产的价差满足均值回复的 Ornstein-Uhlenbeck 过程。另外，Hong and Susmel[10] 将协整模型应用于股票市场，产生了 33%的年化收益，但其收益部分来自汇率升值[11]。Dunisetal[12]将协整法推进到了股票“更高频”交易中，标的选择为欧洲斯托克 50 成分股，配对股票两两之间的资金配比使用卡尔曼滤波进行估计，但该模型忽略了例如配对公司的杠杆比例差异带来的分离风险。Caldeira and Moura[13]使用 Engle-Granger 两步法，在巴西股票市场做了测试，并利用 Johansen 法则检验协整关系，但其选择配对股票的依据是样本内套利收益的夏普率而不是相关系数等。这一模型也有缺点，即 Engle-Granger 两步法及 Johansen 法则具有相关性，并没有起到两重检验保护的效果。Gutierrez and Tse[14]使用 3 支自来水公司股票做协整检验，发现收益来自于 Granger-follower，而不是 Granger-leader；在国内市场上，崔方达，吴亮[15]以上证 50 指数的成分股为样本，通过协整理论组合的配对股票对，取得了超过大盘的收益，并且最大回撤风险也远远低于大盘。在其他资产中，仇中群，程希骏[16]、戴进[17]分别印证了这种理论方法在股指期货和 ETF 上的可行性。尽管协整理论在股票、期货、ETF 等领域获得了较大程度的成功应用，但是，在其他领域，尤其是新兴的资产市场，应用该策略的相对较少。

针对问题二，最常用的方法是基于协整回归残差正态分布的模型。Frazzo[18]以两支股票配的残差的一倍标准差为交易信号，确定进场和出场的时机，以二倍的标准差作为止损的信号，其中假设残差的均值为 0。但这种假设存在很大的局限性，因为交易阈值是固定不变的，Cummins[19]通过对爱尔兰股票交易的数据研究发现，价差均值为零的假设在实际运用中过高地估计了模型的期望收益，过低的估计了模型的持续时间，造成了很大的误差。因此，基于残差非正态分布的假设，Boguslavsky 和 Boguslavskaya[20]，Mudchanatongsuk 等[21]，Montana 等[22]主要采用 ARMA、隐含马尔可夫的 ARMA 和非参数的方法来对交易信号进行建模，根据历史的数据，动态调整交易信号的阈值，使得交易信号动态化，来提高模型在实际操作的稳定性。在后一段时间内，对于交易信号动态化的主要研究转移到 Coupla 方法上，这种方法主要在形成期内计算配对的相关系数或协整标准，然

后计算配对股票收益序列的边际分布函数。对于收益边际分布函数，Stander 等[23]讨论了参数和非参数法两种方法来估计边际分布，Ferreira[24]和 Liew and Wu[25]则偏向于拟合参数分布函数。在得到边际分布函数后，即可确定合适的 copula 函数。Ferreira 仅使用了一个 Copula 函数，参数来自经典最大似然估计。Stander 等[26] 基于 22 个阿基米德 copula，运用 Kolmogorov-Smirnov 拟合度测试选出最佳 copula。Liew 和 Wu 则是从 5 个金融领域常见的 copula 中选择，对选出的 copula 函数计算条件边际分布，如果条件概率高（低）于 0.5，则认为该股票被高（低）估。当条件分布函数超过 5%时进行交易，一般一周后平仓，或者条件分布值回复到 0.5 时平仓。Copula 是一种很好的模拟复杂依存关系的模型，可以很好地确定交易时机，实现交易信号的动态化。但这种方法忽略了数据的时间结构。而近年来，人工智能的方法与时间序列分析相结合，开始应用到配对交易的信号建模中，例如神经网络，遗传算法，这些新的算法主要是用来调整交易信号的动态化，使其能够根据市场的变化，自动调节进场信号、出场信号和止损信号。Thomaidis[27]等运用 NN-GARCH 模型来实现参数的动态性，其主要思想，是根据前 N 个数据获取 M 步长置信区间内的价格，当价格低于置信区间的最小值时，选择买入资产，当价格高于置信区间的最大值，则选择卖出该资产。Stock、Watson[28]引入神经网络算法来动态确定 ARMA 的滞后阶数，使得模型能够动态适应当前市场状况。如何能更有效的引入机器学习算法并结合数据来实现交易信号的动态化是目前对于该领域研究的热点

1.3 研究内容

本文共分为六个部分。

第一部分是绪论，论证研究的可行性，并说明研究的背景和意义，在此基础上，对国内外学者的相关研究做出文献综述，指出本文将要研究的主要内容、结构和方法，并阐述了文章的创新点和不足；

第二部分系统介绍了配对交易的相关理论基础，包括配对交易的基本内涵和配对交易的步骤，为接下来的章节提供方法上的理论基础，同时指出传统方法在交易信号动态化方面研究的不足，为之后的部分做铺垫；

第三部分介绍了将交易信号进行动态化的一种方法——遗传算法，介绍了遗传算法的发展背景、历程以及基本原理和基本内涵，针对经典遗传算法中存在的收敛速度、实时性问题，介绍了几种主要的优化模型，包括：自学习遗传算法、基于 Metropolis 准则的遗传算法等。

第四部分，主要阐述本文收益率定义、风险控制和交易成本等，在此基础上，阐述了本文中程序化交易的整个流程，最后，本文以经典的遗传算法、自学习算

法、基于 Metropolis 准则的遗传算法为基础，建立交易信号模型，同时，结合上述方法，本文将这些算法和 Stacking 算法相结合，提出一种新的交易信号动态化的方法。

第五部分，主要针对数字货币市场的比特币和比特币现金、商品期货市场的焦煤和焦炭，分别从基本面和统计意义上论证了二者能够进行配对交易的合理性，进而建立回归模型，采用遗传算法对交易信号进行动态化调整，通过累计利润、累计交易次数和夏普比率来对不同指标进行纵向对比。

第六部分，对论文的整体思路、实证和结论进行总结、分析和展望。

1.4 创新点

本文主要研究是配对交易中交易信号的动态化问题，传统的方法在处理交易信号问题上主要是基于均值回复，一个重要的假设是回归方程的残差满足正态分布，将正负 0.75 倍的标准差加均值作为阈值的范围，本文在处理该问题时，主要引入多种不同的遗传算法，结合 Stacking 算法，生成多个弱学习器，最终合并成一个强学习器，每个弱学习器都有自己在传统遗传算法上的改进，通过将多种弱学习器结合，来将交易信号进行动态化，避免了单个模型的缺点，从而提高交易的频次和资金的利用率，实证结果显示，优化后的模型具有较好的效果。

第 2 章 配对交易的相关理论基础

2.1 配对交易的基本内涵

配对交易最早是在天体物理学家塔塔西里的指导下，由一群擅长量化研究的科研团队创造的一种用匹配组合的方式买卖股票的程序，其主要思想是价格回归均值，正如大多数人类的活动一样，均值回归的反转现象是一种强大的力量，它驱动着系统和市场的自我平衡，从 20 世纪 80 年代开始，在追求利润行为的驱动下，人们希望将行为模型化，而配对交易正是这一思想的成功实践。

配对交易的策略是在两种互为补充的力量驱使下形成的，而这两种相反的力量的主要来源是人类的贪婪和恐惧。在实践层面上，配对交易主要利用两个相关资产的价差具有均值回复性进行对冲，以获取两个资产的α收益，其核心假设是配对资产的价格具有均值回复性。

2.2 配对交易的步骤

2.2.1 构建配对组合

市场上的投资者需要在市场中找到存在相关关系的资产进行配对交易，根据不同的标准对资产进行配对，不同配对的资产对收益有很大影响。如果两个配对的资产不存在相关关系，会导致投资者出现较大的损失，国内外学者对于寻找配对交易的组合，主要是协整的方法。

协整的方法主要根据价格的平稳性来确定配对资产。资产的历史价格是时间序列的数据，对于时间序列的回归需要建立在时间序列平稳性的基础上，否则很容易出现“伪回归”现象，所谓的“伪回归”是指当两个变量本身并无内在的联系，但是随着时间表现出一致的趋势，对这两个变量建立回归方程，发现可决系数很高，但时间并不存在经济含义。在现实生活中，大部分的经济变量是非平稳的，为了保证对非平稳时间序列回归有意义，需要对其进行协整检验，如果两个非平稳时间序列通过了协整检验，说明两个变量之间存在长期稳定的关系，可以进行建模回归，而对于协整关系的检验，首先需要检验变量的平稳性。

2.2.1.1 平稳性检验：

若时间序列$\{x_t\}$满足如下条件，则称该变量是平稳的。

第一，均值$\mathrm{E}(x_t)=u$,其中u是与 t 无关的常数

第二，方差$\mathrm{Var}(\mathrm{x_t})=\sigma^2$，其中$\sigma$是与 t 无关的常数

第三，协方差$\mathrm{cov}(x_t,\ \mathrm{x_{t-1}})=\gamma_k$,即协方差只与时间间隔 k 有关，而与时间

t 无关

满足以上 3 个条件的时间序列称为平稳时间序列。对于时间序列平稳性的检验，一般采用如下方法：

设随机游走序列

$$\triangle x_t = \delta x_{t-1} + \mu_t,$$

其中μ_t是白噪声，因此

$$x_t = x_0 + \mu_1 + \mu_2 + \cdots\cdots + \mu_t \quad (2.1)$$

$$E(x_t) = \mathrm{E}(x_{t-1}) = E(x_0) \quad (2.2)$$

$$Var(x_t) = t\sigma^2 \quad (2.3)$$

由上式可知，方差与时间 t 有关，因此该时间序列不是平稳的，而随机游走的时间序列$x_t = x_{t-1} + \mu_t$可是看做是$x_t = \delta x_{t-1} + \mu_t$中$\delta$=1的特殊情形，这种情况下，我们称随机变量$x_t$有一个单位根。因此我们可以根据$\delta$的值来判断时间序列的平稳性，如果存在单位根，则时间序列是非平稳的，若不存在单位根则说明时间序列是平稳的。主要有以下两种：

第一种是 DF 检验法。这种方法的原理是针对$\triangle x_t = \delta x_{t-1} + \mu_t$，检验$\delta$是否为 0，当$\delta \neq 0$时，说明时间序列是平稳的；当$\delta$=0时，说明时间序列是非平稳的。其中检验统计量 t 满足 DF 分布，当 t 大于临界值是，时间序列存在单位根，为非平稳时间序列。

第二种方法是 ADF 检验。ADF 主要有 3 个模型：

$$\triangle x_t = \delta x_{t-1} + \sum_{t=1}^{m} \beta_t \triangle x_{t-1} + \varepsilon_t \quad (2.4)$$

$$\triangle x_t = \alpha + \delta x_{t-1} + \sum_{t=1}^{m} \beta_t \triangle x_{t-1} + \varepsilon_t \quad (2.5)$$

$$\triangle x_t = \alpha + \beta t + \delta x_{t-1} + \sum_{t=1}^{m} \beta_t \triangle x_{t-1} + \varepsilon_t \quad (2.6)$$

以上 3 个模型的原假设都是δ=0，若 3 个模型都拒绝原假设，则停止检验，该时间序列是平稳的。

2.2.1.2 协整检验

经过变量平稳性检验之后，要确定变量的单整阶数。随机游走序列$x_t = x_{t-1} + \mu_t$，经过差分后变为：$\triangle x_t = \mu_t$，当$\triangle x_t$是平稳序列时，我们把这种经过一次差分后变成平稳的序列叫做 1 阶单整。如果序列经过 d 次差分后变为平稳序列，则称为 d 阶单整序列，记为 I(d)

假设两种资产价格存在线性相关关系：$y_t = \alpha + \beta x_t + \mu_t$，其中$\mu_t$是随机干扰项，$x_t$是一种资产的价格，$y_t$是另一种资产价格。上式表明，如果两种资产的价格相互影响，当两者的价差偏离长期均衡的位置，那么存在一种潜在的均衡机制对其进行调整，使其恢复到均衡的位置。如果从 t-1 时期到 t 时期，x_t的变化为$\triangle x_t$，

那么y_t变化值的期望是$E(\triangle y_t) = \beta \triangle x_t + \nu_t$，其中$\nu_t = \mu_t - \mu_{t-1}$，若$\mu_t$是白噪声序列，根据均衡机制，如果在 t-1 期末，y_{t-1}比正常均值小，则第 t 期$\triangle y_t$大于$E(\triangle y_t)$；如果 t-1 期末，y_{t-1}比正常均值大，则第 t 期$\triangle y_t$小于$E(\triangle y_t)$。这也意味着y_t相对于均值的变化是暂时的。但是前提是μ_t是白噪声序列，如果随着时间μ_t存在趋势变化，则y_t对均值的偏离随着时间会有积累，导致最终偏离均衡点。如果x_t、y_t满足以下条件：

$$y_t = \alpha + \beta x_t + \mu_t \tag{2.7}$$

μ_t是白噪声序列

我们称x_t和y_t存在协整关系，也就是说即便x_t、y_t是非平稳的时间序列，对二者建立回归方程也是具有统计意义的，对于两个变量之间的协整检验，恩格尔和格兰杰仔 1987 年提出 EG 检验法，检验的方法主要分为两步：

