学校代码：10270　　分类号：　F832.51　　学号：192502507

上海师范大学

硕士专业学位论文

基于强化学习算法的商品期货配对交易策略设计

学　　院：商学院

专业学位类别：金融专业硕士学位

专 业 领 域：金融统计与建模

论 文 类 型：交易策略设计

研 究 生 姓 名：李静

指 导 教 师：龚秀芳

完 成 日 期：2021 年 05 月

论文题目：基于强化学习算法的商品期货配对交易策略设计

论文类型：交易策略设计

学科专业：金融统计与建模

学位申请人：李静

指导教师：龚秀芳

摘　要

统计套利作为一种交易策略，可以在产生无风险利润的同时，规避一定的市场风险。而在众多统计套利策略中，配对交易策略则受到了广泛的重视和应用，也是比较主要的一种交易策略。通常来讲，配对交易是先构建配对资产的多空头寸，从而获得一些资产价差收敛的收益。这种交易策略的一个比较明显的优势，就是可以通过对冲机制的形式，规避一部分投资时的系统性风险，这也就意味着，就算在金融市场整体呈下行状态的时候，配对交易也能盈利。配对交易策略运行的特点之一就是要求证券市场具备做空机制，但是我们国家的股票金融市场只允许做多不允许做空，也就是单边运行模式。虽然我国2010年开始逐渐推行起融资融券业务，但是融券业务需要较高的成本，使得我国金融市场下的配对交易策略很难有所作为。而期货市场具有着良好的做空机制，并且近几年来，我国金融期货市场的产品逐渐丰富，使得配对交易策略更可能适用于期货市场。基于以上，本文采取国内期货市场为研究对象，对配对交易策略进行实证研究。

本文将强化学习算法与传统配对交易模型结合起来，主要改进点为，使用动态参数优化法替代传统配对交易模型中的固定参数法，并将其运用到中国的期货市场，使得本文的交易策略增加一些获利机会。采用上海期货交易所中2010年1月至2020年1月的日收盘价数据，首先对训练期的期货数据进行流动性分析，筛选出流动性较好的品种；然后将其两两组合进行相关性分析，接着对相关系数大于阈值（本文将相关系数阈值设定为0.85）的匹配对进行协整性分析，最终筛选出了

6个协整性较好的匹配对，这6组最终配对组合即为本文的研究对象。然后对最终的配对组合计算配对合约比率后，制定传统的静态配对交易模型的交易策略信号，并且对基于强化学习的动态配对交易策略的初始参数进行初始化等，采用两个模型分别对研究对象进行实证研究，并对训练集、测试集的数据分别计算出累计收益率、索提诺比率、信息比率等衡量盈利能力的指标。

通过研究结果我们可以发现，基于强化学习算法的配对交易模型在6组配对组合中的结果都有着不错的绩效，并且与传统模型相比较下，新模型显著促进了收益率的提升，投资风险也有所减少，具备着持续学习的能力。本文把强化学习算法和配对交易策略两者相结合，设计出基于Sarsa算法的新型配对交易策略，从而达到针对交易模型进行有效调整与优化的目的，有助于改善传统配对交易策略获利能力下降的缺点，并为投资者提供较为有效的套利工具。

关键词：配对交易；商品期货；动态参数；强化学习；Sarsa算法

Abstract

As a trading strategy, statistical arbitrage can avoid certain market risks while generating risk-free profits. Among many statistical arbitrage strategies, the paired trading strategy has received extensive attention and application, and it is also a major trading strategy. Generally speaking, pair trading is to construct long and short positions of paired assets first, so as to obtain some income from asset spread convergence. One of the obvious advantages of this trading strategy is that it can avoid some of the systemic risks of investing in the form of a hedging mechanism, which means that even when the financial market as a whole is in a downward state, paired trading can be profitable. . One of the characteristics of the operation of the paired trading strategy is that the securities market is required to have a short-selling mechanism, but the stock financial market in our country only allows longs and does not allow shorts, which is a unilateral operation mode. Although my country began to gradually implement the margin trading and securities lending business in 2010, the securities lending business requires a higher cost, which makes it difficult for the matching trading strategy in my country's financial market to make a difference. The futures market has a good short-selling mechanism, and in recent years, my country's financial futures market has gradually enriched products, making the matching trading strategy more likely to be applied to the futures market. Based on the above, this article takes the domestic futures market as the research object and conducts an empirical study on the paired trading strategy.

This article combines the reinforcement learning algorithm with the traditional pair trading model. The main improvement is to use the dynamic parameter optimization method to replace the fixed parameter method in the traditional pair trading model, and apply it to the Chinese futures market, making the trading strategy of this article increase Some profit opportunities. Using the daily closing price data from January 2010

to January 2020 in the Shanghai Futures Exchange, first conduct a liquidity analysis on the futures data during the training period to screen out products with better liquidity; then correlate the pairwise combinations Cointegration analysis is performed on matching pairs whose correlation coefficient is greater than the threshold (the correlation coefficient threshold is set to 0.85 in this article), and finally 6 matching pairs with better cointegration are screened out, and these 6 final pairing combinations That is the research object of this article. Then, after calculating the matching contract ratio for the final pairing combination, the traditional static pairing trading model's trading strategy signals are formulated, and the initial parameters of the dynamic pairing trading strategy based on reinforcement learning are initialized. The two models are used to conduct research on the research objects. Empirical research, and calculate the indicators of profitability, such as the cumulative return rate, the Sotino rate, and the information rate, based on the data of the training set and the test set.

Through the research results, we can find that the paired trading model based on the reinforcement learning algorithm has good performance in the results of the six paired combinations, and compared with the traditional model, the new model significantly promotes the increase in the rate of return, and there are also investment risks. The reduction has the ability to continue learning. This article combines the reinforcement learning algorithm and the pairing trading strategy to design a new pairing trading strategy based on the Sarsa algorithm, so as to achieve the purpose of effective adjustment and optimization of the trading model, and help to improve the profitability of the traditional pairing trading strategy. The shortcomings, and provide investors with more effective arbitrage tools.

Keywords: Paired trading; Commodity futures; Dynamic parameters;

Reinforcement learning; Sarsa algorithm

目 录

第 1 章 绪论

1.1 研究的背景

这些年以来，伴随着我国经济全球化进程的加快，以及改革开放的不断推动，国内的经济形式得到了跨越式的发展进步，同时金融市场也逐渐发展完善起来、金融产品也得到了创新，在此基础上，量化投资策略也有了更为丰富的研究对象，这些策略在我国的探索与发展进程也进一步加快。并从国际金融市场的视角来看，市场运行中的不利因素和不确定性因素在不断地增加，从一定程度上加剧了国内证券市场的价格波动，导致一些机构投资者很少能仅仅通过做多股票的方式来得到收益。所以，投资者们比较迫切地想要寻找一种能够承担低风险的同时又获得收益的交易策略。

配对交易策略的思想最早出现在20世纪80年代中期的华尔街，由著名的交易员Jesse Livermore提出的这个概念，并且经过数量分析师、同时也是天体物理学家的Nunzio Tartaglia及其领导的团队将这个概念投入了实际操作，而且在1987年取得了较明显的成功。之后此策略在美国、欧洲等一些相对来说更成熟的金融市场得到广泛的使用，一些机构投资者，如国外的对冲基金等，也利用此策略来得到巨额的低风险收益。长久以来我国的股票金融市场一直是只允许做多而不允许做空，即采用的是典型的单边市场行情。而配对交易策略运行的特点之一就是要求证券市场具备做空机制，所以配对交易在我国股票市场上并不能得到较好实施。

从2010年3月31日开始，中国证监会正式开始了融资融券业务的试点工作，具体包括接受并办理部分试点证券公司的融资融券交易申报等。融资融券（也叫做“证券信用交易”）代表的具体意思就是：对一些符合融资融券业务申报要求的公司，投资者向其交付一定担保物，然后可以从券商公司借到一些资金用来购买证券（融资交易），或者可以从券商公司借到证券并将其卖出（融券交易）。融资融券工作的进行之后，我国的证券市场终于告别了一直以来的单边市场制度，也就是只能做多不能做空，同时也给证券市场融入了做空机制。同时，这项业务的运行也对完善股票合理的定价机制是有利的，比如投资者可以通过融资买入被市场低估的股票，这会导致股票价格上涨；也可以利用融券卖出被市场高估的股票，这会导致股票价格下跌。到2018年12月31日为止，我国的A股股票市场的全部融资融券标的股一共948只，覆盖个股达到27%，融资融券余额累计总值达到

了7557.04亿元，跟业务试推行当天沪深两市649万的余额总值相比较，扩容了将近12万倍。融资融券业务从试点推行开始到今天，一直都在持续、稳步、快速的、发展着，这不仅一定程度提高了A股市场活跃度，也从另一个角度上满足了机构投资者的对冲需求，对促进完善多层次的证券市场制度十分有利。

但在实际操作中，我国的股市T+1交易的情况下，仍以做多为主，并且融资融券业务的标的数量有限、有较高的准入门槛，而期货市场具备着做空机制，因此期货也是进行相关套利研究的较为符合的金融产品之一。进行套利的投资者可以通过一些操作使得金融产品的价格更快地趋于真实的价格，这一定程度上对于金融市场的稳定有利，套利交易的增多也会提高商品期货市场的效率，更好地服务于实体经济。

1.2 研究的目的和意义

随着我国金融市场的不断改善与各种金融产品的不断创新，我们国家正在稳步朝着建立多层次、多维度的金融市场的目标迈进。而在这整个过程里，对各种金融产品进行深入的研究与合适的应用是十分关键的。不仅基础的股票、债券产品占据着重要的市场地位，金融衍生产品也将在未来占据一席之地。在众多金融衍生品之中，期货作为标准化的产品之一，在我国的稳步发展中表现出了对实体经济的积极作用，很大程度上帮助了大宗商品企业对冲掉可能存在的价格风险。同时，一个活跃且有效的金融二级市场，需要各种市场参与者的积极参与，包括套期保值者、套利者和投机者等。而期货市场具备着可进行日内多空交易的特性，此特性为相关套利交易提供了较合适的制度条件。目前国内的期权市场还没有完全成型，并且股票市场为T+1交易制度，在此情况下，若想进行相关的套利研究，期货便是最佳的金融产品研究对象。投资者可以通过套利交易来使得金融产品的价格逐渐趋于金融产品的真实价值，促进金融市场的稳定发展与期货市场效率的提高。所以，以商品期货为对象的套利研究一直是学术界和业界的学者们研究的一大重点，本文也将研究商品期货领域的套利交易。

而配对交易属于量化投资策略的一个分支，其通过计算机发出的投资策略可以克服很多人性的弱点，在西方发达的金融市场一直倍受机构投资者的青睐。如何准确的选择交易模型参数成为了学者研究的重点，在现有的交易模型中，很多方法都取得了一定的成效，但是都需要有各自特定的数据样本，比如有的数据需

要符合GARCH模型或O-U模型，有的数据需预设经验性参数等专家系统，也就是受到一些新的使用条件的约束，致使配对交易在学者研究的途中受到了比较多的限制条件。因此，后续学者将强化学习思想成功引入了配对交易策略中，将两者结合，使得交易模型的参数实现自适应动态优化。这种改进的交易模型有很大的优点，比如在提升盈利能力、提升执行效率两方面都取得了明显的成效。本文将在此研究的基础上，更详尽地阐述强化学习算法的基本原理，以及在配对交易模型中的具体使用步骤，并将这种新型的改进模型引入中国期货市场，有助于提升在商品期货领域的配对交易的盈利能力和执行效率等。

1.3 研究的内容、方法和技术路线

1.3.1 研究内容

本文的研究内容主要包括：

（一） 配对组合的挑选

本文采用2010年1月至2020年1月上海期货交易所的收盘价数据对配对交易进行实证研究。因为进行配对交易的数据要有比较强的流动性，所以我们去除掉交易量不是很活跃的线材和燃料油期货，选择剩余的6个交易品种。并且将2010年1月至2019年1月的数据作为训练集，将2019年1月至2020年1月的数据作为测试集。将我们选取的期货合约按照相关性配对法进行配对，保留下相关系数较大的几对期货，将其作为研究对象。

（二） 配对期货之间的协整关系检验

本文关于协整关系的检验，采取的方法是E-G两步法，在预配对期货确定之后，我们先采用ADF检验，对各配对组的两个期货进行平稳性检验，接着判断其价格序列是否平稳；然后再对其进行回归分析，来寻找期货价格序列之间存在的数学关系，最后得到我们需要的价差序列，对其进行平稳性检验；若残差序列也平稳，那么协整检验就通过，同样就说明该配对组的价格波动存在着长期的均衡关系，最终选取这组配对组合。

（三） 利用强化学习对配对组合进行模型建立

本文将把Sarsa强化学习算法与传统的模型两者相结合，对模型参数的确定方法进行改变，从传统的主观经验法、固定参数法改进成动态参数优化法，接着将新模型与传统配对交易模型进行对比研究，最终对两个交易模型的实证结果进

行总结性的分析。

1.3.2 研究方法

本策略的难点在于“配对资产的选择”、“基于强化学习的配对交易模型设计”。主要用到以下几种方法：

（一） 文献研究法

两支股票（或期货）的价差序列之间存在稳定关系，此理论基于一定的经济原理,而这些立足理论是我们进行经济、金融研究的第一步,也是最重要的一步。本文将对历史文献进行研究，通过对文献的阅读，可以从经济理论出发，归纳、总结股票（或期货）间同向性、相关性、平稳性的内在原因和其中的一些影响因素，并将这些作为本文实证部分“选择配对资产组合”的理论基础。通过学习和分析强化学习模型,并构建价差技术指标，为基于强化学习的配对交易策略模型做准备。

（二） 相关性配对法、协整检验法

相关性配对法，就是两个收盘价序列计算相关系数，并选出相关系数较大的几组期货合约作为配对组合。

两种商品期货的相关系数为：

$$\rho_{X,Y}=\mathrm{cov}(X,Y)/(\sigma_X\sigma_Y) \tag{1-1}$$

式（1-1）中 X 和 Y 就是两种商品期货的收盘价序列。我们挑选各个配对组合中相关性最大的组合，并把他们当作是配对交易的预配对组合。

协整检验：我们挑选出一些合约配对组合后，对这些组合的对数时间序列进行单位根检验，若在5%的显著性水平上，其对数时间序列显示为非平稳序列，那么就要对他们的一阶差分序列再进行平稳性检验。最终再用E-G两步法对这些配对组合的价格序列进行协整性检验。然后将通过协整检验的预备配对组合作为最终配对组合。

（三） Sarsa算法

Sarsa算法是目前使用较广的强化学习算法，虽然这种算法对状态空间的数量和离散型上都具有一定的限制,但是在配对交易中涉及到的状态空间时比较容易满足这些限制条件的。在模型中为了避免陷入出现局部最优的情况，我们使用 $\varepsilon-greedy$ 探索策略。在选取动作的时候增加了一定的随机变化，以此来处理开发与利用之间的平衡性问题。在Sarsa算法的编制过程中，首先需要设定参数的

初始值，参数主要包括评估时间窗口、交易时间窗口、开仓阈值和平仓阈值共4个；接着采用足够数量的迭代来对智能体进行训练；最后用 $\varepsilon-greedy$ 策略优化上述的几个参数，把这一步当作智能体的动作。此外，本文将索提诺比率当作计算奖赏值的指标，该指标在整个学习过程中不断地通过环境来反馈到智能体，最后该比率作为最终的数据输出。

