

清 华 大 学

综 合 论 文 训 练

题目：基于平稳过程的CTA策略

系 别：数学科学系

专 业：数学与应用数学

姓 名：李博扬

指导教师：唐宏岩 副教授

2018 年 5 月 21 日

关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：学校有权保留学位论文的复印件，允许该论文被查阅和借阅；学校可以公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存该论文。

签名： 李博扬 导师签名： 陈良 日期： 2018.6.5

中文摘要

本文研究了经典协整配对模型，指出了其弊端，即配对价格之间的长期均衡关系可能被打破，一成不变的协整模型可能失效。同时本文基于平稳过程的均值回复性质，提出了改进方法，即将配对价差序列分解成随时间缓慢变化的均值项以及具有均值回复特性的平稳波动项，利用前 n 期加权配对价差均值拟合均值项，同时配对的系数也随时间变化，由此建立了动态的价差配对模型。在实证当中，本文利用了交叉验证法对模型进行了参数优化，在对商品期货焦炭和焦煤的主力连续合约验证集的回测中，确定了策略较优的 5min 数据频率以及开仓阈值参数 $k=5$ 。在测试集上验证策略的泛化能力，得到了相对较好的策略绩效，初步确认了改进模型的有效性。在结论与展望中提到了未来模型的改进空间，以便后来者的研究。

关键词：平稳过程；协整配对；交叉验证；CTA

ABSTRACT

In this paper, the classical co-integration pairing model is studied, and its disadvantages are pointed out, that is, the long-term equilibrium relationship between pair prices may be broken, and the invariable co-integration model may be invalid. Based on mean reversion property of stationary process, this paper puts forward the improved method, that is, the pairing spread-price sequence is decomposed into the mean item which changes slowly with time and the stationary fluctuation item with mean reversion property. Then the method uses the last n period average weighted mean to fit the mean item, meanwhile, the coefficient of pairing also changes over time. Finally, the dynamic price-gap pairing model is established. In empirical, this paper utilizes the cross validation method to optimize the parameters of the model. When backtesting the validation set of the main coking coal and coke data in commodities futures contract, the better fitting 5min data frequency and open-threshold parameter $k=5$ are determined. The generalization ability of the strategy was verified on the test set, and relatively good performance was obtained, and the effectiveness of the improved model was preliminarily confirmed. In conclusion and prospect, the potential improvement space of the model is mentioned, so as to facilitate the later research.

Keywords: stationary process; co-integration pairing; cross validation; CTA

目录

第 1 章	引言.....	1
1.1	研究背景.....	1
1.1.1	CTA 策略相关背景.....	1
1.1.2	引入平稳过程的相关背景	1
1.2	研究现状.....	2
1.3	本文研究内容.....	3
第 2 章	协整理论.....	4
2.1	时间序列平稳性.....	4
2.2	单整.....	4
2.3	单位根检验.....	4
2.4	协整检验.....	5
第 3 章	策略构建.....	7
3.1	经典协整配对交易.....	7
3.2	经典协整配对交易的弊端.....	7
3.3	模型改进.....	9
3.3.1	理论部分	9
3.3.2	交易策略	11
3.3.3	交叉验证法优化参数	12
第 4 章	绩效评估方法.....	13
4.1	年化收益率.....	13
4.2	Sharpe 比率	13
4.3	最大回撤率.....	13
第 5 章	套利模型实证案例.....	15
5.1	数据准备.....	15
5.2	参数优化.....	16
5.3	泛化测试.....	19

5.4 结果分析.....	20
第 6 章 结论与展望.....	21
插图索引	22
表格索引	23
参考文献	24
致 谢	25
声 明	26
附录 A 外文资料的书面翻译.....	27

第1章 引言

1.1 研究背景

1.1.1 CTA 策略相关背景

CTA 全称为 Commodity Trading Advisor，直译为商品交易顾问，也称为管理期货。CTA 策略主要是指指标的为大宗商品或者某些金融产品和衍生品（债券、利率、货币和指数等）的期货合约的投资策略。

如今的 CTA 策略主要在基本面和技术分析的基础上，引入量化模型，并且借助计算机编程，根据既定量化模型实时产生买卖信号对标的资产进行投资。CTA 基金策略通过多种或跨期期货合约多空头寸进行投资，一开始只在商品期货市场交易，之后逐渐发展到股指、外汇、债券、利率等金融衍生品期货市场。在中国，由于股票市场无法做空，CTA 策略基本不会涉足于股票市场，所以本文主要探究的是基于平稳过程的配对交易策略在商品期货上的表现。

CTA 策略可分为趋势型策略和反趋势型策略。其中趋势型策略目前应用最为广泛。趋势跟踪型策略主要应用各种市场指标预测当前市场走向，然后建立交易信号。信号建立方式由传统的移动均线、周期理论到如今引入机器学习算法一直在不断更新当中。值得一提的是，有时传统的统计套利方式比新颖的机器学习算法更加有效。本文所讨论的基于平稳过程的 CTA 套利正是基于统计套利中的协整配对模型。

1.1.2 引入平稳过程的相关背景

众所周知，很多行业的盈利都依赖于大数定律。例如赌场，由于赌徒每一局赌局获利的期望为 0，由大数定律可以知道，赌场的经营者几乎不会因为赌局而亏损，于是收取的固定的手续费称为“免费的午餐”。类似的行业还有保险等。大数定律的中心思想为：若一个事件能带来正的期望收益，则大量类似的独立事件可以抵消最终收益的不确定性，从而使得期望利润不变的情况下，利润波动大大降低，从而获取平稳的盈利。

但是，由于大数定律成立的条件——事件之间相互独立在金融时间序列中实在太过苛刻，导致大数定律并不能为 CTA 策略带来“免费的午餐”。以股票市场为例，由金融界普遍认同的有效市场假说，哪怕是市场的弱有效性，都表明实时的

股票价格会反映所有历史股价表达的信息。因此在股票价格时间序列中，要做到每个时刻价格互相独立是不现实的。于是我们需要引入大数定律之外的理论。

我们注意到平稳过程具有一个极其优秀的性质——遍历性，几乎就相当于时间序列领域的大数定律。具有遍历性的平稳时间序列，随着时间不断推移，样本的平均值会不断逼近样本期望。这样一来，我们可以预见到时间序列会不断在它的均值附近来回上下穿梭，这样就提供了套利的机会。

1.2 研究现状

本文主要应用协整理论获取具有可套利平稳期货时间序列，在统计套利领域，最早引入协整模型的是 John Board 和 Charles Sutcliffe^[1]，他们发现大阪、新加坡以及芝加哥市场的日经 225 指数存在协整关系，并证明了其中协整套利的可能性。

A. N. Burgess^[2]介绍了协整分析在统计套利上应用的两种方向：第一种是用来产生“错误定价关系”的方法，利用逐步回归而不是标准回归。第二种是基于协整检验的性质，在他的实证中是基于错误定价时间序列的方差比。他的套利方法在 FTSE100 指数上很成功，在考虑交易手续费的情况下，年化夏普率达到了 7.8。

Carol Alexander 和 Anca Dimitriu^[3]比较了协整优化和跟踪误差统计套利策略的表现。基于统计套利投资组合标准数据的性质，两种策略的年化收益与市场基准相比波动更小，市场相关性低且接近于正态收益分布。但是他们发现协整优化统计套利策略在样本外的表现更好，因此更具有泛化能力，保证了统计套利的可行性，减小了过度对冲和关联交易成本的风险。

陈实^[4]基于经典的协整套利模型，分析了其弊端并提出了均值滤波的方法解决长期均衡改变的问题。基于这一新模型，他又考虑了盘口价差对交易绩效的影响，对交易策略进行了更细致的描述，并给出了股指期货跨期和跨品种交易的实例，绩效表现优异。

仇中群^[5]分别在期现和跨期两种交易策略上对协整方法进行了研究。在期限策略中，他发现两 ETF 基金与沪深 300 存在长期均衡关系，并利用两 ETF 基金对沪深 300 进行了很好的模拟。在跨期策略中，通过对同一标的资产下的不同期货进行长期均衡配对交易，取得了不错的绩效。且协整回归的残差项的波动在可接受值域内，交易的风险较小。

金恺^[6]对沪深 300 指数中的股票按所属行业进行了分划，分别在各个类中进

行相关性检验，在具有一定关联的股票中利用 ADF 检验出同为一阶单整的股票时间序列，然后利用 EG 两步法进行了协整检验，遴选出了具有长期均衡关系的股票配对。在选出的 100 对股票中，有 99 对发出了交易信号，超额收益平均为 29.95%。同时还发现相关系数在 0.85 左右的股票对表现最佳。

张河生和闻岳春^[7]通过改变交易区间与止损范围，大大改善了配对策略绩效。王峥明^[8]采取了利用同向性检验和利润预测与排序的两种配对策略，并比较了优劣。得出了在中国 A 股市场上，基于预测与排序的配对策略总体上优于趋同性配对策略的结论。

陈守东等^[9]利用长期均衡的统计分析方法，基于世界上各大重要的二级市场数据进行了相关性分析。研究表明我国的二级市场发展主要的参考指标与国际市场不存在协整关系，从而得出了我国市场与国外市场相分离的结论。

袁象等^[10]研究了在有效市场中，期货与现货价格存在长期均衡关系。套保投资者需充分考虑这一关系来提高套保有效性。且当均衡关系越强时，越需要考虑协整模型的应用来减少损失。

1.3 本文研究内容

本文主要基于协整模型得到具有均值回归性质的平稳时间序列构建 CTA 套利策略。标的资产将在中国商品期货市场选取，本文将探讨协整参数的最佳选取，包括协整系数、协整均值、开仓阈值以及数据频率，期望得到在某一特定市场上具有稳定盈利能力的策略

第2章 协整理论

2.1 时间序列平稳性

平稳时间序列具有随着时间推移，其统计特征和规律不会变化的优秀性质。本文基于协整理论要得到的就是具有均值回复性质的平稳时间序列，以下是时间序列平稳性的定义：

(1) 严平稳

时间序列 $\{x_t, t \in T\}$ 是严平稳的，等价于其联合概率分布不随时间推移而变化。数学表达如下：

$$P\{x_{t_1} \leq b_1, x_{t_2} \leq b_2, \dots, x_{t_n} \leq b_n\} = P\{x_{t_1+k} \leq b_1, x_{t_2+k} \leq b_2, \dots, x_{t_n+k} \leq b_n\}$$

其中 n, k 为任意正整数， $t_1 < t_2 < \dots < t_n, t_i \in T, b_1, b_2, \dots, b_n$ 为实数。

(2) 宽平稳

时间序列 $\{x_t, t \in T\}$ 是宽平稳的，等价于其期望和方差与时间无关，序列中任意两项的协方差只于它们的时间间隔相关。数学表达如下：

- a. $E(x_t) = \mu, t = 1, 2, \dots, \mu$ 为常数
- b. $\text{Var}(x_t) = \sigma^2, t = 1, 2, \dots, \sigma$ 为常数
- c. $\text{Cov}(x_t, x_{t+k}) = r_k, t = 1, 2, \dots,$

由于严平稳时间序列成立条件太过苛刻，在研究 CTA 策略时，序列达到宽平稳已经满足构建策略所需性质。于是本文提到的时间序列平稳均默认为宽平稳。

2.2 单整

定义：对于时间序列 $\{x_t\}$ ，若 $\{x_t\}$ 非平稳且 $\{\Delta x_t, \Delta x_t = x_t - x_{t-1}\}$ 平稳，则称 $\{x_t\}$ 为一阶单整时间序列，记为 $x_t \sim I(1)$ 。

推广定义：对于时间序列 $\{x_t\}$ ，若 $\{x_t\}$ 经过 $d-1$ 阶差分仍非平稳，但通过 d 阶差分后平稳，则称 $\{x_t\}$ 为 d 阶单整时间序列，记为 $x_t \sim I(d)$ 。

2.3 单位根检验

单位根检验是用于检验时间序列是否平稳的检验。本文使用的是 ADF 检验。

设存在时间序列 $\{y_t\}$ ，考虑以下三种回归方程：

$$\Delta y_t = \omega y_{t-1} + \sum_{i=1}^k \beta_i \Delta y_{t-i} + \varepsilon_t \quad t = 1, 2, \dots, T \quad (2-1)$$

$$\Delta y_t = a + \omega y_{t-1} + \sum_{i=1}^k \beta_i \Delta y_{t-i} + \varepsilon_t \quad t = 1, 2, \dots, T \quad (2-2)$$

$$\Delta y_t = a + \delta_t \omega y_{t-1} + \sum_{i=1}^k \beta_i \Delta y_{t-i} + \varepsilon_t \quad t = 1, 2, \dots, T \quad (2-3)$$

其中 a 为常数， δ_t 为时间趋势函数， $\varepsilon_t \sim iid N(0, \sigma^2)$ ， Δ 为一阶差分算子， k 为滞后阶数。

式(2-1)没有时间趋势、常数项，称为模型 1.

式(2-2)没有时间趋势，有常数项，称为模型 2.

式(2-3)有时间趋势、常数项，称为模型 3.

假设检验过程：

(1) 原假设： $H_0: \omega = 0$;

备择假设： $H_1: \omega < 0$

(2) 首先检验模型 3，若检验结果拒绝 0 假设，停止检验，若不拒绝，则开始检验模型 2，类推最后检验模型 1。

(3) ADF 检验的统计量 < 临界值，则拒绝 H_0 ; 否则不拒绝 H_0

2.4 协整检验

在现实应用时间序列分析时，很容易出现“伪回归”现象。即回归结果的统计量十分显著，但是回归残差却不平稳。一般来说，有限次差分可以使时间序列平稳化，但是差分之后的变量无法直接构建套利策略，所以当差分之后的变量不满足策略构建需求的时候，就无法使用这一技巧了。

所谓协整关系，就是指若干不平稳的时间序列通过乘以系数相加减后得到一个平稳过程的关系。协整理论一方面可以解决如上问题，另一方面可直接得到一种经典的长期价差套利方法。

协整定义如下：

若 k 维时间序列 $y_t = (y_{1t}, y_{2t}, \dots, y_{kt})'$, $t = 1, 2, \dots, T$ 满足如下两个条件：

a. $y_t \sim I(d)$ ，即 $y_{1t}, y_{2t}, \dots, y_{kt}$ 均为 d 阶单整；

b. 存在非零向量 $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)'$ ，使得 $\alpha' y_t \sim I(d - b)$, $d \geq b > 0$.