第一，采用普通最小二乘法估计下列方程

$$E(y_t) = \alpha + \beta x_t \tag{2.8}$$

第二，计算残差

$$\varepsilon_t = y_t - E(y_t) \tag{2.9}$$

然后，检验ε_t的平稳性，采用 ADF 法对ε_t进行检验。如果ε_t是平稳序列，则说明，二者之间存在协整关系。

综上，检验变量之间的协整关系第一步是检验变量的平稳性，如果变量是平稳的，可直接采用普通最小二乘法建立回归方程，如果变量不平稳，则需要进一步检验单整阶数，当单整阶数相同时，可以对变量进行协整关系检验，如果满足协整关系，则二者存在长期均衡关系，可以建立回归方程，否则不能建立回归模型。

2.2.2 制定交易标准及交易头寸

在选择配对资产之后，投资者需要进行投资交易。一般而言，主要分为以下几个过程：追踪两个资产价格之间的价差，当价差过大，达到交易阈值时，进行操作，买入价格被低估的资产，卖出价格被高估的资产。当价格差异减小到一定范围时，反向操作，完成交易；对于触发交易价格差的选择会直接对交易的收益率产生影响，因此，选取一个合适的触发信号十分重要。除此之外，若价格差没有收敛的趋势，需要设置止损点。

假如两种资产x、y已经通过了协整检验，即两种资产的价格满足方程$y_t - \beta x_t = \alpha + \varepsilon_t$，将两种资产的价格差用$\text{spread}_t = y_t - \beta x_t$表示，将价格差进行归

一化处理，用$ms\text{spread}_t$表示。设置一个阈值λσ交易信号，其中σ是价差的标准差，当$ms\text{spread}_t > \lambda\sigma$或$ms\text{spread}_t < \lambda\sigma$时，交易信号出发，具体而言：当$ms\text{spread}_t > \lambda\sigma$时，多头$\beta$单位 x，空头 1 单位 y，当$ms\text{spread}_t$回复到 0 左右时平仓，结束头寸。当λ的设置过小时，交易频率过快，可能导致收益率下降，当λ的设置过大时，可能导致很多交易机会流失，最终导致亏损。根据 Vidyamurthy 的研究，当价差序列ε_t符合正态分布时，将0.75倍的标准差作为交易信号的触发条件是最优的，能够获取到最大的利润。基于此，传统的交易信号，以$[\text{mean}(\varepsilon_t) - 0.75\text{std}(\varepsilon_t), \text{mean}(\varepsilon_t) + 0.75\text{std}(\varepsilon_t)]$作为交易信号，其中$\text{mean}(\varepsilon_t) - 0.75\text{std}(\varepsilon_t)$被称为下阈值，$\text{mean}(\varepsilon_t) + 0.75\text{std}(\varepsilon_t)$被称为上阈值。

由于价差序列$ms\text{spread}_t$是依概率收敛的，不一定是必然收敛的，因此，在交易的过程中，存在发散的可能，比如重大的政治事件或者自然灾害，因此，在交易的过程中，也要注重风险控制。回顾已有的研究，大部分文献在价差达到标准差的两倍时选择止损，即上止损点是3σ，下止损点是-3σ

2.3 小结

本章主要介绍了传统配对交易的基本方法，在保证金充足，时间足够长的情况下，该方法被市场证明是有效的，但中短期来看，并不总是有效，因为对于中短期而言，价差并不总是回归，另外，空头由于存在杠杆，理论上存在强平的风险，因此，改进的一个重要方向在于交易信号的动态调整。本文，以传统方法确认的最优交易信号作为遗传算法的初始范围，在中短期内，通过引入改进的遗传算法对交易信号实现动态调整，与传统方法相比，动态化捕捉短期内其他因素对交易的影响，从而增加收益。

第 3 章 遗传算法及其优化

本章主要介绍交易信号动态化采用的理论模型—遗传算法，介绍基本的遗传算法原理、基本求解过程和特点。

3.1 遗传算法的产生和发展

从 20 世界 50 年代，就有了关于“人工进化系统”的研究。这些研究大多用计算机来模拟生物系统，从生物的角度进行进化模拟、遗传模拟等方面的研究工作。这些早期的研究形成了遗传算法的雏形。

进入 20 世纪 60 年代后，美国 Michigan 大学的 J.Holland 教授在研究和设计自然和人工系统的自适应行为时，受费希尔的《自然选择的遗传理论》的启发，意识到可以模拟生物进化过程来进行遗传策略的设计。

1967 年，Holland 的学生探索了遗传算法在博弈论中的应用，并采用复制、交叉、变异、倒位等遗传操作手段进行国际象棋的对弈研究。1975 年，Holland 教授提出遗传算法基本定理—模式定理，奠定了该算法的理论基础。

进入 20 世纪 80 年代后，遗传算法成为智能优化算法研究的一个热点，研究人员在理论和应用方面对遗传算法进行了大量的研究。1991 年，Davis 出版的《遗传算法手册》，介绍了遗传算法在科学计算、工程技术和社会经济领域中的大量应用实例。自此，遗传算法被广泛应用在生物、工程技术、模式识别和社会科学等领域中。

3.2 遗传算法基本原理

作为演化算法之一，遗传算法的基本思想是将各类实际问题抽象为染色体，通过二进制表示，并采用复制、交叉和变异等遗传算子的操作，根据目标函数搜索全局最优解。

遗传算法中关键的步骤是编码方法和遗传算子的设计。通过编码将求解问题映射为染色体结构，根据该结构进行复制、交叉、变异等操作。

3.2.1 复制算子

复制算子用于从一个旧种群中选择优秀的个体产生新的种群。在算子的整个过程中，需要根据适应值函数，也就是目标函数计算种群中每个个体的适应度，依据概率从这些满足条件的个体中，确定优良个体。其中，复制中的主要分配方法包括：比例适应度分配、排序适应度分配。而优良个体的选择包括：轮盘赌选择法、最佳个体保留法、排序选择法和锦标赛选择法。

（1）轮盘赌选择法:假设群体规模为n,个体i的适应值为f_i,则个体被选择的概率P_{si}是：

$$P_{si} = \frac{f_i}{\sum_{j=1}^{n} f_i} \tag{3.1}$$

该方法的思想是根据选择概率将圆盘分成 n 份，若某个参照点落入第 i 份，则选择个体 i 进入下一代。

（2）最佳个体保留法：该方法是将群体中适应度最高的个体直接复制到下一代，保留最优的个体

（3）排序选择法：计算每个个体的适应值后，根据适应值大小在群体中对个体进行排序，选择前 n 个个体进行下一代

（4）锦标赛选择法:从群体中任意选择一定数目的个体作为锦标赛规模，其中适应度最高的个体保留到下一代，不断循环，直到下一代的数目达到预先设定的值

3.2.2 交叉算子

交叉过程则是结合父代信息产生新个体的过程：交叉算子的操作过程主要是按照交叉概率从群体中随机选择两个个体进行交叉操作，从而产生新的基因组合。交叉操作包括单点交叉和多点交叉。

3.2.3 变异算子

变异算子则是以较小的概率改变某些个体的基因位，一般分为二进制变异和其他变异，其中，二进制变异主要是将基因进行移位，其他变异包括插入、移位等操作。

⑴插入变异：随机从染色体随机选择一个基因，插入到染色体的其它位置。

⑵移位变异：随机选择一个序列，移动到染色体的其它位置。

3.3 遗传算法的求解过程

遗传算法求解过程的基本步骤如下：

（1） 初始化一组候选解群体作为第 0 代。

（2） 使用复制算子生成后代。

（3） 根据交叉概率，将候选解群体作用交叉算子，生成新的候选解。

（4） 根据变异概率，作用候选解。

（5） 计算群体中候选解的适应度，如果满足条件，跳出解，否则返回步骤（2）

遗传算法的操作流程如图 3.1

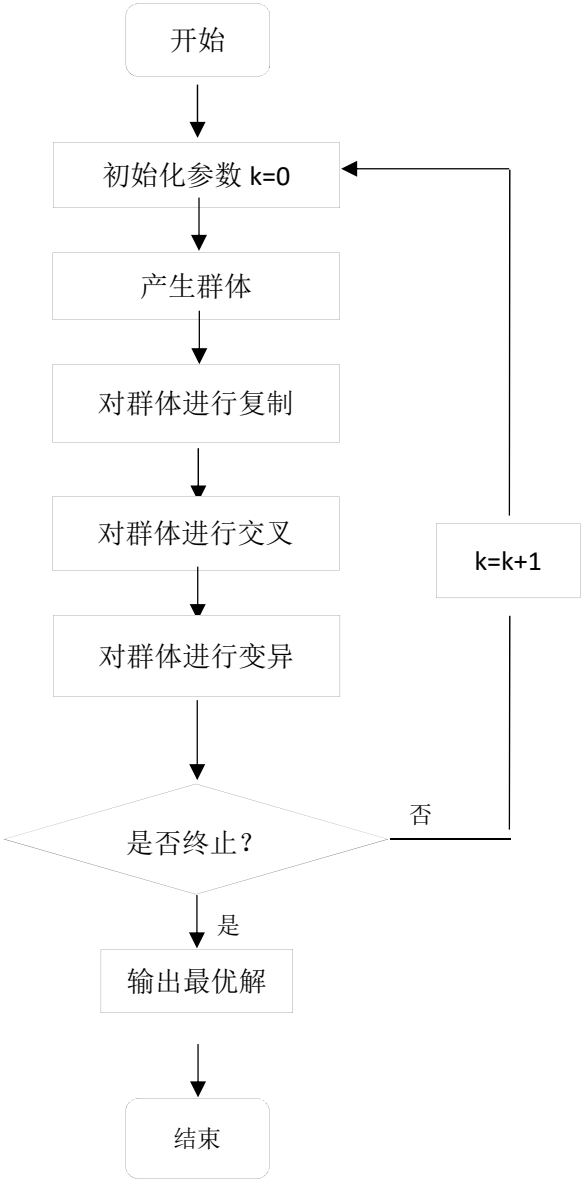

图 3.1 遗传算法流程图

3.4 遗传算法基本特点

从遗传算法的过程可以看出，遗传算法与传统优化算法有所不同，现将差异列如表 1。

表 3.1 遗传算法与传统优化算法的差异

算法 / 名称	传统优化算法	遗传算法
表示方法	解空间	编码
起始点	一个点	一组解
利用信息	导数等	目标函数

转移规则	确定性	随机性

由表 3.1 可知，遗传算法具有如下特点：

（1） 自适应性、自学习性、自组织性：算法能够组织搜索，适应度大，保存到下一代的概率就大，能够适应环境的变化。

（2） 整个求解过程只需要目标函数，对于目标函数，可以是隐性表达式。

（3） 强调概率的转移，并不借助导数。

（4） 有利于解决多个解的复杂问题。

与此同时，和其他优化算法相比，遗传算法具有并行性，适用于目前的分布式计算，同时易于和其它技术，如神经网络相结合，形成性能更优的求解方法。虽然，遗传算法有很多优点，但是，存在两个比较显著的问题：

（1） 寻优的过程中可能收敛的速度过慢，导致整个迭代时间过长，陷入局部。

（2） 过早收敛，前期出现超级个体，导致后期停滞不前。

针对以上问题，对遗传算法有了改进的算法。

3.5 自学习遗传算法

根据遗传原理可知，优秀的父代以较大概率产生优秀的子代。若父代通过一定的方式进行学习以提高自身的性能，经过复制后以较大的概率使子代的性能提高。结合该原理，针对传统遗传算法存在收敛慢的问题，在最优保存传统遗传算法的基础上引入学习算子。

3.5.1 基本概念

定义 4.1 对于串x，如果满足条件$f(x) > \bar{f}$,其中$f(x)$是串x的适应度，$\bar{f}$是群体平均适应度，则称串x是群体$P(K)$的最优串。

定义 4.2 对于优良串$x_1, x_2, \dots, x_k$,如果满足：

$$x_{im} = x_{jm} \quad \forall i, j \in \{1,2, \dots, k\} \text{ 且 } i \neq j \tag{3.2}$$

其中，x_{im}为串x_i的第m位。则称m是K个优良串$x_1, x_2, \dots, x_k$的一个优良位。

定义 4.3 对于模式H,如果满足$\bar{f}(H) > \bar{f}(\hat{H})$,其中$\hat{H}$是群体中任意一个模式，$\bar{f}(H)$是模式$H$的平均适应度值。

