

配对交易的最优阈值

于晓雨¹, 毕秀春², 张曙光¹

(1. 中国科学技术大学管理学院, 安徽合肥 230026; 2. 贵州财经大学数学与统计学院, 贵州贵阳 550025)

摘要: 考虑到市场的波动性和不确定性, 如何在保持稳定收益的基础上有效控制风险是亟待解决的问题. 通过遗传算法求解带止损条件的配对交易最优阈值, 在协整和部分协整条件下的沪深300指数和中证500指数分行业配对股票中进行实证检验, 实证结果表明, 与固定10%止损和不止损条件下的最优阈值相比, 带止损条件的交易阈值设置使配对交易在保持较高水平的收益的基础上能有效控制风险和损失.

关键词: 止损; 遗传算法; 配对交易; 最优阈值; 部分协整检验

中图分类号: C812 **文献标识码:** A **doi:** 10.3969/j.issn.0253-2778.2020.06.010

引用格式: 于晓雨, 毕秀春, 张曙光. 配对交易的最优阈值[J]. 中国科学技术大学学报, 2020, 50(6): 784-792.

YU Xiaoyu, BI Xiuchun, ZHANG Shuguang. Optimal threshold of pairs trading[J]. Journal of University of Science and Technology of China, 2020, 50(6): 784-792.

Optimal threshold of pairs trading

YU Xiaoyu¹, BI Xiuchun², ZHANG Shuguang¹

(1. School of Management, University of Science and Technology of China, Hefei 230026, China;

2. School of Mathematics and Statistics, Guizhou University of Finance and Economics, Guiyang 550025, China)

Abstract: Considering the volatility and uncertainty of the market, how to effectively control the risk on the basis of maintaining stable return is an urgent problem to be solved. Here genetic algorithm was used to solve the optimal threshold of pairs trading with stop loss condition. Empirical test was carried out in the matching stocks of CSI 300 and CSI 500 Indices industries under the condition of cointegration and partial cointegration. Results show that the presented method controls risk and loss more effectively on high return than 10% stop loss and no stop loss.

Key words: stop loss; genetic algorithm; pairs trading; optimal threshold; partial cointegration

0 引言

配对交易是从二十世纪八十年代中期开始兴起的一类统计套利策略^[1]. 其基本思想是通过统计的方法寻找配对股票价格差稳定在固定阈值内的两支股票, 在其价格差扩大超出固定阈值时, 买入被低估的股票, 卖出被高估的股票, 因此, 寻找最优阈值在配对交易中占有重要的地位.

目前针对配对交易的阈值研究主要涉及三类研究方法: 随机控制方法、时间序列方法和智能算法. Ornstein-Uhlenbeck(O-U)模型是随机控制方法中最常见的模型, 其主要思想为通过假设股票的对数价格服从 O-U 过程, 通过求解 HJB 方程得到最优效用下的交易阈值. 这方面的研究参见文献[2-6]. 另外, 还有 geometric Brownian motion (GBM) 模型^[7,8]、Levy 跳模型^[9]、机会成本模型^[10]等. 时间序列方法主要是通过 GARCH 模型来预测

股票波动率, 利用波动率来决定交易阈值. 杨艳军和陈思岑^[11]基于 GARCH 模型预测国债期货价差序列残差的方差, 提出基于开平仓阈值的配对交易策略. Chen 等^[12]通过 TAR-GARCH 模型求解配对交易的开平仓阈值.

智能算法主要是通过神经网络、机器学习、仿生寻优算法等智能预测和优化算法来预测价差走势、决定交易阈值. Huck^[13,14]通过 Elman 中性网络预测未来一周的收益, 根据预测收益的不同进行配对交易. 龙奥明等^[15]基于 LSTM 神经网络对价差增量进行预测. 胡文伟等^[16]基于强化学习算法提出了自适应模式的动态参数优化方法, 有效提高配对交易策略的效率. 与传统的随机控制、时间序列方法相比, 智能算法得到的结果更贴合实际, 能够灵活有效地计算和更新参数.

以上基于配对交易的最优阈值研究大多只考虑开平仓两个阈值或设置固定止损, 未将最优止损

收稿日期: 2020-02-14; 修回日期: 2020-04-26

基金项目: 国家自然科学基金(11401556, 11471304), 贵州财经大学科研项目(2020YJ026)资助.