1.3.3 研究技术路线图

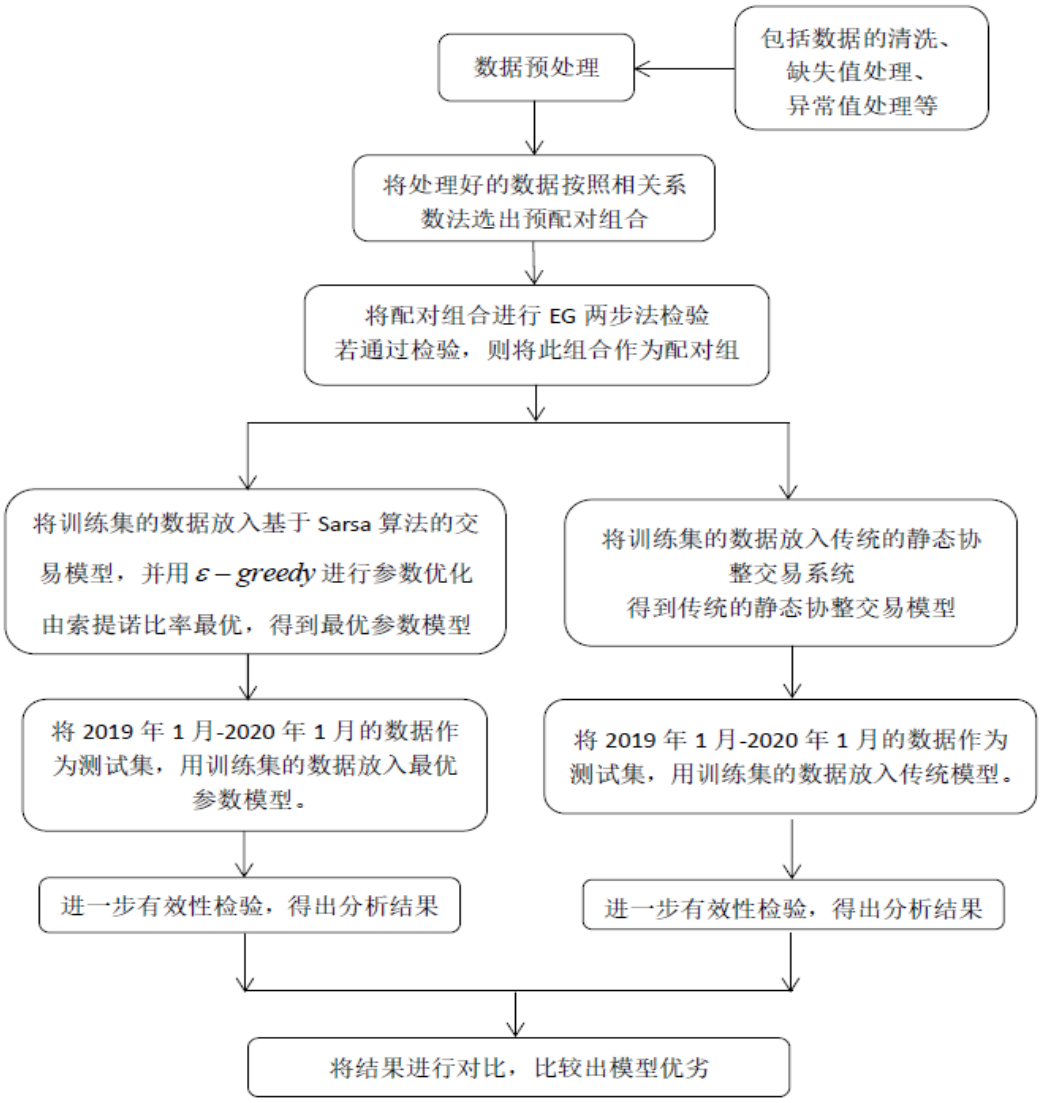

图 1-1 技术路线图

1.4 本文的主要创新

（一） 研究对象和处理方法方面的创新

虽然说之前国内的一些学者已经运用相关配对交易模型对我国股票市场进行了交易策略的实证，但是我国股票市场做空机制并不完善，并且目前国内关于配对交易策略的研究文献中，在期货市场领域的实证研究相对来说较少。伴随着我国商品期货市场的成交规模逐渐扩大，并且商品期货的流动性近几年来逐渐转好，因此本文以国内商品期货市场作为研究对象。在分析处理方法上，特别是关于匹配组合对的选择方面，以往的研究都是从基本面进行分析，然后选择相关性较大的期货品种，而本文的分析处理方式完全以数据为基础，首先从期货流动性开始考虑，挑选出年日均成交量在2万手以上的期货品种，也就是流动性较大的期货品种，对这些品种分别两两配对计算相关性系数，并且筛选出相关性系数大于0.85的期货对进行协整性检验，对结果进行分析，然后得到适合配对交易的最终配对组合，对最终配对组合进行实证研究。

（二） 在研究方法方面的创新

本文在交易时机的选择上有所创新，主要与传统方法不同的点在于：传统的配对交易关于开仓、平仓时机的选择是静态的，属于静态策略，当价差序列的偏离程度达到几倍的标准差时就进行开仓操作，当价差序列回到合理区间的时候就进行平仓操作。本文在前人研究的基础上，对模型参数的确定方法进行改变，由传统的主观经验法和固定参数法改进为自适应模式的动态参数优化法，并且更加详细地阐述基于强化学习算法的新型配对交易模型原理。将其应用到国内期货市场，采用了较新的期货市场数据，所得到的实证结果有一定的参考价值，也提高国内期货市场配对交易效率和绩效。

（三） 评价分析方面的创新

作为一种交易策略，本文更加关注此策略的“实战能力”，分别基于传统静态模型与强化学习动态模型设置了配对交易策略，然后对样本内和样本外的数据都进行了回测分析，得到一些实证结果，对结果进行分析对比了两种交易策略的收益能力。在评价指标的选取方面，本文在计算累计收益率等指标外，还加入了更关注下行风险的索提诺比率，以及可以描述超额风险所带来超额收益的信息比率，更全面地比较出了两种模型所构造的交易策略的交易结果、获利能力、抗风险能力等。

第 2 章 相关理论回顾与文献综述

2.1 相关理论回顾

2.1.1 配对交易理论

在股票市场上，投资者通常认为进行交易的规律就是买入估值低的股票、卖出估值高的股票，从中可以获取收益。但是实际上投资者看到的都是股票的表面价格，其真实价值是很难判断的，而配对交易通常就被用来解决这个难题，其本质是使用相对定价的理论去评定股票真实价值。首先探讨金融学中的套利定价理论（APT理论），假设有两只类似的股票，那这两个股票的价格一定是近似的，若它们的价格不一样，那通常是因为其中一只价格被高估而另一只价格被低估。因此配对交易的具体操作，简单来说就是卖出高价的股票、买入低价的股票，这样在未来错误定价得到自我纠正的时候，投资者可以从中获得收益。这里比较重要的一点是，若我们想要进行配对交易，并不需要在意选择的股票对的真实价值是否一致，尽管投资者观察到的价格可能是不正确的，但这两只股票观察到的价格必须是一致的。在配对交易的过程中，观察到的两只股票价格之间的缺口通常被定义为价差，这个数值越大，那么误定价的程度也越大，股票之间潜在的利润也可能越高。基于以上谈到交易思想，便可以得出，要进行股票之间配对交易的最关键的就是寻找众多股票里的“匹配组合”。如果我们仅从历史经验来寻找具有一定可交易性的股票对，尽管也有一定的可行性，但这种方式的主观性较强，很难保证长期来说股票对之间仍有可交易性。相反，也可以利用一些经济学的模型来对历史数据进行挖掘分析，这种方法的优点就是可以找到更优、更适合交易的匹配对，并对价差序列进行量化的度量，然后可以更准确地知道股票定价偏差的大小。

我们首先假设有两支股票，P_{it} 代表着在时间 t 我们观察到的股票 i 的价格，$P_{\mathrm{it}}=\ln\left(P_{it}\right)$ 就是其相应的对数价格，从很多文献中可以得出，P_{it} 是单位根非平稳过程，并且服从随机游走模型：$P_{\mathrm{it}}=P_{it-1}+r_{it}$，式子中的 $\{r_{it}\}$ 代表着收益率，并且形成了 P_{it} 的不相关新序列。由套利定价理论可以得到，若两支股票的风险因子是相似的，那这两支股票的收益率应该也是相似的。所以，$P_{1\mathrm{t}}$ 与 $P_{2\mathrm{t}}$ 可能被一个公共成分所驱动，并且二者具有协整关系。也就是说，就是 $P_{1\mathrm{t}}$ 和 $P_{2\mathrm{t}}$ 之间可能存在一个线性组合，这个线性组合使得 $\omega_t=P_{1\mathrm{t}}-\gamma p_{2t}$ 成为了单位根平稳过程，因为一个典型的平稳过程是属于均值回复过程的，所以该线性组合也具有均值回复性

质。

我们假设两个价格序列$\{P_{1t}\}$和$\{P_{2t}\}$有误差修正形式为：

$$\begin{bmatrix} P_{1t}-P_{1,t-1} \\ P_{2t}-P_{2,t-1} \end{bmatrix}=\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}\left(\omega_{t-1}-\mu_{\omega}\right)+\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \tag{2-1}$$

从式（2-1）其中$\mu_{\omega}=E\left(\omega_t\right)$表示$\omega_t$的均值，上述中参数$\gamma$、$\mu_{\omega}$、$a_1$、$a_2$的值可采用最大似然估计或最小二乘法进行估计。

方程的左边是由两支股票的对数收益率构成的，这个方程代表着收益率依赖于ω_{t-1}，ω_{t-1}是一个平稳序列。从方程式的具体来看，$\omega_{t-1}-\mu_{\omega}$代表着偏离两支股票长期均衡的偏差值。这个方程式也代表对一支协整股票，其过去偏离均衡的程度决定这其收益率的大小。其中系数a_1和a_2代表着过去序列的偏差分别对收益率r_{1t}和r_{2t}的影响程度，并且在实际应用过程中a_1和a_2的符号应该是相反的，代表的含义为向均衡回返。

然后我们假设有一个投资组合，这个投资组合是由一股股票1的多头和γ股股票2的空头所构成的，我们给定一个时间期限t时，这个投资组合的收益率为：

$$\begin{aligned} r_{p,t+1} &= \left(p_{1,t+i}-p_{1,t}\right)+\gamma\left(p_{2,t+i}-p_{2,t}\right) \\ &= \left(p_{1,t+i}-\gamma p_{2,t+i}\right)-\left(p_{1,t}-\gamma p_{2,t}\right) \\ &= \omega_{t+i}-\omega_t \end{aligned} \tag{2-2}$$

从式（2-2）我们可以得出，这个投资组合在持有期间的收益率$r_{p,t+i}$为价差的增量，这也从另一个角度证明，投资组合的收益率并不会依赖于ω_t的均值。

2.1.2 强化学习理论

强化学习其实是一个学习的过程，但是在人工智能的视角来看，强化学习就是一个从环境到行为的映射，在这个映射中，不同的行为会通过学习逐渐在环境中得到比较高的奖赏值。伴随着国际和国内的金融市场的发展进步，学术界关于人工智能的研究逐渐广泛，其中涉及强化学习的学术研究也慢慢增多。而强化学习其实也是另一种形式的机器学习，并且具有一大优点，就是不用提供另外的训练信息，并且可以正常地参加系统的工作，因此其在别的些领域中也受到了极高的重视，并得到了广泛的应用，主要包括运筹学、控制工程、神经科学、决策理论和心理学等。强化学习在实际研究中涉及的范围较广，在很多行业、领域都有所应用。

（一）强化学习算法的基本概念

（1） 强化学习算法的基本结构和映射

强化学习算法主要包括学习、学习系统以及外部环境这几个基本结构，如图2-1所示。这几个结构中的“学习”，是指计算机程序在运行时，会增加自身经历的次数，进而可以提升系统性能；其中的“学习系统”，是指一个系统可以对未知的环境有比较好的适应力；而在强化学习中，智能体就相当于另一种形式的学习系统，因为两者都有很多相似的特点，比如智能体的一个明显的属性为学习能力，就是通常来说，一个具备学习能力的智能体也会被当作学习系统，因此在后文中，我们把智能体和学习系统当作等价的概念；而这几个结构中的“外部环境”，通常是与智能体交互的一个对象。

接下来主要介绍强化学习算法的映射，其主要包括状态、行为、策略和强化信号。所谓的“状态”，也就是外部环境，其集合也被叫做环境的状态空间。映射中的“行为”，也可以叫做“动作”，在学习算法中也被认为是系统做出的决策或控制行为，代表学习系统在某一个状态会选择不同的控制行为来作用于环境。在实际应用过程中，当前的状态往往会制约行为变量，使得其不能随意取值，只能被限制在一定的范围内，而这个范围被叫做允许行为空间。而“策略”的含义，就是系统做出的不同决策构成的一个按顺序生成的序列，也可以叫做是决策序列。因此策略可以当作从状态空间到行为空间的一个映射。最后，“强化信号”是在系统工作过程中，外部环境针对学习系统的行为给出的奖赏值或者惩罚值。一个标准的强化学习的系统，其基本结构如图2-1所示，其中单智能体就是学习系统，从环境中得到当前状态的信息s，同时把试探行为u反馈给环境，进而得到环境对该行为u的评价值r，以及环境状态s'。在整个过程中，假如这个行为u使环境输出了正的奖赏值，那智能体在之后会更倾向于选择这个试探行为；假如这个行为使环境输出了负的惩罚值，那智能体在之后的循环中会更避免选择这个试探行为。在迭代过程中，由状态至动作的映射会不断改变，最终系统的性能会得到优化。

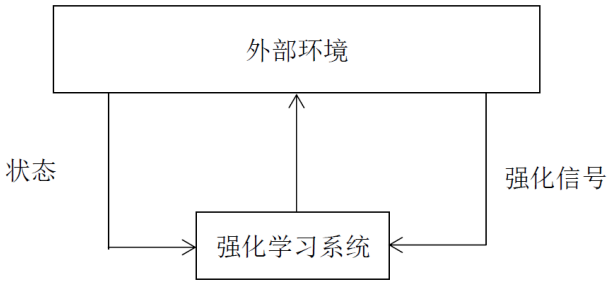

图 2-1 强化学习系统原理

（二）理论背景与基础模型

在强化学习的众多模型中有简单也有复杂的，较为复杂的模型主要体现在，外部环境不仅仅只有单状态，而且后续报酬受后续状态影响；较为简单的模型主要体现在，外部环境中仅有单状态。其中目前主要的研究对象是复杂强化学习算法模型，在强化学习研究领域，复杂强化学习模型影响后续报酬的特性也被称为延迟报酬。接下来我们将基于复杂强化学习算法模型，来简单讲述一下强化学习模型的理论背景，并且以这些为基础来详细介绍一些强化学习的基本模型。

（1） 理论背景

强化学习模型的主要理论背景包括马尔可夫决策过程和半马尔可夫决策过程。通过研究过去的文献，我们发现大部分学者的研究都是基于马尔可夫决策过程的。然而需重点提出，强化学习并不受限于马尔可夫决策过程，只是这个过程可以用来处理随机序贯决策的问题，并且这个过程是离散时间有限状态，换句话说要求时间状态是离散且有限的，在研究强化学习的基本算法时，此特点提供了最简单的框架。

马尔可夫决策过程经常被用来解决一些随机性序贯决策问题。但是实际在运行过程中，系统在每一个观察点上都得做出决策行为，并且在事前并不会准确了解系统状态的转移的概率值，因此做这种决定有一定的随机性。所以系统运行时在每一个观察点上，决策者会依据从历史数据中观察到的系统状态，从众多方案中选一个决策来执行，这一步也叫“做出决策”。接下来决策者会继续观察得到新的状态，然后选择新的决策行为，这样循环进行，最终系统会到最优状态。在整个序贯决策的过程中，当系统状态的转移跟之前系统的发展无关时，我们以称其为无后效性或者马尔可夫性，并且马尔可夫决策过程可以处理这种问题。

（2）基础模型

像上文所提到的，强化学习模型是以马尔可夫决策过程和半马尔可夫决策过

程为背景的，因此，强化学习模型主要包括以几个基本的元素：报酬函数、值函数、策略π、行为空间U、状态空间S。

下文中将重点介绍强化学习中的值函数，本文将会分别在马尔可夫决策过程和半马尔可夫决策过程背景上给出其定义。然后基于此定义，进一步分析得出强化学习模型。

在有了马尔可夫决策过程背景下的强化学习模型中的值函数以后，在这个值函数的基础上对马尔可夫决策过程状态值函数进行修改，或者用原本马尔可夫决策过程中对状态值函数的定义，来定义状态-行为的值函数$Q^{\pi}(s,u)$。$Q^{\pi}(s,u)$的定义就是：

$$Q^{\pi}(s,u)=E\left(r_{t+1}+\gamma r_{t+2}+\gamma^{2}r_{t+3}+\cdots\middle|s_{t}=s,u_{t}=u,\pi\right) \tag{2-3}$$

式（2-3）中Q^{π}代表的是状态-行为的值函数。并且值函数$Q*$代表的是最优策略$\pi*$下对应的状态-行为的值函数，因此也被叫做最优化值函数。在强化学习算法中的一些算法里，$Q*$发挥着十分重要的作用。$Q*$代表在任意一个状态-行为(s,u)下，在之后阶段通过使用最优策略而取得的报酬的和。$Q*$跟$V*$之间的关系为：

$$V^{*}=\max Q^{*} \tag{2-4}$$

因此，得出

$$Q^{*}(s,u)=R(s,u)+\gamma\sum_{S'\in S}P(s'|s,u)\max_{u'\in U_{S'}}Q*(s',u') \tag{2-5}$$