则称序列 y_t 是 (d, b) 阶协整的，记为： $y_t \sim CI(d, b)$ ， α 为协整向量。特别的，对两个变量而言，必须是同阶单整才可能存在协整关系。

本文将使用 JJ 检验方法对协整关系进行检验。JJ 检验基于 Vector Auto-

Regression 模型，是进行多变量协整的优选。

设存在 k 维时间序列 $y_t = (y_{1t}, y_{2t}, \dots, y_{kt})' \sim I(1)$, k 为正整数, $t=1, 2, \dots, T$, 考虑 p 阶 VAR 模型:

$$y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + B x_t + \varepsilon_t \quad (2-4)$$

其中 A_1, A_2, \dots, A_p, B 为系数矩阵, x_t 为确定的 k 维向量, $\varepsilon_t \sim iid N(0, \sigma^2)$ 。

此时差分后模型如下:

$$\Delta y_t = \Pi y_{t-1} + \sum_{j=1}^p \Gamma_j \Delta y_{t-j} + B x_t + \varepsilon_t \quad (2-5)$$

其中 $\Pi = \sum_{j=1}^{p-1} A_j - I$, $\Gamma_j = -\sum_{i=j+1}^p A_i$ 。

由于 Δy_t 与 Δy_{t-j} 均 $\sim I(0)$, 于是只需考虑 Πy_{t-1} 是否 $I(0)$ 。Granger 定理告诉我们: 若系数矩阵 Π 的秩 $r < m$, 那么存在秩均为 r 的 $m \times r$ 阶矩阵 α 和 β , 使得 $\Pi = \alpha \beta'$, 且 $\beta y_{t-1} \sim I(0)$ 。其中 r 为协整秩, β' 为协整向量矩阵, 描述了 r 个协整关系, 矩阵 α 反映了 r 个协整关系对 Δy_t 的影响。

JJ 检验的 0 假设为: 最多存在 s 个协整关系, $s=0, 1, \dots, k-1$ 。

第3章 策略构建

3.1 经典协整配对交易

在经典的资产价格理论下，资产价格时间序列符合几何布朗运动，资产对数价格为一阶单整。在实际数据中，Vidyamurthy^[11]研究发现资产的对数价格确实为一阶单整。对于两个一阶单整的资产对数价格时间序列 $\{p_t^A\}$ 和 $\{p_t^B\}$ ，我们利用线性回归拟合下式：

$$p_t^A - \beta p_t^B = \mu_t$$

其中 β 为协整系数， μ_t 为残差。然后利用 ADF 检验验证 $\{\mu_t\}$ 的平稳性，若 $\{\mu_t\}$ 平稳，则回归方程可写为下式：

$$p_t^A - \beta p_t^B = \mu + \epsilon_t \quad (3-1)$$

其中 μ 为常数， ϵ_t 为均值为0的平稳过程。此时资产A与B之间对数价格存在长期均衡关系。这意味着，若某时刻 $\epsilon_{t_1} > 0$ ，则可视为在 t_1 时刻资产A的价格被高估，资产B价格被低估，于是卖出1单位A，买入 β 单位B。等到 t_2 时刻残差回归均值0时（即 $\epsilon_{t_2} = 0$ ），进行平仓，则可获取约 ϵ_{t_1} 的利润；反之，若某时刻 $\epsilon_{t_1} < 0$ ，则可视为在 t_1 时刻资产A的价格被低估，资产B价格被高估，于是买入1单位A，卖出 β 单位B。等到 t_2 时刻残差回归均值0时（即 $\epsilon_{t_2} = 0$ ），进行平仓，则可获取约 $-\epsilon_{t_1}$ 的利润。

由此可知，只要残差项偏离均值之后，在未来可以回归均值，那么按如上方方法可套利，且偏离均值越远，利润越高。

3.2 经典协整配对交易的弊端

由于经典的资产定价理论所需的前提假设在现实中十分苛刻，我们或许可以在统计量检验上得到两个相关资产内在存在长期均衡的协整关系，但是这一协整关系是否会随时间而变化就不得而知了。换句话说，多个资产之间是否真实存在协整关系，且这一协整关系是否长期不变，还不得而知。

由3.1可知，我们建立在平稳过程上的套利策略往往需要这一平稳时间序列在均值附近来回波动，且波动越大，收益越大。策略在平稳时间序列远离其期望时开仓，在回复到均值时平仓，以此获取超额收益。但是，在长期均衡，也即协

整关系不稳定的情况下，残差序列可能长期偏离均衡位置而无法回复，此时一笔交易周期过长势必导致平仓亏损，且这样的亏损往往是不能接受的。

举一个例子，中国商品期货市场上焦炭和焦煤往往存在价格上的关联，这是由它们的内在性质决定的，因此我们认为它们可能存在长期均衡关系。我们用焦炭主力连续合约（简称 J）和焦煤主力连续合约（简称 Jm）从 2013 年 3 月 22 日到 2018 年 4 月 22 日的 5 分钟频率对数价格数据作为样本，并以 2013 年 3 月 22 日到 2016 年 12 月 30 日的样本作为训练样本拟合回归方程（3-1）得到：

$$p_t^{Jm} - 0.8696p_t^J = 0.6118 + \epsilon_t$$

其中， p_t^{Jm} 为 t 时刻焦煤主力连续合约对数价格， p_t^J 为 t 时刻焦炭主力连续合约对数价格。

图 3.1 为训练样本残差序列 $\{\epsilon_t\}$ 的图像，表 3.1 为序列 $\{p_t^{Jm}\}$ 和 $\{p_t^J\}$ 以及它们的差分序列 $\{\Delta p_t^{Jm}\}$ 和 $\{\Delta p_t^J\}$ 的 ADF 检验结果和协整检验结果。

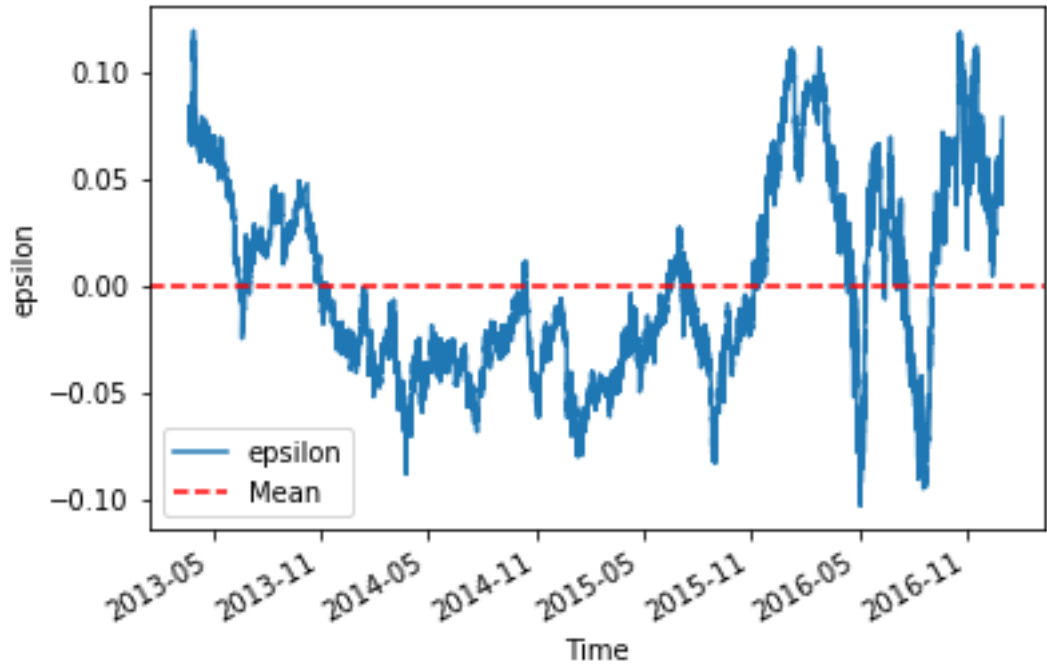


图 3.1 训练样本残差序列 $\{\epsilon_t\}$

表 3.1 序列统计检验结果

序列	统计量值	显著性临界值			假设检验	结论
		1%	5%	10%		
p_t^{Jm}	-1.1774	-3.43	-2.86	-2.57	不拒绝 H_0	单位根

p_t^l	-1.0918	-3.43	-2.86	-2.57	不拒绝 H_0	单位根
Δp_t^{lm}	-51.2264	-3.43	-2.86	-2.57	拒绝 H_0	平稳
Δp_t^l	-49.3338	-3.43	-2.86	-2.57	拒绝 H_0	平稳
ϵ_t	-2.9609	-3.43	-2.86	-2.57	拒绝 H_0	平稳

可以看到序列 $\{p_t^{lm}\}$ 和 $\{p_t^l\}$ 均为一阶单整序列，且残差项平稳，因此我们可以认为在训练集上序列 $\{p_t^{lm}\}$ 和 $\{p_t^l\}$ 存在协整关系，即长期均衡的关系。

但是由图 3.2 可以发现，在 2017 年-2018 年的测试集样本上，前述训练得到的模型已经失效。残差序列几乎稳定的将至均线以下，丧失了回复均值 0 的能力。如果还按照经典协整配对交易策略进行交易，则会在 2017 年初开仓后，几乎一年的时间内无法平仓，这对期货交易员来说存在极大亏损风险，因此策略失效。此时我们应当认为，焦炭与焦煤期货的长期均衡关系已经改变，这促使我们对原有策略进行改进。

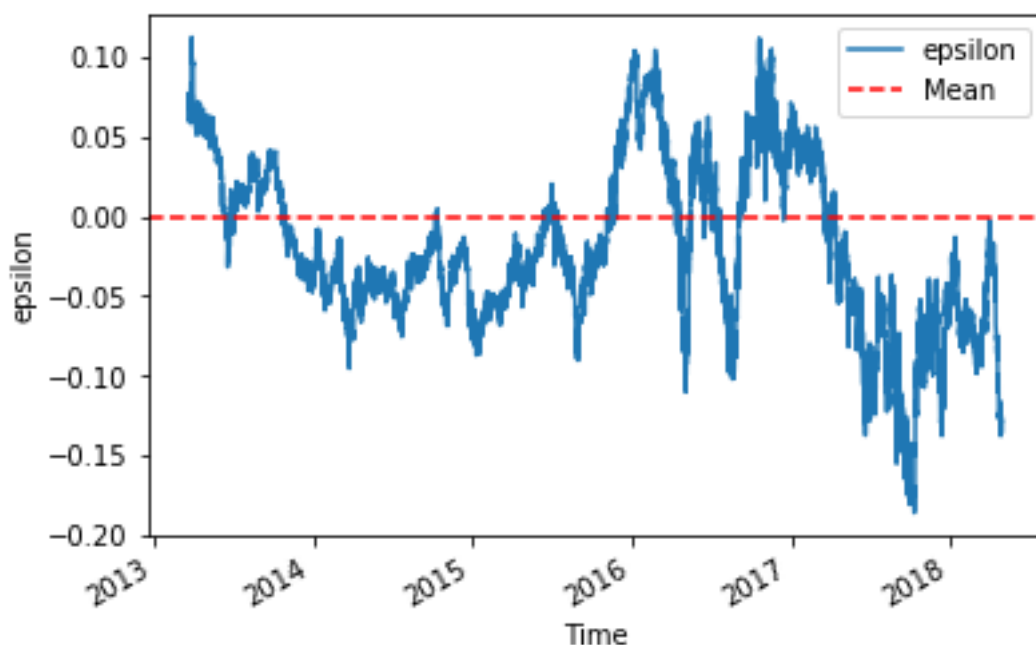


图 3.2 训练集与测试集样本残差序列 $\{\epsilon_t\}$

3.3 模型改进

3.3.1 理论部分

为了避免 3.2 节中长期均衡被打破而造成大量亏损的情况发生，本文对经典协整模型进行了改进。我们允许原始模型（3-1）式中常数项随时间缓慢变化，得

到下式:

$$\Delta_t = p_t^A - \beta p_t^B = \mu_t + \epsilon_t \quad (3-2)$$

此时, 将资产 A 与 B 做协整回归之后, 残差项被分解为一个随时间缓慢变化的均值项和一个具有显著均值回复性质的平稳波动项。基于 (3-2) 式, 我们同样可以给出类似的套利策略: 当 t_1 时刻波动项 ϵ_{t_1} 偏离均衡 μ_{t_1} 过远时触发开仓信号, 然后持续持仓至某时刻 t_2 波动项 ϵ_{t_2} 回复到均衡 μ_{t_2} 触发平仓信号。下面将给出可行的 $\{\mu_t\}$ 选择, 使得波动项 $\{\epsilon_t\}$ 满足平稳且具有均值回复性质。

基于 (3-2) 式, 我们不再需要资产 A 和 B 对数价格满足长期均衡的协整关系, 在陈实^[4]的研究中指出, 只需要两个资产对数价格满足它们的对数差分 $\{\Delta p_t^A, \Delta p_t^B\}$ 满足联合弱平稳过程即可推导出: 对于任意的系数 β , 序列 $\{\Delta_t\}$ 均为平稳过程。且当均值序列 $\{\mu_t\}$ 选择为前 n 期 $\{\Delta_t\}$ 序列的移动平均时, 波动项序列 $\{\epsilon_t\}$ 为平稳过程。

本文创新性地提出: 在时刻 t 时, 我们认为越接近 t 时刻的 $\{\Delta_t\}$ 值越能反映均值项的走势, 于是均值序列 $\{\mu_t\}$ 的选取应该为前 n 期 $\{\Delta_t\}$ 序列的加权移动平均, 赋予离 t 时刻近的 $\{\Delta_t\}$ 值更高的权重。即为下式:

$$\mu_t = \frac{\sum_{i=1}^n (n-i+1) \Delta_{t-i+1}}{1+2+\dots+n} \quad (3-3)$$

由下引理 3.1 可以保证在均值序列 $\{\mu_t\}$ 如上取值时, 波动项序列 $\{\epsilon_t\}$ 为平稳过程。以此来保证基于平稳过程均值回复性质的套利策略的有效性。

引理 3.1 当资产 A 与 B 的价格对数增量 $\{r_t: r_t = \Delta_t - \Delta_{t-1}\}$ 为平稳过程, 且均值序列 $\{\mu_t\}$ 按照式 (3-3) 取值时, 波动项序列 $\{\epsilon_t\}$ 为平稳过程。

$$\begin{aligned} \text{证明} \quad \epsilon_t &= \Delta_t - \mu_t \\ &= \frac{\sum_{i=1}^n (n-i+1) (\Delta_t - \Delta_{t-i+1})}{1+2+\dots+n} \\ &= \frac{\sum_{i=2}^n (n-i+1) \sum_{j=0}^{i-2} (\Delta_{t-j} - \Delta_{t-j-1})}{1+2+\dots+n} \\ &= \frac{\sum_{i=2}^n (n-i+1) \sum_{j=0}^{i-2} r_{t-j}}{1+2+\dots+n} \end{aligned}$$

由于序列 $\{r_t\}$ 为平稳过程, 其线性组合也为平稳过程, 于是序列 $\{\epsilon_t\}$ 为平稳过程。证毕。

综上所述, 我们可以对配对对数价差序列 $\{\Delta_t\}$ 进行分解, 其中随时间缓慢变化的 $\{\mu_t\}$ 序列称为均值项序列, 另一部分具有很强均值回复能力的平稳序列 $\{\epsilon_t\}$ 称为波动项序列。在此基础上, 我们可以套用经典协整配对模型的交易策略, 在 ϵ_t 偏离均衡时开仓, 在之后回归均衡时平仓。但是此时, 交易的收益由 $\{\epsilon_t\}$ 的均值回

复特性带来的确定收益和均值项 $\{\mu_t\}$ 的趋势带了的不可预测收益共同提供。正是由于 $\{\mu_t\}$ 的不可预见性带来了交易风险，我们希望最小化这一风险。

陈实^[4]在研究中指出：由于两种配对资产无需再满足协整关系，因此无论 β 取何值，(3-2)式均能被拟合，于是可以通过对 β 的合理取值，使得由均值项 $\{\mu_t\}$ 的不确定趋势带来的风险最小化。最后他证明了， β 的最优值即是通过协整回归得到的协整参数。

这样我们可以基于过去 n 期的资产 A 与 B 的对数价格数据 $\{p_t^A\}$ 和 $\{p_t^B\}$ ，回归得到 t 时刻的 β 的最优估计 $\hat{\beta}_t$ 如下：

$$\hat{\beta}_t = \frac{\sum_{i=0}^{n-1} \left(p_{t-i}^A - \frac{\sum_{j=0}^{n-1} p_{t-j}^A}{n} \right) \left(p_{t-i}^B - \frac{\sum_{j=0}^{n-1} p_{t-j}^B}{n} \right)}{\sum_{i=0}^{n-1} \left(p_{t-i}^B - \frac{\sum_{j=0}^{n-1} p_{t-j}^B}{n} \right)^2} \quad (3-4)$$