作者简介: 于晓雨, 女, 1995年生, 硕士. 研究方向: 金融工程, 风险管理. E-mail: yuxiaoyu@mail.ustc.edu.cn

通讯作者: 毕秀春, 博士/副教授. E-mail: xcbl@ustc.edu.cn

条件融入模型. 考虑到市场的波动性和不确定性, 尤其是中国股市历年的波动十分剧烈, 设置止损是非常必要和重要的. 而目前国内外研究止损条件下的配对交易阈值的相关文献较少, 毕秀春等^[17], Zhang^[18], Lindberg^[19]通过随机控制方法求解带止损条件的最优阈值, 止损的加入使得收益有一定的损失, 但有效降低了风险.

近年来, 遗传算法被广泛应用于经济和金融领域^[20-26]. 但目前国内外基于遗传算法研究配对交易最优阈值的文献较少, Huang等^[27]将遗传算法应用于优化协整配对权重、平均移动时间和布林带带宽三个参数指标, 并在十个不同市场检测, 研究结果表明遗传算法得到的收益率显著优于基准线收益率, 体现了遗传算法的适用性. 遗传算法在解决不同非线性问题时表现出很好的鲁棒性、全局最优性和高效率.

本文通过遗传算法求解带止损条件的配对交易最优阈值, 并与文献^[28-30]设置的10%固定止损以及无止损设置下的最优阈值进行比较. 目前的配对交易大多以协整性为配对标准, Clegg和Krauss^[30]提出部分协整检验的概念, 并以固定开平仓阈值、10%固定止损的设置, 在标普500夏普比率排名前10支同行业股票对中进行了实证检验, 本文将在常用的协整性检验的基础上, 验证符合部分协整性的股票对在沪深300指数和中证500指数分行业成分股中的表现.

1 理论与模型设置

1.1 遗传算法理论及参数设置

遗传算法是一类借鉴生物界的进化规律即适者生存、优胜劣汰遗传机制演化而来的随机化搜索方法, 其主要特点是直接对结构对象进行操作, 不存在求导和函数连续性的限定; 具有内在的隐并行性和更好的全局寻优能力; 采用概率化的寻优方法, 能自动获取和指导优化的搜索空间, 自适应地调整搜索方向, 不需要确定的规则. 遗传算法已被广泛地应用于组合优化、机器学习、信号处理、自适应控制和人工生命等领域, 是现代有关智能计算中的关键技术. 遗传算法使用二进制遗传编码, 其执行过程如下:

(I) 初始化. 设置种群规模 N , 交叉概率 P_c , 变异概率 P_m , 最大进化代数 T , 进化代数计数器 $t=0$, 随机生成 N 个个体作为初始种群 $P(0)$.

(II) 个体评价. 计算种群 $P(t)$ 中各个个体的适应度.

(III) 种群进化. 分为选择母体、交叉、变异和选择子代四个步骤.

①选择母体: 从 $P(t)$ 中通过选择算子选择出 $M/2$ 对母体 ($M \geq N$).

②交叉: 对所选的 $M/2$ 对母体, 以概率 P_c 交叉形成 M 个中间个体.

③变异: 对 M 个中间个体分别以概率 P_m 变异, 形成 M 个候选个体.

④选择子代: 从上述形成的 M 个候选个体中

通过适应度的排序选择出 N 个个体组成新一代种群 $P(t+1)$.

(IV) 终止检验. 若 $t=T$, 则以进化过程中所得到的具有最大适应度个体作为最优解输出, 终止计算.

由执行过程可见, 遗传算法主要包括三个基本操作: 选择、交叉和变异.

选择: 即根据适应度进行串赋值. 当在一个种群中进行自然选择时, 适应度越大的母体被选择的概率也就越大. 个体 I_i 的采样概率如下定义:

$$P(I_i) = \frac{f(I_i)}{\sum_{i=1}^M f(I_i)},$$

其中, $f(I_i)$ 为个体 I_i 的适应度, M 为候选个体的数量.

交叉: 即模拟两条染色体交换基因片段. 与人类的双倍染色体不同, 遗传算法的染色体是单倍体, 如图1所示, 交叉时两条染色体之间交换以编码表示的等位基因.

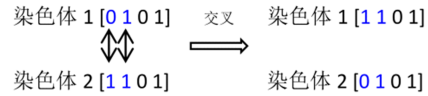


图1 遗传算法交叉示意图

Fig. 1 Crossover diagram of genetic algorithm



图2 遗传算法变异示意图

Fig. 2 Variation diagram of genetic algorithm

变异: 即某个基因位发生变化. 表示在算法中即某个位置的编码发生突变, 如图2所示, 蓝色位置原先编码为1, 突变后该位置的编码变为0.

为分析遗传算法的工作原理, Holland^[31]提出模式定理.