3.5.2 优良位搜索算法

根据定义可知，优良模式是群体中性能最优模式，为了寻找最优模式，需要首先找到优良位，最优保存传统遗传算法通过在传统算法加入保存机制，来寻找最优位，其结构如下：

BEGIN

初始化参数P_c, P_m, N;

产生初始种群$P(K); K = K + 1$;

根据适应度函数，计算$P(K)$中串的适应度;

计算当前串中最优适应度的串，作为 solution;

REPEAT

将复制算子、交叉算子、变异算子作用到$P(K)$;

计算$P(K)$中串的适应度;

IF $solution < P(K)$中最优串的适应度;

$solution := P(K)$中最优串的适应度

$K = K + 1$

UNTIL(满足条件);

输出最优的串和 solution

END

3.5.3 优良模式搜索算法

根据定义 4.1、定义 4.2 能够找到编码群体中的优良串和优良串对应的优良位，进而能够确认一种优良模式。例如，G1=10100, G2=00101 为优良串，则第 2 位～第 4 位是串G1、G2的优良位，*010*为优良模式。为了寻找优良模式，算法结构如下：

BEGIN

假设群体规模为$N, x_1, x_2, \ldots, x_n$是群体中N个串;

将优良模式的数组初始化为-1;

REPEAT

IF 存在优良位

BEGIN

将优良位上的值替换初始数组中相应的位;

RETURN;

END

ELSE K=K-1

UNTIL(满足条件)

END

3.5.4 自学习算法

通过优良模式搜索算法得到的优良模式是在群体中含有一定数目的优良串中所有的优良位确定的模式，并且这些模式是群体中最优测串，进而保证了找到的模式是最优的模式。在数组中，如果某一位不是－1，则说明该位置是确定的，否则该位置不确定，如果找不到优良模式，我们认为群体中个体的性能差不多，已经没有改进的空间。为了让群体具有自学习功能，定义学习算子如下：

设x是群体中低于平均适应度的串，p_l是学习概率，则学习算子S表示如下：

$$S: (x, H) \rightarrow y$$

$$y_i = \begin{cases} x_i & rand(0,1) > p_l \\ H_i & rand(0,1) \leq p_l \end{cases} \tag{3.3}$$

其中，rand(0,1)是(0,1)之间的随机数。

根据学习算子的定义可知，x向模式H学习之后变成y，y以概率p_l进入优良模式，使得算法的效率得到提高，其算法结构如下：

```
BEGIN
    初始化参数Pc, Pm, P1, N;
    产生初始种群P(K); K := 0;
    确定P(K)中串的适应度;
    解solution := P(K)中最优串的适应度;
    REPEAT
        得到P(K)中的优良模式 H;
        IF(H 存在)
            对群体P(K)中性能较差较差的串作用学习算子;
        对群体P(K)作用复制算子、交叉算子和变异算子
        K: K + 1
        确定P(K)中串的适应度;
        IF (solution) < P(K)中最好串的适应度
            solution := P(K)中最好串的适应度
    UNTIL(满足条件);
    输出solution
END
```

3.6 基于 Metropolis 准则的遗传算法

自学习算法为算法的收敛速度慢提供了一种解决思路，但除此之外，过早收

敛也是遗传算法面对的另一个挑战。造成过早收敛的一个重要原因是种群中个体的一致性，因此如何增加种群的多样性是解决该问题的一个重要思路，基于此，该部分根据 Metropolis 判别准则来提高选择的多样性。

3.6.1 Metropolis 准则的基本内涵

1953 年，Metropolis 根据固体在一定温度下达到热平衡的过程，提出一种新的采样方法，该方法称为 Metropolis 准则。该方法源于统计物理学中对固体退火的研究，给定粒子的初始状态i,记此时的能量是E_i,随机改变该粒子的位置信息，达到一个新的状态j,此时该粒子的能量是E_j。如果$E_j < E_i$，则称新的状态为重要状态，因此若$E_j > E_i$，新状态是否重要，可以根据以下进行判断：

$$Z = e^{\frac{E_i - E_j}{kT}} \tag{3.4}$$

其中，Z小于 1，T是温度，k是常数。

随机产生一个介于 0 和 1 之间的随机数r，若$Z < r$,则认为新的状态是重要的。若新状态j是重要状态，则更新当前的状态为j，否则不变，重复以上过程。在大量的随机改变后，系统趋向于稳定的状态，该状态的概率分布依概率收敛到Gibbs正则分布。

上述的接受准则被称为 Metropolis 准则。

3.6.2 Metropolis 准则下的复制算子

在传统的遗传算法中，复制策略是根据适应度进行的，适应度高的个体入选下一代的概率很高，而性能差的个体几乎不会进入下一代，这样导致的结果是种群可能过早的产生单一性的问题，陷入局部解，为了尽量避免该问题的产生，保证算法的收敛到全局最优解，本文设计了一类改进的复制算子。

设群体的数量是N,经交叉、变异后产生的种群数目为2N,而且前 N 个是父代，后 N 个是子代，复制过程分为以下两步：

将中间过程中产生的最优个体进行保存，保证最优解不会丢失。

在父代群体和子代群体中，随机选取个体i和个体j,根据 Metropolis 准则，则个体i和个体j,进入下一代的概率是：

$$f(i) = \begin{cases} 1, & f(i) \geq f(j) \\ e^{\frac{f(i)-f(j)}{T}}, & others \end{cases}$$

$$f(j) = \begin{cases} 0, & f(i) \geq f(j) \\ 1 - e^{\frac{f(i)-f(j)}{T}}, & others \end{cases}$$

其中，f(i)、f(j)是个体的适应度，在作用复制算子之后，T乘以衰减系数α，使T值降低。

上式表明，如果$f(i) \geq f(j)$，则个体i进入下一代，j被淘汰；如果$f(i) < f(j)$，则个体i以概率$e^{\frac{f(i)-f(j)}{T}}$进入下一代，个体j以概率$e^{\frac{f(i)-f(j)}{T}}$被淘汰，当迭代足够多的次数时，

$$e^{\frac{f(i)-f(j)}{T}} \to 0$$

此时，个体i和个体j进入下一代的概率是：

$$f(i) = \begin{cases} 1, & f(i) \geq f(j) \\ 0, & others \end{cases}$$

$$f(j) = \begin{cases} 0, & f(i) \geq f(j) \\ 1, & others \end{cases}$$

综上所述：该过程的算法如下：

```
BEGIN
    将最优的个体直接进入下一代;
    在父代和子代随机选择个体i和个体j
    IF f(i) ≥ f(j)
        BEGIN 个体i入选下一代
        END
        T = T × α
    ELSE
        IF rand(0,1) ≤ min(1, exp(f(i) − f(j))/T)
            BEGIN 个体i进入下一代
            END
        ELSE
            BEGIN 个体j进入下一代
            END
        T = T × α
END
```

3.7 两级递阶遗传算法

在人类进化的整个过程中，地理因素是形成不同种群的重要因素，正如我们看到的不同国家、不同民族，由于气候、环境等因素，导致不同的群体在不同的区域发展，形成不同的民族特色，促进着社会的发展和进步，由于各个地区资源的不同，导致其进化程度有很大区别，例如世界上有发达国家和发展中国家一样，但不同的地区也并不是完全隔断的，例如，可以通过交通工具进行交流和学习，从而促进本地的发展。

根据以上过程，设计了两级递阶遗传算法，其基本原理：在搜索的过程中，不再采用单一的一个种群，而是采用多个种群同时进行，每个种群相当于一个地区，每个种群在自己的地区单独发展和进化，每个种群进化一代之后，会相互交流，选择最好的解来帮助其它种群进行进化，其基本的框架如下：

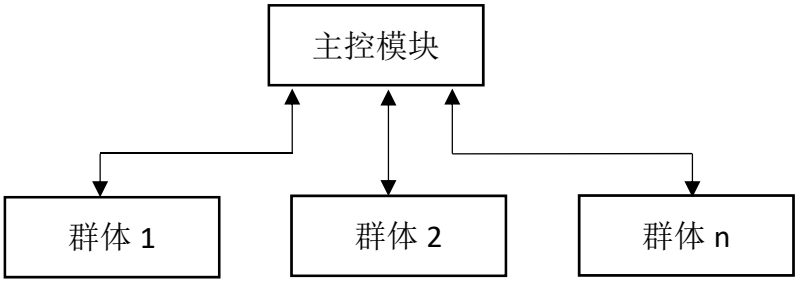

图 3.2 两级递阶遗传算法框架

算法的整个过程分为两步：

第一步：群体单独进行，群体 n 分别用$P_1, P_2, \dots, P_n$表示，n 个群体进化一代后得到的最好的解是$X_1, X_2, \dots, X_n$。适应值分别是$f(X_1), f(X_2), \dots, f(X_n)$,每个群体进化之后进行一次遗传操作。

第二步：在第一步的基础之后，主控模块选择出每个群体中的最优解，将最优解随机替换到每个种群中，进行下一次进化。

重复以上两步，直到达到迭代停止条件。通常而言迭代停止有 3 种条件：

（1） 迭代次数达到预先设置的次数

（2） 迭代误差小于设定的误差范围

（3） 寻找到全局最优解

3.8 小结

本章主要介绍了遗传算法的相关概念，以及经典的遗传算法、自学习遗传算法和基于 Metropolis 准则的遗传算法，其中，自学习遗传算法和基于 Metropolis 准则的遗传算法主要用来解决经典遗传算法在逼近最优解的过程中，存在的收敛速度慢和过早收敛这两个问题。本章通过介绍这几种遗传算法，主要是为下一章

节交易信号动态化部分提供理论基础，通过引入遗传算法来实现交易信号动态化是本文的主要创新点和工作。

第 4 章 交易信号动态化的优化模型

在本章中，主要根据现有的方法提出一种交易信号动态化的优化模型，同时，建立该方法下程序化交易的一般流程。

4.1 风险控制和收益率

4.1.1 风险控制

在整个交易系统中为了加强风险控制，需要将价格偏离控制在一定的范围之内，一旦出现较大的偏离，需要及时止损。目前，止损信号更多的还是采用 3 倍标准差的思想，即当残差满足正态分布时，有 95%的概率保证其偏离均值的程度不会超过 1.96 倍的标准差，当价差近似正态分布时，价差偏离均值 3 倍标准差是小概率事件，如果出现这样的信号，可能是受到外部结构性影响因素或者出现较大的经济事件，导致短期内价差很难向均值回复，如果资金存在杠杆，发生爆仓的概率会增加，从资金安全性考虑，此时，应该进行止损操作。

4.1.2 收益率计算

本节采用夏普比率作为收益率－风险度量指标，夏普比率是衡量金融资产安全性和盈利性的综合指标，其核心思想是：在市场有效的前提下，理性投资者将持有有效的投资组合，即在风险最小的约束下，期望收益最大，其定义为：

$$SR = \frac{E(R_P) - R_f}{\sigma_p}$$

其中，$\mathrm{E}(R_P)$:投资组合的预期报酬率

R_f: 无风险利率

σ_p投资组合的标准差

该公式能够反映出投资组合每承担一单位的总风险，能够产生多少的超额报酬。

4.2 交易信号动态化模型

4.2.1 理论依据

每一种遗传算法都有自己的优缺点，应用在不同的数据类型中会产生不同的

效果，如何有效的结合各种模型的特点，进行模型融合，是本文对遗传算法优化的一个主要思路。

进行模型融合的主要理论基础是集成学习，集成学习是统计学中一个非常重要的分支，其基本思想是用多个弱学习器构成一个强学习器。之所以需要进行集成，是因为每个弱学习器都存在一定的差异性（例如参数空间，目标函数等），这就会导致最终的决策结果有所不同，也就是说弱学习器本身也会犯错误，最终导致预测效果的准确度大大降低。因此，根据集成学习的理论，将多个弱学习器进行整合，能有效地在一定的准确率范围内，提高模型的泛化能力，但是并不是所有模型都可以进行集成学习，一般情况下，要求每个弱分类器的错误率低于0.5。

常见的集成学习算法包括：Bagging、Boosting、Stacking。

Bagging 方法是采用有放回的方式进行抽样，用抽取的样本建立子模型，对子模型进行迭代和训练，重复这个过程多次，最后进行模型融合，具体过程是在一个包含 m 个样本的数据中，随机采样，则每次样本被采集的概率是$\frac{1}{m}$，不被采集的概率是$1-\frac{1}{m}$。如果 m 次采样都没有采集中的概率是$(1-\frac{1}{m})^m$。当$m \to \infty$时，$(1-\frac{1}{m})^m \to 0.368$，也就是说大约有 36.8%的数据，没有被采样集采中，因此，通过有放回的重复采样过程来提高模型的泛化能力。