式（2-5）代表的是状态－行为值函数的Bellman方程。所以可以看出，对在状态S下的行为u评价，然后得到的评价值即为$Q^{\pi}(s,u)$，并且与V^{*}等价，也就是说通过求解V^{*}得出的最优策略，这跟通过求解$Q^{\pi}(s,u)$的方式间接得出的最优策略，具备着同样的效果。以上就是在马尔可夫决策过程背景下，Q^{*}的定义、Q^{*}的Bellman方程以及值迭代公式。

并且，状态值函数与状态-行为值函数两者在半马尔可夫决策过程背景下，Bellman方程如下：

$$V*(s)=\max_{u\in U_{s}}\left[R(s,u)+\sum_{s'\in S、\tau}P(s',\tau|s,u)V*s'\right] \tag{2-6}$$

$$V*(s)=R(s,u)+\sum_{s'\in S、\tau}\gamma^{\tau}P(s',\tau|s,u)\max_{u'\in U_{s'}}Q*(s',u') \tag{2-7}$$

式(2-7)中τ表示状态s以及行为u保持的时间；$P(s',\tau|s,u)$表示联合概率。

在强化学习模型里面，学习系统中关于如何做出动作，也就是某一时刻的行

为方式是由策略来决定。在某些时候，一个简单的函数或表格也可以表达出策略，换句话说，策略是从状态到行为的一个映射，其中的“状态”为环境的状态，“行为”是不同状态下采取的行为。而当策略变得比较复杂时，就会涉及到大量的计算。因此可以得出，强化学习算法的核心便是策略。

在强化学习问题中，一般来说，是把状态或状态-行为映射到一个数值，其中的状态为感知环境的状态。而策略是否改变是由所获得的报酬值的大小而决定的，如果说策略在做出了行为后得到的报酬很低，那么策略在下次就会去选择另外的行为，此时策略就会因为得到了比较低的报酬而变化。

报酬函数和值函数两者都表明了系统采取的行为重哪个最有利，只是报酬函数使从当前来看的，而值函数则是从长远来看的。通常在实际进行决策和评估决策的时候，选择行为的过程要基于值判断才可以执行，而且受到关注最多的就是值函数。但从环境可以直接得到报酬，值函数却要从智能体整个生存期的数据中得到，所以跟报酬函数相比起来，值函数更加不好确定。

（三）强化学习算法的典型算法

强化学习从提出到现在一直在不断发展进步，同时也从经典的标准强化学习算法中衍生出不同类型的算法。所以先研究典型算法，再进行其他研究就会容易很多。

强化学习的很多算法着相似的核心部分，就是都会对状态值函数$V*$或者状态-行为值函数$Q*$进行迭代，进行迭代的基本公式为：

$$V(s)\leftarrow(1-a)V(s)+a(r(s,u)+\gamma V(s')) \tag{2-8}$$

式（2-8）中a是随时间衰减的学习率。

（1） TD算法

TD学习算法属于强化学习算法中最主要的算法之一。是由Richard S. Sutton在1988年首次提出。其迭代公式为：

$$V(s)\leftarrow V(s_t)+a(r_{t+1}+\gamma V(s_{t+1})-V(s_t)) \tag{2-9}$$

此算法的具体学习过程如以下几步：第一步先对V值进行初始化；接下来在状态s_t下，学习智能体会依据策略来确定当下的行为u_t，在这个过程中获得经验$(s_t,u_t,r_{t+1},s_{t+1})$；接着智能体根据获得的经验，来改变当前的状态，直到成为目标状态的时候算法停止。接下来系统会重新从初始状态进行迭代，直到学习结束。这个算法个特点就是会把蒙特卡罗和动态规划思想两者结合，也就是说，TD算法

既不需要系统模型，通过系统的经验进行学习，又能像动态规划思想一样，利用迭代来估计最优值函数。

（2） Q-learning算法

其实Q-learning算法的本质也是另一种形式的离策略TD算法，也就是只取最大值，不管之前策略带来的影响。Q学习跟TD算法不同的点在于，Q学习在迭代中采用的是状态-行为值函数。而与之也有相同的点，就是Q学习的最简单形式也是一步算法，也称为一步Q-learning。Q学习算法的迭代规则为：

$$Q(s_t,u_t)\leftarrow Q(s_t,u_t)+a\left[\gamma_{t+1}+\gamma\max Q\left(s_{t+!},\underset{u\in U_s}{u_{t+1}}\right)-Q(s_t,u_t)\right] \quad (2\text{-}10)$$

在式（2-10）中的行为值函数Q为函数Q^*的估计值，也可以说是期望值的最优行为值函数，与前面所采取的策略是相互独立的。这样一来，可以很大程度上将算法的计算量减少。

接下来介绍Q-learning算法的具体学习过程：首先将$Q(s,u)$，状态s，进行初始化；并且在状态s_t下利用Q得到的策略来选择下一步的 行为u_t，执行u_t，并观察得到r_{t+1}， s_{t+1}，然后可以得到经验$\langle s_t,u_t,r_{t+1},s_{t+1}\rangle$，接下来由Q-learning的算法迭代规则来更新$\langle s_t,u_t\rangle$；直到达到目标状态的时候，结束循环。

从强化学习算法的整体研究流程上来看，提出Q学习算法可以算得上是取得了阶段性的胜利，并为后面的研究打下坚实的基础。

（3） Sarsa算法

在Sarsa算法刚被提出时，被称作时改进版的Q-learning算法，Sarsa算法采用的仍然是Q值迭代，其可用式（2-11）表示：

$$Q(s_t,u_t)\leftarrow Q(s_t,u_t)+a\left[\gamma_{t+1}+\gamma Q(s_{t+!},u_{t+!})-Q(s_t,u_t)\right] \quad (2\text{-}11)$$

智能体在每个学习步骤中，策略要先由当前的Q值来决定行为u_t，然后得到经验$\langle s_t,u_t,r_{t+1},s_{t+1}\rangle$,接着策略由当前的$Q$值来决定$s_t$状态下的行为$u_t$，由此可以得到一个五元组$\langle s_t,u_t,r_{t+1},s_{t+1},u_{t+1}\rangle$，并且由公式对$Q(s_t,u_t)$进行更新，然后把行为$u_{t+!}$当作是智能体的下一个行为去执行。因此，Sarsa算法与Q学习算法的主要不同在于，后者是利用后续状态的最大行为值来进行迭代，而前者是利用实际的后续状态-行为值来更新当前状态。由此可以看出，Sarsa算法算是一种在策略的TD算法，并且也符合算法收敛性的理论。

2.2 文献综述

2.2.1 商品期货的文献综述

参考近些年的研究文献，我们可以发现，众多学者的量化交易研究都是以我国的股票市场为基础，但是实际上与国外股市相比，我国股市仍然存在着很多的缺点，比如国内股市缺乏良好的做空交易机制，由于股票市场的T+1交易，实际操作中无法实现高频交易模式等。而相对于股票市场来说，我国期货市场在配对交易策略的使用上具有显著的优势。因此本文选择期货市场作为研究对象。

随着国内商品期货市场中投资规模的日渐扩大和投资者数目的不断增加，该市场在最近这些来年取得了快速的发展。其中丁秀玲和华仁（2007）[1]以大连商品交易所的交易所的豆粕期货和大豆期货作为研究对象，对其进行跨品种套利研究，研究结果发现这两个品种之间的套利机会不是非常明显，主要原因是我国商品期货在当时的品种数量比较少。之后随着国内的股指期货上市，仇中群和程希骏（2008）[2]以股指期货为研究对象，进行了跨期套利的研究，研究结果显示基于协整的跨期套利可以保证价差在合理的区间内波动，并且发现基于这个原理建立的配对交易套利模型可以有良好的获利效果。何树红等（2013）[3]是通过GARCH模型对股指期货构建配对交易策略，通过实证结果论证了日内跨期套利可以得到较高的收益。扈文秀等（2013）[4]选取了5分钟高频数据，通过在商品期货指数和大宗商品类股票之间]以协整理论为基础建立套利策略，证实了此方案下套利策略具有可行性。覃良文等（2015）[5]通过使用HP滤波方法，将两个满足协整理论的沪铜期货价格序列，分解为长期趋势和短期周期波动的两个部分，得到的套利结果显示出平滑指数较小，并且利润率比较高。于孝建和邹倩倩（2018）[6]对商品期货构建了基于O-U过程的配对交易策略，结果表示配对交易策略在我国商品期货市场有一定的有效性。丁纯（2020）[7]以沪深300ETF期权与股指期货为研究对象，构建了在不同市场的相同标的资产的金融产品的套利模型，证明其虽受宏观因素影响影响短期价格有所背离，但经市场套利行为价差终收敛回归至均衡状态。路旭洋（2020）[8]以商品期货为研究标的，对传统的商品期货量化套利模型使用和声搜索算法进行了优化，并于传统的遗传算法优化进行了对比，结果显示基于和声搜索算法的套利模型寻在找最优解效果和收敛速度上都有所提升。

2.2.2 配对交易的文献综述

配对交易理论最早出现于20世纪80年代的美国，并且这个理论属于统计套利和量化投资领域中比较重要、比较主流的一种投资策略。GRANGER（1981）[9]最早提出该思想；Board（1996）[10]对不同价差序列通过使用协整理论，得到的分析结果显示其存在协整关系的同时，也有着套利空间；GATEV et al.（2006）[11]也对配对交易的基本原理做出了详尽的解释；王伟峰、刘阳（2007）[12]和仇中群、程希骏（2008）[13]以沪深300股指期货为研究对象，并且基于协整模型对其进行模拟交易高频数据分析，实证结果检验了基于协整的套利模型具有有效性，并且很有效率。这些众多的研究都表明一个事实，配对交易策略同样适用于中国市场。

在涉及到的研究方法上，HUCK（2010）[14]将神经网络以及多步预测引入了配对交易，基于标普100指数股票的应用结果也令人满意，为配对组合的选择和检验提供了理论和方法；李世伟（2011）[15]在考虑异方差和ARCH效应的基础上，通过建立基于GARCH模型的协整套利策略，证明了改进的模型在套利效果上优于传统的套利策略；SONG et al.（2013）[16]通过采用HJB方程刻画价值函数的方式进行研究，得到的结果显示出最优平仓问题是可以利用一系列quasi-algebrai方程来解决的，并且还给出了一些数值分析的案例。ZENG et al.（2014）[17]研究了统计套利中的许多问题，包括资产组合选择、参数边界寻优和最优交易策略设计等。王利斌(2014)[18]在套利交易中引入了变结构协整理论，从不同的角度来证实在股指期货的跨期套利中，变结构协整比传统方法更优。也有一些学者把其他领域的方法比如人工蜂群算法（2014）[19]、神经网络（2014）[20]等运用进来。

除了这些基于投资组合理论和统计分析方法，在挑选不同配对组合，以及证实配对组合的有效性方面，也有很多研究。比如赵胜民等（2015）[21]对内地的一些主要指数成份股进行实证分析交易，并验证了其有效性。此外，还有一些运用随机控制（2015）[22]、遗传算法（2015）[23][等方法的研究。朱丽蓉等（2015）[24]也以我国期货市场为研究对象，采用协整模型、误差修正模型及基于协整关系的GARCH模型来进行实证研究，结果发现基于协整的GARCH模型最优。邢恩泉（2015）[25]利用计算机快速循环运算的特点，来将传统的协整配对交易策略进行一定的改进，通过对经验型选择参数进行遍历性研究，来循环查找最优配对组合以及确定建仓阈值，进而使得模型可以根据数据变化来进行自我动态修正。胡伦超等（2016）[26]以融资融券标的股数据为研究对象，证明了配对交易的有效性。Rad（2016）

[27]以美股为研究对象，分别运用价差、协整、Copula策略三种模型进行研究，结果显示在动荡的市场条件下，协整策略是最优策略；NGO et al.（2016）[28]把配对交易的规则进行简化，转化成3种组合结构之间的最优切换问题，也就是A和B都空仓、A长仓B短仓、A短仓B长仓三种，结果证实了存在最优切换点，而且用一些数值仿真方法做了举例证明。

随着参数研究的不断深入，学者们开始注意到固定参数以及静态模型的局限性。刘阳（2016）[29]等创造性地把神经网络和动态GARCH模型结合起来，并且挖掘出价格偏差中存在的非线性特征，使得构造的动态GARCH模型可以更及时地感知到波动性的变动，来达到降低传统静态模型预测偏差的目的；张波和刘晓倩(2017)[30]在前人构造EGARCH-M模型的基础上，通过提出新的协整关系——修正的协整模型来进行套利研究；毕秀春、于晓雨（2020）[31]等人将遗传算法与部分协整理论相结合，利用遗传算法求得了最优阈值，克服了部分协整理论在阈值选取粗糙等方面的问题。

2.2.3 强化学习的文献综述

（一）强化学习算法的发展

强化学习属于机器学习的一种主要的模式，在没有知识背景和预定义的情况下，强化学的相关算法可以通过数值化处理，来表现出其强大的学习能力；并且能够通过在学习模型与所处环境之间不断进行交互，利用反馈得到的信息来一步步改进决策能力。在强化学习算法中的众多算法里面，Q-learning学习和Sarsa学习是其中两个较为重要的，Q学习是一种离策略，而Sarsa是一种在策略，并且效果通常好于Q学习算法，但是标准Sarsa算法对状态空间的限制条件较多，要求其必须离散并且空间数较小。

这两种算法中Q学习也是最早的在线强化学习算法。Moore（1993）[32]提出若使用Q学习做函数逼近，在某些特定况下马尔可夫决策过程并不收敛，同时一些微小的噪音可能也会使得无法选择出最优策略。Jaakkola（1994）[33]认为Q学习其实是一种离策略的学习算法，只有当存在着一定的限制条件、且在理论上可以收敛的情况下，才可以利用Q学习求得最优控制策略；在我国Sarsa算法也被应用在很多新兴产业中，其中包括组织运作过程控制（2004）[34]、网络建模（2009）[35]、机器人控制（2014）[36]和交通信号控制（2015）[37]等不同的领域。

（二）强化学习算法在金融领域的应用

金融交易中的市场环境有其独特的特点，也就是这个市场环境是高度动态嘈杂的，而强化学习方法的特殊之处也在于其能与动态环境相互作用，并且利用交互得到的经验或者奖励值来提高系统的决策能力，因此，学术界的很多学者便开始将强化学习与金融市场结合起来。SUTTON et al.(1999)[38]提出若要求解金融相关的问题，那么一定存在不确定性和动态性，其认为强化学习算法比较适合求解这类问题。

目前为之，虽然强化学习方法在金融相关领域的运用较多，但实际上主要集中在证券交易尤其是高频交易和投资组合管理方面。Moody和Wu(1998)[39] 不仅对RRL的理论依据、组织构成进行了详细地解释,另外还分别对比了信息比率和斯特林比率各自作为目标函数时的收益结果，结果显示，以标准普尔500指数和部分美股为研究对象的实证中,以斯特林比率作为目标函数的模型收益相对比较高；Jangmin(2005)[40]等人在2005年构建了基于RRL的自适应投资组合策略,可以比较有效利用一些来自特定股票、基金的时序信息来进行训练,并且基于这种新型策略，可以在投资组合中对高风险资产、无风险资产各自的份额进行比较合理地分配,另外此新型策略策略在韩国的股市的表现也优于一些经典的资产配置策略；Bekiros （2010）[41]创新性地把自适应网络模糊推理系统和强化学习两者进行结合，从而构建出一种非套利的高频交易系统；TAN et al.（2011）[42]在强化学习算法地基础上，使用了自适应网络模糊推理系统的人工智能模型，进而构建出了一个非套利型的高频交易系统。Yang等人（2012）[43]在股票交易中应用了Q学习地方法，Fallahpour（2016）[44]等首次把强化学习与协整配对交易策略两者相结合，首先为了满足算法的条件，先把估计窗口、交易窗口、交易阀值与止损阀值这几个参数进行离散化地处理，把索提诺比率当作模型中的奖励函数，使用梯度策略方法对交易时间窗口、交易阈值等参数进行动态调整，从而使得模型可以自适应于金融市场的环境变化；在此之后，胡文伟等（2017）[45]以我国债券市场上交易量最大的债券产品为研究对象，以强化学习模式的参数动态优化为基础，构建出了自适应配对交易模型,实证分析的结果表现出此自适应配对交易模型显著地提升了交易的收益率,另外这个模型还可以使得累计收益率不断增加直到收敛于最大值，具备着持续学习的能力。赵珊珊（2018）[46]在此基础上将Sarsa强化学习方法与配对交易相结合，实现了模型的动态参数调整，并将此模型成功地应用在银行业的股票市场中。Chen（2018）[47]等在Agent强化学习算法的基础上，