于是对于每一个时刻 t ，我们都有对应的 $\hat{\beta}_t$ 。每一个 $\hat{\beta}_t$ 均对应一个价差序列 $\{\Delta_t\}$ 、均值项序列 $\{\mu_t\}$ 和波动项序列 $\{\epsilon_t\}$ 。于是套用经典协整配对模型策略可得到改进模型的交易策略。

3.3.2 交易策略

本部分基于模型 $\Delta_t = p_t^A - \beta_t p_t^B = \mu_t + \epsilon_t$ 构建套利策略。基本套利过程如下：

- 1、在 t 时刻，按照 (3-4) 式，利用过去 n 期 A、B 资产的对数价格序列 $\{p_t^A\}$ 和 $\{p_t^B\}$ 估计出 t 时刻的系数 $\hat{\beta}_t$ 。
- 2、根据估计量 $\hat{\beta}_t$ 得到 Δ_t ，根据 (3-3) 式计算均值项 μ_t ，根据 $\epsilon_t = \Delta_t - \mu_t$ 计算出构建策略所需的波动项序列 ϵ_t 。
- 3、根据 t 时刻 ϵ_t 是否在交易域发出交易信号，交易资产为单位金额的 A 资产和 $\hat{\beta}_t$ 金额的 B 资产。一旦开仓信号被触发，固定此时刻的系数值 $\hat{\beta}_t$ ，直到之后的 $t + \Delta t$ 时刻波动项触发平仓信号。

对于交易信号应当何时被触发的问题，如果当 $\epsilon_t > 0$ 就马上触发开仓信号，当之后波动项第一次 ≤ 0 的时刻马上平仓，则获取的收益十分有限，且当波动项在均衡位置来回高频波动时易造成频繁开平仓，收益极小且浪费大量手续费，无疑会造成亏损。由于商品期货交易自带杠杆效应，于是对开平仓阈值的优化在 CTA 策略中尤为重要，否则将造成大量损失。

在一些经典的 CTA 策略中，例如布林通道突破策略，开仓信号大多选择当前价格 \pm 近期价格标准差，本文在此基础上希望更优化开仓信号。考虑 t 时刻开仓阈值为 $k * \text{std}_n^t(\epsilon)$ （这里 $\text{std}_n^t(\epsilon)$ 为 t 时刻波动项序列近 n 期的标准差），平仓阈值

为 0。即：

开仓信号：在 t 时刻，若当前无仓位，当 $\epsilon_{t-1} \geq -k * \text{std}_n^t(\epsilon)$ 且 $\epsilon_t < -k * \text{std}_n^t(\epsilon)$ 时，认为波动项序列 $\{\epsilon_t\}$ 较均衡位置向下偏离，即资产 A 被低估，资产 B 被高估，买入单位资金资产 A，卖空 $\hat{\beta}_t$ 资金资产 B；当 $\epsilon_{t-1} < k * \text{std}_n^t(\epsilon)$ 且 $\epsilon_t > k * \text{std}_n^t(\epsilon)$ 时，认为波动项序列 $\{\epsilon_t\}$ 较均衡位置向上偏离，即资产 A 被高估，资产 B 被低估，卖空单位资金 A，买入 $\hat{\beta}_t$ 资金资产 B。

平仓信号：在 t 时刻，若当前有仓位，则计算开仓时刻的 $\hat{\beta}$ 对应的 $\epsilon_t^{\hat{\beta}}$ 。若目前持有 A 的多头和 B 的空头，则当 $\epsilon_t^{\hat{\beta}} > 0$ 时平仓。若目前持有 A 的空头和 B 的多头，则当 $\epsilon_t^{\hat{\beta}} < 0$ 时平仓。

3.3.3 交叉验证法优化参数

本文涉及到的可自选参数有计算周期 n ，开仓阈值系数 k 。对于计算周期 n ，本文通过固定 $n=100$ ，在不同频率的数据上实现策略来达到优化开仓大小的目的，同时得到策略最适合的数据频率。本文考虑 1s、5min 和 1h 三种不同频率数据，考虑 $k=0.2, 0.5, 1, 5$ 四个开仓阈值系数，期望获取一种较优的策略实现，具体的绩效评估方法见第 4 章。

所谓交叉验证法，即是通过将样本分为训练集、验证集和测试集，在训练集上训练各种参数下的模型，在验证集上比较各种参数模型的优劣，并筛选出最优模型在测试集上测试其泛化能力。交叉验证法广泛应用于机器学习算法中，是一种筛选参数，优化模型的有效方法。由于本文的策略模型已经构建完毕，所以不需要训练集生成模型，所以我们只需将资产对数价格数据分为验证集和测试集即可。验证集用于最优化开仓阈值系数 k 和数据频率两个参数，测试集用于测试策略的泛化能力，即在样本外的数据中是否也有预期表现。

第4章 绩效评估方法

4.1 年化收益率

年化收益率 (Annualized Return) 是一个最基本的绩效评估指标, 一个好的 CTA 策略必须具有可观的年化收益率。计算公式如下:

$$AR_t = \left(\frac{V_t}{V_0} \right)^{\frac{1}{T}} - 1$$

其中 AR_t 表示长度为 t 天的投资年化收益率, V_0 表示期初资产, V_t 表示期末资产, T 表示一年内有效投资时间, 一般股票和期货取为年内交易日天数约为 250。

4.2 Sharpe 比率

Sharpe 比率也成为夏普率, 是一个衡量收益风险比的最常用指标。同时夏普率也是个相对指标, 一个好的 CTA 策略要具有相对于其他策略更高的夏普率。计算公式如下:

$$S = \frac{R - R_f}{\sqrt{Var(R)}}$$

其中 R 为策略的年化收益率, R_f 为无风险收益率, $Var(R)$ 为收益率的年化方差。其中分子部分又称为超额收益率。为了计算方便, 我们统一省略无风险收益率部分, 于是 S 的计算如下:

$$S = \frac{R}{\sqrt{Var(R)}}$$

4.3 最大回撤率

最大回撤率 (Max Drawdown) 是一个重要的风险度量指标, 它能反映买入资产组合后至今出现过的最差的情况。计算公式如下:

$$MD_t = \max_{0 \leq i < j \leq t} \frac{V_i - V_j}{V_i}$$

其中 MD_t 为前 t 期的最大回撤率， V_i 表示第 i 期的投资净值。最大回撤率可以用来衡量投资者可能遇到的最大亏损。一个好的 CTA 策略应该具有较小的最大回撤率。

第5章 套利模型实证案例

5.1 数据准备

本文将中国商品期货焦炭与焦煤进行配对，焦炭记为 J，焦煤记为 Jm。在产业链上，由于焦煤是焦炭的上游产业，简单的来说，焦炭由焦煤炼成，所以二者的价格很大概率存在着均衡关系。本文数据来自 TradeBlazer 交易软件，分别准备了频率分别为 1s、5min 和 1h 的 J 和 Jm 主力连续合约价格，时间区间为 2013 年 3 月 22 日到 2018 年 4 月 22 日。策略运行时间取为 J 和 Jm 交易时间的交集。图 5-1 为 J 和 Jm 主力连续合约对数价格曲线，可见两组数据存在着一定的均衡关系。

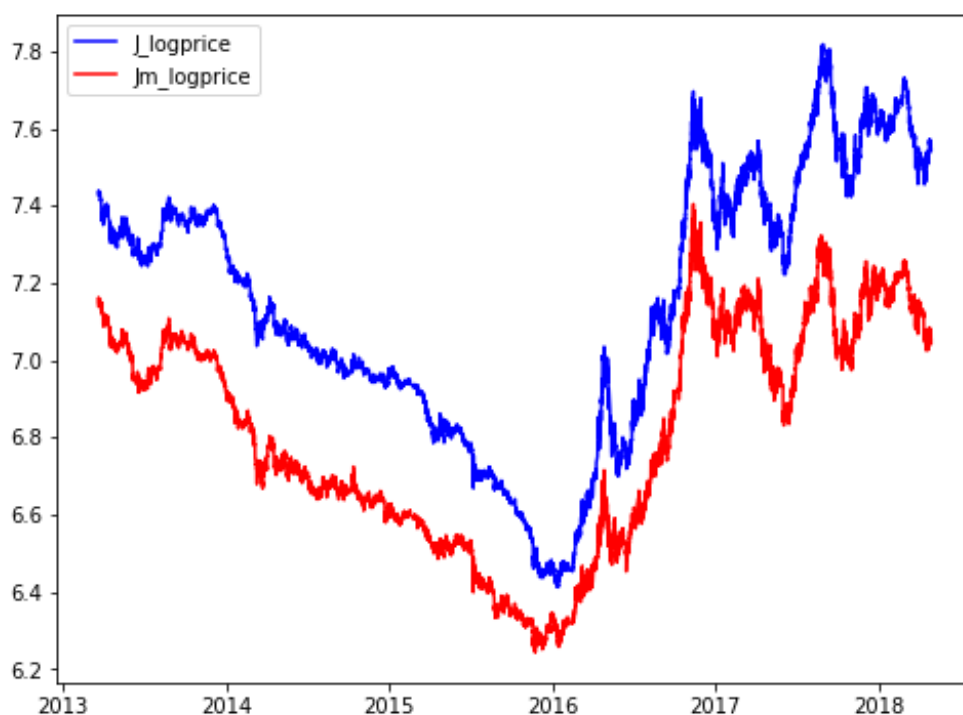
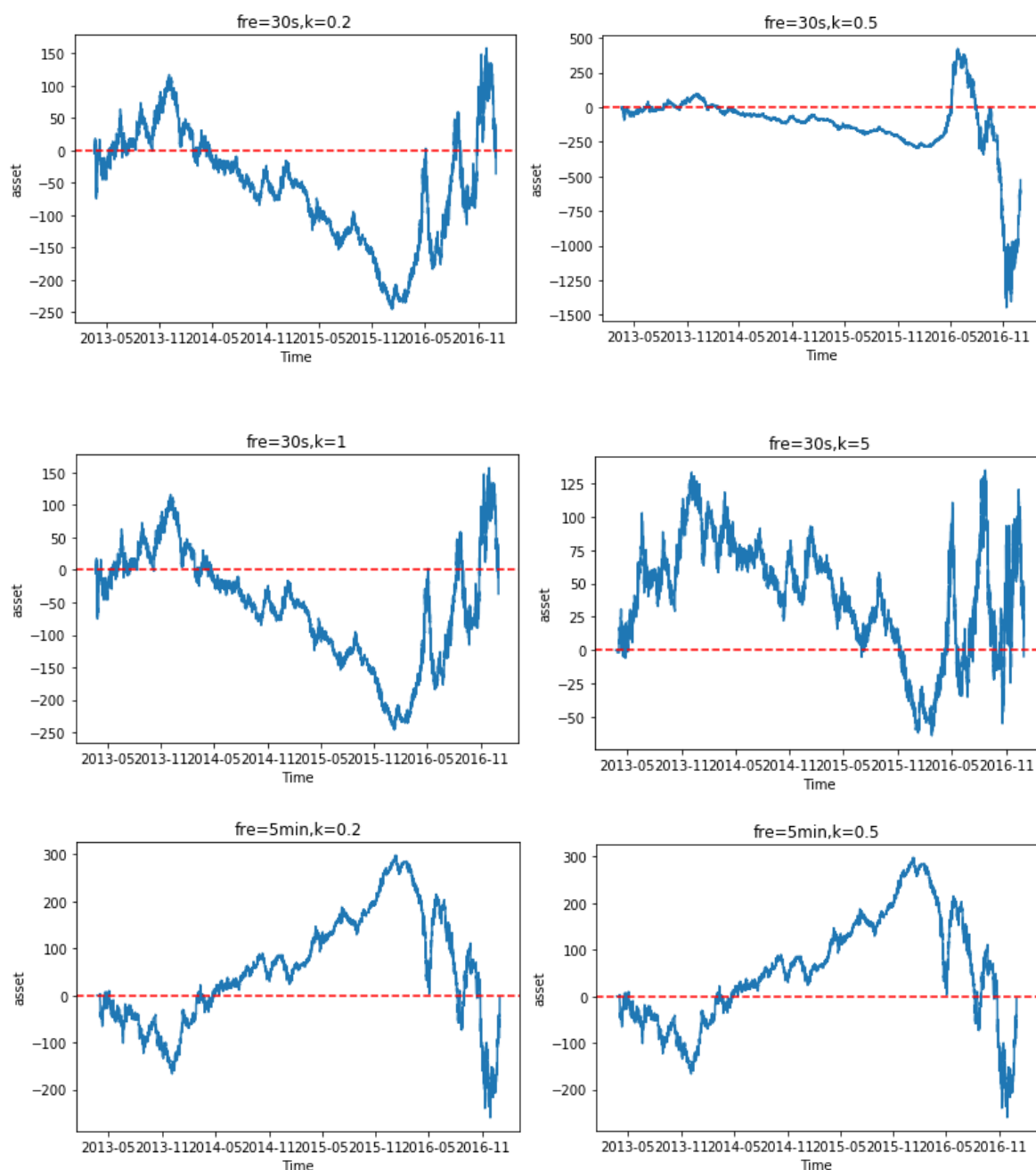


图 5.1 焦炭与焦煤样本对数价格曲线

交易手续费设置参考国泰君安期货 2018 年 5 月 17 日发布的《期货品种手续费标准》，J 与 Jm 每手交易手续费为 0.024%。由于期货交易存在保证金制度，即交易存在杠杆，所以本文考虑策略绩效时，同时计算了初始资金为保证金以及初始资金为单位资产价格两种情况。焦炭和焦煤的保证金比例均为 15%。

5.2 参数优化

本节采取 3.3.3 中交叉验证法优化参数，将原始数据分为验证集和测试集。验证集范围为 2013 年 3 月 22 日到 2016 年 12 月 30 日。测试集范围为 2017 年 1 月 3 日到 2018 年 4 月 22 日。本文首先在验证集上进行参数筛选，以第 4 章绩效评估方法和净资产曲线图作为评价依据，选出相对表现最优策略作为最终策略，并在 5.3 节中在测试集上验证策略的泛化能力。