Boosting 方法能够用来提高弱分类算法的准确度。这种方法主要构建一组预测函数集合，通过一定方式将集合的函数进行组合，构建一个新的预测函数。它是一种框架算法，通过对样本集的操作来获取样本子集，然后采用弱分类算法在样本集合上生成一系列基分类器。通过基分类器来提高其他弱分类器的准确度，采用 Boosting 算法得到不同训练样本集，用该样本子集去训练生成基分类器，针对每一个得到的样本用该基分类算法在该样本上产生一个基分类器，训练样本 n 次后，就可产生 n 个不同的基分类器，在 Boosting 算法之下，对这些不同的基分类器分配不同的权重，产生最后一个决策分类器，在这不同的 n 个基分类器中，每一个分类器的准确度可能很低，但是将他们的结果联合起来之后，可能显著的提高模型最终的准确度，通过该方法，可以提高模型整体的效果。

Stacking 方法是将多种分类器组合在一起，进而取得更好表现的一种集成学习模型。一般情况下，该模型分成两层，第一层是训练多个不同的模型，然后以第一层训练的各个模型输出作为输入来对第二层模型进行训练，以得到一个最终的输出结果。假设给定包含n个样本的数据集$\theta = \{(x_i, y_i), i = 1, 2, \dots, n\}$，其中，$x_i$是第 i个样本的属性向量，y_i是第 i个样本的真实值，一共有 t 个不同算法的学习器。为了避免过拟合问题，通过交叉验证的方法构建第二层数据集，其具体做

法是将数据集按照一定的比例分成训练集和测试集，在训练集训练 t 个算法，得到 t 个模型，然后在测试集进行预测，重复该过程，则 t 个样本，并存在对应的预测值和真实值，通过这些预测值和真实值建立第二层数据集合，形成最终的决策模型。

本文主要采用 Stacking 的方法对各个遗传算法进行整合，形成多层遗传算法，最终形成对交易信号动态化的优化。

4.2.2 模型框架

本节在传统方法的基础上，通过 Stacking 方法将常见的传统遗传算法、自学习遗传算法、基于 Metropolis 遗传算法进行有效整合，在原有模型的基础上对交易信号模型进行优化，通过不断迭代，在避免模型过拟合的前提下，最大化预测的准确度。

其基本框架如下：

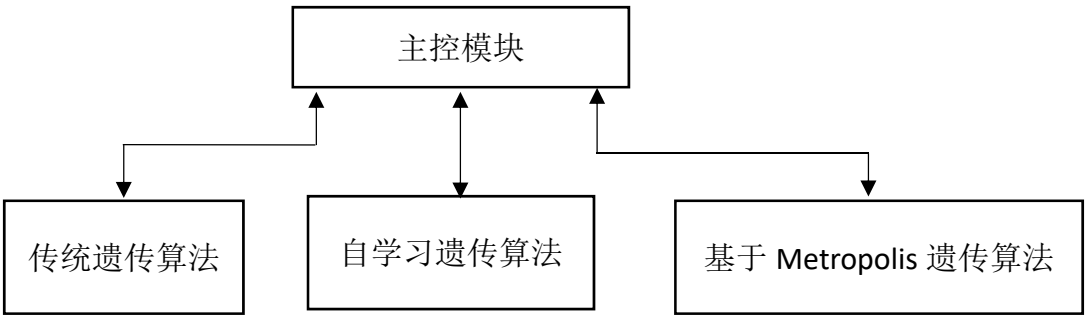

图 4.1 优化模型框架

4.2.2.1 主控模块

主控模块采用 Stacking 的方法进行模型融合，模型采用二层结构，为了避免模型的过拟合问题，采用 5 折交叉验证，其过程如下：

输入向量为 3 个模型预测的最优信号值:$\hat{y} = (\varepsilon_{t1}, \varepsilon_{t2}, \varepsilon_{t3})$,

其中：

ε_{t1}：传统遗传算法对第 t 期最优信号的预测；

ε_{t2}：自学习遗传算法对第 t 期最优信号的预测；

ε_{t3}：基于 Metropolis 遗传算法对第 t 期最优信号的预测；

输出值为当期的最优信号值:$y = \varepsilon_t$

其中，当期最优信号值根据过去一个小时的市场深度信息来定义，使得利润最大化的交易信号阈值,该值采用遍历的方法获取。

令：

$$\mathrm{Y}=\begin{pmatrix}\varepsilon_t\\ \varepsilon_{t-1}\\ \vdots\\ \varepsilon_{t-n}\end{pmatrix}\quad \mathrm{X}=\begin{pmatrix}\varepsilon_{t1} & \varepsilon_{t2} & \varepsilon_{t3}\\ \varepsilon_{t-1,1} & \varepsilon_{t-1,2} & \varepsilon_{t-1,3}\\ \vdots & \vdots & \vdots\\ \varepsilon_{t-n,1} & \varepsilon_{t-n,2} & \varepsilon_{t-n,3}\end{pmatrix}$$

将数据集按照 7:3 的比例划分训练集和测试集，对训练集采用 5 折交叉验证建立模型，其过程如下：

将$\boldsymbol{X}$的训练集随机分成 5 个子样本，定义为 train1、train2、train3、train4、train5

$$\underbrace{\begin{pmatrix}\varepsilon_{t1} & \varepsilon_{t2} & \varepsilon_{t3}\\ \vdots & \vdots & \vdots\\ \varepsilon_{t-int\left(\frac{5}{n}\right),1} & \varepsilon_{t-int\left(\frac{5}{n}\right),2} & \varepsilon_{t-int\left(\frac{5}{n}\right),3}\end{pmatrix}}_{train1}$$

.

.

.

$$\underbrace{\begin{pmatrix}\varepsilon_{t-4int\left(\frac{5}{n}\right)+1,1} & \varepsilon_{t-4int\left(\frac{5}{n}\right)+1,2} & \varepsilon_{t-4int\left(\frac{5}{n}\right)+1,3}\\ \vdots & \vdots & \vdots\\ \varepsilon_{t-n,1} & \varepsilon_{t-n,2} & \varepsilon_{t-n,3}\end{pmatrix}}_{train5}$$

用 train1、train2、train3、train4 训练传统遗传算法模型，将模型应用到 train5 上，获得预测值:

$$\tilde{y}=(\varepsilon_{t-4int\left(\frac{5}{n}\right)+1,1},\varepsilon_{t-4int\left(\frac{5}{n}\right)+2,1},\varepsilon_{t-4int\left(\frac{5}{n}\right)+3,1},\varepsilon_{t-4int\left(\frac{5}{n}\right)+4,1},\varepsilon_{t-4int\left(\frac{5}{n}\right)+5,1}),$$

用 train1、train2、train3、train5 训练传统遗传算法模型，将模型应用到 train4 上，获得对应的预测值，重复以上过程 5 次，获得传统遗传算法下，对应的预测值，记为：

$$\tilde{\mathrm{Y}}_{1,1}=\begin{pmatrix}\varepsilon_{t1}\\ \varepsilon_{t-1,1}\\ \vdots\\ \varepsilon_{t-n,1}\end{pmatrix}$$

重复以上过程 100 次，获得$\tilde{\mathrm{Y}}_{1,1},..,\tilde{\mathrm{Y}}_{100,1}$,记

$$\tilde{Y}_1 = \frac{1}{n+1}\sum \tilde{Y}_{1,i} \qquad i = 0,1,..n$$

同理，针对自学习遗传算法和基于 Metropolis 遗传算法采用同样的过程，得到$\tilde{Y}_2$和$\tilde{Y}_3$，则最终预测值为 3 个弱分类器在各个位置出预测的均值，记为：

$$\tilde{Y} = \frac{1}{3}\sum \tilde{Y}_i \qquad i = 0,1,3$$

4.2.2.2 遗传算法模块

遗传算法的目标函数是利润最大化，限制条件包括：价格偏差控制在 3 倍标准差之内、总资金限制。

其中，在传统遗传算法下复制算子，采用轮赌选择法，变异算子和交叉算子采用概率值；自学习遗传算法在此基础上引入学习算子，在每一次进行复制算子、变异算子和交叉算子之前，首先在适应度值上作用学习算子，使得整个模型继承上一次迭代的优良特性，更好的逼近最优值，其中学习算子的数学表达式为：

$$S:(x,H) \to y$$

$$y_i = \begin{cases} x_i & rand(0,1) > p_l \\ H_i & rand(0,1) \le p_l \end{cases} \qquad (4.1)$$

其中，$rand(0,1)$是$(0,1)$之间的随机数，p_l是y进入优良模式的概率。

而基于 Metropolis 准则下的遗传算法，主要将复制算子进行改进，引入 Metropolis 准则，来将一些性能表现没那么突出的个体也引入下一代，进一步提高泛化能力，最大程度的获取全局最优解。

4.2.2.3 遗传算法流程

算法流程如下：

（1） 初始化一组候选解群体作为第 0 代。

（2） 采用 Stacking 算法，随机采样一部分数据，针对该部分数据分别作用传统的遗传算法、自学习遗传算法和基于 Metropolis 遗传算法。

（3） 采用每种遗传算法下的种群单独进化，产生优良的个体。

（4） 针对每一代进化的结果，通过主控模块，选择最优的个体，将最优个体的编码随机替换成进入下一代的个体。

（5） 针对其它个体，进行变异和重组，作用候选解。

（6） 计算群体中候选解的适应度，如果满足条件，跳出解，否则返回步骤（2）

4.3 程序化交易流程

第一步：检测整体仓位的保证金率，若低于 20%，则强行按照回归系数平掉

一部分仓位；直到保证率在 50%以上，检测价差是否在 3 倍标准差之内，若大于 3 倍标准差，则强行平仓。

第二步：检测账户中资产的持仓比例，若二者比例不是回归系数比例，则跳出程序，发送报警邮件。

第三步：采用多线程同时获取资产的行情数据，包括：二者的成交价格和十档盘口深度数据。

第四步：计算行情获取的时间，若大于 2 秒，返回第一步，否则继续。

第五步：根据成交价产生交易信号。

第六步：根据资产 1 的买一价，计算在阈值范围内资产 2 在买一挡位的安全垫的手数a_1；根据资产 1 的卖一价，计算在阈值范围内资产 2 在卖一挡位的安全垫的手数a_2；根据资产 2 的买一价，计算在阈值范围内资产 1 在买一挡位的安全垫的手数a_3；根据资产 2 的卖一价，计算在阈值范围内资产 1 在卖一挡位的安全垫的手数a_4；

第七步：当安全垫的手数，大于预先设定的水平，则在一种品种上挂限价单，数量为当前可开量的 10%，停留 0.2 秒，获取成交手数，根据之前回归方程计算的比例，按照市价单再另外一种品种上挂对应比例手数的单子，完成开仓操作。

例如：经过比较a_1最大，而且大于预先设定的安全垫，则在该资产上根据买一的价格（实际情况下，会按照买一价加 0.01 元）挂单，停留 0.2 秒，取消订单，获取已成交数目，若已成交 7.62 手，则按照配对资产买一的价格市价做空 1 手该配对资产。

第八步：当交易过程中出现委托单时，将委托单信息传递到第二步中的账户资产平衡检测中，同时对委托的资产 1 和资产 2 进行移动止损。

4.4 小结

本章基于传统配对交易理论和交易信号动态化模型，提出了经过优化后的模型，其主要思路是：以经典配对配对交易理论获得的阈值范围作为遗传算法的初始搜索值，通过建立经典遗传算法、自学习遗传算法和基于 Metropolis 准则的遗传算法，采用 Stacking 算法将这些弱分类器进行组合，实现模型的融合，从理论上证明了这种融合对于实时性和准确性的提高，同时又避免了过拟合现象，最后，提出了完整的程序化交易流程，方便进行操作。

第 5 章 实证分析

在本章中，将以数字货币市场中的比特币和比特币现金以及我国商品期货市场中的焦煤和焦炭为例，论证本文建立模型的有效性。

5.1 数字货币市场—以比特币和比特币现金为例

5.1.1 配对条件分析

5.1.1.1 基本面分析

2017 年 8 月，比特币由之前的单一加密数字货币分裂成现在的比特币和比特币现金，其二者的主要区别在于存储容量。比特币依然坚持 1M 的存储上限，而比特币现金将这一上限提高到 8M，所以说从基本面上来看，二者存在很强的相关性，当一种数字货币价格增长时，另外一种数字货币的价格也应该出现相同的趋势，从二者的成交价数据也能很好的印证这一点。