开发出了一种交易系统，并且此系统可以模仿专家系统的交易策略。王欣等（2019）[48]提出了未来智能动态定价的研究方向，以便将强化学习技术应用在动态定价领域。王现磊等人（2019）[49]构建出了基于模拟退火策略的Sarsa算法，并与传统的Sarsa算法、Q学习算法进行比较，在测试表现中结果显示优于后两者，也表明了交易算法需要良好的自适应性。

目前为止，尽管强化学习模型已经在在金融领域有很多应用，但是跟其他领域比较起来，其在金融领域的应用才刚刚处在起步的阶段，尤其是在金融领域的配对交易、统计套利上的应用更是比较少。

2.2.4 文献述评

目前国内对于传统静态协整模型有着许多的改进方案，尽管都有着各自特定的样本条件，但都取得了一些成效，相对于传统协整模型来说也有一定进步，但是其实这些改进方法都受到了一些使用条件的限制，有些要求数据符合GARCH模型或者O-U过程等等，有些要先预设经验性参数等专家系统，并且当环境发生超预计变化的时候，这些改进方案都不能及时应对，所以其仍然有着局限性。

因此，用强化学习模型中的Sarsa算法与$\varepsilon-greedy$算法结合，对传统的模型进行改进，可以实现模型参数的自适应动态调整，进而开发出一种可以随着环境变化进行自适应调整，并且不需要知识背景、不需要进行预定义的动态优化策略，是十分有意义的。并且国内的研究中还未将此思想用于中国期货市场，因此本文的研究可以提高国内期货市场配对交易效率、提升配对交易的绩效，是研究的一大突破口。

第 3 章 基于强化学习的配对交易策略的理论框架

本章将介绍基于强化学习的配对交易策略所需要的相关研究方法和理论框架。首先介绍配对交易中选择交易对时需要的的相关性分析、平稳性检验、协整关系和E-G两步检验法的具体理论与步骤；其次介绍构建强化学习交易模型系统的思想，包括交易模型系统的整体框架、工作流程、奖赏值选取等；最后介绍商品期货配对交易模型的评价分析指标，以及各个指标的意义和计算方式等。

3.1 最佳期货组合挑选

在配对交易策略的形成期内，要利用许多计量模型，逐步筛选出配对交易模型需要的最佳期货组合。第一步，通过相关性分析，筛选出大于相关系数阈值的期货对；接下来，将剩下的期货对进行单位根检验，若满足一阶单整就保留，否则删掉；然后，对保留的期货组合进行协整检验，从而筛选出具有长期稳定的协整关系的期货组合，并且作为最终期货组合。以下将依次介绍相关性分析的具体原理、单位根检验的具体原理以及E-G协整检验的具体步骤。

3.1.1 相关性分析

相关性分析是对两个或多个随机变量间的相关关系进行分析，可以衡量变量间的相关程度。设两个序列变量为 $X=(x_1,x_2,\cdots,x_\tau)$，$Y=(y_1,y_2,\cdots,y_\tau)$，则其 R 系数可以表示为：

$$R=\frac{\sum_{t=1}^{T}\left(x_t-\overline{X}\right)\left(y_t-\overline{Y}\right)}{\sqrt{\sum_{t+1}^{T}\left(x_t-\overline{X}\right)^2\sum_{t=1}^{T}\left(y_t-\overline{Y}\right)^2}} \tag{3-1}$$

式（3-1）中，$\overline{X}$ 和 $\overline{Y}$ 分别为变量 X 和 Y 的均值，R 的范围在[-1,1]之间，如果 R 的绝对值 $|R|$ 越大，说明两个变量之间的相关性越强。特别地，当 $R=1$ 的时候，表示着两个变量值间完全正相关；当 $R=-1$ 的时候，表示两个变量值间完全负相关。所以说相关系数数值的大小，一定程度上可以代表变量之间相关性的大小。

3.1.2 平稳性检验

（一）平稳性的概念

分析时间序列，就是为了使用变量的过去预测变量的未来，其前提就是要求历史可以重演，也就是过去的数据有代表性和可延续性，并且必须能贯穿整个时期。不然的话，就不可以根据历史来预测未来的想法。接下来介绍平稳的概念，若一个时间序列的基本特性维持不变，那它就是平稳的；而如果其基本特性只存在于所发生的当期，不能维持不变，那么我们也给这样的时间序列叫做非平稳的时间序列，并且不能通过其预测未来。在金融领域中的有很多变量经常发生突变，从而形成一种不平稳的状态，所以在金融领域的很多变量都难以估计。我们对时间序列的进行有关分析，首先需要这个序列为平稳时间序列，这样的条件下对其研究和预测，才能取得我们想要的结果。并且平稳时间序列也分为强平稳和弱平稳，其中强平稳的条件比较严格，在实际应用中不容易满足，所以通常来说，平稳的时间序列也就是弱平稳时间序列。

（二）强平稳和弱平稳

存在一个时间序列 $\{X_t\}$，那么对于任何一个正整数n，任意取$t_1,t_2,\cdots,t_n \in T$，同时对于任意一个整数τ 都存在：

$$F_X\left(x_{t_1},x_{t_2},\cdots,x_{t_n}\right)=F_X\left(x_{t_{1+\tau}},x_{t_{2+\tau}},\cdots,x_{t_{n+\tau}}\right) \quad (3\text{-}2)$$

式（3-2）的意义也就是，$X_{t_1},X_{t_2},\cdots,X_{t_n}$ 的联合分布$F_X\left(x_{t_1},x_{t_2},\cdots,x_{t_n}\right)$与$X_{t_{1+\tau}},X_{t_{2+\tau}},\cdots,X_{t_{n+\tau}}$ 的联合分布$F_X\left(x_{t_{1+\tau}},x_{t_{2+\tau}},\cdots,x_{t_{n+\tau}}\right)$等价，所以可以认为序列$\{X_t\}$满足强平稳的条件。因此也可以看出，这个序列需要满足条件比较严格，在理论上不容易证明，同时在实际应用时也不容易检验。

跟强平稳需要的前提条件比起来，弱平稳更宽松，所以也可以把弱平稳性叫做宽平稳性。接着来详细介绍弱平稳，若平稳是用序列的低阶矩来表示其主要性质的，因此只需要保证一个序列的低阶矩平稳，就可以把这个序列归属为弱平稳的范畴。我们假设$\{X_t\}$是一个具体的时间序列，当$\{X_t\}$符合以下三个条件：任意取$t\in T$，都有$E\left(X_t\right)=\mu$，即时间序列的均值是固定的常数；任意取$t\in T$，都有对应的$E\left(X_t^2\right)<\infty$，即时间序列存在着二阶矩；任意取$t\in T$，对于任意一个整数$h$，阶数$m$，存在$v_m\left(t\right)=v_m\left(t+h\right)$，也就是在$t$时点的$m$阶自协方差$v_m\left(t\right)$和$t+h$时点的$m$阶自协方差$v_m\left(t+h\right)$是相等的，换句话说，$m$阶自协方差$v_m\left(t\right)$不会因为时间点$t$的变化而改变，而是只跟阶数$m$的大小有关。这时我们就把序列$\{X_t\}$称作是弱平稳的时间序列。

（三）ADF 检验

单位根检验作为一种统计检验方法，可以用来检验时间序列平稳性，其特点为比较客观，而且具有一定的说服力。在单位根检验中，用的比较多的有DF检验、ADF检验以及PP检验。在本文接下来的研究中，是要对两个变量进行单位根检验，因此将使用ADF检验方法来检验变量之间的平稳性。ADF检验的公式可以表述为：

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \cdots + \delta_p \Delta y_{t-p} + \varepsilon_t \quad (3\text{-}3)$$

在式（3-3）中，α 是截距项；β 为对应的趋势项；$\delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \cdots + \delta_p \Delta y_{t-p}$ 则为ADF检验的增广项，可以通过AIC和BIC准则来指定具体期数。ADF检验的原假设和备择假设分别是：

$$H_0 : \gamma = 0 \text{；} \ H_1 : \gamma < 0 \quad (3\text{-}4)$$

该检验对应的统计量为：

$$DF = \frac{\hat{\gamma}}{SE\left(\hat{\gamma}\right)} \quad (3\text{-}5)$$

在式（3-5）中，SE是标准误。

3.1.3 E-G两步协整检验

（一）协整关系

通常来说，如果有两个或者多个序列自身非平稳，但是存在着一个它们之间的线性组合为平稳，那我们可以称它们之间是具有协整关系的。我们首先来介绍单整的概念，就是如果一个时间序列 Y_t 自身非平稳，但经过 d 次差分以后变为平稳序列，那么我们就可以把序列 Y_t 称为是 d 阶单整，一般记作 $Y_t \sim I(d)$。接着给出协整关系的定义：假设时间序列 X_t 为 d 阶单整，若同时存在其它的时间序列 $Y_t = \beta X_t \sim I(d-b)$，那么一般称时间序列 X_t 是具有 d、b 阶协整关系的，一般记作 $X_t \sim CI(d,b)$，其中 $b>0$，β 则是一个非零向量，把 β 称作是协整向量。

特别地，当时间序列 y_t，$x_t \sim I(1)$ 并且 $y_t = \alpha + \beta x_t \sim I(0)$ 的时候，称 x_t 和 y_t 都是协整的，α 与 β 则是协整系数。

（二）E-G两步协整检验法

两个时间序列存在协整关系需要满足两个条件，首先它们必须为一阶单整向量，也就是说本身非平稳，但一阶差分后的序列平稳；其次是这两个序列必须存

在着一个线性组合是平稳的，也就是残差平稳。所以，在对两个时间序列建立线性关系之前，要先检验它们两个之间的协整性。

若想要进行协整检验，也就是要检验两个时间序列变量是否存在协整关系。通常来讲有两种方法，第一种是E-G两步检验法，具体来说是对线性回归的残差来进行协整检验；第二种是J检验，是从线性回归的回归系数的角度来进行协整检验。在本文中，将采用的主要方式是E-G两步检验法。接下来详细介绍此检验方法，最早是在1987年由Engle和Granger两位学者提出的，具体思路如下：若是两个变量满足协整的条件，那么就证明这两个变量之间存在着协整关系，那么在很大程度上，由自变量所组成的线性组合（比如一元线性回归方程）可以来解释因变量；并且，因变量和该线性组合的残差序列之间应该是平稳的。因此接下来要检验这两个变量之间是不是协整，也可以说是要检验线性回归方程中的残差序列是不是平稳。

E-G两步协整检验法的详细步骤如下：第一步要先用最小二乘法(OLS)来对序列 x_t 与序列 y_t 之间的关系系数进行估计，第二步要用单位根检验法来检验这两个序列之间的残差项是不是平稳的；接着先用ADF检验统计量来检验变量 x_t 与 y_t 的平稳性，再利用模型 $y_t=\alpha+\beta x_t+\varepsilon_t$，采用OLS回归估计出参数 α 和 β 的值，最终再通过ADF检验统计量对残差 ε_t 估计值的平稳性进行检验，整体的检验流程如图3-1所示。

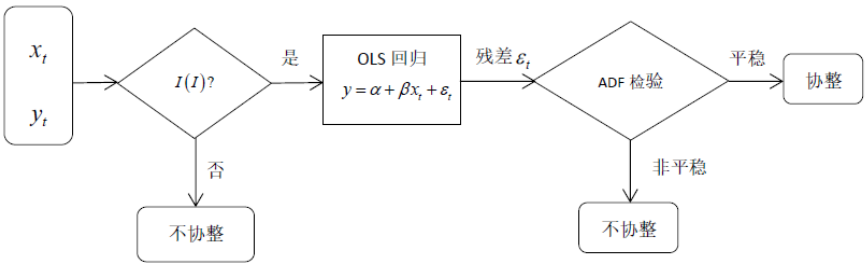

图 3-1 EG 两步协整检验方法示意图

3.2 构建强化学习交易模型系统

由上文得到最终配对组合后，将利用基于强化学习构建的配对交易模型，对最终配对组合的训练集数据进行参数的不断迭代与优化，得到最优参数后利用测试集数据进行回测检验。以下将介绍基于强化学习的配对交易模型的一些具体原

理。

将强化学习算法的思想与配对交易模型相结合，形成的改进的新型交易模型系统如图3-2所示。在这个配对交易系统中，最核心的就是交易决策系统，要作的任务就是进行交易指令的决策和执行，在强化学习系统中相当于是智能体；可以把证券市场当作是交易系统中的环境，证券的价格当作是配对交易系统中的环境状态；把投资绩效的某种评估指标当作是配对交易模型中的奖赏值，也是决策系统与环境交互的关键；其中评估时间窗口、交易时间窗口、开仓阈值、平仓阈值4个参数，就可以当作是智能体的具体行为，这些要素交互进行，使交易系统动态地工作。

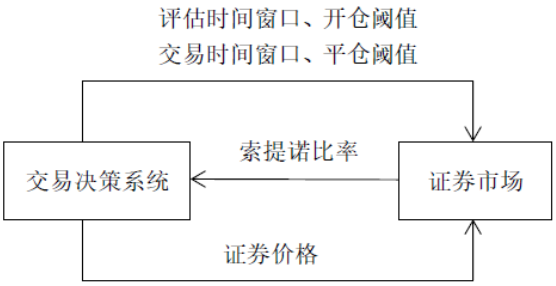

图 3-2 交易模型系统

我们从上文可以看出，整个交易系统的工作流程将从预设4个参数的初始值开始，接下来智能体会去跟踪配对期货的价差序列，并且密切监控着环境的状态；如果说配对期货的价差偏离了长期的均值，同时超过开仓阀值的时候，智能体就可以根据环境状态的改变，而对配对期货组合下达开仓指令，也就是买入走势比较强的期货，卖出一些走势比较弱的期货；当持仓成功建立之后，智能体就会继续地监控和评估环境状态，持续地跟踪配对期货的价差序列，同时在整个过程中实时动态地对参数值进行调整。若配对期货的价差序列开始下降并且回归到正常值以后，或价差序列不断扩大并且超过平仓阀值，这两种情况都会使得智能体对配对资产期货组合下达平仓指令，也就是说卖出一些走势比较强的期货，买入一些走势比较弱的期货，并输出交易系统中的奖赏值作为奖励；接着随着当前的信息和值函数都被更新，整个算法会重新进行迭代，同时智能体也会继续关注着环境的状态，并持续跟踪配对期货的价差序列，直至下一次期货投资组合重新建仓；这样重复很多次，直到整个投资期结束。在决策系统运行的整个过程中，智能体会不停地基于每次期货投资组合开仓、平仓获得的奖赏值和环境状态的变化，对最优参数进行动态地调整。

参考过去的很多研究，较多学者都使用相对来说比较简单的夏普比率来当作奖赏函数，来评定投资组合的绩效，但在后来的许多研究中，学者们逐渐对风险的考量增加，然而因为夏普比率本身对这方面的考量较为缺乏，因此胡文伟等人在参考前人研究的基础上，采用索提诺比率作为江山该函数。索提诺比率的定义为：

$$SortinoRatio = \frac{R-T}{TDD} \tag{3-6}$$

式（3-6）中 R 指的是平均收益率，TDD代表的是最低要求收益率（本文设为0），TDD代表下行风险方差，表示为：

$$TDD = \sqrt{\frac{1}{N}\left[\min\left(0, r_t - T\right)\right]^2} \tag{3-7}$$

式（3-7）中，r 代表着第 i 期的收益率，N 代表总期数。因此在交易策略的目的主要是实现收益最大化的时候，索提诺比率可以当作是一个较合适的性能评定指标。并且在评估配对交易模型的性能时，与夏普比率比起来，索提诺比率更关注于策略的下行风险。

为了避免寻找最优参数式陷入局部最优的情况，本文将使用 $\varepsilon - greedy$ 探索策略，此策略在选取动作的时候，会引入一些随机变化来处理开发与利用之间的平衡问题，换句话说就是以概率1-$\varepsilon\left(\varepsilon \in [0,1]\right)$来利用已有策略，以概率 ε 来搜索新的策略。在学习过程中的初期，ε 可以挑选较大的数值，慢慢地随着迭代次数增加，随着时间推移，智能体的学习程度不断加深，经验不断变得丰富，随机性就可以逐渐降低，ε 值就会逐渐变小。

基于以上的论述，尽管Sarsa算法对状态空间在数量和离散性上的要求更高，但是若要进行金融领域的配对交易，其涉及到的状态空间相对比较容易满足这些要求。并且在策略的实际效果一般来说会优于离策略，并且本文主的目的是验证模型的有效性，所以本文决定采用Sarsa算法进行实证研究。

3.3 商品期货配对交易的评价分析

将每一个期货对都进行回测检验后，我们需要得到一系列的评价指标，来量化模型的“好坏”程度。

因为不同期货品种的合约价值不同，如果仅从总收益而言，不足以表现不同

配对交易策略的差异，因此本文采用了如表3-1所示的5个指标，对策略在不同交易对上的表现进行衡量。

表 3-1 评价指标

评价指标	累计收益率
	年化收益率
	交易次数
	平均每笔回报
	索提诺比率
	最大回撤
	信息比率

现将表3-1中的各指标解释如下：累计收益率，也就是总收益率，在后文的研究中，这个值就是把每次交易的收益率累计相加的和。年化收益率，即将累计收益率年化，代表该笔投资一年收益的比率。交易次数，即一共进行的交易次数（我们将开仓并平仓当作是一次完整的交易）。其中平均每笔回报=累计收益率/交易次数。

索提诺比率，代表着投资在承担着相同单位下行风险的时候，可以得到的超额回报率。这个比率的值越高，说明投资组合在承担相同单位的下行风险时，得到的超额回报率越高，这个比率的计算公式如式（3-6）所示.