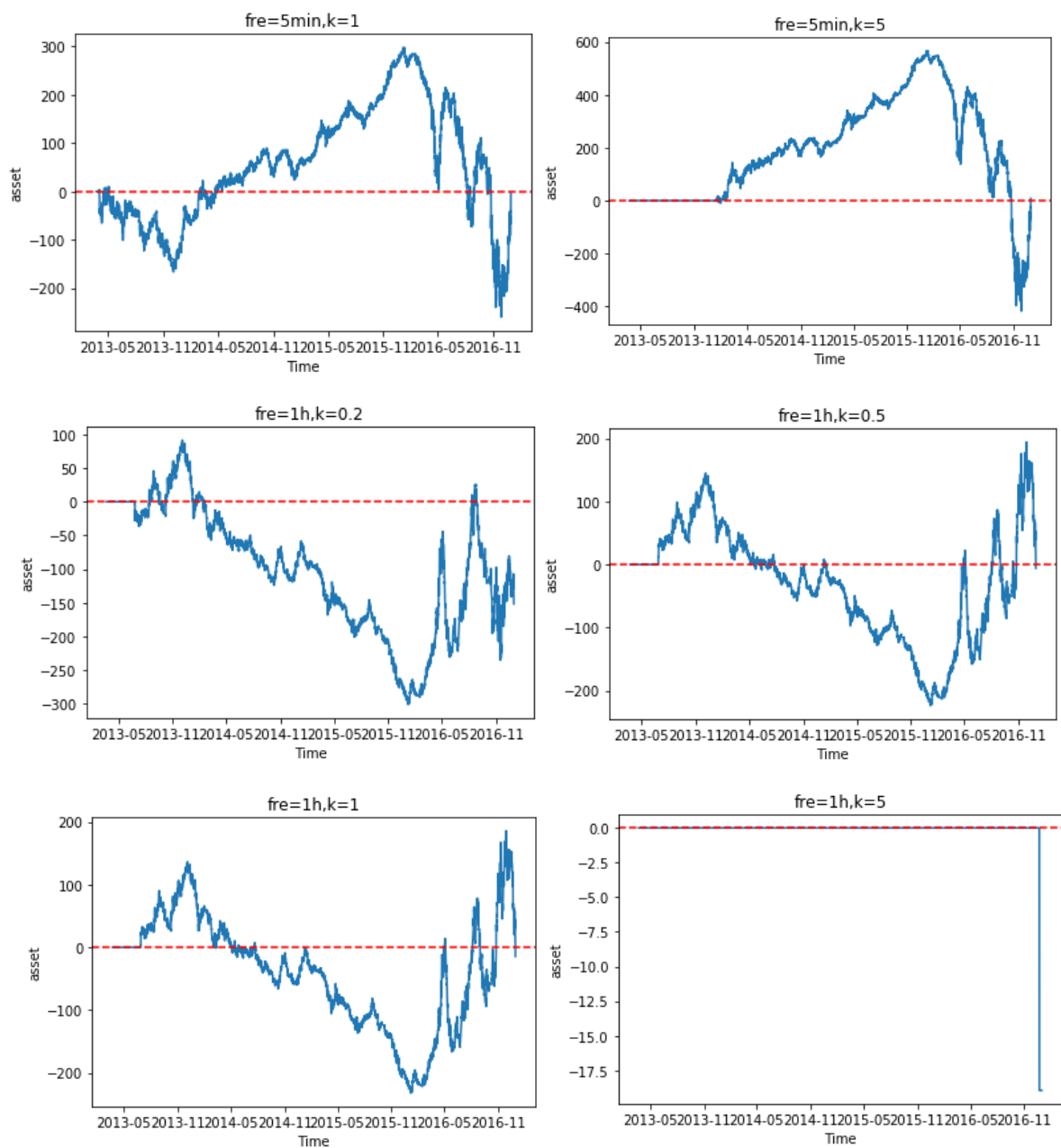


图 5.2 验证集上不同参数组合策略净资产曲线

表 5.1 验证集上不同参数组合策略绩效表现（初始资金为单位资产价格）

数据频率 fre	k	年化收益率 (%)	夏普率	最大回撤率 (%)
30s	0.2	-0.58	-0.01	-20.08
30s	0.5	-9.48	-0.05	-88.71
30s	1	-0.59	-0.01	-20.09
30s	5	-0.07	0.00	-10.83

5min	0.2	-0.06	0.00	-28.09
5min	0.5	-0.06	0.00	-28.09
5min	1	-0.06	0.00	-28.09
5min	5	0.13	0.00	-43.71
1h	0.2	-2.48	-0.14	-22.10
1h	0.5	-0.10	-0.01	-20.16
1h	1	-0.02	-0.01	-20.25
1h	5	-0.30	-0.27	-1.12

表 5.2 验证集上不同参数组合策略绩效表现（初始资金为单位资产保证金）

数据频率 fre	k	年化收益率 (%)	夏普率	最大回撤率 (%)
30s	0.2	-4.05	-0.01	-97.94
30s	0.5	-9.97	-0.01	-276.60
30s	1	-4.10	-0.01	-98.05
30s	5	-0.48	0.00	-51.04
5min	0.2	-0.43	0.00	-101.28
5min	0.5	-0.43	0.00	-101.28
5min	1	-0.43	0.00	-101.28
5min	5	0.86	0.00	-120.11
1h	0.2	-21.69	0.00	-114.11
1h	0.5	-0.65	0.00	-92.75
1h	1	-1.56	-0.01	-94.76
1h	5	-2.06	-0.28	-7.50

图 5.2 展示了数据频率分别为 30s、5min、1h，开仓阈值系数分别为 0.2、0.5、1、5 的策略净资产变化情况。表 5.1 为初始资金为一手交易价值时的策略表现。表 5.2 为初始资金为一手保证金价值时的策略表现。综合来看，策略在验证集上表现很差，但是验证集只是为了筛选可行的参数组合，于是本节着重于选出相对稳定可控的参数组合，以便于未来优化策略。

由图 5.2 可知，当数据频率取 30s 或 1h 时，除了 k=0.5 时出现异常图像外，策略在验证集前段可以盈利，但是中后段波动太大，策略几乎失效。而数据频率为 5min 时，无论 k 取何值，净资产变化趋于一致，且中段 2013 年 11 月到 2015 年 12 月表现良好，这对于一个保持参数近两年不变的策略来说已经非常可观。尽管后段同样遭遇大幅波动下降的问题，但是相比之下更有规则，未来更具有可优化性。同时，5min 频率下的累积收益的最高值均高于 30s 与 1h 频率，策略的盈利能力更优。于是本节基本确定数据频率应选择 5min。

由表 5.1 或表 5.2 可以同样看出 5min 频率下策略的稳定性，且在 5min 频率下取各种不同开仓阈值参数 k 时，仅由 $k=5$ 时策略能得到正收益。回到图 5.1， $k=5$ 时策略似乎回避了前期所有亏损的交易，从约 2014 年才开始交易，可见 $k=5$ 时策略确实得到了优化。且对比 $k=0.2$ 、 0.5 以及 1 时的累积最高收益， $k=5$ 时几乎是它们的两倍，于是本节选择 $k=5$ 作为较优参数。

综上所述，本文选择数据频率 5min，开仓阈值参数 $k=5$ 作为最优模型。5.3 节将展示此模型在测试集上的表现。

5.3 泛化测试

本节利用 5min 频率数据与开仓阈值参数 $k=5$ 的改进策略在测试集上进行交易。交易结果如下：

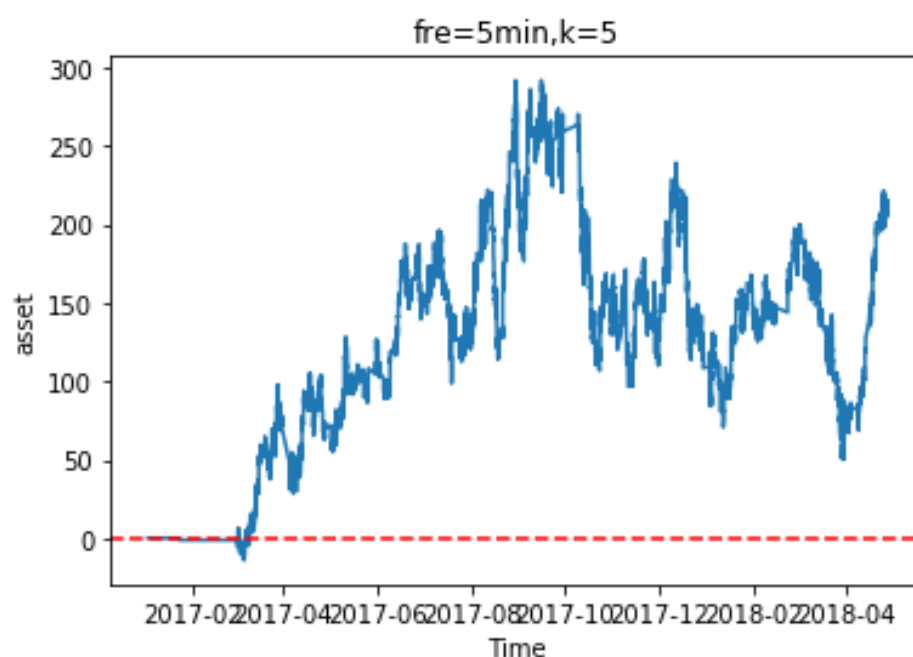


图 5.3 5min 数据频率， $k=5$ 时测试集回测净资产曲线

表 5.3 测试集上 5min 数据频率， $k=5$ 时策略绩效表现（初始资金为单位资产价格）

数据频率 fre	k	年化收益率 (%)	夏普率	最大回撤率 (%)
5min	5	9.00	0.34	-12.22

表 5.3 测试集上 5min 数据频率，k=5 时策略绩效表现（初始资金为单位资产保证金）

数据频率 fre	K	年化收益率 (%)	夏普率	最大回撤率 (%)
5min	5	56.18	0.45	-44.36

由图 5.3 和表 5.3、5.4 可见，优化参数后的策略在测试集上表现有了大幅度提升。若初始资金设为一手资产价值，则年化收益达到 9%，这优于大部分市场上的公募基金产品。若考虑杠杆，初始资金设为一手资产所需保证金，则年化收益更是达到 56.18%，这对于私募基金都是很客观的收益。但是可以发现夏普率都不高，意味着收益率存在很大波动。且最大回撤率也偏高，策略潜在巨额亏损的风险。

尽管如此，一个近一年不调参数，不加止损止盈，没有仓位控制的策略能跑出这样的效果已实属不易，已经可以证明策略的有效性。

5.4 结果分析

虽然改进模型在验证集上表现不佳，但是依然没有妨碍筛选出表现最稳定，最具有改进前景的策略组合。下面给出了一些策略在验证集上表现不佳可能的原因：

- 1、验证集时间跨度将近 4 年，在这 4 年内每一种策略均未曾进行调参、止损止盈、仓位控制等优化，所以表现不佳实为正常。由于 5.2 节的目的只是为了优化参数，所以策略的表现并不能反映策略是否合理、有效。
- 2、由图 5-1，J 和 Jm 对数价差变化频繁，特别在 2016 年前后出现急剧变化，策略反应过慢造成绩效大幅震荡。
- 3、由于交易存在手续费，由于信号本身波动较大，较低的开仓阈值可能造成过于频繁的交易，浪费大量交易费用。由图 5-2 也可以发现，当交易阈值参数越大时，交易结果似乎会更好。

而策略在测试集上能获取正收益，很大部分原因在于，由图 5-1，数据在 2017 到 2018 年波动趋同，价差变化不大，所以策略能套取相对稳定的利润。总体而言，改进模型是有效的，只是缺乏动态调参、止损止盈、仓位管理等优化操作。

第6章 结论与展望

本文基于平稳过程的均值回复性质,改进了以协整模型为基础的CTA配对套利策略。本文通过设置随时间缓慢变化的均值项以及配对系数 β ,实现了模型的动态优化,以适应长期均衡不成立的情况。同时本文利用交叉验证法筛选策略最适合的数据频率以及开仓阈值系数,使得策略对于参数的敏感性下降,表现更加稳定。尽管策略在训练集上表现不佳,但是仍然足够判断合适的参数组合。交易策略在测试集上的表现有了明显的提升,策略具有初步的盈利能力,不考虑杠杆的情况下收益率超过大多数公募基金产品,考虑杠杆的情况下可以媲美表现较好的私募基金。但是策略在风险控制方面有本质性缺陷,收益率波动过大,潜在随时巨大亏损,因此此交易策略还需要进一步改进才能进行实盘交易。以下是可行的进一步研究方向:

- 1、本文尚未对窗口大小 n 进行优化,实际上,当市场行情变化剧烈时,窗口应该适当缩小以应对加剧的价格波动。进一步的研究应当考虑开窗大小的优化。
- 2、策略还未加入止损止盈,通过合理的止损信号可以避免过大的损失,减小最大回撤率。
- 3、策略并未进行仓位控制,每一次触发交易信号都只进行单位资产组合的交易。实际上,如果能够判断交易信号的胜率以及强弱,可以在信号足够强时进行加仓处理,在信号不强时不进行开仓,这样可以增强收益并且降低收益的不确定性。事实上,如今国内外有很多研究机器学习算法在交易择时方面的应用,所以 CTA 策略仓位控制在未来有很大的发展前景,能大大优化策略绩效。
- 4、研究参数的动态调整也可对策略表现起到至关重要的作用。在实盘应用上,大多数量化投资者都会实时盯盘,对于市场出现的宏观事件作出判断同时调整仓位或者设置参数。本文由于以历史数据回测为主,几乎不考虑宏观因素的影响,这在实盘投资中是十分危险的。本文只是提供了一种可行的 CTA 策略,具体的实施还需要实时的调整。

插图索引

图 3.1 训练样本残差序列 $\{\epsilon_t\}$	8
图 3.2 训练集与测试集样本残差序列 $\{\epsilon_t\}$	9
图 5.1 焦炭与焦煤样本对数价格曲线.....	15
图 5.2 验证集上不同参数组合策略净资产曲线	17
图 5.3 5min 数据频率, $k=5$ 时测试集回测净资产曲线	19

表格索引

表 3.1 序列统计检验结果	8
表 5.1 验证集上不同参数组合策略绩效表现（初始资金为单位资产价格）	17
表 5.2 验证集上不同参数组合策略绩效表现（初始资金为单位资产保证金）	18
表 5.3 测试集上 5min 数据频率, k=5 时策略绩效表现（初始资金为单位资产价格）	19
表 5.3 测试集上 5min 数据频率, k=5 时策略绩效表现（初始资金为单位资产保证金）	20

参考文献

- [1] Board J, Sutcliffe C. The dual listing of stock index futures: Arbitrage, spread arbitrage, and currency risk[J]. Journal of Futures Markets, 1993, 16(93-76):29-54.
- [2] Burgess A N. Statistical Arbitrage Models of FTSE 100[J]. 1999.
- [3] Alexander C, Dimitriu A. A Comparison of Cointegration & Tracking Error Models for Mutual Funds & Hedge Funds[J]. Icm Centre Discussion Papers in Finance, 2004.
- [4] 陈实. 基于平稳过程和技术分析的交易策略研究[D]. 华东师范大学, 2017.
- [5] 仇中群. 基于协整的股指期货套利研究[D]. 中国科学技术大学, 2009.
- [6] 金恺. 基于协整方法的沪深 300 成分股配对交易研究[D]. 浙江大学, 2013.
- [7] 张河生, 闻岳春. 基于参数调整的协整配对交易策略: 理论模型及应用[J]. 西部金融, 2013(01):11-16.
- [8] 王峥明. 中国 A 股配对交易策略实证研究[D]. 哈尔滨工业大学, 2010.
- [9] 陈守东, 韩广哲, 荆伟. 主要股票市场指数与我国股票市场指数间的协整分析[J]. 数量经济技术经济研究, 2003(05):124-129.
- [10] 袁象, 王方华, 曹范愚. 协整关系对期货套期保值策略的影响[J]. 数理统计与管理, 2003(02):44-47.