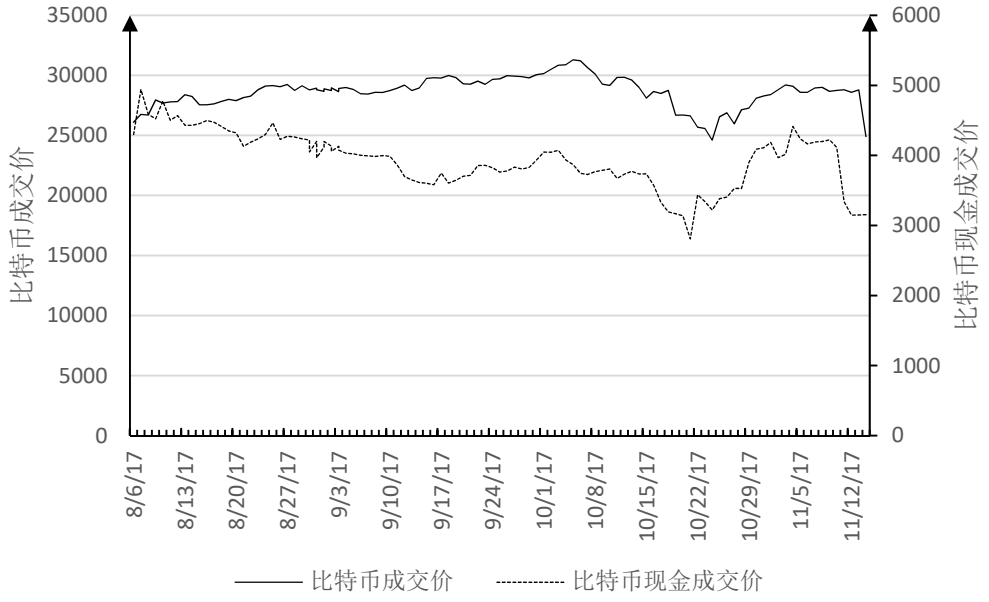

图 5.1 比特币和比特币现金成交价

如图 5.1，主坐标轴表示比特币成交价序列，次坐标轴表示比特币现金成交价序列，从图中的走势可以发现，二者存在很强的相关性。

5.1.1.2 统计分析

（一）单位根检验

根据协整理论，判断两个时间序列是否存在协整关系，首先需要进行平稳性

检验，其中单位根检验是一种主要方法。

令比特币现金成交价序列的自然对数是x，比特币成交价序列的自然对数是y，采用 ADF 检验，表 5.1 为比特币现金单位根检验的结果。

表 5.1 比特币现金成交价序列单位根检验

	置信水平	t 统计量	P 值
ADF 检验值		-2.65401	0.3215
	1%	-3.45786	
	5%	-2.87354	
	10%	-2.57324	

如表所示，该价格序列在 3 个置信水平下没有通过平稳性检验，即比特币现金成交价序列存在单位根。

表 5.2 比特币成交价序列单位根检验

	置信水平	t 统计量	P 值
ADF 检验值		-2.01493	0.2208
	1%	-3.45786	
	5%	-2.87354	
	10%	-2.57324	

如表所示，该价格序列在 3 个置信水平下也没有通过平稳性检验，即比特币成交价序列存在单位根。

考虑到两者都存在单位根，再对比特币现金成交价格和比特币成交价格序列进行一阶差分，再进行单位根检验。表 5.3 和表 5.4 分别是二者进行 ADF 检验的结果。

表 5.3 比特币现金成交价序列一阶差分单位根检验

	置信水平	t 统计量	P 值
ADF 检验值		-15.68493	0.0000

	1%	-2.574797	
	5%	-1.942176	
	10%	-1.615803	

表 5.4 比特币成交价序列一阶差分单位根检验

	置信水平	t 统计量	P 值
ADF 检验值		-19.91404	0.0000
	1%	-2.574797	
	5%	-1.942176	
	10%	-1.615803	

从上表可以看出，在 1%置信水平下，比特币现金和比特币成交价格的一阶差分序列均通过了 ADF 检验，这表明在 1%的置信水平下，二者的价格序列是一阶单整的，理论上存在长期协整。

（二）协整检验

根据上表的结果，采用回归方法建立比特币现金价格x和比特币价格y的回归方程，对方程残差进行协整检验。结果如表 5.5。

表 5.5 价格序列的回归分析

变量	系数	标准差	P 值
常数	1312.10112	25.02	0.0000
比特币现金	7. 6225173	0.1653	0.0000

其中，拟合优度为：74.85%

两个价格序列的回归方程是：

$$y_t = 1312.1 + 7.62x_t + \varepsilon_t$$

对回归方程的残差序列进行 ADF 检验，结果如下：

表 5.6 回归方程残差的 ADF 检验

	置信水平	t 统计量	P 值

ADF 检验值		-2.78403	0.0030
	1%	-2.574797	
	5%	-1.942176	
	10%	-1.615803	

残差序列在 99%的置信区间下是平稳的，因此,比特币现金和比特币成交价格序列存在协整关系。

因此，从基本面分析的结果和统计检验的结果来看，比特币和比特币现金之间满足配对交易的前提。

5.1.2 不同模型下的交易信号构建

在数字货币的现货市场中，并不存在直接的做空机制，一般而言，可以通过间接的方式实现做空，具体如下：

（1） 在交易平台 1 和交易平台 2 分别注册用户,分别充入 1 个比特币和 7.62 比特币现金等值的人民币到两个账户。

（2） 在平台 1 购买 1 个比特币，在平台 2 购买 7.62 个比特币现金

（3） 当残差超过上阈值，卖出一定数量的比特币，同时买入 7.62 倍的比特币现金；

当残差低于下阈值，买入一定数量的比特币，同时卖出 7.62 倍的比特币现金。

（一）经典配对交易模型的交易信号

通过经典配对模型,可以计算下阈值是 14.5,上阈值是 17.5,其中$\mathbf{mean}(\boldsymbol{\varepsilon_t}) = \mathbf{16}, \mathbf{std}(\boldsymbol{\varepsilon_t}) = \mathbf{2}$

本文以 2017 年 8 月 6 日～2018 年 2 月 7 日为样本数据进行回测，假设每一次限价单的成交比例为 30%，每天进行一次统计，初始资金是 14 万元，回测的结果如下：

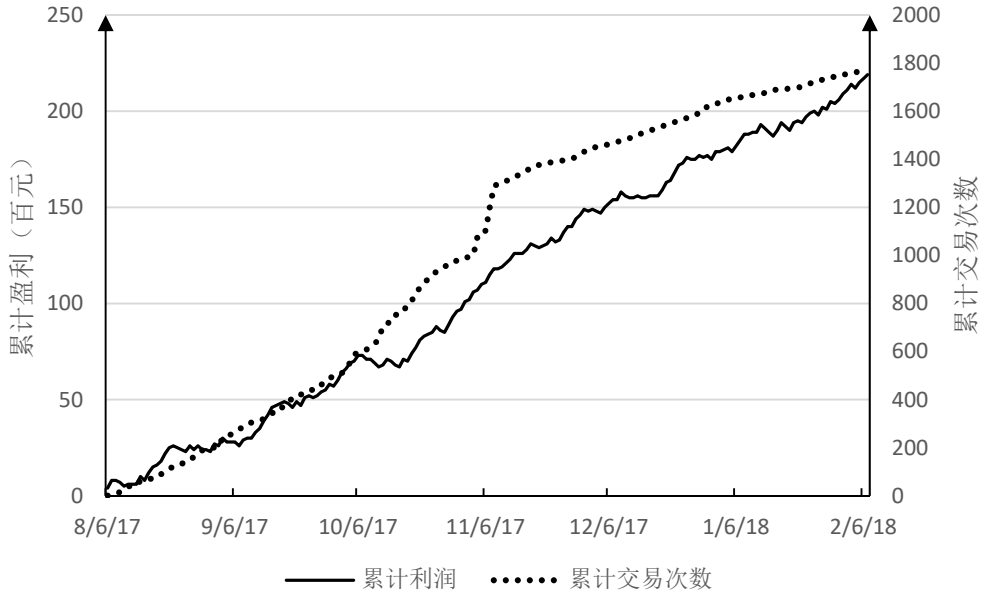

图 5.2 经典配对算法回测结果

如图所示，经典配对交易模型的年化收益率大约是 35.7%，每个月累计交易次数在 150 次左右。

（二）基于遗传算法的交易信号

（1）个体编码：对交易信号的上阈值用无符号的二进制整数表示，并且将交易信号的下阈值与上阈值之间的差值大于手续费作为约束条件。

（2）初始种群：选取种群的规模是 500，即群体中由 500 个体组成。每个个体的取值范围在(12，20)，采用 16 位表示。

（3）适应度计算：比特币的平仓价*比特币平仓的数目+比特币现金的平仓价*比特币现金的平仓数目-比特币的开仓价*比特币的开仓数目-比特币现金的开仓价*比特币现金开仓的数目-手续费。

其中，比特币和比特币现金的手续费是双向的，买入和卖出同时收取，每次的费率是 0.1%

考虑到如果频繁改变阈值，可能导致亏损的现象，本文采取的交易规则如下：

每 1 个小时统计之前 3 个小时的成交量，如果比特币成交量大于初始购买量的 1 倍，则上阈值加 0.1，下阈值减 0.1；如果比特币成交量小于初始购买量的 1 倍，同时大于 0，则阈值不变；否则，根据之前 12 个小时的数据采用不同的交易信号动态化算法，获取新的上阈值和下阈值，作为开仓和平仓信号。本文以 2017 年 8 月 6 日～2018 年 2 月 7 日的数据为例，每 2 秒通过 API 接口采集火币网和币行的实时行情数据，建立不同的交易模型。

初始设置保持和传统配对交易策略一样，结果如下：

1、 经典遗传算法

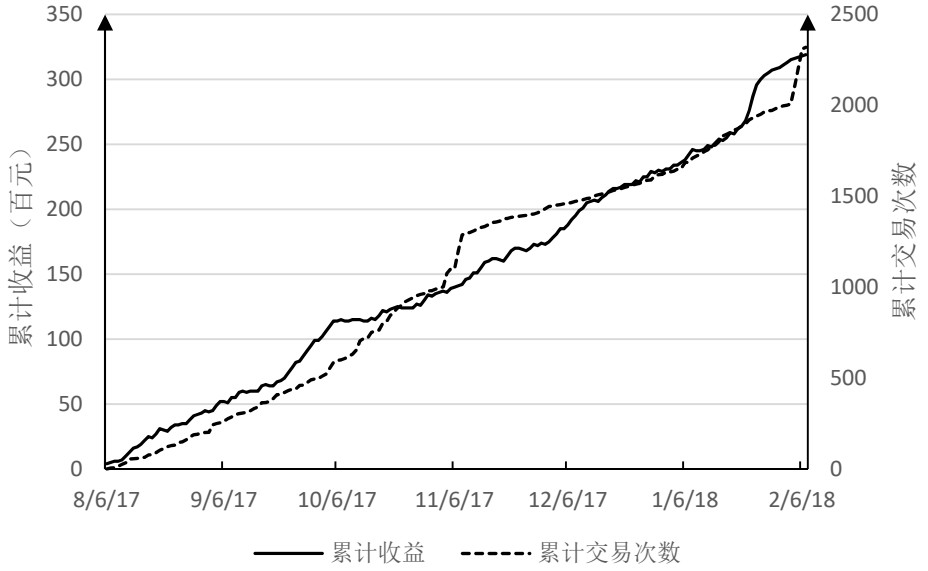

图 5.3 经典遗传算法回测结果

如图所示，基于经典遗传算法，进行交易信号的动态化调整，在回测区间内能够实现 30%的年化收益率，其中，30 天内的累计交易次数在 230 次左右，最大回撤控制在 2%以内。