回撤是指在给定的某一段时间内资产组合的净值由最大降到最小的幅度。我们假定 D_i 是第 i 天的组合净值， D_j 就是第 j 天的组合净值， j 在数值上大于 i，那么最大回撤（max_drawdown）的计算公式就是：

$$\max\_drawdown = \max\left(\frac{D_i - D_j}{D_i}\right) \tag{3-8}$$

简要来说，最大回撤代表的时计量单位在短时间内，从高到低下降的最大值。这个指标可以表明投资者持续获利能力的大小，并且当存在特殊事件时，也可以展现出投资者可以承担的极限亏损值，可以衡量投资产品的抗风险能力。在我们对量化策略进行回测时，如果其最大回车率很大，就说明这个策略的盈利能力不太稳定，要对其优化。

信息比率的缩写为IR，用来反映承担超额风险所带来的超额收益的高低。而这里提到的“超额收益”，表示收益相较于特定参照指标的差额。其计算公式为：

$$IR = \frac{\alpha}{\theta} \tag{3-9}$$

式（3-9）中α代表着投资组合的超额收益，也就是真实预期收益率与定价模型计算得到的收益率（或者大盘报酬率）之间的插值；θ是主动风险,可以通过超额收益的标准差计算得到。

从公式我们可以看出，这个指标不是以对应的无风险利率当作参照，反而我们可以自主选择对应的参照标准，并且也可以得到调整以后的收益。如果一个投资组合有较高的信息比率,那么其在特定误差之下获得的超额收益也会相对比较大。因此，在实际应用过程中，当一个投资组合有着比较高的信息比率的时候，那么它的业绩也会更加理想,并且这个投资组合的实际表现会更加持续地优于参照指标。本研究将会选择沪深300指数收益率作为我们的标的资产收益。

第 4 章 配对交易策略的设计方案

本文前三章介绍了目前商品期货配对交易的研究现状，以及本文研究的基于强化学习算法的商品期货配对交易策略的理论基础与框架。本章将基于第二章所介绍的理论基础，利用第三章所提到的配对交易系统进行实证，分别构建基于协整的静态配对交易策略与基于强化学习的动态配对交易策略。

4.1 商品期货合约概况

期货种类主要分为商品期货和金融期货，而相较于金融期货，商品期货的种类更丰富，交易所也更分散，因此影响因素也更加的广泛。另外较为重要的一点是，从市场平均值来看，商品期货市场的交易活跃度基本低于金融期货市场的交易活跃度。所以，像第二章文献综述中提到的，目前我国对于期货交易的研究中，基本都是在金融期货市场中选择研究样本的范畴，而对于商品期货市场样本的研究是比较少的。因此本文想要研究的关于商品期货的配对交易，则可在一些交易量较为活跃的商品期货品种上得以实现。

到目前为之，我国主要的期货交易所一共有四个，分别是中国金融期货交易所、上海期货交易所、大连商品交易所和郑州商品交易所。其中中国金融期货交易所主要从事的交易为金融期货，包括沪深 300、上证 50、中证 500 股指期货和十年国债、五年国债期货。而另外的三个交易所关于商品期货的交易则比较多。截至2021年1月，三大交易所共有商品期货55种，行业上涵盖了农产品、贵金属、基本金属和能源化工等多个方面，形成了比较全面的商品期货品种体系，而且可以服务于实体经济，方便进行风险对冲。在这几个交易所中，大连商品交易所覆盖的种类最广泛，包含的期货品种包括农产品、基本金属和能源化等工，目前已上市的共计16种；上海期货交易所中比较多的种类则是基本金属以及部分化工产品，目前已上市的共计11种期货合约；而郑州商品交易所的商品期货基本主是农产品，同时也有着少量的能源化工合约，共计23种期货合约。

因为不同的商品期货都对应着不同的交易品种，并且不同交易品种都有着不同的特点，因此交易所会根据这些不同的特点，来设置不一样的合约交易单位、报价单位和最小变动价位，这样一来，投资者之间交易与对冲会更加方便。另外，对于不同种类的商品期货而言，尽管其交易活跃性很大程度上受到其对应标的的基本面的影响，但是合约的参数设置等一些因素也会对商品期货交易的活跃度产

生一些影响。第一个方面，一份期货合约的账面价值由合约乘数所决定，也就是由每份合约的交易单位决定着。比如说豆粕合约的交易单位是10吨，但铁矿石的交易单位却是100吨，一来这跟产品的单位质量是有关系的，二来也受日常市场交易的平均合约单位所影响。第二个方面，成交价格的波动幅度会被最小变动价位所影响，最小变动价位比较小的合约，其交易可能会更加的频繁。而我国对不同品种商品期货的最小变动价位而言，之间的差距也比较明显，最大的为10元，最小的仅为0.05元。

因此，对不同期货来说，其合约参数都有比较大的差别，那么要保证交易策略研究可实现性和有效性的关键之处，就是如何选择恰当的标的品种，如何确定合适的样本数据集。

4.2 数据获取与预处理

本文将选用上海期货交易所的数据进行配对交易策略的设计。若是同一品种的期货合约，那么一般其至少会有6个同时上线的不同到期日的合约，这些不同到期日的合约的活跃程度都是不一样的，且数据长度均不同；而主力合约就是这些合约中交易量最大的，并且处于最活跃的月份的合约，也是最容易成交的合约，主连合约便是治理合约的连续，因此本文采用的期货合约均为主力连续合约。

通常来说，在新的期货品种在刚上市的时候，这些期货的定价机制不会十分的完善，因此，它们的价格波动率也会起伏比较大，进而其价格有可能很大程度上与其真实价值不符。例如2015年3月上市的锡期货，从主连合约的数据来看，在上市三个月后，跟最高价相比，下降了22900个点，近20%的幅度。在这三个月中，尽管也有合适套利机会存在，但并没有代表性，也不能用来预测未来的市场。本文选用的时间跨度为2010年1月-2020年1月，所以为了数据具备代表性以及完整性，选用2010年以前上市的期货合约，分别是沪铜、沪铝、橡胶、燃油、沪锌、沪金、螺纹钢、线材共计8种期货合约。

本文首先对市场内期货品种的流动性进行简单地分析，若流动性太差，配对交易就很容易产生单边成交从而增加一定的风险。从表4-1中可以看出，2010年以前上市的期货合约中，交易最活跃的螺纹钢，年日军交易量为3130443手，最不活跃的为线材期货，年日均成交量为192手。本文以日均成交量2万手为标准，去掉了不活跃的燃油、线材，留下了沪铜、沪铝、橡胶、沪锌、沪金、螺纹六种期货合约，有利于提高配对交易策略实际交易时的成功率，并且可以降低冲击成

本。

本文按照训练集（形成期）、测试集（交易期）的形式划分数据集，以前70%作为训练集（一共1702个数据），后30%作为测试集（一共729个数据）。

表 4-1 基础筛选期货品种样本

NO.	交易品种	标准合约上市日	年日均成交量
1	螺纹钢(Rb)	2009-03-27	3130443
2	橡胶（Ru）	2002-01-07	809770
3	沪铜（Cu）	2002-01-07	532251
4	沪锌（Zn）	2007-03-26	460899
5	沪金（Au）	2008-01-09	141718
6	沪铝（Al)	2002-01-07	135677
7	燃油（Fu）	2004-08-25	14942
8	线材（Wr）	2009-03-27	192

4.3 配对组合挑选

4.3.1 相关性分析

本文将以训练集（约为2010年至2016年）的数据来选取配对期货，把该段时间内各期货的收盘价格按照前文的式子，并利用Python软件计算得到其相关性，各个期货的相关系数如表4-2所示，为了可以更加值观地展示结果，图4-1给出了相关性热力图,图右侧的不同颜色刻度条代表着不同的颜色对应的相关系数。

表 4-2 相关系数图

相关性	Cu	Al	Ru	Zn	Au	Rb
cu	1.000000	0.940722	0.957618	0.424173	0.604870	0.944991
al	0.940722	1.000000	0.908300	0.449574	0.645579	0.945885
ru	0.957618	0.908300	1.000000	0.431841	0.674595	0.927841
zn	0.424173	0.449574	0.431841	1.000000	0.059736	0.401864
au	0.604870	0.645579	0.674595	0.059736	1.000000	0.605308
rb	0.944991	0.945885	0.927841	0.401864	0.605308	1.000000

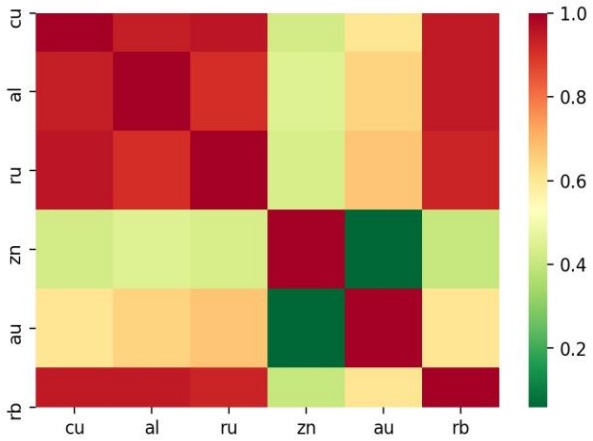

图 4-1 相关系数热力图

从表4-2和图4-1中可以看出，不同期货对的相关系数相差较大，本文将相关系数的阈值设为0.85，也就是在交易对预选择中，筛选出相关系数大于0.85的配对组合。如表4-3所示，得到6组预配对组合。

表 4-3 预配对组合

配对组合		相关系数
Al(x0)	Rb(y0)	0.945885
Cu(x1)	Ru(y1)	0.957618
Cu(x2)	Rb(y2)	0.944991
Al(x3)	Ru(y3)	0.9083
Cu(x4)	Al(y4)	0.940722
Ru(x5)	Rb(y5)	0.927841

4.3.2 协整检验

接下来再采用E-G两步法对与配对组合进行协整检验，在此，我们以Al（x0）期货与Rb（y0）期货为例，进行协整关系研究的展示，首先，做单位根检验，检验得到的结果如表4-4所示。

表 4-4 单位根检验结果

变量	临界值			T统计量	P值	平稳性
	1% 显著性水平	5% 显著性水平	10% 显著性水平			
X0	-3.221579	-2.645267	-2.347883	-1.771318	0.3211	非平稳
Y0	-3.221579	-2.645267	-2.347883	-1.362723	0.5268	非平稳
△x0	-3.221707	-2.645279	-2.347889	-24.24385	0.0000	平稳
△y0	-3.221707	-2.645279	-2.347889	-22.68279	0.0000	平稳

根据以上检验结果显示，Al(x0)期货和Rb(y0)期货的收盘价时间序列在经过一阶差分之后，都在显著性水平1%之下通过了平稳性检验，所以我们可以认为这两个期货之间存在着长期的协整关系。

然后，本文将采用E-G两步法进行协整检验，先通过最小二乘估计对两支期货的时间序列进行一元回归分析，得到的回归方程为：

$$y0_t=\beta_0+\beta_1 x0_t \tag{4-1}$$

分析结果如表4-5（1）、表4-5（2）所示

表 4-5 回归分析结果（1）

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.150384	0.056837	-16.29438	0.0000
X	1.486173	0.013203	79.86097	0.0000

表 4-5 回归分析结果（2）

R-squared	0.946282	Mean.dependent var	4.683561
Adjusted R-squared	0.946036	S.D. dependent var	0.235266

根据表4-5可知，两支期货的协整关系为：

$$y0_t=-1.150384+1.486173x0_t \tag{4-2}$$

然后对其残差序列

$$\mu_t = y0_t + 1.150384 - 1.486173x0_t \quad (4-3)$$

进行平稳性检验，检验结果如表4-6所示。

表 4-6 平稳性检验结果

		t-统计量	P值
ADF统计量		-5.762540	0.0000
临界值	1%临界值	-3.657245	
	5%临界值	-2.874566	
	10%临界值	-2.341235	

综上所述，残差序列为在1%的置信水平下通过检验。这说明了A1期货与Rb期货之间的长期协整关系存在。

另外对于本文中另外五对配对期货同样使用上述方法，检验结果均证明五对期货对通过了协整检验，最终我们得到如表4-3所示的6对配对组合。

4.3.3 价差序列可视化

由上文得到最终的6对期货配对组合后，我们将计算得出其价差序列。假设两只期货A和B的收盘价为x、y，满足协整检验的条件。将x、y作为两个变量进行一元线性回归，得到一元回归方程 $y^* = \alpha + \beta x + \varepsilon_i$：由此得到残差序列：$Spread = y^* - x$。将残差序列标准化，$Z\_Spread = \left(Spread - \mu_{spread}\right) / \sigma_{spread}$ 得到的即为价差时间序列。图4-2至图4-13为分别为表4-3所示的6对配对组合分别在训练集、测试集的价差时间序列。