致 谢

感谢唐宏岩老师对本次论文训练的悉心指导，明确了论文训练的目的和答辩的框架，使得本次论文综合训练进展顺利，成果颇丰。

感谢答辩组老师对本次论文训练的意见和建议，使得本次研究更加完善。

同时感谢 TradeBlazer 软件提供的金融数据支持。

声 明

本人郑重声明：所提交的学位论文，是本人在导师指导下，独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。

签名： 李博扬 日期： 2018.6.5

附录 A 外文资料的书面翻译

简单平稳性技术分析在中国指数期货高频数据上的盈利能力

摘要:

技术交易规则已经被金融市场从业人员广泛使用了很长一段时间。它的盈利能力仍然存在争议，并且很少有人考虑交易规则所利用到的技术指标的平稳性。我们将 MA, KDJ 和布林带转变成平稳过程然后利用 2012 年 1 月 4 日到 2016 年 12 月 31 日的 3 种不同频率 (15s, 30s, 60s) 的 CSI300 股指期货数据研究这些交易规则的盈利能力。我们利用交易绩效和风险度量来评估交易规则的实际价值，同时利用 ADF-test 来验证平稳性以及 SPA 检验来弄清交易规则的优良表现是由于本身的优越性还是纯粹的运气。结果表明，在不考虑交易成本得情况下，每个指标都有几个重要的参数组合。当考虑交易成本时，交易利润将完全消失。我们同时也提出了一个方法来降低技术分析的风险。

1. 介绍
2. 技术指标
 - 2.1 传统技术指标
 - 2.1.1 移动平均线

移动平均线是最受欢迎的技术指标之一，它能在短期移动均线穿过长期移动均线时触发交易信号。一般来说，移动均线时一种趋势跟踪指标。当短期移动均线高于（低于）长期移动均线时，价格更可能升高（降低）。记价格序列 $\{P_t\}$ 的 n 期移动平均如下：

$$MA(n, P)_t = \frac{P_t + P_{t-1} + \dots + P_{t-n+1}}{n} = \frac{\sum_{i=t-n+1}^t P_i}{n}$$

若 $n_l > n_s$ ， $MA(n_s, P)_t$ 表示短期移动均线， $MA(n_l, P)_t$ 表示长期移动均线。更具体的说，若 $MA(n_s, P)_t$ 向上（向下）穿过 $MA(n_l, P)_t$ ，将触发一个买入（卖出）信号。然而在短线与长线在短时间内互相频繁穿过时会产生许多虚假信号，这几乎没有带来任何利润并且增加了交易成本。为了避免这种情况，交易者经常设定一个有上下轨道的过滤区域，在这个区域内不会产生任何交易信号。于是基于移动平均和过滤区域的交易规则如下定义：

- 开 多 仓 : $MA(n_s, P)_{t-1} - MA(n_l, P)_{t-1} \leq up - band$ & $MA(n_s, P)_t - MA(n_l, P)_t > up - band$

- 平多仓： $MA(n_s, P)_{t-1} - MA(n_l, P)_{t-1} > up - band$ & $MA(n_s, P)_t - MA(n_l, P)_t \leq up - band$
- 开空仓： $MA(n_s, P)_{t-1} - MA(n_l, P)_{t-1} > low - band$ & $MA(n_s, P)_t - MA(n_l, P)_t \leq low - band$
- 平空仓： $MA(n_s, P)_{t-1} - MA(n_l, P)_{t-1} \leq low - band$ & $MA(n_s, P)_t - MA(n_l, P)_t > low - band$

2.1.2 随机指数

George Lane 在上世纪 50 年代推广了随机振动指标，这是一个利用了支持与阻力水平的动量指标。随机振动技术指标的值由当前价位与一段时间内的价格范围的关系决定。随机指数由两条线表示，其中主线叫做%K。第二条线叫做%D，是%K的迭代指数移动平均。有时交易者也会考虑另外一条叫做%J的线，它是%K与%D的线性组合。由于这三条线的名称，随机指数也被叫做 KDJ。算法如下，记 L_t 和 H_t 为 t 时刻之前包括 t 时刻的 n 期价格的最小和最大值。

$$L_t = \min\{P_t, P_{t-1}, \dots, P_{t-n+1}\},$$

$$H_t = \max\{P_t, P_{t-1}, \dots, P_{t-n+1}\}.$$

然后我们定义 $RSV_t, \%K_t, \%D_t, \%J_t$ ，这里 $m, k < n$ ，通常我们取 $m = 3, k = 3, n = 5, 9, 14$ 。

$$RSV_t = 100 * \frac{P_t - L_t}{H_t - L_t},$$

$$\%K_t = \frac{m-1}{m+1} \%K_{t-1} + \frac{1}{m+1} RSV_t,$$

$$\%D_t = \frac{k-1}{k+1} \%D_{t-1} + \frac{1}{k+1} \%K_t$$

$$\%J_t = 3\%K_t - 2\%D_t.$$

有很多解释随机指数的方法，三种常用方法如下：

- 当振动指标（%K 或者%D）低于某个水平（例如 20）然后又升到该水平以上时买入，当振动指标高于某个水平（例如 80）然后又降至该水平以下时卖出。
- 当%K 线上穿%D 线时买入，下穿时卖出。
- 寻找分歧。例如：价格屡创新高但是随机指标没有超过先前的高点。

2.1.3 布林带

布林带是由 John Bollinger 在上世纪 80 年代发明的一种技术分析工具，它由一条上轨一条下轨组成。

记 n 期指数平均如下:

$$\text{EMA}(n, P)_t = \frac{nP_t + (n-1)P_{t-1} + \cdots + P_{t-n+1}}{n + (n-1) + \cdots + 1} = \frac{2}{n(n+1)} \sum_{i=t-n+1}^t (n-t+i)P_i$$

考虑 $\text{EMA}(n, P)_t$ 为布林带的中线, 记上下轨分别为 $\text{EMA}(n, P)_t + K\sigma_t$ 和 $\text{EMA}(n, P)_t - K\sigma_t$, 这里 K 为参数, σ 为价格和 $\text{EMA}(n, P)_t$ 的标准差, 定义如下:

$$\sigma(n, P)_t = \sqrt{\frac{1}{n} \sum_{i=t-n+1}^t (P_i - \text{EMA}(n, P)_t)^2}$$

当价格高于上轨, 我们认为市场处于牛市。相反的, 当价格低于下轨, 我们认为市场为熊市。因此, 当价格上穿上轨时触发买入信号, 当价格下穿下轨时触发卖出信号。也就是说, 当价格处于上下轨之间时不触发任何信号。

2.2 平稳技术指标

2.2.1 平稳过程简介

定义 2.1. 随机过程 $\{X_t\}$ 为平稳过程, 若对任意 $a > 0$ 和任何有限时刻 $t_1 < t_2 \dots < t_n$, 随机向量 $\{X_{t_1}, X_{t_2}, \dots, X_{t_n}\}$ 与 $\{X_{t_1+a}, X_{t_2+a}, \dots, X_{t_n+a}\}$ 有相同的概率分布。

由定义 2.1. 可知, 平稳过程的分布不会随时间改变。很容易看到, 若 $\{X_t\}$ 是平稳过程, $f(x)$ 是一个使得 $\{f(X_t)\}$ 为随机过程的函数, 则 $\{f(X_t)\}$ 为平稳过程。

平稳过程的定义很完美且太难满足。我们很难检验一个随机过程是否平稳。在实际应用中, 人们更多关心的是随机过程的均值和方差。在这个意义下, 弱平稳过程更加实用。

定义 2.2. 弱平稳过程满足一下两个条件: 对每个 $c > 0$,

- $E[X_t] = E[X_{t+c}]$;
- $E[X_s X_t] = E[X_{s+c} X_{t+c}]$.

很容易看到均值与方差存在的强平稳过程一定是弱平稳过程。

2.2.2 对数收益是平稳的

记 P_t 为某金融标的在 t 时刻的价格, p_t 为对数价格, 即 $p_t = \log P_t$. 对任何正数 a , $\Delta p_t = p_t - p_{t-a}$ 为 a 期的对数收益。

在参考文献[18]中, 作者指出我们不能拒绝零假设: 对数收益时间序列在统计检验下是弱平稳的。尽管真正的假设是对数收益时间序列是强平稳的, 但是理论与统计检验的差距在应用中并不很重要。所以我们承认这一结论。相应的平稳性检验会在 4.3.1 提到。

在对数收益 $\{\Delta p_t\}$ 平稳的前提下，按照参考文献[18]的表述，当 N 足够大时，我们可以假设在时间区间 $[\frac{j}{N}, \frac{j+1}{N}]$ 内有至多不超过一次交易。

$$\left(\frac{P_{\frac{j+1}{N}}}{P_{\frac{j}{N}}}\right)^{I_{\frac{j}{N}}} = \exp\left\{I_{\frac{j}{N}}\left(p_{\frac{j+1}{N}} - p_{\frac{j}{N}}\right)\right\}$$

其中 I_n 为 n 时刻的仓位。 $I = 1$ 表示多头， $I = -1$ 表示空头， $I = 0$ 表示平仓。因此 T 时刻之前实现的总对数收益（不考虑交易成本）为

$$\sum_{j < NT-1} I_{\frac{j}{N}} \left(p_{\frac{j+1}{N}} - p_{\frac{j}{N}}\right)$$

在中国金融期货交易所中，一些产品的交易成本只占交易总价值很小的比例。假设单边交易成本百分比为 c ，于是 T 时刻前总实现的对数收益为

$$\sum_{j < NT-1} \left[I_{\frac{j}{N}} \left(p_{\frac{j+1}{N}} - p_{\frac{j}{N}}\right) + \log \frac{1-c}{1+c} * \left| I_{\frac{j+1}{N}} - I_{\frac{j}{N}} \right| \right]$$

当 $N \rightarrow \infty$ 时，上述求和收敛到

$$\int_0^T I_s dp_s + \log \frac{1-c}{1+c} K_T = \sum_{i: T(i) < T} (p_{T_i} - p_{S_i})(I_{T_i} - I_{S_i}) + \log \frac{1-c}{1+c} K_T \quad (1)$$

K_T 为 T 时刻之前的交易次数， S_i 为第 i 次交易开始的时刻， T_i 为第 i 次交易结束的时刻。当 $\{(I_t, \Delta p_t)\}$ 平稳时，方程(1)中每一项也都平稳。因此由强遍历性定理，平均对数收益收敛到

$$\frac{1}{T} \left[\sum_{i: T(i) < T} (p_{T_i} - p_{S_i})(I_{T_i} - I_{S_i}) + \log \frac{1-c}{1+c} K_T \right]$$

2.2.3 平稳移动均线

根据参考文献[18]， P_t 与 MA_t 都不平稳。记 n 期平稳移动平均如下

$$\begin{aligned} SMA(n, P)_t &= \frac{MA(n, P)_t}{P_t} \\ &= \frac{1 + \frac{P_{t-1}}{P_t} + \frac{P_{t-2}}{P_t} + \dots + \frac{P_{t-n+1}}{P_t}}{n} \\ &= \frac{1 + e^{p_{t-1}-p_t} + e^{p_{t-2}-p_t} + \dots + e^{p_{t-n+1}-p_t}}{n} \end{aligned}$$

N 期平稳移动平均是对数收益 $\{\Delta p_t\}$ 的函数。 $SMA(n_s, P)_t$ 与 $SMA(n_l, P)_t$ 的比值，

记作 $R(n_s, n_l, P)_t$ ，也是 $\{\Delta p_t\}$ 的函数。当 $\{\Delta p_t\}$ 平稳时， $SMA(n, P)_t$ 和这一比值也平稳。

$$R(n_s, n_l, P)_t = \frac{SMA(n_s, P)_t}{SMA(n_l, P)_t} = \frac{MA(n_s, P)_t}{MA(n_l, P)_t}$$

于是 $MA(n_s, P)_t$ 上穿（下穿） $MA(n_l, P)_t$ 相当于 $R(n_s, n_l, P)_t$ 上穿（下穿）1。由2.1.1，我们同样设定一个有上下轨的过滤区域，在其中不会有任何交易信号被触发。于是SMA的交易规则和过滤区域如下定义

- 开多仓： $R(n_s, n_l, P)_{t-1} \leq 1 + up - band$ & $R(n_s, n_l, P)_t > 1 + up - band$
- 平多仓： $R(n_s, n_l, P)_{t-1} > 1 + up - band$ & $R(n_s, n_l, P)_t \leq 1 + up - band$
- 开空仓： $R(n_s, n_l, P)_{t-1} \geq 1 + up - band$ & $R(n_s, n_l, P)_t < 1 + up - band$
- 平空仓： $R(n_s, n_l, P)_{t-1} < 1 + up - band$ & $R(n_s, n_l, P)_t \geq 1 + up - band$

令下轨为上轨的相反数，于是它们就可以被一个参数 b 表示。因此，我们设定了参数 (n_s, n_l, b) 。其中 n_s 一般取为1, 5, 10, 15。 n_l 可取为20, 30, 60, 120。 b 在 $0.1 * 10^{-3}, 0.5 * 10^{-3}, 1 * 10^{-3}, 1.5 * 10^{-3}$ 中取值。这些参数的选择得到了64种平稳移动均线规则。因为我们需要测试3种不同频率的数据，所以总共有192种平稳移动均线策略。

2.2.4 平稳 KDJ

由KDJ的定义，

$$RSV_t = 100 * \frac{e^{p_t} - e^{\min\{p_t, p_{t-1}, \dots, p_{t-n+1}\}}}{e^{\max\{p_t, p_{t-1}, \dots, p_{t-n+1}\}} - e^{\min\{p_t, p_{t-1}, \dots, p_{t-n+1}\}}}$$

$$= 100 * \frac{e^{p_t - p_{t-n}} - e^{\min\{p_t - p_{t-n}, p_{t-1} - p_{t-n}, \dots, p_{t-n+1} - p_{t-n}\}}}{e^{\max\{p_t - p_{t-n}, p_{t-1} - p_{t-n}, \dots, p_{t-n+1} - p_{t-n}\}} - e^{\min\{p_t - p_{t-n}, p_{t-1} - p_{t-n}, \dots, p_{t-n+1} - p_{t-n}\}}}$$

当 $\{\Delta p_t\}$ 平稳时， $\{RSV_t\}$ 也平稳。 $\%K$ 是 RSV 的函数，而 $\%D$ 是 $\%K$ 的函数，因此若 $\{\Delta p_t\}$ 平稳，则 $\{\%K_t\}$ 与 $\{\%D_t\}$ 均平稳。

对于交易策略，我们选择如下在交易者中很受欢迎的策略

- 开多仓： $\%K_{t-1} < \%D_{t-1}$ & $\%K_t \geq \%D_t$ & $20 \leq \%K_t \leq 80$;
- 平多仓： $\%K_{t-1} > \%D_{t-1}$ & $\%K_t \leq \%D_t$;
- 开空仓： $\%K_{t-1} > \%D_{t-1}$ & $\%K_t \leq \%D_t$ & $20 \leq \%K_t \leq 80$;
- 平空仓： $\%K_{t-1} < \%D_{t-1}$ & $\%K_t \geq \%D_t$;

KDJ(m, n, k)有3个参数。对每组高频数据，我们从(5, 1, 3), (5, 3, 3), (9, 3, 3), (14, 3, 3)和(19, 3, 3)中选择参数，产生15种KDJ策略。

2.2.5 平稳布林带

显然，原始的布林带是不平稳的。我们可以把布林带转化成如下形式：

$$\begin{aligned}
 S \text{ Boll}(n, P)_t &= \frac{P_t - EMA(n, P)_t}{\sigma_t} \\
 P_t - \frac{nP_t + (n-1)P_{t-1} + \dots + P_{t-n+1}}{\frac{n(n+1)}{2}} \\
 &= \frac{e^{p_t} - \frac{ne^{p_t} + (n-1)e^{p_{t-1}} + \dots + e^{p_{t-n+1}}}{\frac{n(n+1)}{2}}}{\sqrt{\frac{1}{n-1} \sum_{i=t-n+1}^t (P_i - EMA(n, P)_t)^2}} \\
 &= \frac{e^{p_t} - \frac{ne^{p_t} + (n-1)e^{p_{t-1}} + \dots + e^{p_{t-n+1}}}{\frac{n(n+1)}{2}}}{\sqrt{\frac{1}{n-1} \left[e^{p_t} - \frac{2}{n(n+1)} (ne^{p_t} + \dots + e^{p_{t-n+1}}) \right]^2 + \dots + \frac{1}{n-1} \left[e^{p_{t-n+1}} - \frac{2}{n(n+1)} (ne^{p_t} + \dots + e^{p_{t-n+1}}) \right]^2}} \\
 &= \frac{e^{p_t - p_{t-n}} - \frac{2}{n(n+1)} (ne^{p_t - p_{t-n}} + (n-1)e^{p_{t-1} - p_{t-n}} + \dots + e^{p_{t-n+1} - p_{t-n}})}{\sqrt{\frac{1}{n-1} \sum_{i=t-n+1}^t \left[e^{p_t - p_{t-n}} - \frac{2}{n(n+1)} (ne^{p_t - p_{t-n}} + \dots + e^{p_{t-n+1} - p_{t-n}}) \right]^2}} \quad (2)
 \end{aligned}$$