2、 自学习算法

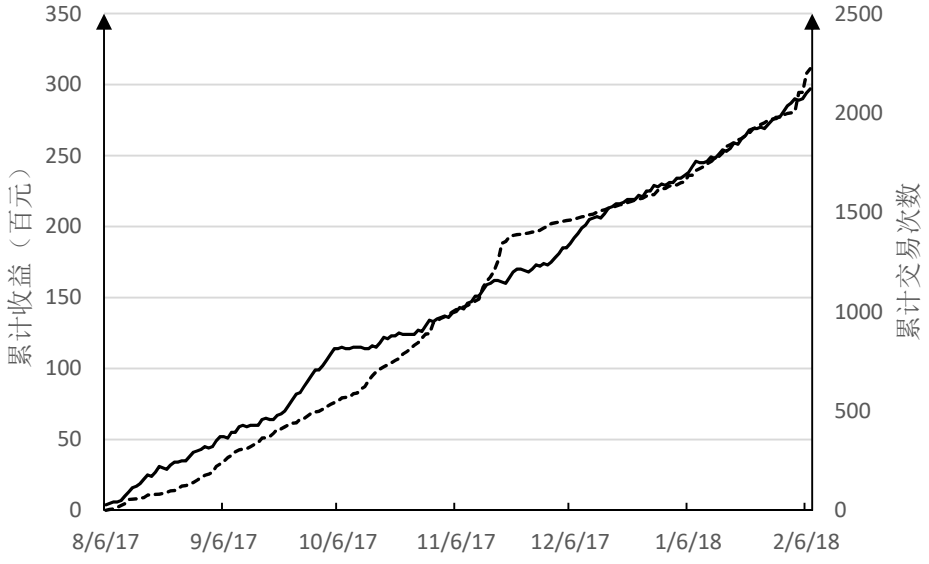

图 5.4 自学习遗传算法回测结果

如图所示，基于自学习算法，对交易信号进行动态调整后，能够实现 35%的年化收益率，30 天内的累计交易次数在 220 次左右。

3、 基于 Metropolis 准则

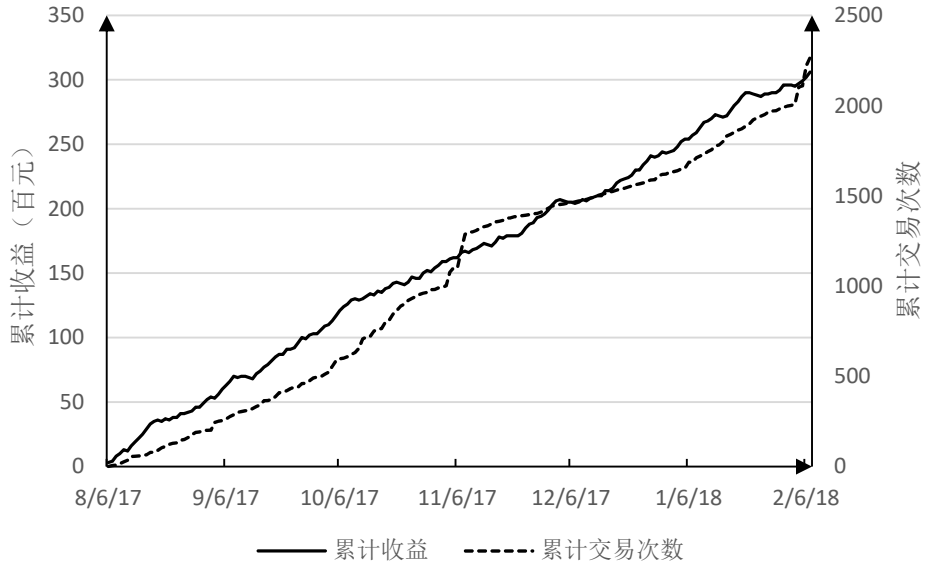

图 5.5 基于 Metropolis 准则遗传算法回测结果

如图所示，基于 Metropolis 准则对交易信号进行动态调整后，能够实现 35%的年化收益率，30 天内的累计交易次数在 260 次左右，最大回撤同样控制在 2%以内。

4、 优化模型

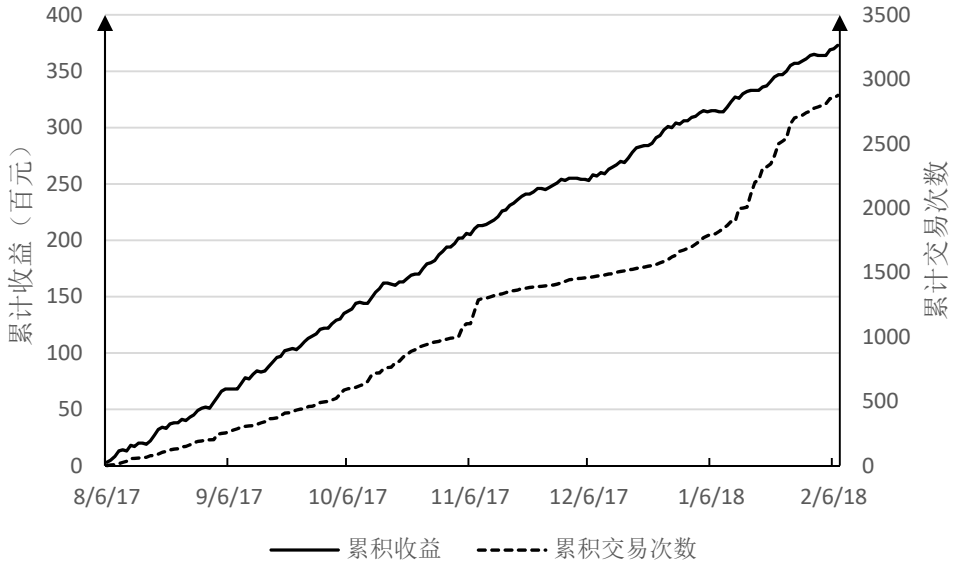

图 5.6 优化模型回测结果

如图所示，基于优化算法对交易信号进行动态化调整后，能够实现 45%的年化收益率，30 天内的累计交易次数在 340 次左右，模型整体效果与单独使用某

一类型的遗传算法对交易信号进行调整相比，在收益率方面有所提高。

5.1.3 结果分析

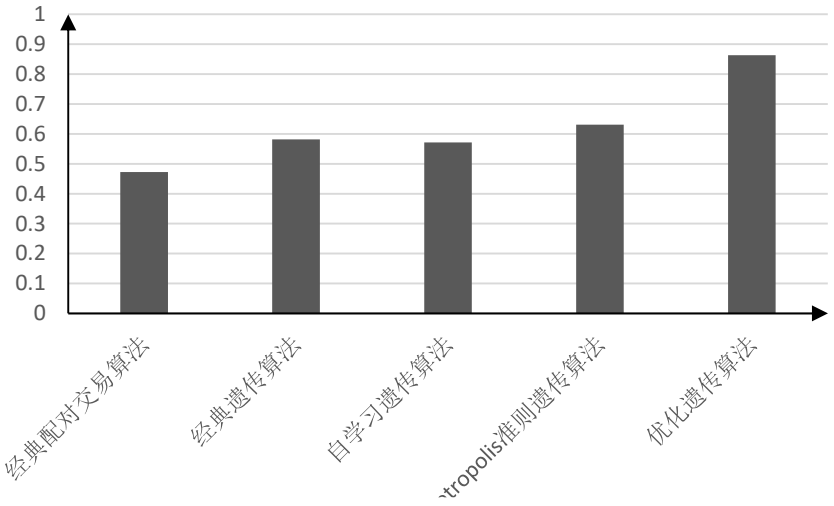

图 5.7 模型夏普比率

如图所示，在回测区间内，经典配对交易算法的夏普比率是 0.4722；经典遗传算法的夏普比率是 0.5819；自学习遗传算法的夏普比率是 0.5714；Metropolis 准则下遗传算法的夏普比率是 0.6311；优化遗传算法的夏普比率是 0.8631。

因此，经过优化之后，模型能够根据过去一段时间的变化进行动态调整，从而提高交易频次，提高资金的利用率，获取更高的利润。

5.2 商品期货市场—以焦煤期货和焦炭期货为例

5.2.1 配对条件分析

5.2.1.1 基本面分析

焦煤(JM)也称冶金煤，是中等及低挥发成分的中等粘结性及强粘结性的一种烟煤。焦炭(J)亦称“焦块”、“焦渣”。它是锅炉炉内加热到 850℃以上时，随着温度升高，煤中的有机物分解，其中挥发性产物溢出后，残留下的不挥发产物。焦煤的主要用途是炼制焦炭，而焦炭的最终用途是冶炼钢铁，因此其下游企业以焦炭加工及自己拥有焦炭加工厂的钢铁企业为主。究其用途，其 90%用于炼铁、炼钢上，因此，在基本面上，二者存在很强的相关性。

从焦煤和焦炭期货的主力合约的历史价格走势上（数据截取 2015 年 1 月 1 日至 2017 年 7 月 1 日），也能很好的印证这一点。

从图中可以看出，焦煤和焦炭之间存在很强的相关性。

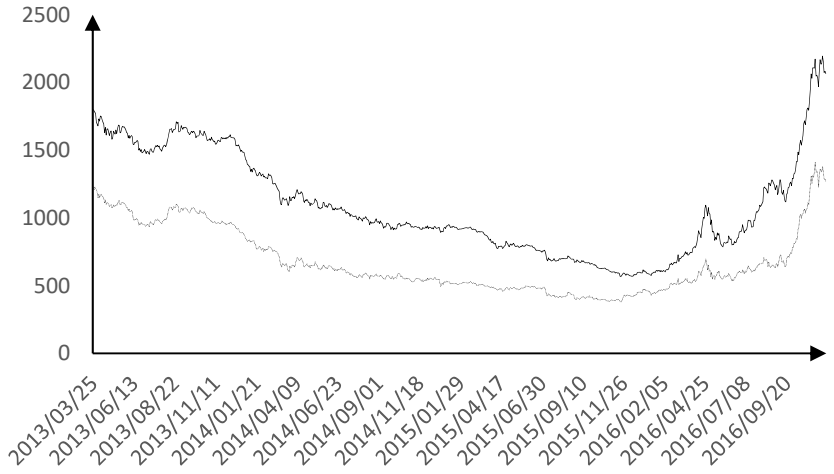

图 5.8 焦煤和焦炭成交价序列

5.2.1.2 统计分析

令焦煤(JM)成交价序列的自然对数是x，焦炭(J)成交价序列的自然对数是y，采用 ADF 检验，表 5.7 为焦煤单位根检验的结果。

表 5.7 焦煤成交价序列单位根检验

	置信水平	t 统计量	P 值
ADF 检验值		-1.706415	0.4278
	1%	-3.457865	
	5%	-2.873543	
	10%	-2.573242	

如表所示，该价格序列在 3 个置信水平下没有通过平稳性检验，即焦煤(JM)成交价序列存在单位根。

表 5.8 焦炭成交价序列单位根检验

	置信水平	t 统计量	P 值
ADF 检验值		-1.01493	0.4208
	1%	-3.457865	
	5%	-2.873543	

	10%	-2.573242	

如表所示，该价格序列在 3 个置信水平下也没有通过平稳性检验，即焦碳(J)成交价序列也存在单位根。

考虑到两者都存在单位根，再对焦煤和焦炭成交价格序列进行一阶差分，再进行单位根检验。表 5.9 和表 5.10 分别是二者进行 ADF 检验的结果。

表 5.9 焦煤成交价序列一阶差分单位根检验

	置信水平	t 统计量	P 值
ADF 检验值		-5.644478	0.0000
	1%	-2.574797	
	5%	-1.942176	
	10%	-1.615803	

表 5.10 焦炭成交价序列一阶差分单位根检验

	置信水平	t 统计量	P 值
ADF 检验值		-35.34211	0.0000
	1%	-2.574797	
	5%	-1.942176	
	10%	-1.615803	

从上表可以看出，在 1%置信水平下，焦煤(JM)和焦炭(J)成交价格的一阶差分序列均通过了 ADF 检验，这表明在 1%的置信水平下，二者的价格序列是一阶单整的，存在长期协整的可能性。

（二）协整检验

根据上表的结果，采用回归方法建立焦煤x和焦炭y的回归方程，对方程残差进行协整检验,结果如表 5.11。

表 5.11 价格序列的回归分析

变量	系数	标准差	P 值

常数	-352.1	8.918282	0.0000
焦煤	2.41	0.236121	0.0000

其中，拟合优度是 88.85%

两个价格序列的回归方程是：

$$y_t = 2.41x_t - 352 + \varepsilon_t$$

对回归方程的残差序列进行 ADF 检验，结果如下：

表 5.12 回归方程残差的 ADF 检验

	置信水平	t 统计量	P 值
ADF 检验值		-5.644745	0.0000
	1%	-2.574797	
	5%	-1.942176	
	10%	-1.615803	

残差序列在 99%的置信区间下是平稳的，因此,焦煤(JM)和焦炭(J)的主力合约成交价格序列存在协整关系。

5.2.2 不同模型下的交易信号构建

与数字货币市场不同，在商品期货市场中，存在直接的做空机制，因此，可以直接进行配对交易，具体如下：

（1） 在大连商品交易所汇入一定数量的人民币，根据回归方程，检测二者的残差。

（2） 当残差超过上阈值，做空一定数量的焦炭，同时做多 2.41 倍的焦煤；当残差低于下阈值，做多一定数量的焦炭，同时做空 2.41 倍的焦煤。

（一）经典配对交易模型的交易信号

根据历史价格数据计算得知：$\text{mean}(\varepsilon_t) = 13, \text{std}(\varepsilon_t) = 3$，因此，根据经典配对交易模型，交易信号的上阈值是 15.25，下阈值是 10.75。

本文以 2017 年 8 月 6 日～2018 年 2 月 7 日为样本(时间间隔为 2 秒)，进行回测，假设每一次限价单的成交比例为 30%，初始资金是 14 万元，统计结果如下：