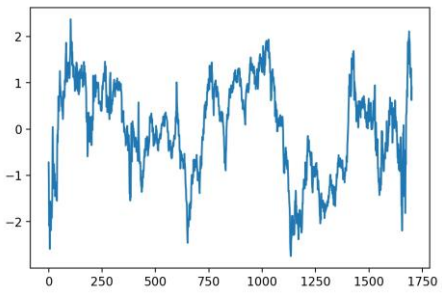

图 4-2 x0-y0 价差序列（形成期）

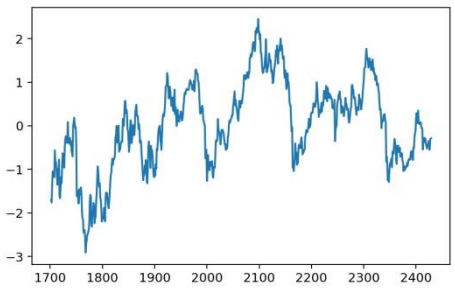

图 4-3 x0-y0 价差序列（交易期）

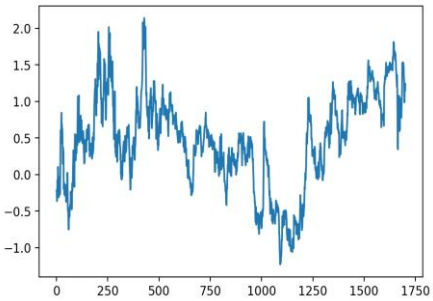

图 4-4 x1-y1 价差序列（形成期）

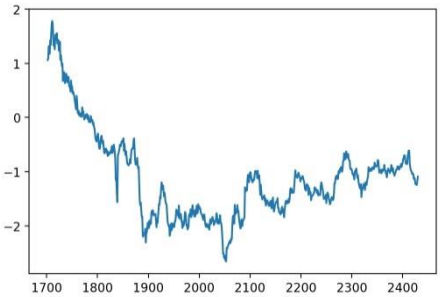

图 4-5 x1-y1 价差序列（交易期）

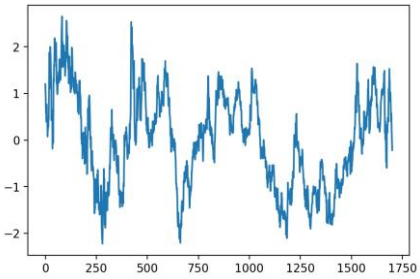

图 4-6 x2-y2 价差序列（形成期）

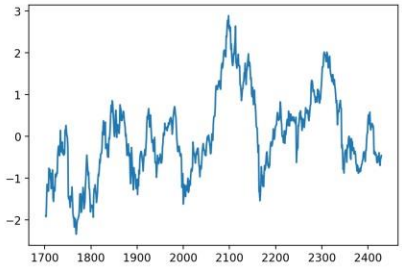

图 4-7 x2-y2 价差序列（交易期）

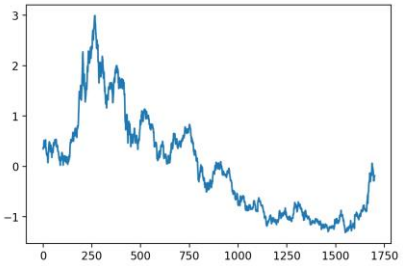

图 4-8 x3-y3 价差序列（形成期）

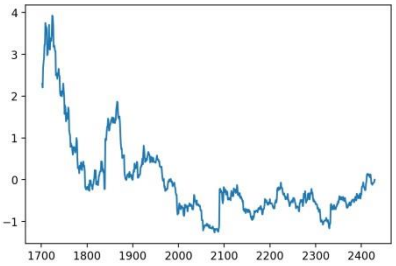

图 4-9 x3-y3 价差序列（交易期）

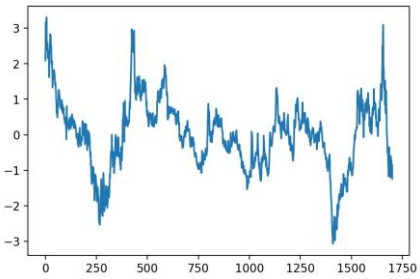

图 4-10 x4-y4 价差序列（形成期）

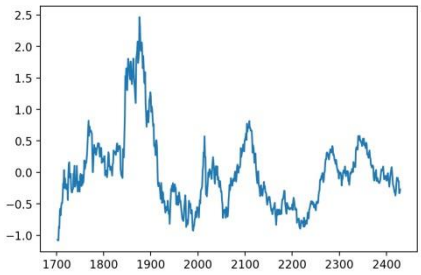

图 4-11 x4-y4 价差序列（交易期

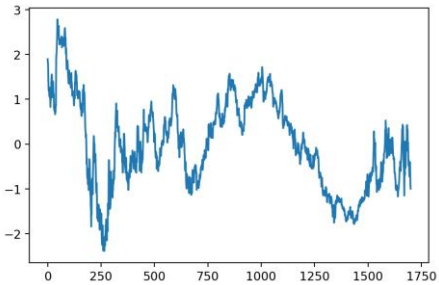

图 4-12 x5-y5 价差序列（形成期）

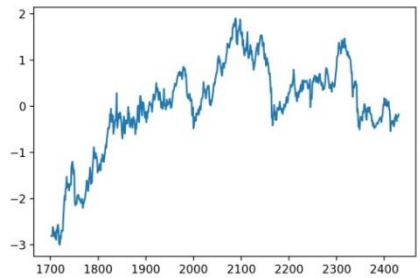

图 4-13 x5-y5 价差序列（交易期）

4.4 配对组合配比

配对交易实际上是风险中性策略，其本质就是利用配对组商品期货之间不合理的价差，进行买入或卖出两种商品，并且在价差恢复合理时或者反向不合理时进行平仓操作或者反向开仓操作，从而达到套利的目的。但是在实际的配对交易策略操作过程中，为了满足风险中性，我们应该以何种比率对配对组的合约进行建仓呢？由上文的分析得到，对于协整性比较好的配对期货组合，这两种商品的对数价格之间的关系可以用线性方程：$\ln(Y_t)=\alpha+\beta\ln(X_t)+\varepsilon_t$ 来表示。若不考虑残差项的话，可以对方程进行一些变换，可以得到：

$$\frac{Y_t}{X_t}=e^{\alpha}*X_t^{\beta-1} \tag{4-4}$$

通过计算出各期货合约对的拟合结果，从而由式（4-4）计算得出各期货合约对的合约配比，结果如表4-7所示，

表 4-7 合约配比结果

配对组合		β 值	配比（约）
Al(x0)	Rb(y0)	0.4895316	2：1
Cu(x1)	Ru(y1)	0.73774428	5：4
Cu(x2)	Rb(y2)	0.08976982	11：1
Al(x3)	Ru(y3)	3.812258	1：4
Cu(x4)	Al(y4)	0.1726719	6：1
Ru(x5)	Rb(y5)	0.11440974	8：1

确定好最终的配对交易匹配对，并且计算好匹配对合约配比之后，下文将要以x0（Al），y0（Rb）为例来设计配对交易的主体策略，展示交易的具体过程。

4.5 基于协整的静态配对交易策略的设计方案

4.5.1 交易阈值设定

在确定好最终的配对交易匹配对之后，要做的便是进行交易策略信号的设置，这将直接决定着交易策略收益的数目。交易信号主要包含三个，买入（卖出）价差建仓信号、买入（卖出）价差平仓信号以及止损信号。其中每个信号的触发条

件都不同，比如建仓信号会因为价差序列偏离均衡值而发生，平仓信号会因为价差序列重新回到均衡值而产生，止损信号是为了合理控制风险而建立的，具体来讲就是当价差序列偏离均衡值太多时，就会产生止损信号以防损失继续扩大。

在本节配对交易策略的研究中，参考以往的文献，将采用静态阈值设定开仓点和平仓点。具体来讲，本文首先由训练集数据，首先先算出价差时间序列的标准差，这个值是一个固定常数，然后将固定的倍数的标准差作为我们建仓、平仓和止损的信号。用μ和σ分别表示训练集数据的均值和标准差，即μ=0，σ=1。本节由以往经验设定开仓触发阈值为$\mu\pm1.5\sigma$，也就是1.5，平仓阈值为$\mu\pm0.2\sigma$，也就是0.2，止损阈值为$\mu\pm2.8\sigma$，也就是2.8。

当标准残差线从下到上穿过卖空价差阈值的时候，也就是Z_Sprea>1.5时，说明资产B价格高于资产A价格，价差开仓信号就会触发，卖出资产B，买入资产A；当标准残差线从下到上穿过买回价差平仓阈值的时候，也就是Z_Sprea<0.2时，价差平仓信号就会触发，买入资产B，卖出资产A。这样就可以完整地进行一次做空套利交易的操作。同样的，当标准残差线从下到上穿过买入价差开仓阈值的时候，也就是Z_Sprea<-1.5时，说明资产A价格高于资产B价格，开仓信号就会触发，买入资产B，卖出资产A；当标准残差从下到上穿过卖出价差平仓阈值的时候，也就是Z_Sprea>-0.2时，价差平仓信号就会触发，到此也完成了一次做多套利交易的操作。

4.5.2 交易信号生成

上文中提到的标准残差在0值附近来回震荡，当每次穿过阈值线时，就会触发对应交易，如图4-14为训练集（形成期）的数据形成的交易信号，图4-15为测试集（交易期）的数据形成的交易信号，中间的黑色虚线为平仓信号，最上面和最下面的红色实线为止损阈值，剩下的两条绿色实线为建仓阈值。

也就是当标准残差上穿绿色实线μ+1.5σ时，做空配对期货，反向建仓（即买入x0-AL期货。卖出y0-Rb期货，二者资金配比如表X）；标准残差下穿黑色虚线μ+0.2σ时，反向平仓止盈；标准残差下穿绿色实现μ-1.5σ时，做多配对期货，正向建仓（即买入y0-Rb期货，卖出x0-Al期货，二者资金配比如表X）；标准残差上穿黑色虚线μ-0.2σ时，正向平仓止盈；标准残差突破红色实线$\mu\pm2.8\sigma$时，及时止损，此时不开仓。

图 4-14 价差序列交易信号（协整配对）（形成期）

图 4-15 价差序列交易信号（协整配对）（交易期）

4.6 基于强化学习的动态配对交易策略的设计方案

4.6.1 交易流程

本文选取了交易对的收盘价作为特征，不同品种期货的价差序列有时是正数，有时是负数，并且数值大小也不一样，为了模型可以更好地训练，对特征指标进行归一化处理。

将我们选取的历史数据处理完以后，首先需要设置的包括参数初始值、迭代次数 M 、精度和记忆库的大小 R_N 。首先对历史数据中的训练集（形成期）进行协整检验；如果这一部分数据通过了协整检验，那我们就会选择这部分训练集数据，对比到强化学习模型，就是模型中的状态 S_{t-n} 到 S_t ，同时把它们当作是强化

学习模型的输入值，之后模型会产生一系列相应的动作 $a_t \in \{buy, sell, hold, stop\}$。

当动作 a_t 是 *buy* 的时候，系统将会买入资产A，同时卖出资产B，但若想要卖出资产B，需要先检查一下现在的仓位中是不是还存在着资产B，因此交易系统会先检验资产B的仓位状态，若资产B的仓位不为0，就代表着投资者拥有B资产，此时就会买入资产A卖出资产B，若资产B的仓位为0，就代表着投资者此时没有资产B，因此不能卖出资产B，那么就会去买入资产A。同样，当动作 a_t 是 *sell* 的时候，交易系统会先检验资产A的仓位状态，若资产A的仓位不为0，此时就会卖出资产A买入资产B，否则的话就不存在资产A，仅仅买入资产B。当动作 a_t 是 *hold* 的时候，系统会使资产A和B的仓位保持不变。当动作 a_t 是 *stop* 的时候，系统会将资产A、B全部卖空，它们的仓位就会都变成0，在整个过程中，不同动作的不同返回会使得环境产生不同的回报也就是奖励 r_t。

当环境返回了奖励 r_t 之后，就会把状态 s_t、s_{t+1}、动作 a_t 和奖励 r_t 都存入到忆库 R 中，此时会先判断代码的执行次数 n，加入这个数值 n 大于记忆库大小 R_N 时，系统就会将记忆库更新，具体来讲就是删掉最初存入的状态、奖励和回报值，取而代之的是会把最新的状态放进 R 里，如此就可以根据最新时间的信息来取样更新模型参数。最后判断 n 与训练集数据长度的大小，若 n= 训练集数据长度，则表明此时已经对所有训练集数据进行了训练。

在测试阶段，由当前状态 S_t，调用训练集中得到的最优参数集合，在测试周期的每一天，当前状态都依照greedy策略通过最优参数集合来选取动作 a_t，得到当即奖励 r_t，以及下一时刻的状态 S_{t+1}，此时更新记忆库 R，直至 S_{t+1} 达到测试期的最后一个交易日。

4.6.2 参数设置

在智能体的动作空间中，首先需要对4个参数的初始值进行预设，具体来讲就是评估时间窗口、交易时间窗口、开仓阈值、平仓阈值这几个参数。评估时间窗口初始值设为从6天到60天，每6天为一步；交易时间窗口初始值设为从12天到120天，每6天为一步；开仓阈值为增加或减少的访问是1到5，每步是1；平仓阈值为开仓阈值的基础上加减1-2，每步是0.5。而开仓阀值和平仓阀值两者均为连续型参数，因此在实际研究中要对其先进行离散化处理，在后文中将通过0.1单位来抽取数值的均分方式，使其满足离散化处理的要求。

本文中将使用的 $\varepsilon - greedy$ 探索策略，其特点是，系统在选择动作的时候会

加入一些的随机变化，以便更好地处理开发与利用之间的平衡问题。在训练阶段，我们将ε初始值设为1，随着迭代次数的增加，智能体的学习程度会不断加深，使得随机性会逐渐开始降低，从而ε值也会逐渐减小。并且将参数的精度α初始值设为1。

我们设置整个训练过程迭代1000次，并且会通过$\varepsilon-greedy$策略进行选择不同的动作集合，所谓的动作集合也就是上文提到的4个参数的集合，选择不同的动作之后会产生不同的状态，同时需要根据这些状态来更新对应的值函数。之后智能体将会选择最优参数进行回测交易。

另外，保证金比例为合约价值的10%，不考虑交易手续费及滑点.

4.6.3 交易环境

表4-8给出对的是本研究进行模拟交易的实验环境，主要包括电脑配置、编程语言和所用机器学习框架三部分。

表 4-8 交易环境

操作系统（OS）	Windows 64位
处理器（CPU）	Inter(R)Core(TM)i7-4720HQ @ 2.60GHz
主存（RAM）	16GB
编程语言	Python 3.7.1
机器学习框架	Tensorflow

第 5 章 配对交易策略的有效性评价

本章将对前文分别构建的基于传统协整模型和强化学习模型的配对交易策略的回测结果进行有效性评价。首先分别对两个模型在训练集、测试集的回测结果进行展示并简要分析，然后将根据两种策略回测结果计算出不同评价指标，来对两种模型进行对比分析。

5.1 实证结果分析

5.1.1 传统协整模型

根据前文的静态交易策略，运用Python软件编写程序，分别对如表X所示的交易对测试集的数据，即2017年1月至2020年1月的数据，进行回测检验，获得实证结果。以x0（Al）-y0（Rb）为例，图5-1表示日数据下，x0（Al）期货与y0（Rb）期货对交易获得的累计总利润。图5-2表示日数据下，x0（Al）期货与y0（Rb）期货对交易的日收益率。

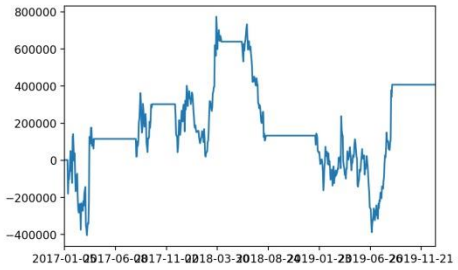

图 5-1 X0-y0 累计总利润图

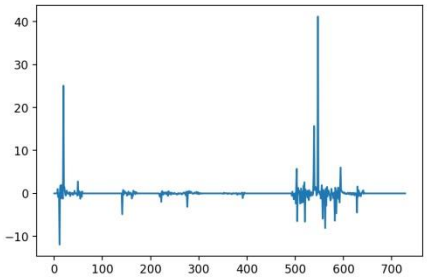

图 5-2 x0-y-日收益率图

5.1.2 强化学习模型

根据前文的强化学习模型交易策略，运用Python软件编写程序，分别对如表4-3所示的六组期货对进行仿真交易，测试在训练集和测试集中的效果。图5-3

至图5-8展示出了新模型在训练集（形成期）中的交易信号，图中最中间的黑线条为平仓信号线，最上面和最下面的红线条为止损信号线，其余两条绿线条为建仓信号线。

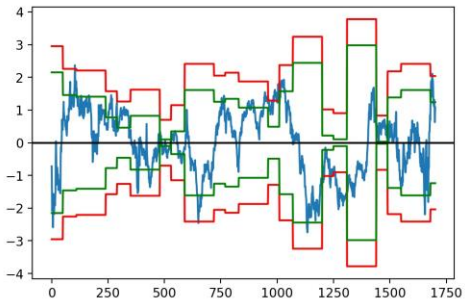

图 5-3 x0-y0 交易信号(强化学习-形成期)

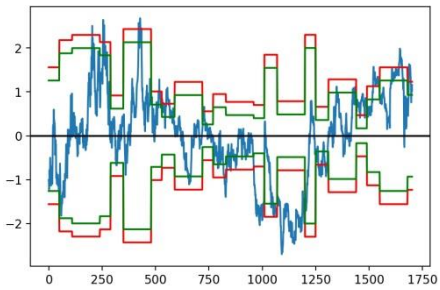

图 5-4 x1-y1 交易信号(强化学习-形成期)

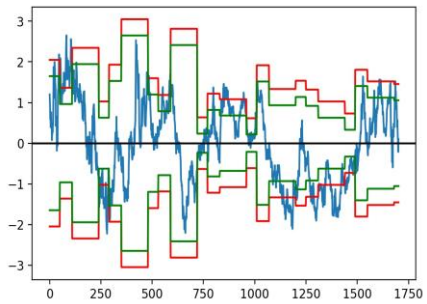

图 5-5 x2-y2 交易信号(强化学习-形成期)

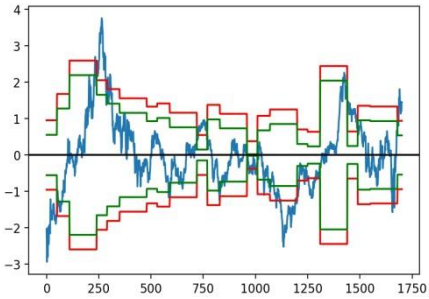

图 5-6 x3-y3 交易信号(强化学习-形成期)

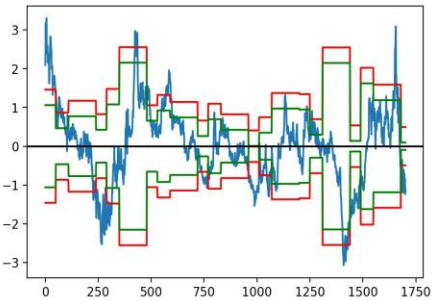

图 5-7 x4-y4 交易信号（强化学习-形成期）

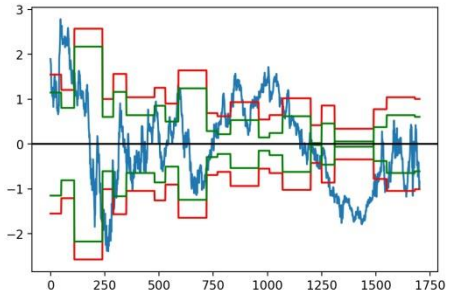

图 5-8 x5-y5 交易信号（强化学习-形成期）

接下来对测试集（交易期）的数据，即2017年1月至2020年1月的数据，进行回测检验，获得实证结果。如图5-9至图5-14所示，为新模型在测试集（交易期）中的交易信号。