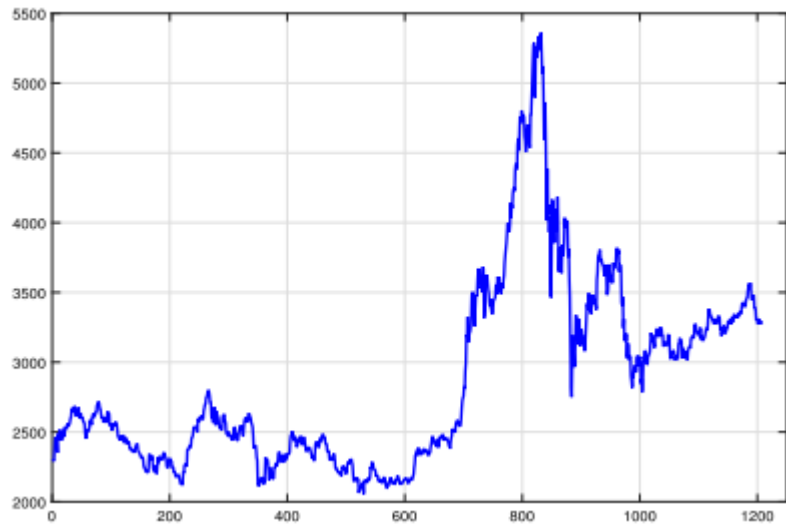


图 1. 2012. 1. 4-2016. 12. 30 IF 主力合约每日数据

$S \text{ Boll}(n, P)_t$ 是 $\{e^{p_{t-i} - p_{t-n}}; i = 1, 2, \dots, n-1\}$ 的函数。当 $\{p_t\}$ 是强平稳增长时， $\{S \text{ Boll}(n, P)_t\}$ 是平稳过程。布林带之间的价格满足不等式 $-K < S \text{ Boll}(n, P)_t < K$ 。

于是交易信号由以下规则触发：

- 开多仓： $S \text{ Boll}(n, P)_{t-1} \leq K$ & $S \text{ Boll}(n, P)_t > K$

- 平多仓: $S\text{ Boll}(n, P)_{t-1} \geq K \ \& \ S\text{ Boll}(n, P)_t < K$
- 开空仓: $S\text{ Boll}(n, P)_{t-1} \geq -K \ \& \ S\text{ Boll}(n, P)_t < -K$
- 平空仓: $S\text{ Boll}(n, P)_{t-1} \leq -K \ \& \ S\text{ Boll}(n, P)_t > -K$

由平稳布林带策略生成的交易信号与两个参数有关。一个与 $\text{EMA}(n, P)_t$ 有关, 另一个为 σ 系数。n 的取值可以为 20, 30, 60 和 120, K 可以取为 0.1, 0.5, 1, 1.5, 2 和 2.5。因此, 3 组高频数据共有 72 种平稳布林带策略。

以上所有高频平稳策略每单位时间只交易最多一手, 每个交易日结束后强制平仓。

3. 数据描述

我们在 CSI300 (中国证券指数 300) 股指期货主力合约上应用上述策略中的交易规则来探究哪个技术策略有最好的表现。我们选择这一投资标的的原因如下。

CSI300 股指期货是中国金融期货交易所于 2010 年 4 月 16 日推出的中国金融市场第一支股指期货。CSI300 包括 300 支上海和深圳证券交易所股票。这些样本覆盖了大约 60% 上海和深圳市值, 表现出了很好的代表性。因此, CSI300 股指期货有很好的市场流动性, 反映了市场的总体趋势。选择 CSI300 的另一个好处是, 尽管中国 A 股市场禁止卖空, 但在期货市场是允许的, 这为我们的研究带来了便利和完整性。

为了使我们的研究更全面, 我们选择了 3 种不同的高频数据(15s, 30s 和 60s) 来测试所有交易规则。CSI300 的所有 3 种高频数据涵盖的时间为 2012 年 1 月 4 日到 2016 年 12 月 30 日, 分别有 1276799 (15s)、638399 (30s)、319199 (60s) 个数据点。因为数据量太庞大, 人们很难再一张图中区分高频数据和每日数据。因此, 图 1 展示了每日数据的趋势。需要提到的一点是 2016 年 1 月 1 日之前 CSI300 股指期货交易时间为每个交易日的 9:15 到 11:30, 13:00 到 15:15。但是从 2016 年 1 月 1 日开始时间变为 9:30, 结束时间变为 15:00。我们用到的数据符合这一规则。

我们分别估计 3 种高频数据的对数收益, 即对数差收益, 然后找到 15s、30s、60s 数据中最大的对数收益分别为 8.02%、7.84%、7.71%, 相应的最小对数收益分别为 -6.35%、-5.70%、-5.84%。表 1 列出了这三种高频数据对数收益的基本统计量。当与正态分布比较时, 三种收益分布有更高的峰度和右偏态。

频率	样本数	最大值	最小值	均值	标准差	峰度	偏度
15s	1276799	8.0172%	-6.3485%	2.5267E-07	0.0006	986.0586	0.7322
30s	638399	7.8433%	-5.7044%	5.0467E-07	0.0008	512.5485	1.3490

60s	319199	7.1733%	-5.8385%	1.0122E-06	0.0012	208.6592	0.9192
-----	--------	---------	----------	------------	--------	----------	--------

表 1. CSI300 股指期货统计描述

4. 方法论

4.1 绩效度量

4.1.1 绩效与年化绩效

策略的绩效定义为在最后给定的时刻，即回测期间，策略基于初始资金产生的净盈利或者净损失。不同的策略可能有不同的回测期间，为了便于比较不同的策略，年化绩效是一个更好的选择。年化绩效计算了一年的绩效，等于绩效乘以 250（一年的交易日天数）然后除以回测期间的天数。

4.1.2 PnL 盈亏

盈亏指数就，简称 PnL，表示净收益与盈利或亏损交易数之比。如下为定义公式。

$$\text{PnL} = \frac{\text{TradeProfit} - \text{TradeLoss}}{\text{Max}(\text{TradeProfit}, \text{TradeLoss})}$$

PnL 取值范围为-100 到 100. 正值代表策略总体带来正的收益，负值表示带来净亏损。

4.1.3 平均利润/平均亏损

这一指标是盈利交易的平均利润与未盈利交易的平均亏损之比。若一个策略时好策略，它会让能盈利的交易发生，让亏损尽快停止。这就是说，如果一个策略在上述意义下是好的，那么这一比值会比较高。

考虑到横向比较的便利性，我们在文章中将通过每日收益计算平均利润与亏损。

4.1.4 盈利交易占比

这个指标给出了策略的胜率，表示可盈利的交易次数占总交易次数的比例。高的比值表明策略策略有高的概率能正确预测价格变动的方向。

一个盈利的策略不一定有高的胜率。比如，如果一共交易 4 次，一次交易盈利 5 个点，三次交易亏损 1 个点，总胜率只有 25%但是总利润有 2 个点。

4.2 风险度量

4.2.1 夏普率

夏普率是一种估计每单位风险所能获取的超额收益的指标。超额收益是策略收益与无风险收益率之差。风险由策略收益的标准差度量。详细公式如下。

$$\text{Sharp ratio} = \frac{E(r) - r_f}{\sqrt{\text{Var}(r)}}$$

其中 r_f 为无风险收益率， r 为目标策略的收益。

值得一提的是，我们计算每日的夏普率然后转换成年化夏普率来度量我们的策略。因为高频夏普率在比较中不具备足够代表性，然而计算年化夏普率需要 5 年以上的数据才有意义。 r_f 在每个交易日结束强制清算计算夏普率时设置为 0。转换公式如下。

$$\text{Annual Sharp ratio} = \text{Daily Sharp ratio} * \sqrt{250}$$

其中 250 表示一年中平均交易天数。

4.2.2 最大回撤率

最大回撤率是另一个度量风险的常用指标。它表示在一段给定时期内策略最坏的表现。假设有 n 个交易日。记 V_i 为第 i 天的资产净值， $W_i = \max_{j=1,2,\dots,i} \{V_j\}$ 。

$$\text{最大回撤率} = \max_{i=1,2,\dots,n} \left\{ 1 - \frac{V_i}{W_i} \right\}$$

这一指标越小，策略的鲁棒性越好。同样我们也是基于每日利润考虑最大回撤率。

4.2.3 年化收益/最大回撤率

正如指标名的含义，这一指标是年化收益与最大回撤率的比值。这一指标是衡量策略稳定性的常用的且重要的综合标准。

4.3 检验统计量

4.3.1 ADF 检验

ADF 检验是增强 Dickey-Fuller 检验的简写。尽管这一检验只能测试一个单位根过程，而不是一个弱平稳过程。这两者的相似之处让 ADF 检验能够检验弱平稳过程。有很多关于 ADF 检验的论文已经发表。Dickey-Fuller 检验是一个用于评估经济与金融数据时间序列性质的常用单位根检验。MacKinnon, Harris 和 Cheung&Lai 关注的是检验参数的使用，其中说明了滞后顺序和样本大小都可能影响有限样本的测试表现。

单位根的 ADF 检验利用模型 $y_t = c + \delta t + \phi y_{t-1} + \beta_1 \Delta y_{t-1} + \beta_2 \Delta y_{t-2} + \dots + \beta_p \Delta y_{t-p} + \varepsilon_t$ 来验证零假设。这里 Δ 为差分算子， p 为滞后差分数， ε_t 为 0 均值更新过程。单位根的零假设为 $H_0: \phi = 1$ ；在备择假设下， $\phi < 1$ 。模型变量允许有不同的增长特征。 $\delta = 0$ 的模型没有趋势项， $\delta = 0, c = 0$ 的模型没有趋势项和漂移。

统计学家采用的逻辑为：如果你不能证明当前的样本轨迹不太可能由一个弱平稳过程生成，那么我们简单的保持弱平稳的假设。

我们将采取 ADF 检验来验证高频数据的对数收益的平稳性。

4.3.2 改进预测能力检验

正如 White[21]提到的，当一个给定的数据集被重复利用多次，一些碰运气的策略可能很重要，而不是任何内在的优点。这种现象被称为数据窥探偏见。White[21]创建了现实检查（WRC）测试来修正数据窥探效果。Hansen[25]提出了改进的预测能力（SPA）检验，它由 WRC 改进而来，在大多数情况下更加有效。因此，我们选择 SPA 检验来检查数据窥探偏见。

假设有 K 种不同的策略和 T 个交易日，记 $\overline{d_{k,t}} = \sum_{t=1}^T d_{k,t}$ 为策略 k 的绩效，这里 $d_{k,t} (k \in 1, 2, \dots, K, t \in 1, 2, \dots, T)$ 表示策略 k 在第 t 天的绩效度量。假设在第 t 天有 N 个时间点，单边交易成本为 c ，于是绩效度量 $d_{k,t}$ 可写成如下形式：

$$\begin{aligned} d_{k,t} &= \sum_{n=1}^{N-1} \left(\log \frac{P_{n+1}}{P_n} * I_n + \log \frac{1-c}{1+c} * |I_{n+1} - I_n| \right) \\ &= \sum_{n=1}^{N-1} \left(\Delta p_{n+1} * I_n + \log \frac{1-c}{1+c} * |I_{n+1} - I_n| \right) \end{aligned}$$

这里 I_n 为 n 时刻的仓位。 $I = 1$ 表示多仓。 $I = -1$ 表示空仓。 $I = 0$ 表示平仓。

记 $\mu_k = E(d_{k,t})$. SPA 检验的零假设如下：

$$H_0: \mu_k \leq 0; k \in 1, 2, \dots, K$$

Politis 和 Romano[33]提出了一个平稳的辅助程序。我们按照这个方法来得得到 SPA 检验的结果。首先，生成一个 $B \times T$ 的随机矩阵 $\{R_{b,t}\}$ 。对每一行，例如第 b 行， $P^*(R_{b,1} = t, t = 1, 2, \dots, T) = 1/T$ 。当 $t > 1$ 时， $P^*(R_{b,t} = R_{b,t-1} + 1) = Q$ 且 $P^*(R_{b,t} = s, s = 1, 2, \dots, T) = (1 - Q)/T$ 。

然后，记 $\{d_{k,t}^*(b) := d_{k,R_{b,t}}^*, t = 1, 2, \dots, T\}$ 为第 b 次重新取样。第 b 次重新取样的绩效为 $\overline{d_k^*(b)} = \sum_{t=1}^T d_{k,t}^*(b) / T$ 。记 P^* 为辅助概率测度，SPA 检验在显著性水平为 α 的临界值可如下描述：

$$\widehat{q}_\alpha^* = \max(0, \widehat{q}_\alpha), \widehat{q}_\alpha = \inf\{q | P^*[\sqrt{n} \max_{k=1,2,\dots,K} \frac{\overline{d_k^*} - \overline{d_k} + \widehat{\mu}_k}{\widehat{w}_k} \leq q] \geq 1 - \alpha\}$$

其中， $\widehat{\mu}_k = \overline{d_k} \chi\{\sqrt{nd_k} \leq -\widehat{w}_k \sqrt{2 \log \log n}\}$ ， \widehat{w}_k 表示 $w_k := \text{var}(\sqrt{nd_k})$ 的一致估计量。由 Hansen[25] 的介绍，本文使用 \widehat{w}_k 的核估计量。当 $\max_{k=1,2,\dots,K} \sqrt{nd_k} / \widehat{w}_k > \widehat{q}_\alpha^*$ 时拒绝零假设。

进一步来说，Hsu 等人[26]拓展了 SPA 检验到了 Step-SPA 检验，让研究人员能够识别重要的模型。Step-SPA 检验可由以下步骤实现：

- 第一步：将 $\overline{d_k}/\widehat{w_k}$ 以倒序排列。
- 第二步：假设第 k 个模型有最大的 $\overline{d_k}/\widehat{w_k}$ 。当 $\sqrt{nd_k}/\widehat{w_k} > \widehat{q}_\alpha^*$ 时停止，否则跳到第三步。
- 第三步：从总的模型中移除第二步中的第 k 个模型，然后继续对剩下的模型执行第一步和第二步。
- 第四步：重复第三步知道没有模型被拒绝。所有被移除的模型被认为是显著的模型。

由 Hsu 等人[26]，我们取 $Q=0.9$, $B=500$, 显著性水平设为 5% 和 10%。

5. 实证结果

5.1 绩效

我们设置初始资金为 100 万，一次最多交易一手。我们检验平稳移动均线在 3 组高频数据上的 64 组参数，即共有 192 种不同策略。如果短线与长线的参数固定，当过滤带的宽度增大时，交易的次数有减少的趋势。对每组高频数据，超过五分之四的策略有正的年化收益，具体的数字是 48/64, 52/64, 56/64. 然而，夏普率超过 1.5 的策略并不是很多，总共只有 30 个。需要注意的另外一点是，只有 16 个策略胜率超过 50%。平稳移动均线策略是一个趋势跟踪策略，导致胜率比较低。对趋势方向的错误判断或者反复震荡可能导致一些损失。但是捕捉到大趋势将会缓解这一情况甚至带来一些利润。