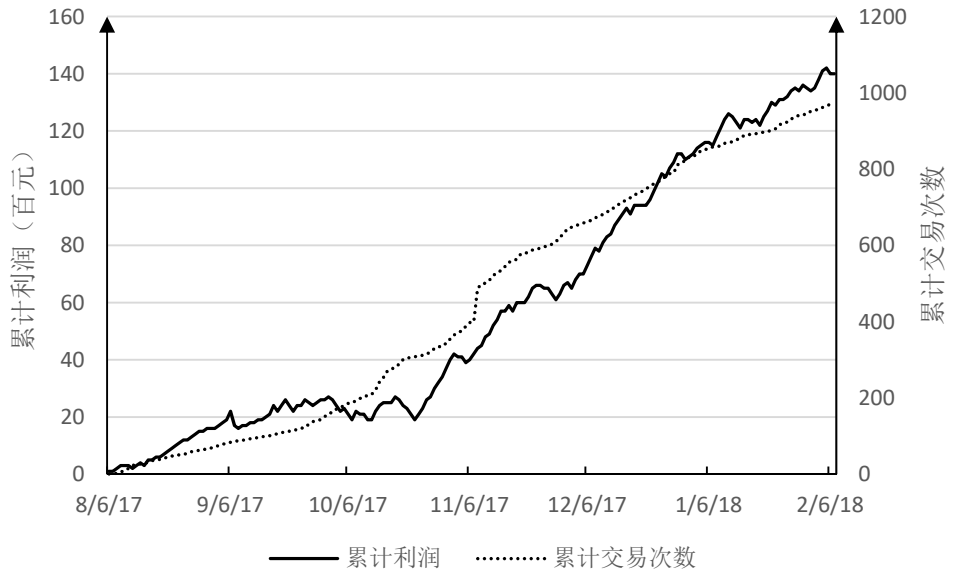

图 5.9 经典配对算法回测结果

如图所示，在商品期货市场中，采用经典的配对交易，年化收益率是 24%，30 天内共平均交易次数是 240，最大回撤 5%。

（二）基于遗传算法的交易信号

（1）个体编码：对交易信号的上阈值用无符号的二进制整数表示，并且将交易信号的下阈值与上阈值之间的差值大于手续费作为约束条件。

（2）初始种群：选取种群的规模是 500，即群体中由 500 个体组成。每个个体的取值范围在(7，19)，采用 16 位表示。

（3）适应度计算：焦煤的平仓价*焦煤平仓的手数+焦炭的平仓价*焦炭的平仓手数-焦煤的开仓价*焦煤的开仓手数-焦炭的开仓价*焦炭开仓的手数-手续费。

其中，焦煤交易 1 手大约 154 元，焦炭交易 1 手大约 67 元。

考虑到如果频繁改变阈值，可能导致亏损的现象，采取如下规则：

每 1 个小时统计之前 3 个小时的成交量，存在交易，则阈值不变，否则，根据之前 12 个小时的数据采用不同的交易信号动态化算法，获取新的上阈值和下阈值，作为开仓和平仓信号，本文以 2017 年 8 月 15 日—2017 年 9 月 1 日的数据为例，每 2 秒通过 API 接口采集大连商品交易所的实时行情数据，建立不同的交易模型。

1、经典遗传算法

图 5.10 经典遗传算法

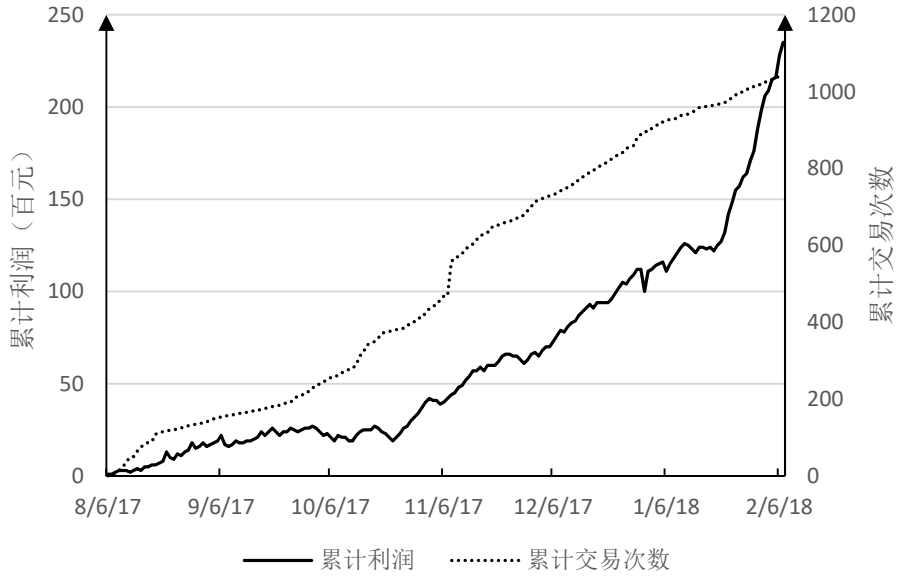

图 5.10 经典遗传算法回测结果

如图所示，在商品期货市场中，采用经典的遗传算法对交易信号进行调整，年化收益率可以达到 35%左右，30 天内平均交易约 260 次，最大回撤 4%。

2、自学习算法

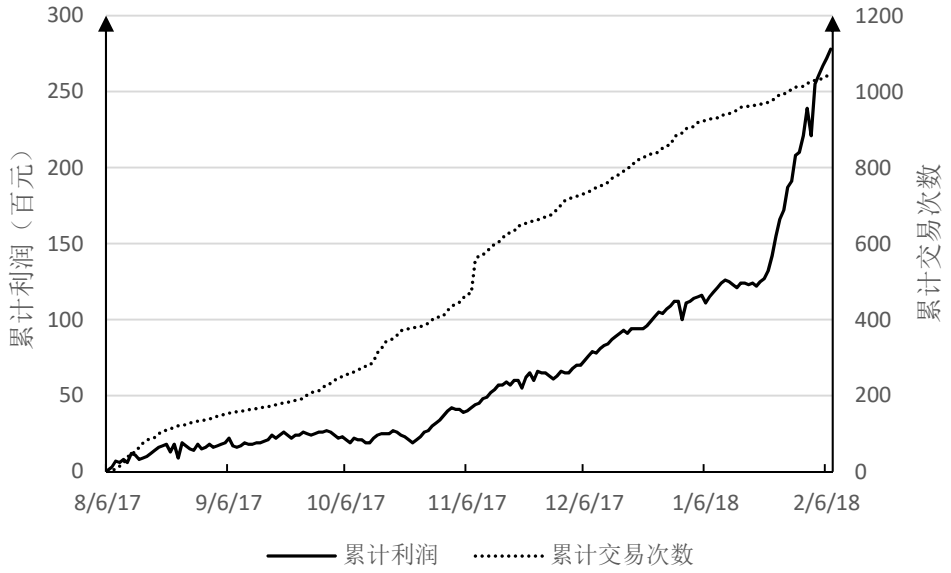

图 5.11 自学习算法回测结果

如图所示，采用自学习算法对交易信号进行动态调整，年化收益率可以达到 38%，30 天内平均交易约 280 次，最大回撤 4.5%。

3、基于 Metropolis 准则

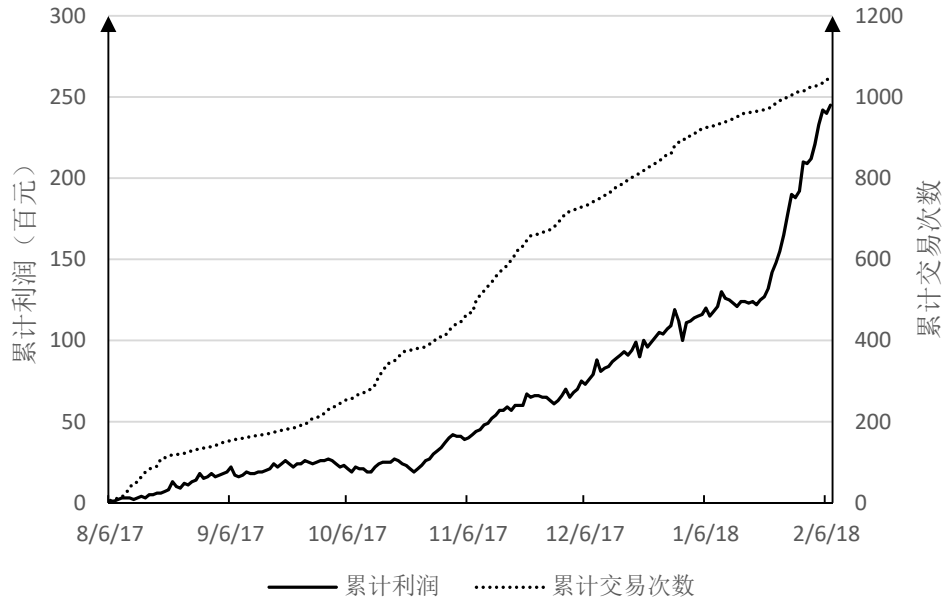

图 5.12 Metroplis 准则回测结果

如图所示，采用基于 Metroplis 准则的算法对交易信号进行动态调整，年化收益率可以达到 36%，30 天内共交易 270 次，最大回撤 4.2%。

4、优化模型

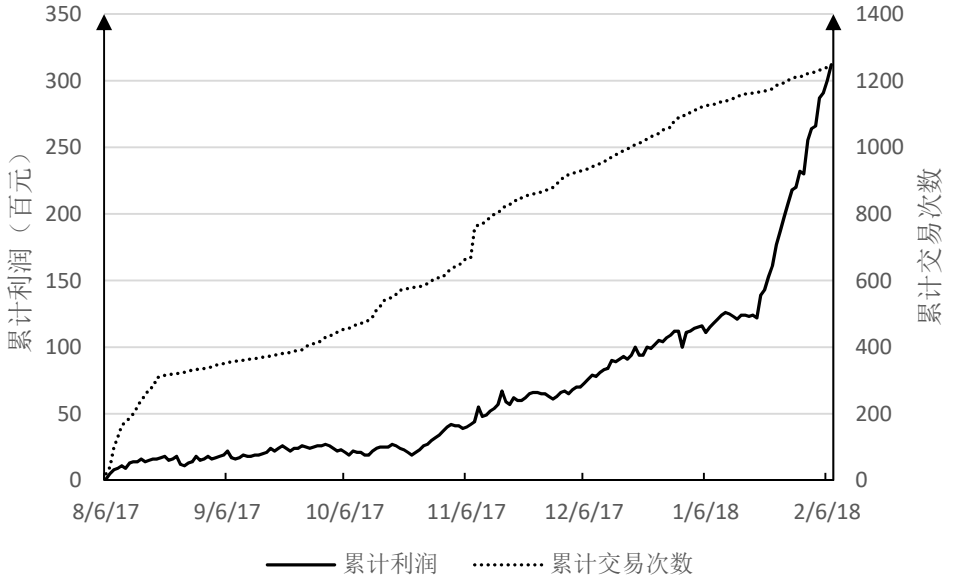

图 5.13 优化模型回测结果

如图所示，采用优化后的模型对交易信号进行动态调整，年化收益率能够实现 43%，30 天内平均交易次数约为 380 次，最大回撤 3%。

5.2.3 结果分析

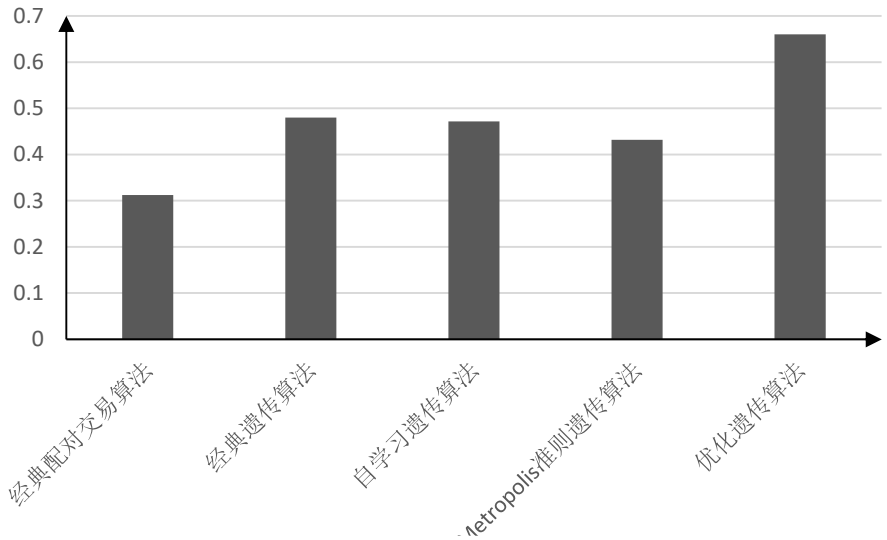

图 5.14 模型夏普比率

如图所示，在回测区间内，经典配对交易算法下，夏普比率是 0.3122；经典遗传算法下对交易信号的动态调整，能够实现的夏普比率是 0.4801；自学习遗传算法下对交易信号进行动态调整，能够实现的夏普比率是 0.4715；基于 Metropolis 准则下的遗传算法对交易信号进行动态调整，能够实现的夏普比率是 0.4319；优化后遗传算法对交易信号进行动态调整，能够实现的夏普比率是