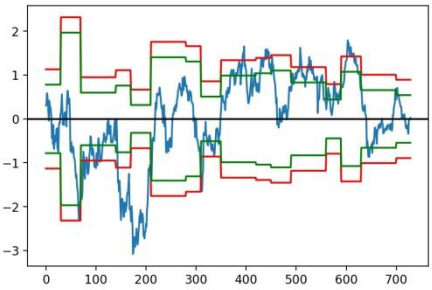

图 5-9 x0-y0 交易信号（强化学习-交易期）

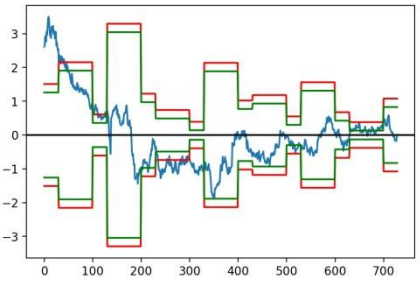

图 5-10 x1-y1 交易信号（强化学习-交易期）

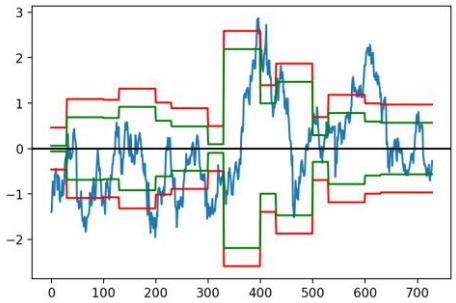

图 5-11 x2-y2 交易信号(强化学习-交易期)

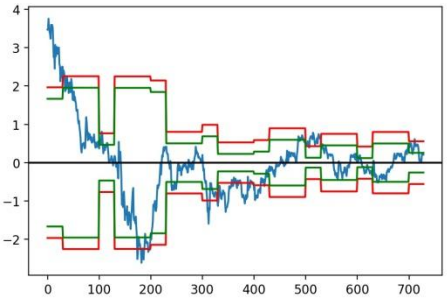

图 5-12 x3-y3 交易信号(强化学习-交易期)

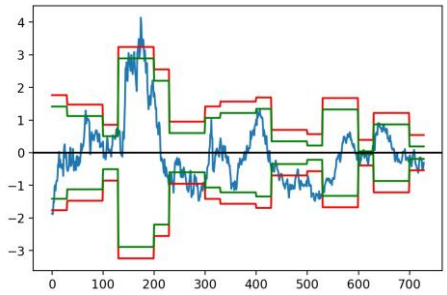

图 5-13 x4-y4 交易信号(强化学习-交易期)

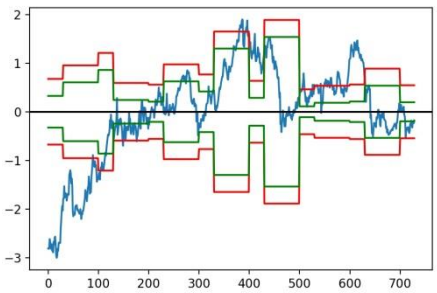

图 5-14 x5-y5 交易信号(强化学习-交易期)

图5-15至图5-20表示日数据下，各期货对交易获得的累计总利润。图5-21至图5-26表示日数据下，各期货对交易的日收益率。从图中可以看出，基于强化学习模型的交易策略，在各期货对上都取得了不错的收益，如期货对x0（Al）-y0（Rb），在回测期间的累计总利润达到约25000元的收益，期货对x1（Cu）-y1（Ru），在回测期间的累计总利润达到40000元的收益。但是由于每种期货对的合约价值不同，仅从总利润与日收益率来看，不足以比较其表现差异。因此本文在5.2中展示了不同交易策略下，各期货对的累计收益率、年化收益率等指标。

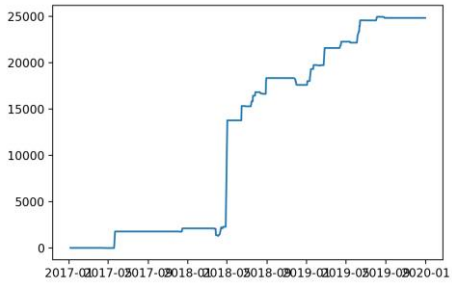

图 5-15 x0-y0 累计总利润图

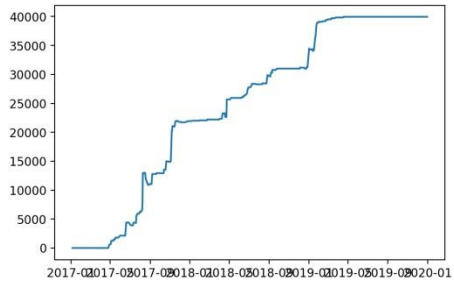

图 5-16 x1-y1 累计总利润图

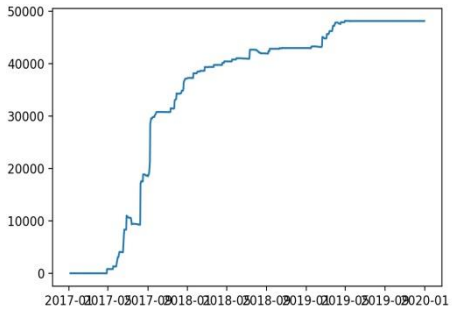

图 5-17 X2-y2 累计总利润图

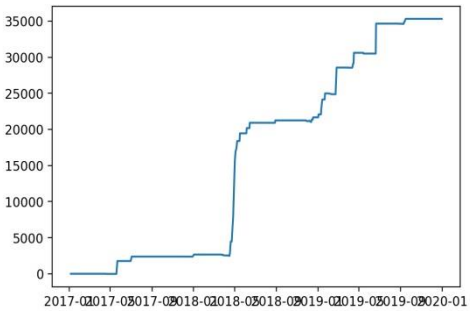

图 5-18 X3-y3 累计总利润图

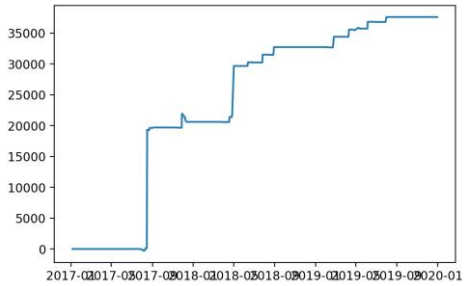

图 5-19 X4-y4 累计总利润图

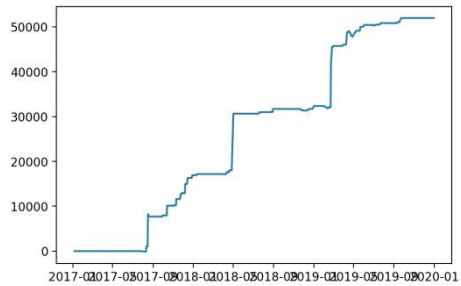

图 5-20 X5-y5 累计总利润图

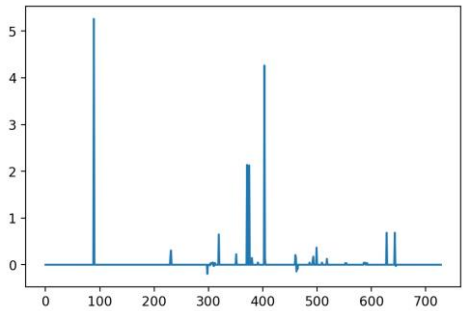

图 5-21 X0-y0 日收益率图

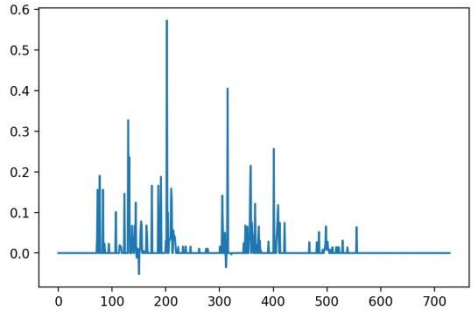

图 5-22 X1-y1 日收益率图

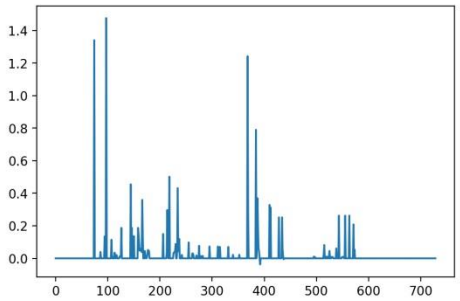

图 5-23 X2-y2 日收益率图

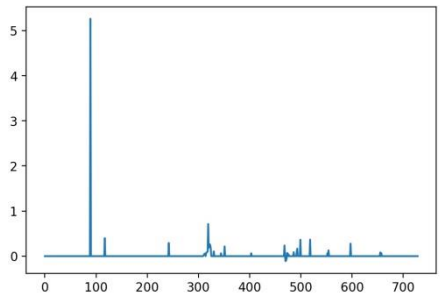

图 5-24 X3-y3 日收益率图

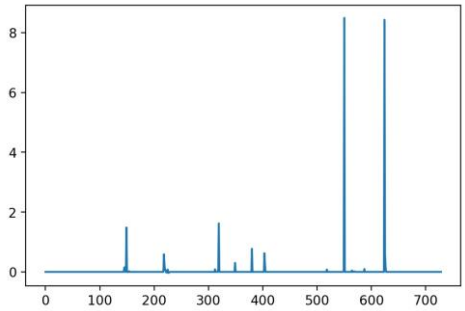

图 5-25 X4-y4 日收益率图

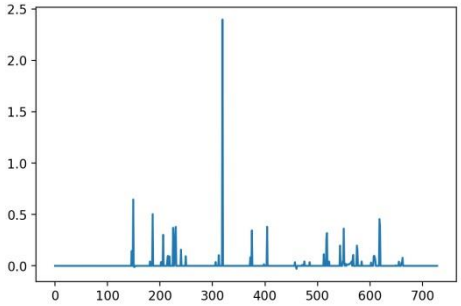

图 5-26 X5-y5 日收益率图

5.2 模型之间的比较分析

为了更好地反映策略地表现，表5-1至表5-6汇总了6对期货对的训练集与测试集，分别在传统模型与新模型中的各项性能表现。

（一） 从图表中可以看出，本文提出的基于强化学习的配对交易都取得了比传统静态配对交易更高的收益。

以x0（A1）-y0（Rb）为例，在训练集上，新模型在累计收益率、年化复合收益率上的表现都已经全面超越传统模型，分别从47.10%增加到132.13%、1.20%增加到2.08%，平均每笔回报也从0.84%增加到了2.03%；反映市场风险的最大回撤从-1.43%下降至-1.57%，有小幅度下降；索提诺比率从0.82%增加到2.88%，表明交易对在承担相同单位下行风险能获得更高的超额回报率；信息比率从0.47%增加至0.86%，由于本文以沪深300指数为标的资产，因此其相对沪深300指数的超额收益风险比增加。在测试集上，累计收益率从7.30%增加到18.18%，年化复合收益率从0.18%增加到0.46%，平均每笔回报从0.37%增加到0.76%，最大回撤从-1.98%下降到-2.24%，交易次数从20次增加到了24次，索提诺比率从0.60增加至2.38，信息比率由0.24提升至0.49。相对训练集来说，造成新模型的性能相对于训练集来说提高程度则较小一些的原因可能是由于测试集相比于训练集时期较短。

（二）从各模型的回测结果来看，新模型对这六组交易策略的性能都有着不同程度的改善。

就累计收益率而言，新模型使得x2（Cu）-y2（Rb）期货交易对的累计收益率从1.49%提高至13.88%，使得x4（Cu）-y4(Al)期货交易对的累计收益率从7.84%增加至24.47%，使得x5（Ru）-y5（Rb）期货交易对的累计收益率从-4.78%增加至10.41%，提高程度十分明显；而对x1（Cu）-y1（Ru）、x3（Al）-y3（Ru）期货交易对的影响程度较小；并且新模型在x3（Al）-y3（Ru）交易对的累计收益为10.56%，在x5（Ru）-y5（Rb）期货对上的累计收益为10.41%，其收益相对较低，从图4-9和图4-13可以看出，x3（Al）-y3（Ru）在交易期期间价差大幅度上升，x5（Ru）-y5（Rb）在交易期期间价差大幅度降低，即使之后回复到了均值水平，但也导致在该期货对上的收益相对较低。

就最大回撤而言，新模型对x2-y2期货交易对的影响较为显著，从-1.56%降低至-8.88%，使得交易策略的风险大幅度降低；但是对x3-y3期货交易对而言，新模型却使得其最大回撤从-0.26增加到了-0.18。从索提诺比率的值来分析，新模型对x1（Cu）-y1（Ru）、x2（Cu）-y2（Rb）期货交易对的影响十分显著，分别使索提诺比率从0.39增加至6.92、从0.48增加至10.00，说明新模型使得这两组交易策略在承担相同下行风险的情况下，能获得更高的超额回报率。就信息比率而言，新模型对x2（Cu）-y2（Rb）、x4（Cu）-y4(Al)、x5（Ru）-y5（Rb）期货交易对的提升较为明显，分别从0.03提升至0.36、从0.04提升至1.21、从-0.08提升至0.25。

从表5-1至表5-6可以看出，综合而言，基于强化学习的配对交易策略在x2（Cu）-y2（Rb）期货交易对上的表现最优，而在x3（Al）-y3（Ru）期货交易对上的表现较差，对评价指标的提升较弱。此外，新模型在其余的x1（Cu）-y1（Ru）、x4（Cu）-y4(Al)、x5（Ru）-y5（Rb）期货交易对上的表现也较为优越，这些交易获得的累计收益率较高，且索提诺比率、信息比率均有所增加。

表 5-1 x0-y0 模型结果对比

	传统静态		强化学习	
	训练集	测试集	训练集	测试集
累计收益率/%	47.10	7.30	132.13	18.18
年化复合收益率/%	1.20	0.18	2.08	0.46
交易次数	56	20	65	24
平均每笔回报/%	0.84	0.37	2.03	0.76
索提诺比率	0.82	0.60	2.95	2.38
最大回撤/%	-1.43	-1.98	-1.57	-2.24
信息比率	0.47	0.38	0.86	0.49

表 5-2 x1-y1 模型结果对比

	传统静态		强化学习	
	训练集	测试集	训练集	测试集
累计收益率/%	15.62	9.66	24.83	17.21
年化复合收益率/%	0.17	0.11	0.27	0.19
交易次数	63	49	77	58
平均每笔回报/%	0.25	0.20	0.32	0.30
索提诺比率	0.57	0.39	7.17	6.92
最大回撤/%	-1.04	-246.59	-4.70	-266.45
信息比率	0.32	0.18	0.49	0.21

表 5-3 x2-y2 模型结果对比

	传统静态		强化学习	
	训练集	测试集	训练集	测试集
累计收益率/%	24.80	1.49	28.50	13.88
年化复合收益率/%	2.88	0.17	2.14	1.61
交易次数	73	23	67	34
平均每笔回报/%	0.34	0.06	0.43	0.41
索提诺比率	1.57	0.48	13.42	10.00
最大回撤/%	-4.86	-1.56	-3.33	-8.88
信息比率	0.15	0.03	0.78	0.36

表 5-4 x3-y3 模型结果对比

	传统静态		强化学习	
	训练集	测试集	训练集	测试集
累计收益率/%	19.42	8.61	24.53	10.56
年化复合收益率/%	1.89	1.04	2.30	0.98
交易次数	29	22	27	18
平均每笔回报/%	0.67	0.39	0.91	0.59
索提诺比率	1.56	0.95	2.01	1.34
最大回撤/%	-6.51	-0.26	-8.52	-0.18
信息比率	0.59	0.29	0.63	0.27

表 5-5 x4-y4 模型结果对比

	传统静态		强化学习	
	训练集	测试集	训练集	测试集
累计收益率/%	18.86	7.84	33.19	24.47
年化复合收益率/%	2.73	1.62	7.12	5.25
交易次数	31	17	42	35
平均每笔回报/%	0.61	0.46	0.79	0.70
索提诺比率	0.49	0.24	0.73	0.44
最大回撤/%	-82.84	-2.41	-3.79	-3.40
信息比率	0.85	0.04	1.79	1.21