KDJ 常用的参数并没有移动均线那么多。这里有 15 个 KDJ 策略。15s 的 KDJ 策略表现不好，没有能盈利的。5 个 30s 策略中的 3 个 and 所有 5 个 60s 策略都有正的年化收益。

对于平稳布林带策略，72 个中的 24 个有盈利，而且最好的一个达到了 59.31% 一年。但是只有 3 个策略有超过 1.5 的夏普率，只有 1 个策略胜率超过 50%。不够高的夏普率表明稳定性应该要加强。布林带生成的交易规则属于通道突破策略。类似于趋势跟踪策略，它的盈利取决于高利润率而不是高胜率。请注意，279 种策略的平均利润与平均亏损之比都大于 1。

Strategy	LDT ^a	SDT ^a	ASP ^b	ADP ^b	AR ^c	MDP ^c	AR/MDP ^c	SR ^d	PnL ^d	WR ^e	AP/AL ^e
MA_15(10,120,0.0001)	10353	10384	113.90	1953.70	48.84%	6.33%	7.71	2.17	0.36	51.94%	1.45
KDJ_60(5,1,3)	20035	19648	41.31	1355.88	33.90%	11.68%	2.90	2.24	0.38	51.94%	1.49
Boll_30(120,0.1)	22234	22325	64.37	2372.41	59.31%	6.64%	8.93	2.33	0.40	50.79%	1.62

- a. LDT 为长时间交易，SDT 为短时间交易
- b. ASP 表示平均单利润，ADP 表示平均日利润
- c. AR 表示年化利润，MDP 表示最大回撤率
- d. SR 表示夏普率指标，PnL 表示利润与损失指标
- e. WR 表示胜率，AP/AL 表示平均利润与平均损失之比

表 2. 三种策略绩效

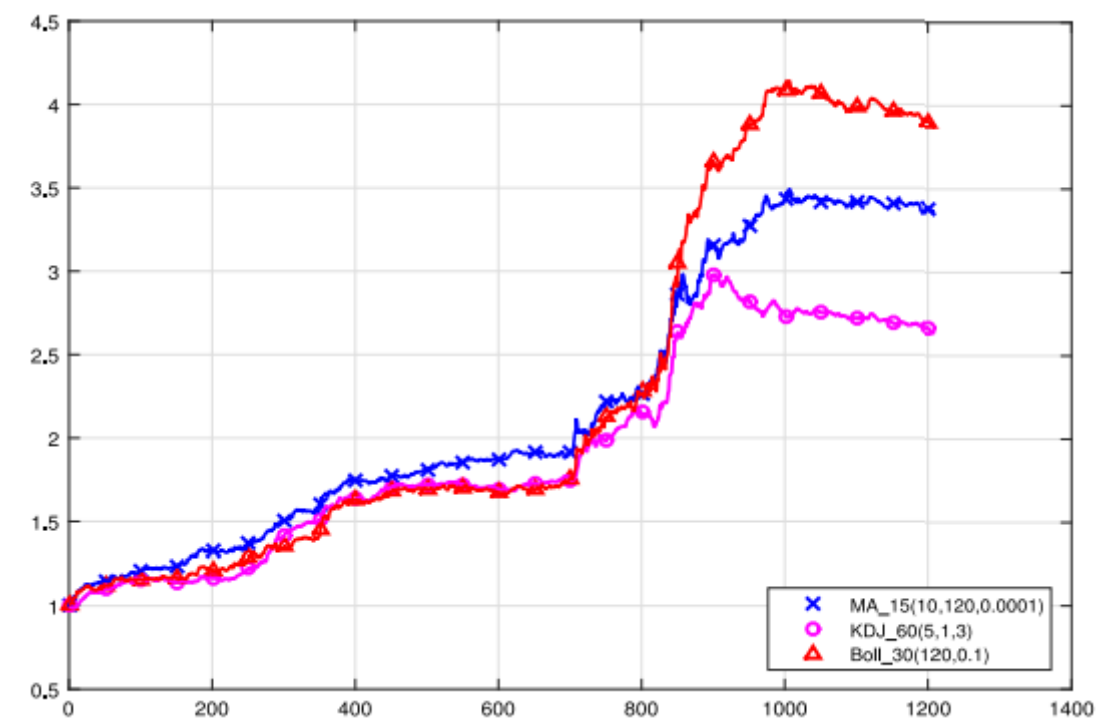


图 2. 2012. 1. 4-2016. 12. 30 三种策略净值

Cost	0.5%%	0.35%%	0.25%%	0.23%%	1.15%%	23%%
------	-------	--------	--------	--------	--------	------

表 3. 2012. 1. 4-2016. 12. 30 交易成本

每种技术分析策略中最赚钱的策略在表 2 中单独列出，图 2 展示了 3 种策略的净值曲线。为了总体的简洁，具体的 279 中策略结果在附录的参考文献[34]的表 11-19 展示。

以上显示了交易成本未被考虑的情况。一旦成本被包括在内，没有任何策略可以带来利润。有两个可能的原因。如表 3 所示，交易成本在 2015 年下半年被调高，特别是 2015 年 9 月 7 日的调整导致了新的成本比 2012 下半年到 2015 上半年高了百倍。这一变化不仅提高了成本，同时也使市场特征产生动荡。好消息是

交易成本已经被调整至 9%，这给投资者未来将会更低的希望。另一个解释这一结果的原因是市场有效性。有一个大胆的假设，交易成本可能具有确保市场效率的功能。值得注意的是，在这篇论文中，我们仅仅讨论了简单平稳技术交易规则。它们是基本的规则且有进一步发展的潜力。只是就像参考文献[17, 19, 20]给出的策略，有一些可以覆盖成本并盈利的方法。但是这不是本文的领域。

5.2 ADF 检验

首先，我们检验原始时间序列对数收益的平稳性，即 $\{\Delta p_t\}$ 。表 4-6 简化后的 15s, 30s 和 60s 数据的测试结果。不管采用的滞后差分项数，最后的测试结果仍然是“1”，意味着对数收益是一个弱平稳过程。另一方面，在 Black-Scholes 框架 [18] 中，股价可以写成 $P_t = P_0 \exp\{\sigma B_t + (r - \frac{\sigma^2}{2})t\}$ ；因此， $\Delta p_t = \sigma(B_t - B_{t-1}) + (r - \frac{\sigma^2}{2})$ ，这里 $B_t - B_{t-1}$ 是两个独立布朗运动之差，因此它是平稳的。总的来说，根据经典金融模型和样本数据验证(ADF 检验)， $\{\Delta p_t\}$ 是平稳过程。图 3(a) (b) 和 (c) 展示了 15s, 30s 和 60s 数据相应的对数收益。

其次，我们验证三个平稳指标的平稳性，发现所有的都拒绝了零假设。换句话说，我们不能拒绝这些指标是平稳的。简单起见，这里表 7 只展示 15s 数据的 S Boll(120,2.5)结果。图 4 表示 IF1611 的 15s 数据在 2016. 10. 20-2016. 11. 16 中 S Boll(120,2.5)和 Boll(120,2.5)的形状。由图 4(a)，原始布林带有明显的趋势。但是，图 4(b)中平稳布林带在 0 附近震荡。

表 4. 15s ADF 检验结果

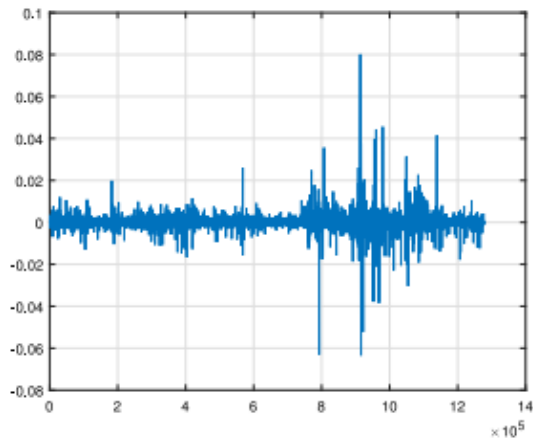
	test - t1, lags - 0	test - t1, lags - 1	test - t1, lags - 2
coeff	-0.0115	[-0.0286; 0.0170]	[-0.04; 0.03; 0.01]
tStats	-12.9587	[-22.7347; 19.1567]	[-0.0414; 0.0296; 0.0125]
FStat	Inf	535.2234	366.8779
AIC	-1.5532e + 07	-1.5532e + 07	-1.5532e + 07
BIC	-1.5531e + 07	-1.5532e + 07	-1.5532e + 07
p-value	1.00e - 03	1.00e - 03	1.00e - 03
H	1	1	1

表 5. 30s ADF 检验结果

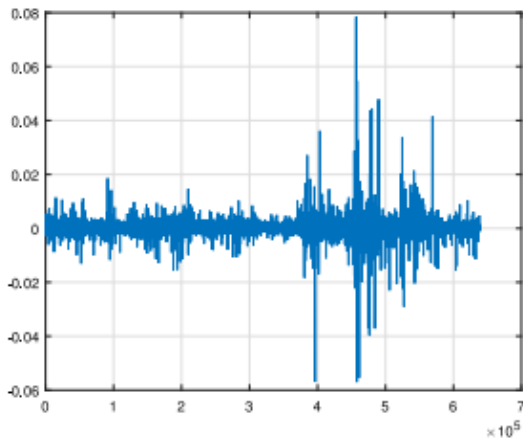
	test - t1, lags - 0	test - t1, lags - 1	test - t1, lags - 2
coeff	-0.0310	[-0.0134; -0.0171]	[-0.0198; -0.0106; 0.0063]
tStats	-24.8163	[-7.4790; -13.6445]	[-8.9914; -5.9158; 4.9963]
FStat	Inf	802.4794	413.7396
AIC	-7.3326e + 06	-7.3327e + 06	-7.3328e + 06
BIC	-7.3326e + 06	-7.3327e + 06	-7.3327e + 06
p-value	1.00e - 03	1.00e - 03	1.00e - 03
H	1	1	1

表 6. 60s ADF 检验结果

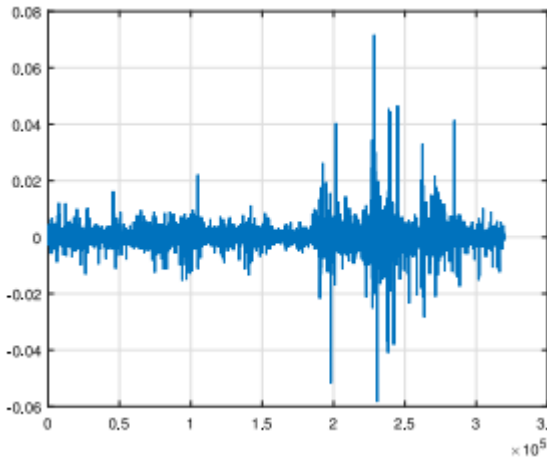
	<i>test - t1, lags - 0</i>	<i>test - t1, lags - 1</i>	<i>test - t1, lags - 2</i>
coeff	0.0047	[0.0134; -0.0088]	[0.0090; -0.0044; 0.0044]
tStats	2.6326	[5.3600; -4.9529]	[2.9693; -1.7582; 2.4835]
FStat	<i>Inf</i>	31.7311	18.9508
AIC	-3.4501e + 06	-3.4501e + 06	-3.4501e + 06
BIC	-3.4501e + 06	-3.4501e + 06	-3.4501e + 06
p-value	1.00e - 03	1.00e - 03	1.00e - 03
H	1	1	1



(a) 15s 数据对数收益



(b) 30s 数据对数收益



(c) 60s 数据对数收益

图 3. IF 主力合约在 2012. 1. 4-2016. 12. 30 的对数收益

	<i>test - t1, lags - 0</i>	<i>test - t1, lags - 1</i>	<i>test - t1, lags - 2</i>
coeff	0.9582	[0.9607; -0.0607]	[0.9621; -0.0643; -0.0356]
tStats	3.7799e + 03	[3.7571e + 03; -68.6809]	[3.7305e + 03; -72.4007; -40.2039]
FStat	<i>Inf</i>	1.4345e + 07	7.1826e + 06
AIC	1.3410e + 06	1.3363e + 06	1.3346e + 06
BIC	1.3410e + 06	1.3363e + 06	1.3347e + 06
p-value	1.00e - 03	1.00e - 03	1.00e - 03
H	1	1	1

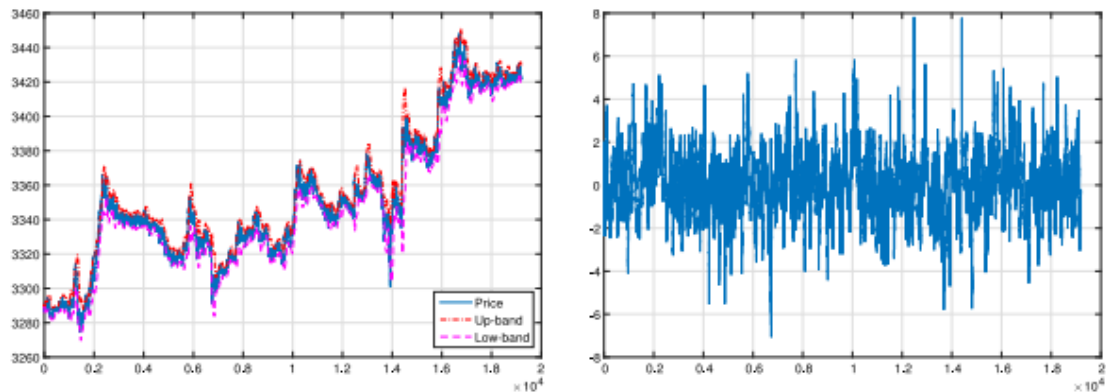
表 7. 15s 平稳布林带 ADF 检验结果

最后，我们检查所有交易规则产生的对数收益是否为平稳的。我们检验所有 279 个策略中的单个对数收益，所有的 ADF 检验都得到“h=1”的结果，即对数收益平稳。图 5 展示了 60s 数据 Boll(120,0.1)的结果。由图 5(a)，曲线有上升趋势。图 5(b)是累积对数收益平均值线。在一开始，这个趋势剧烈震荡；但是随着时间推移，这种趋势开始稳定下来并且收敛到一个高于 $5 * 10^{-5}$ 的特定值。

5.3 数据探测

表 8 展示了对每组高频数据所有 297 个策略中显著策略的数量。54 个策略在显著水平 5%和 70 个策略在显著水平 10%下是显著的。一些可盈利策略没有通过 SPA 检验。两个原因可能可以解释这一现象，第一个是可能偶尔会有一些策略成功，即数据窥探现象。另一个原因是尽管 SPA 测试极大地提高了测试的能力，仍然有可能忽略具有预测能力的方法（第二种错误类型），就如 Hansen[25]所述。然而，根据 SPA 检验显著的策略大多数都是表现最好的策略。

Step-SPA 测试结果表明，在校正数据窥探偏差后，所有平稳指标都有显著预测能力。尽管一些参数组合和数据频率可能偶然胜出，大多数盈利策略仍保持显



著的预测能力。在 3 个平稳指标中，不管数据频率如何，移动均线和布林带在一些参数下都表现较好。然而，KDJ 在 15s 数据上的 Step-SPA 检验不再具备 5%和 10%显著水平。当考虑交易成本时，没有策略是显著的，原因与第 5.1 节提到的相同。