0.6602。

因此，经过优化之后，模型在阈值的动态化方面，相对与经典的配对策略，在交易频次和盈利率方面都有一定程度的提高，同时进一步提高了资金的利用率，从而获得一个更加稳健的收益。

5.3 小结

本章主要以数字货币市场中的比特币和比特币现金、商品期货市场中的焦煤和焦炭为例，系统性比较了经典配对交易的交易信号和基于遗传算法的多种交易信号下，收益率、交易频次，从夏普比率来看，在一定程度上，经过优化的遗传算法模型在交易频次和利润方面，都存在显著的提升。

第 6 章 结论与展望

6.1 结论

6.1.1 交易信号优化模型的主要工作

配对交易策略作为一种常见的统计套利策略，具有收益稳定，风险低等优点，其主要思想是具有相同性质的资产，在市场中的价格走势应趋向一致，基于这种原理，当二者的价格偏差超过一定幅度，可以采用做多一种资产，做空另外一种资产的方式进行套利，通过这种策略，来获取超额收益。

但是，对于“价格偏离”的概念，不同的研究给出了不同的定义。较为经典的方法是将价格的偏离假设为正态分布，从统计学的角度，给出置信区间来定义价格偏离。从长期来看，这种定义是十分合理的，然而，在短期内“价格的偏离”可能呈现不同的形态和统计规律，很难通过某一特定的分布来描述偏离的过程，因此，基于短期内对价格偏离的理解，本文从遗传算法的角度进行阐释。遗传算法是一种解决最优解的方法，传统的方法是通过数学规划的方法来寻求最优解，这要求目标函数是可导的。然而，现实生活中，很多问题的目标函数是不存在导数的，或者很难计算导数，而且，约束条件很难用解析式来表达，遗传算法主要用来解决这种非线性的数学规划问题。与经典方法相比，遗传算法获得的解可能是局部最优解，而不是全局最优解。

因此，本文研究的重点是通过遗传算法来动态调整短期内配对策略中的交易信号。遗传算法的主要流程是用二进制生成一组候选解，通过复制、交叉和变异的过程，根据目标函数进行不断迭代，最终获取最优解。然而，针对经典算法中一些问题，一些学者提出改进的遗传算法，主要包括自学习遗传算法、基于 Metroplis 准则的遗传算法和两阶遗传算法，用来解决遗传算法中优良基因位保留、避免过拟合等问题，从而，针对不同的数据类型和业务逻辑，进一步提高了模型的应用场景和应用的广泛性。

6.1.2 交易信号优化模型的主要内容

本文的主要工作是通过集成学习中的 Stacking 方法，将经典遗传算法、自学习遗传算法和基于 Metroplis 准则的遗传算法相结合，使得不同模型的性能进行融合，对配对交易策略中交易信号实现动态化调整。构建了一种短期内，对价格偏离程度的描述，同时，建立了短期内配对交易策略程序化交易的一般框架。为了验证方法的有效性，本文在数字货币市场和期货市场进行了回测，分别选取比特币和比特币现金；焦煤和焦炭两组资产标的进行模拟，以夏普比率作为评价指

标，统计在一定周期内累积利润和累积交易频次。回测结果显示：同经典的配对交易方法相比，短期内采用遗传算法来确定交易信号，能显著提高收益率和交易频次，优化模型的性能；同时，在遗传算法框架内，通过模型融合对遗传算法进行优化后，与单独使用经典的遗传算法、自学习遗传算法和基于 Metroplis 准则的遗传算法相比，优化后的遗传算法在夏普比率这一评价指标上，要优与其它遗传算法。

因此，通过 Stacking 方法将经典遗传算法、自学习遗传算法和基于 Metroplis 准则的遗传算法相融合得到的优化模型，能够在短期内的提高资金利用率，回测结果表明，该模型能够显著提高收益。

6.1.3 交易信号优化模型的应用

本文以数字货币市场中的比特币和比特币现金的价格、期货市场中的焦煤和焦炭的价格为例，回测优化模型与传统方法在交易频次、累计利润和夏普比率方面的异同。结果表明，无论是在数字货币市场还是在商品期货市场，交易信号优化模型在这三个指标上的表现都要优于经典配对交易模型，同时也优于单独使用经典遗传算法、自学习遗传算法和基于 Metroplis 准则遗传算法确定的交易信号。

整体对比数字货币市场和期货市场，结果表明，数字货币市场在这三个指标上的表现要优于期货市场，其主要原因是数字货币市场中参与的主体是普通投资者，非理性因素较大，而机构参与者较少，市场的无效性比较明显，而期货市场在监管方面比较严格，参与的主体是机构，市场的无效性相对较低，收益率也会相对减少。

总体而言，无论是在数字货币市场，还是期货市场，都在一定程度上验证了模型的有效性，为风险厌恶性投资者选择交易策略提供了借鉴。

6.2 展望

随着中国“一带一路”战略的推进，中国的金融市场需要越来越开放，未来，更多的国际投资者可能参与到中国的金融市场中来，然而，我国的金融对冲工具却并不完善，尤其针对广大投资者，缺少必要的手段。

本文研究的主要工作是短期内配对交易中交易信号动态化问题，配对交易是一种低风险，相对高收益的投资策略，经典的配对交易策略，采取固定阈值，收益相对较低，本文通过引入遗传算法，并对其进行优化，使其交易信号能够在短期内进行动态调整，提高了收益率，在一定程度上可以为专业投资者和普通投资者在进行策略选择时，提供一部分参考，帮助投资者降低风险，同时获取收益。

参考文献

[1]Alexander. Optimal hedging using counteraction[J]. Philosophical Transactions of the Royal Society London ,1999,357:2039–2058.

[2]Alexander, Dimitriu. Indexing and statistical arbitrage. Journal of Portfolio Management [J] 2005,31:50–63.

[3]Alexander ,,Dimitriu,A. Indexing，integration and equity market regimes. [J]International Journal of Financial Economics , 2005,10:1–19.

[4]Avellaneda，M.，Lee， 2010. Statistical arbitrage in the US equities market. Quantitative Finance[J],2010, 10:761–782.

[5] Girma P B. Risk arbitrage opportunities in petroleum futures spreads[J]. Journal of Futures Markets,1999,19(8):931–955.

[6]Simon The many shapes of knowledge[J].Revue NationalPersée,1999,88(1): 23-39.

[7] Liu M,Li C L,Stamatoyannopoulos G,et al. Gammaretroviral vector integration occurs overwhelmingly within and near DNase hypersensitive sites[J]. Human Gene Therapy,2012，23(2):231-7.

[8] Wahab and Cohn. The gold-silver spread: Integration，cointegration， predictability， and ex-ante arbitrage[J]. Journal of Futures Markets,1994,14(6):709–756.

[9]Elliot,R,Van Der Hoek,Malcolm, Pairs trading[J]. Quantitative Finance.2005,102:271–276.

[10] Edwards S,Susmel R. Volatility dependence and contagion in emerging equity markets[J]. Journal of Development Economics,2001,66(2):505-532.

[11] Broumandi S,Reuber T. Statistical arbitrage and FX exposure with South American ADRs listed on the NYSE[J].Financial Assets & Investing,2012,3(2):5-18.

[12] Lahmiri S. An Exploration of Backpropagation Numerical Algorithms in Modeling US Exchange Rates[J]. 2015. 5(6):15-28

[13] Caldeira J，Moura G V. Selection of a Portfolio of Pairs Based on Cointegration: A Statistical Arbitrage Strategy[J]. Ssrn Electronic Journal,2013,11(1):49-80.

[14] Gutierrez J A,Tse Y. Illuminating the Profitability of Pairs Trading: A Test of Relative Pricing Efficiency of Markets for Water Utility Stocks[J]. Journal of Trading,2011,15(4):361-5.

[15]崔方达，吴亮．配对交易的投资策略[J]．统计与决策，2011，23:156-159.

[16]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.

[17]戴进．基于协整的股指期货和 ETF 的统计套利[J].中国证券期货，2012,10:1-2.

[18]Frazzo,Geczy，C.C.&Musto,D.K. Adapt the Basic Vasieck Interest-RateModel to Generate a Mean Reversion Signal for Stock Pairs[J] .Working Paper，2002,10:1-16

[19]CumminsMichel,M.& Chandra,G Loss Protection in Pair trading ThroughMinimum Profit Bounds:A Integration Approach[J].Journal of AppliedMathematics and Decision Sciences 2006,10:5-25

[20]Boguslavsky,Boguslavskaya,Optimal arbitrage trading. Working paper[J].University of Amsterdam.2010,23:10-19

[21]Mudchanatongsuk，S.，Primbs，J.A. Optimal pairs trading: a stochastic control approach[J]. American Control Conference，2007,78:1035–1039.

[22]Montana. Flexible least squares for temporal data mining and statistical arbitrage[J]. Expert

Systems with Applications 2006,36:2819–2830.
[23] Schoeman J P,Goddard A,Leisewitz A L. Biomarkers in canine parvovirus enteritis.[J]. New Zealand Veterinary Journal， 2013， 61(4):217-22.
[24] Laerte Guimar,es Ferreira， Ferreira N C， Ferreira M E. Remote sensing of vegetation: evolution and state of the art[J]. Acta Scientiarum Biological Sciences,2008,30(4):12-34.
[25] Yang J， Liew K M， Wu Y F， et al. Thermo-mechanical post-buckling of FGM cylindrical panels with temperature-dependent properties[J]. International Journal of Solids & Structures，2006,43(2):307-324.
[26] Schoeman J P， Goddard A， Leisewitz A L. Biomarkers in canine parvovirus enteritis.[J]. New Zealand Veterinary Journal,2013,61(4):217-22.
[27]Thomaidis，N.S.， Niek， K.& Dounias， G. An intelligent Statistical ArbitrageTrading System [J]. Advances in Artificial Intelligence,2006,5: 18-22
[28]Stock， J.H.， Watson， M.WTesting for common trends[J]. Journal of the American Statistical Association 83,2:1097–1107.
[29]贾丽平. 比特币的理论、实践与影响[J]. 国际金融研究，2013，12:14-25.
[30]陈道富，王刚. 比特币的发展现状、风险特征和监管建议[J]. 学习与探索，2014，04:88-92.
[31]刘刚，刘娟，唐婉容. 比特币价格波动与虚拟货币风险防范——基于中美政策信息的事件研究法[J]. 广东财经大学学报，2015，03:30-40.

在学期间发表的学术论文及研究成果

1、邸玉娜，由林青,中国对一带一路国家的投资动因、距离因素与区位选择[J].中国软科学,2018(02):168-176.

2、王琴英,由林青,王佳佳,卫士加.互联网金融背景下“软信息”对于违约行为的影响——基于 Logit 模型[J].金融理论与实践,2017(03):60-65.

3、王琴英,由林青,张燕萍,徐程.价格预期组合效应对我国玉米市场价格波动的影响[J].价格理论与实践,2016(04):106-108.

4、“华为杯”全国研究生数学建模竞赛，国家级二等奖，2016.10，由林青,王佳佳，卫士加.

致谢

三年的时间眨眼间过去，我的研究生生涯也即将结束，还没来的及细细回味，我却将要离开这片热土，三年的时间里，有迷茫，有收获，有遗憾，有感动。寄以学位论文结尾之处，对三年以来支持我的亲人、老师、同学表示感谢。

首先我要感谢我的父母，你们含辛茹苦将我养大成人，永远支持我的选择，是你们的无声的爱陪伴我度过研究生这段科研之路。曾经的迷茫，是母亲的关怀让我努力前行，是父亲的鼓励让我不忘初心。同时对一直支持我的亲人表示感谢。

感谢我的导师王琴英老师。王老师科研态度严谨，待人真诚，不仅在论文写作上给予我很大的支持，也在生活中的小事上帮助我，开导我。虽然我科研上没做出大的成就，但王老师一直默默地支持鼓励我。

感谢我的女友吕文明，感谢你的理解与支持，感谢你的陪伴和鼓励，感谢你的监督与安慰，能够认识你我感到很幸运。

感谢魏瀚焘、贺倩在科研上给予我的帮助，感谢陈凌同学营造的轻松的科研环境，一起奋斗的日子我永远不会忘。

感谢胡哲、张新彤、姜玉鹏营造的良好的宿舍氛围，感谢你们的陪伴与帮助，感谢所有 2015 级数量经济学专业的学生，有你们的陪伴，我的生活变的更加精彩。