表 5-6 x5-y5 模型结果对比

	传统静态		强化学习	
	训练集	测试集	训练集	测试集
累计收益率/%	-10.28	-4.78	16.40	10.41
年化复合收益率/%	-1.77	-0.82	1.16	1.79
交易次数	48	35	78	46
平均每笔回报/%	-0.21	-0.14	0.21	0.23
索提诺比率	-0.24	-0.46	1.94	1.02
最大回撤/%	-4.03	-1.55	-5.06	-3.40
信息比率	-0.02	-0.08	0.57	0.25

第 6 章 结论

6.1 总结

配对交易作为统计套利策略中一种，具有市场中性的特点，其核心问题主要有两点，首先是形成期关于配对期货组合的选择问题，其次是交易期间关于交易流程的控制问题。基于这两个核心问题，配对交易策略才可以在情况复杂的经济金融市场得到成功地运用，同时也给投资者提供一种较可能获利的投资策略。

本篇文章依据协整理论、并且将强化学习算法与配对交易策略的思想进行结合，选定在2010年1月至2020年1月的上海期货交易所商品期货的日收盘价进行活跃度检验，将不活跃的期货品种剔除，接着进行相关性检验，挑选相关性大的期货对，我们选取六个期货对作为研究对象；在确定配对组合以后，将进行单位根检验、协整检验等过程来验证配对期货间是否具有长期的协整关系，最终确定满足条件的期货对，采用传统静态模型与强化学习模型对最终期货对进行仿真交易。

通过对两个模型分别进行仿真交易，从测试的结果我们可以得出，从收益和风险的角度来说，构建的改进模型明显优于传统模型，可以显著地提升交易策略的获利能力，具体来讲，绝大部分期货对的交易次数都有所增加，最大回撤有所减少，并且索提诺比率也基本都增加，信息比率基本都增加，累计收益率由也有所增长。回测的结果还表现出，构建的基于强化学习的新型配对交易策略在我国商品期货市场同样具备着有效性，并且可以取得显著的正收益。

目前许多机构投资者在实际进行量化交易时，都采用配对交易策略，但是从他们得到的盈利结果来看，通过此策略得到的利润逐渐开始减少。为了应对这种情况，像本文中的把强化学习的思想和配对交易策略结合起来，构建出可以实现参数的自适应动态调整模型，在我国金融市场中使用这种模型，其获利能力会略高于传统的配对交易策略，并且给市场中的投资者提供一个新型套利的工具。但是此模型在实际应用的时候也有一定的缺点，例如强化学习算法有可能会过拟合、并且系统中的操作过程略微复杂，容易存在黑箱问题等，在未来计算技术的发展进步之后，这些问题或许都可以解决。

6.2 交易策略的优缺点分析

本文的技术贡献主要有：（1）本文构建的新模型给金融市场量化交易领域

添加了一个新的交易策略，可以改善传统的配对交易策略获利能力下降的问题，并且使得进行配对交易的过程中，获利机会增多，有着重要的应用价值；（2）本文对传统的配对交易模型改进的部分比较多，通过把强化学习思想和传统的配对交易结合起来，进而构建出的新型配对交易模型，可以使得模型参数进行自适应动态优化；（3）在我国融资融券和股指期货等做空机制不断发展的背景下，随着期货产品的逐渐丰富，本文改进的新型配对交易模型可作为一种有效的套利手段和风控工具来供投资者使用；（4）本文以我国期货市场中交易量较大的几种期货为研究对象，验证了基于强化学习算法的配对交易在期货市场的可行性、可获利性。

本文的局限性：（1）首先，在匹配对的选择方面，高协整性就代表着变量之间长期走势的一致，但是在短期内也有概率发生偏离，这会使得其相关性降低。因此高相关性并不代表着高协整性，反之也是这样。而本研究在选择匹配对时删掉了对相关系数在0.85以下的期货对，而这些被删掉的期货对中可能也会有一些的协整性会比较高。所以鉴于本文并没有对所有期货对的协整性进行全面检验，在后续的研究中可以适当地降低相关系数，进而全面考虑国内流动性较好的期货，并对其协整关系做更加全面的分析，同时也仍要避免出现的配对组合价格走势不一致；（2）本篇文章是主观设定了交易模型的一些初始参数的，尽管自适应模型的特性在于可以对参数进行自动优化，但如果未来的研究可以更高效地选择初始参数，模型的收敛速度将会大幅增高；（3）最后，本文的实证研究本质上只是基于编程的一次仿真实验，因此对收益率等指标的计算会存在一定的误差，也有可能存在不能按照设定的期货价格来成交等情况。基于复杂的实盘交易环境，还需要考虑更多情况，比如相关交易费用、滑点的设置等因素。在后续的研究中若考虑到交易成本等其他费用，可以更加准确地计算收益率。

6.3 相关展望

配对交易策略属于比较经典的统计套利策略里的其中一种，国内的许多投资者在进行量化交易的时候，都会采用配对交易的策略。然后目前我国的证券市场的趋于饱和，投资者们采用配对交易策略实际所产生的利润开始逐渐下降。为了应对这种情况，减轻传统配对交易策略的负面影响，一个方法从参数优化的模型角度来考虑，本文用到的强化学习算法就是一个很好的例子，另一种方法就是从

配对的对象角度来考虑，比如说在以前的研究中，很多学者都是以股价（或期货价格）价差的均值回复作为配对交易的对象，其实也可以考虑股价（期货价格）波动率和收益率，将其作为配对交易的对象也可以对传统交易策略进行改进。

从配对交易的实证研究方面来说，本文的策略主要评价指标位索提诺比率，也就是在强化学习算法中的以索提诺比率作为奖励函数，在后续的研究中，建议研究者在评价策略的优劣的过程中，也可以采用更加全面、准确的评价指标。另外强化学习算法被用来在金融市场的投资领域进行量化交易的时候，也有一些不足之处，随着计算机技术的快速发展希望得到改善。这些不足之处首先是，我们采用主观设定的方法对参数进行初始化，之后再通过算法对参数优化，因此模型参数的初始值有很大的主观性，其次就是我们采用的算法在参数优化方面的能力，还有很大的改善空间。随着未来的量化研究更加全面、采用的技术不断进步，这两者都有望得到进一步改善。并且之后的研究可以将将数学与统计领域的与时俱进的研究成果引入到模型中来，从而建立更完善的动态参数优化体系。

从国内的金融市场体系来说，尽管目前我国金融市场逐渐饱和，并且对于融资融券等做空机制并没有完全开放，以统计套利作为基础的配对交易的量化投资策略在这种背景下并不十分有利。但是考虑到目前的金融市场发展趋势，鉴于市场上法制监管的情况，金融市场上的很多资金还是会向权益性投资涌入，因此机构投资者将会再一次青睐基于统计套利的配对交易策略。

随着我国期货市场卖空机制的施行以及商品期货市场的发展成熟，配对交易策略将会有更加广阔的应用空间。相信在未来配对交易在我国商品期货市场的交易机会会逐渐增多，配对交易策略也定会在投资领域大放光彩，有着更大的用武之地。反过来讲，套利交易的增多也会影响我国的金融市场，会促进我国商品期货市场的效率提升，从而更好地服务于实体经济。并且随着计算机的相关技术快速发展，量化投资在我国的金融市场上会被更加广泛地应用，我国的金融市场也会慢慢发展成一个更成熟并且理性的国际性金融市场。

参考文献

[1] 丁秀玲,华仁海.大连商品交易所大豆与豆粕期货价格之间的套利研究[J].统计研究,2007(02):55-59.

[2] 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,2008,26(12):26-29.

[3] 何树红,张月秋,张文.基于 GARCH 模型的股指期货协整跨期套利实证研究[J].数学的实践与认识,2013,43(20):274-279.

[4] 扈文秀,牛静,李芳,牛洁.基于统计套利模型的商品指数期货双跨套利方案研究[J].管理评论,2013,25(09):100-107.

[5] 覃良文,唐国强,林静.基于 HP 滤波和协整理论的期货套利研究[J].湖北大学学报(自然科学版),2015,37(06):570-576.

[6] 于孝建,邹倩倩.基于 OU 过程的商品期货市场配对交易策略[J].南方金融,2018(03):52-60.

[7] 丁纯.沪深 300ETF 期权合成标的资产与股指期货套利的研究[J].商讯,2020(34):91-92.

[8] 路旭洋.基于和声搜索算法的商品期货套利策略优化[J].中国集体经济,2020(35):75-77.

[9] Granger C.W.J.. Some properties of time series data and their use in econometric model specification[J]. North-Holland,1981,16(1).

[10] John Board,Charles Sutcliffe. The dual listing of stock index futures: Arbitrage, spread arbitrage, and currency risk[J]. John Wiley & Sons, Ltd,1996,16(1).

[11] Evan Gatev,William N. Goetzmann,K. Geert Rouwenhorst. Pairs Trading: Performance of a Relative-Value Arbitrage Rule[J]. The Review of Financial Studies,2006,19(3).

[12] 王伟峰,刘阳.股指期货的跨期套利研究——模拟股指市场实证[J].金融研究,2007(12):236-241.

[13] 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,2008,26(12):26-29.

[14] Nicolas Huck. Pairs trading and outranking: The multi-step-ahead forecasting case[J]. European Journal of Operational Research,2010,207(3).

[15] 李世伟.基于协整理论的沪深 300 股指期货跨期套利研究[J].中国计量学院学报,2011,22(02):198-202.

[16] Qingshuo Song,Qing Zhang. An optimal pairs-trading rule[J]. Automatica,2013,49(10).

[17] Zhengqin Zeng,Chi-Guhn Lee. Pairs trading: optimal thresholds and profitability[J]. Quantitative Finance,2014,14(11).

[18] 王利斌. 基于变结构协整的股指期货跨期套利研究[D].中国科学技术大学,2014.

[19] 刘永波.投资组合优化的可行性规则人工蜂群算法[J].智能系统学报,2014,9(04):491-498.

[20] 李栋,张文宇.基于 FAM-ELM 股票价格预测研究[J].计算机仿真,2014,31(08):209-212+316.

[21] 赵胜民,闫红蕾.A 股市场统计套利风险实证分析[J].管理科学,2015,28(05):93-105.

[22] K. Charalambous,C. Sophocleous,J. G. O'Hara,P. G. L. Leach. A deductive approach to the solution of the problem of optimal pairs trading from the viewpoint of stochastic control with time‐dependent parameters[J]. Mathematical Methods in the Applied Sciences,2015,38(17).

[23] 陈艳,王宣承.基于变量选择和遗传网络规划的期货高频交易策略研究[J].中国管理科学,2015,23(10):47-56.

[24] 朱丽蓉,苏辛,周勇.基于我国期货市场的统计套利研究[J].数理统计与管理,2015,34(04):730-740.

[25] 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(01):65-69.

[26] 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(04):1-9.

[27] Hossein Rad,Rand Kwong Yew Low,Robert Faff. The profitability of pairs trading strategies: distance, cointegration and copula methods[J]. Quantitative Finance,2016,16(10).

[28] Minh-Man Ngo,Huyên Pham. Optimal switching for the pairs trading rule: A viscosity solutions approach[J]. Journal of Mathematical Analysis and Applications,2016,441(1).

[29] 刘阳,李艳丽,陆贵斌.基于信息更新 NN-GARCH 模型的统计套利研究[J].统计与决策,2016(02):169-171.

[30] 张波,刘晓倩.基于 EGARCH-M 模型的沪深 300 股指期货跨期套利研究——一种修正的协整关系[J].统计与信息论坛,2017,32(04):34-40.

[31] 毕秀春,于晓雨,张曙光.基于遗传算法—部分协整理论的配对交易方法及应用[J].统计研究,2020,37(09):82-94.

[32] Andrew W. Moore,Christopher G. Atkeson. Prioritized sweeping: Reinforcement learning with less data and less time[J]. Machine Learning,1993,13(1).

[33] Tommi Jaakkola,Michael I. Jordan,Satinder P. Singh. On the Convergence of Stochastic Iterative Dynamic Programming Algorithms[J]. Neural Computation,1994,6(6).

[34] 石春生,梁洪松.组织运作过程中的自适应机理[J].管理科学,2004(01):12-16.

[35] 刘小峰,陈国华,李真.零售网络的结构建模与演化分析[J].管理科学,2009,22(04):23-30.

[36] 李静静.基于模糊K均值聚类和Sarsa(λ)算法的自适应爬壁机器人路径规划[J].计算机测量与控制,2014,22(09):2879-2881+2885.

[37] 戈军,周莲英.基于 SARSA(λ)的实时交通信号控制模型[J].计算机工程与应用,2015,51(24):244-248.

[38] P.Read Montague. Reinforcement Learning: An Introduction, by Sutton, R.S. and Barto, A.G.[J]. Trends in Cognitive Sciences,1999,3(9).

[39] John Moody,Lizhong Wu,Yuansong Liao,Matthew Saffell. Performance functions and reinforcement learning for trading systems and portfolios[J]. Journal of Forecasting,1998,17(5‐6).

[40] Jangmin O,Jongwoo Lee,Jae Won Lee,Byoung-Tak Zhang. Adaptive stock trading with dynamic asset allocation using reinforcement learning[J]. Information Sciences,2005,176(15).

[41] Bekiros S D. Heterogeneous trading strategies with adaptive fuzzy actor–critic reinforcement learning: A behavioral approach[J]. Journal of Economic Dynamics and Control, 2010, 34(6): 1153-1170.

[42] Zhiyong Tan,Chai Quek,Philip Y.K. Cheng. Stock trading with cycles: A financial application of ANFIS and reinforcement learning[J]. Expert Systems With Applications,2011,38(5).

[43] Yang S, Paddrik M, Hayes R, et al. Behavior based learning in identifying high frequency trading strategies[C]//Computational Intelligence for Financial Engineering & Economics (CIFEr), 2012 IEEE Conference on. IEEE, 2012: 1-8.

[44] Saeid Fallahpour,Hasan Hakimian,Khalil Taheri,Ehsan Ramezanifar. Pairs trading strategy optimization using the reinforcement learning method: a cointegration approach[J]. Soft Computing,2016,20(12).

[45] 胡文伟,胡建强,李湛,周剑峰.基于强化学习算法的自适应配对交易模型[J].管理科学,2017,30(02):148-160.

[46] 赵珊珊. 基于强化学习算法的配对交易策略研究[D].安徽大学,2018.

[47] Chen C T, Chen A P, Huang S H. Cloning Strategies from Trading Records using Agent-based Reinforcement Learning Algorithm[C]//2018 IEEE International Conference on Agents (ICA). IEEE, 2018: 34-37.

[48] 王欣,王芳.基于强化学习的动态定价策略研究综述[J].计算机应用与软件,2019,36(12):1-6+18.

[49] 王现磊,郝文宁,陈刚,余晓晗.基于模拟退火策略的 Sarsa 强化学习方法[J].计算机仿

真,2019,36(04):219-222+228.

致 谢

入学两年了，一转眼已是快离开的日子。很感谢各位老师们的淳淳教诲，感谢各位老师教会了我们如何做人，如何做事，教会了我们以宽容的心态面对生活，以诚恳的态度面对工作。也感谢各位老师对我们人生的建议，指引着我们朝着正确方向努力。在日常生活和学习的点滴中，帮助我们形成正确且成熟的人生观、价值观、世界观。

在此论文完成之际，在这里要特别要感谢我的指导教师，在整个研究生生涯，我的导师给予了我很大的帮助。她为人亲和，同时治学严谨、学识渊博，在我进行学术研究时，给我创造了良好的精神氛围。在整个论文的写作过程中，从选题时写开题报告，预答辩时写初稿，直到终稿完成，一遍又一遍地指出每一版的具体问题，从格式到内容严格把关，其认真的科研与学习态度值得我一辈子学习。

同时也感谢和我一起学习、写论文的小伙伴们，没有同学们无私的帮助，没有大家一起互相讨论的日子，论文工作也不可能顺利完成。我和小伙伴们一起度过了很多快乐的时光，也感谢小伙伴们在生活中给予的各种帮助。

诚挚地感谢呵护我成长的家人，在我人生遇到困难的时候，父母总给予我鼓励、包容和关爱。20 多年来成长的路上，是父母一直在背后默默支持、照顾我，是我坚实的后盾。家人在精神上和物质上的支持，以及从小树立的榜样，这些都使我更加坚定地去追求人生理想。家人的爱是天下最无私的，作为子女，只有永无止境的奋斗，努力工作，过上自己想要生活，才可以报答父母的养育之恩。

最后，再次向所有关心我的老师、家人、和朋友们表示由衷的谢意，你们的关爱我永远铭记在心，你们的支持是我学习和工作的最大动力。谢谢大家！