(a) 原始布林带

(b) 平稳布林带

图 4. 原始布林带与平稳布林带在 IF1611 的 15s 数据上的比较

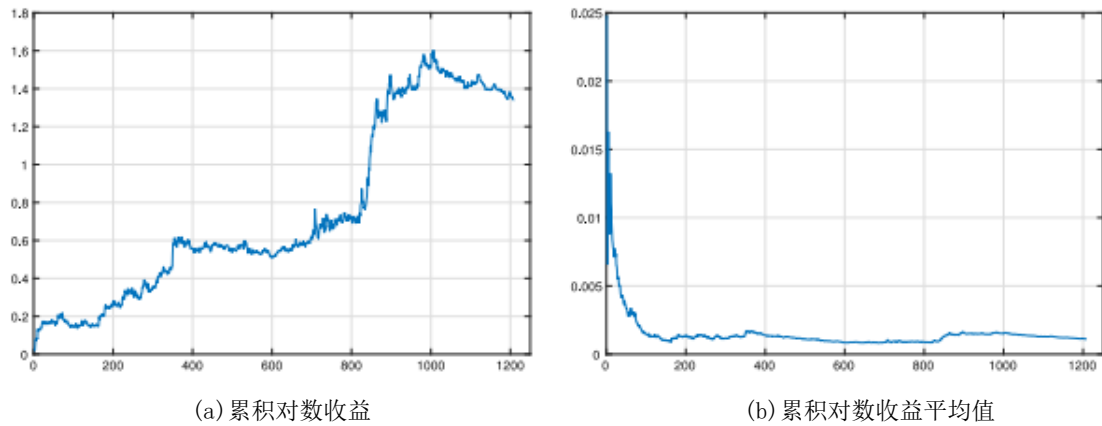


图 5. 60s 数据Boll(120,0.1)结果

Method	15s		30s		60s	
	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$
MA	15	19	22	25	7	14
KDJ	0	0	2	2	3	4
Boll	1	1	3	3	1	2
Total	16	20	27	30	11	20

表 8. Step-SPA 检验结果

- a. AR 表示年化收益, MDP 表示最大回撤率
- b. AR/MDP 为 AR 与 MDP 之比, SR 为夏普率指标

Strategy	AR ^a	MDP ^a	AR/MDP ^b	SR ^b	Strategy	AR ^a	MDP ^a	AR/MDP ^b	SR ^b
Boll_30(120,0.1)	59.31%	6.64%	8.93	2.33	MA_30(1,30,0.0001)	30.74%	9.44%	3.26	1.51
Boll_30(120,0.5)	37.61%	9.69%	3.88	1.63	MA_30(1,60,0.0001)	38.47%	8.31%	4.63	1.92
Boll_60(60,0.1)	44.55%	9.38%	4.75	1.79	MA_30(1,60,0.0005)	37.63%	12.28%	3.06	1.68
KDJR_30(14,3,3)	28.86%	11.76%	2.45	1.99	MA_30(5,60,0.0001)	30.83%	6.10%	5.05	1.64
KDJR_60(5,1,3)	33.90%	11.68%	2.90	2.24	MA_30(5,60,0.0005)	39.89%	6.45%	6.19	1.86
KDJR_60(9,3,3)	29.08%	8.40%	3.46	2.10	MA_30(5,60,0.001)	33.70%	6.14%	5.49	1.51
MA_15(10,30,0.0001)	35.96%	10.48%	3.43	1.74	MA_30(1,120,0.0001)	42.61%	6.61%	6.45	2.09
MA_15(5,60,0.0001)	40.36%	7.76%	5.20	1.90	MA_30(1,120,0.0005)	40.84%	7.12%	5.74	1.84
MA_15(10,60,0.0005)	31.80%	7.99%	3.98	1.50	MA_30(1,120,0.001)	38.20%	8.96%	4.26	1.67
MA_15(15,60,0.0001)	39.46%	9.79%	4.03	1.86	MA_30(1,120,0.0015)	36.60%	9.66%	3.79	1.57
MA_15(5,120,0.0001)	42.61%	6.77%	6.29	2.09	MA_30(5,120,0.0005)	36.27%	10.03%	3.62	1.61
MA_15(5,120,0.0005)	45.67%	4.54%	10.05	1.94	MA_30(5,120,0.001)	41.30%	9.58%	4.31	1.83
MA_15(5,120,0.001)	37.71%	4.63%	8.15	1.59	MA_30(5,120,0.0015)	35.19%	10.18%	3.46	1.62
MA_15(10,120,0.0001)	48.84%	6.33%	7.71	2.17	MA_30(10,120,0.001)	35.79%	10.21%	3.51	1.56
MA_15(10,120,0.0005)	45.25%	5.71%	7.92	2.00	MA_30(10,120,0.0015)	35.19%	9.66%	3.64	1.57
MA_15(10,120,0.001)	34.84%	6.07%	5.74	1.56	MA_30(15,120,0.001)	36.44%	9.51%	3.83	1.55
MA_15(15,120,0.0001)	35.18%	7.23%	4.87	1.70	MA_60(1,60,0.0005)	37.87%	7.76%	4.88	1.70
MA_15(15,120,0.0005)	34.81%	7.56%	4.61	1.52	MA_60(15,60,0.001)	36.75%	11.15%	3.30	1.58

表 9. 策略池

5. 4 策略改进

如上所述, 我们讨论了使用中国指数期货高频数据的简单平稳技术交易规则的盈利能力。但是在实际交易中, 投资者同时关注盈利能力和稳定性。第一个性质保证利润, 但是世界上没有完美的策略每时每刻都能盈利。一次大的损失可以让之前积累的收益在一天内付之一炬。从这点来看, 稳定性可能更加重要。

市场总是在变化，没有策略能永远盈利。这就是为什么现实交易员总是调试他们的策略的原因。在本节中，我们每隔几天从策略池中选一次最佳策略，然后将选中的策略投入市场一段时间。在策略池中比较策略绩效的时间称为训练期，定义选择的策略在市场上的时间为测试期。策略池由“好”策略组成。“好”的标准是夏普率大于 1.5。具体的策略名和相关指标在表 9 中列出。池中所有策略根据 SPA 检验结果都是显著的。在本节中，我们忽略交易成本。

每个策略名的开头字符串表示策略使用的计数指标，“_”后面的数字表示数据频率，括号中数字是策略的参数。例如，Bo11_30(120, 0.1) 表示由平稳布林带产生的策略且 30s 数据参数为 120 和 0.1。

至于训练期和测试期的长度，我们设置步长为 10 的 10-80 天测试期和步长为 10 的 20-80 天训练期，并且要求测试期不长于训练期。所以我们得到 35 组训练-测试期参数。为了说明问题，这里用第一组参数解释了操作。从第 1 天到第 20 天，所有策略池中的策略一起运行，然后我们选择最好的一个，命名为策略 A。将策略 A 在第 21-30 天投入市场。在一段时间内，所有策略在第 11-30 天一块运行来产生下一个最佳策略 B，然后让策略 B 在第 31-40 天投入市场。其他过程以此类推。如果策略表现在第 1 期和第 2 期表现都是最佳，那么策略 A 和 B 可以是同一个策略。有最大年化收益/最大回撤率的策略时一个训练期的最佳策略。

表 10 列出了不同长度训练期和测试期的优化策略的绩效。从夏普率来看，策略池中最好的策略是 Bo11_30(120, 0.1)，它的夏普率为 2.33，年化收益与最大回撤率之比为 8.93。优化策略的最高的夏普率为 2.46，此时参数为 (30, 20)。年化收益与最大回撤率之比为 9.22，比 8.93 更好。图 6(a) 是 Bo11_30(120, 0.1) 的净值曲线，图 6(b) 是参数为 (30, 20) 的优化策略的净值曲线。综上我们可以说优化策略比单一平稳技术指标策略更加稳定。这可能是一个接下来可行的研究方向。

6. 结论

在本文中，我们基于中国股指期货的三组高频数据(15s, 30s, 60s)，利用了几种绩效和风险度量方法和两种统计检验，ADF 检验和 SPA 检验，对简单的平稳技术交易规则的盈利能力进行了测试。交易规则包括平稳移动均线，平稳布林带和 KDJ 策略。不论滞后项数是多少，所有 297 个策略的对数收益都通过了 ADF 检验。对每种平稳技术指标，根据 SPA 检验，都存在显著的参数组合。并且在没有交易成本得情况下，根据绩效考核和风险度量表现优异的策略都通过了 SPA 检验。当考虑交易成本时，最好的策略不再比其他的更有优势，甚至在 2015 年 9 月 2 日之后会有高额损失。最后，我们从风险的角度提出了一种可行的方法来提高我们策略的稳定性。

Train ^a	Test ^a	AR ^b	MDP ^b	AR/MDP ^c	SR ^c	Train ^a	Test ^a	AR ^b	MDP ^b	AR/MDP ^c	SR ^c
20	10	47.68%	6.96%	6.85	2.35	60	50	50.46%	8.28%	6.09	2.30
20	20	44.44%	5.14%	8.65	2.19	60	60	36.96%	9.37%	3.94	1.71
30	10	48.54%	9.17%	5.29	2.30	70	10	42.21%	10.76%	3.92	1.93
30	20	49.74%	5.39%	9.22	2.46	70	20	40.78%	7.50%	5.44	1.87
30	30	47.57%	6.44%	7.39	2.30	70	30	50.08%	5.13%	9.76	2.41
40	10	42.28%	8.44%	5.01	2.08	70	40	46.02%	9.69%	4.75	2.22
40	20	47.26%	3.33%	14.18	2.31	70	50	50.80%	5.81%	8.75	2.42
40	30	46.99%	6.64%	7.08	2.35	70	60	45.60%	9.16%	4.98	2.11
40	40	40.25%	7.34%	5.48	1.93	70	70	53.62%	9.41%	5.70	2.36
50	10	53.58%	3.53%	15.16	2.34	80	10	50.49%	11.38%	4.44	2.26
50	20	46.40%	4.86%	9.55	2.18	80	20	53.16%	5.49%	9.68	2.37
50	30	50.54%	4.28%	11.80	2.43	80	30	51.73%	6.01%	8.61	2.35
50	40	38.97%	6.42%	6.07	1.82	80	40	50.74%	10.02%	5.06	2.21
50	50	38.54%	4.38%	8.79	1.71	80	50	53.65%	7.29%	7.36	2.40
60	10	44.96%	10.56%	4.26	2.01	80	60	47.48%	9.04%	5.25	2.22
60	20	39.57%	5.82%	6.80	1.85	80	70	48.29%	10.17%	4.75	2.21
60	30	40.35%	8.07%	5.00	1.82	80	80	38.22%	10.56%	3.62	1.79
60	40	46.97%	6.21%	7.57	2.25						

- a. Train 表示训练期的长度，Test 表示测试期的长度
b. AR 表示年化收益，MDP 表示最大回撤率
c. AR/MDP 表示年化收益与最大回撤率之比，SR 表示夏普率指标

表 10. 优化策略的绩效

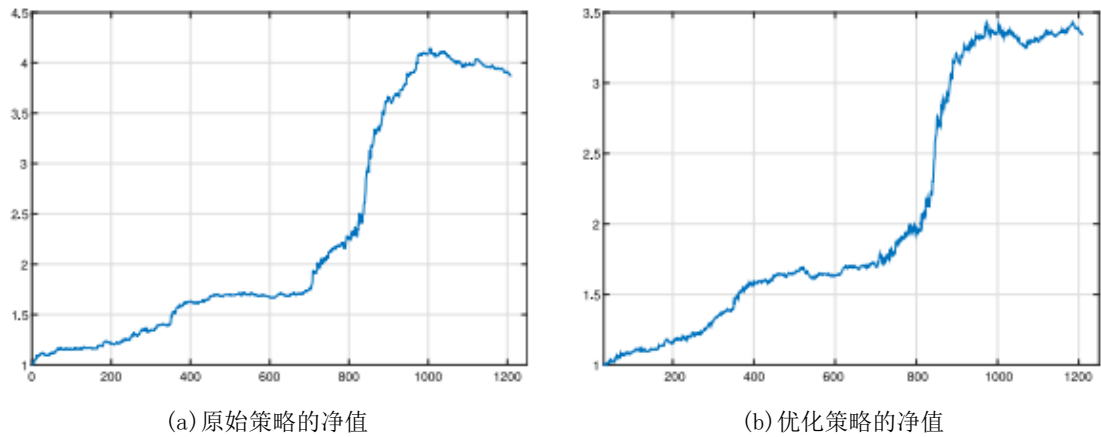
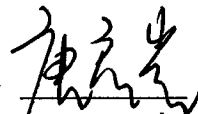
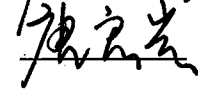
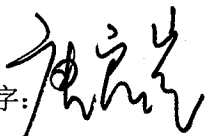


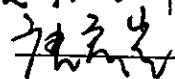
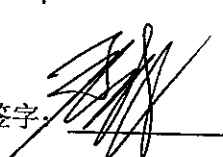

图 6. 原始和优化策略的净值曲线

书面翻译对应的原文索引

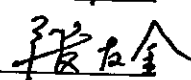
- [1] Chen J C, Zhou Y, Wang X. Profitability of simple stationary technical trading rules with high-frequency data of Chinese Index Futures[J]. 2017.

综合论文训练记录表

学生姓名	李博扬	学号	2014012118	班级	数 41
论文题目	基于平稳过程的 CTA 策略				
主要内容以及进度安排	<p>本文研究了经典协整配对交易模型，指出了其弊端，即配对价格之间的长期均衡关系可能被打破，协整模型可能失效。同时提出了改进方法，即将配对价差序列分解为随时间缓慢变化的均值项和具有均值回复性质的平稳波动项，利用前 n 期数据拟合配对模型，建立动态价差配对模型。然后本文在焦炭焦煤两种商品期货上验证了策略的有效性，并进行了参数优化，得到了泛化能力较好的策略模型。</p> <p>本研究的进度安排如下：三月份，搜集相关文献，完成外文文献翻译工作；四月份，通过研读相关论文，确立研究目标，完成理论模型构建；五月份，搜集相关数据，利用 python 编程实现模型，并进行参数优化，确立最终模型，撰写论文。</p> <p style="text-align: right;">指导教师签字: </p> <p style="text-align: right;">考核组组长签字: </p> <p style="text-align: right;">2018 年 3 月 26 日</p>				
中期考核意见	<p>进展顺利，得到预期成果。</p> <p style="text-align: right;">考核组组长签字: </p> <p style="text-align: right;">2018 年 4 月 23 日</p>				

指导教师评语	<p>论文基于平稳过程建立了动态的价差配对模型,并利用交叉验证对模型参数进行了优化,进行了实证分析,逻辑清晰,写作规范清晰,表明作者具有扎实的专业基础和较强的创新能力。同意其参加答辩并答辩。</p> <p>指导教师签字: </p> <p>2018年 6 月 5 日</p>
评阅教师评语	<p>本文基于平稳过程的性质,对协整配对模型给出了改进,建立了动态的价差配对模型,并进行了实证分析,做到了较好的策略效果。论文层次分明,条理清楚,写作规范,达到毕业要求。</p> <p>评阅教师签字: </p> <p>2018 年 6 月 6 日</p>
答辩小组评语	<p>一致同意,通过答辩。</p> <p>答辩小组组长签字: </p> <p>2018 年 6 月 16 日</p>

总成绩: 89

教学负责人签字: 

2018 年 6 月 18 日