定理 1.1 在典型的遗传算法中, 如下公式成立:

$$P(H)^{t+1} \geq P(H)^t \frac{\bar{f}(H)}{\bar{f}} (1 - P_m \cdot O(H)) \cdot \left(1 - P_c \cdot \frac{L(H)}{L-1}\right),$$

其中, H 表示模式, 指编码空间中相似的模块, 例如 $H_0 = [0, *, *, 1]$ 为一个模式, $[0, 0, 1, 1]$ 和 $[0, 1, 1, 1]$ 都包含该模式, $L(H)$ 表示模式的长度, 指第一固定基因位和最后一个固定基因位的距离, $O(H)$ 指模式的阶, 即固定基因位的个数, $\bar{f}(H)$ 为模式平均适应度, 即种群中包含该模式的个体适应度的平均值, $P(H)^{t+1}$ 表示在 $t+1$ 代种群中模式出现的概率.

本文通过遗传算法求解配对交易的最优阈值. 其中参数设置如下: 定义上下两方向的开平阈值分别为 $L_{10}, L_{1c}, L_{20}, L_{2c}$, 止损阈值为 L_{loss} . 由于 Z -score(下文中式(1)和(2))是标准化后的均值回归值, 为与常见的开平仓阈值范围^[30] $[-2.5, 2.5]$ 保

持一致,同时预留更多的选择空间,设置上开仓阈值 L_{10} . 搜索范围为 $[0,3]$,上平仓阈值 L_{1c} 搜索范围为 $L_{10} + [-6,0]$;下开仓阈值 L_{20} 搜索范围为 $[-3,0]$,下平仓阈值搜索范围为 $L_{20} + [0,6]$;止损阈值 L_{loss} 搜索范围为 $[0,0.1]$. 为保证计算充分,设置最大进化代数 $T=500$ 为建议值 $T \in [100,500]$ 的最高值;设置种群规模 $N=100$,变异概率为 $P_m=0.1\%$. 为防止过拟合,遗传算法的参数没有经过调试,都为默认参数或建议参数.

1.2 协整理论方法

现有协整方法中应用最广泛的是 Engle-Granger(EG)检验法^[32]和 Johansen 检验法^[33]. 本文使用 EG 检验法进行协整性检验. 下面给出单整和协整的定义:

定义 1.1 若一个时间序列 $\{y_t\}$ 在成为稳定序列之前必须经过 d 次差分,则称该时间序列是 d 阶单整. 记为 $y_t \sim I(d)$.

定义 1.2 设随机向量 $\{X_t\}$ 中所含分量均为 d 阶单整,记为 $X_t \sim I(d)$. 如果存在一个非零向量 β ,使得随机向量 $Y_t = \beta X_t \sim I(d-b)$, $b > 0$,则称随机向量 X_t 具有 d, b 阶协整关系,记为 $X_t \sim CI(d, b)$,向量 β 被称为协整向量.

本文分两步选择符合协整条件的股票对并计算得到 Z-score 值.

步骤 1 平稳性检验:对配对股票每日复权收盘价时间序列进行 ADF 平稳性检验,若序列不平稳,则进行一阶差分后再进行平稳性检验,若满足平稳性,则该时间序列为一阶单整. 这里使用的是 python statsmodels. tsa. stattools 包中的 adfuller 函数.

步骤 2 协整性检验:由于符合协整性的配对较多,因此本文只对 1 阶单整的时间序列进行两两协整性检验,这里使用的是 python statsmodels. tsa. stattools 包中的 coint 函数. 将通过协整性检验的股票对加入配对股票池中.

步骤 3 计算得到 Z-score 值:通过 OLS 法计算符合协整性股票对线性回归方程 $y_t = \alpha + \beta \cdot x_t$ 的 α, β 值,残差为 $e_t = y_t - (\alpha + \beta \cdot x_t)$. Z-score 为标准化后的残差序列,即

$$Z\text{-score}_t = \frac{e_t - \text{mean}(e_t)}{sd(e_t)} \quad (1)$$

1.3 部分协整理论方法

部分协整理论是协整理论的弱化,它允许残差序列同时具有均值回归和随机游走性质. 假设 $X_1 = (X_{1,t})_{t \in T}$ 和 $X_2 = (X_{2,t})_{t \in T}$ 为两个时间序列,代表不同股票价格,称 X_1 和 X_2 是部分协整的,如果:

$$X_{2,t} = \beta X_{1,t} + W_t,$$

$$W_t = M_t + R_t,$$

$$M_t = \rho M_{t-1} + \epsilon_{M,t}, \epsilon_{M,t} \sim N(0, \sigma_M^2),$$

$$R_t = R_{t-1} + \epsilon_{R,t}, \epsilon_{R,t} \sim N(0, \sigma_R^2),$$

其中,参数 β 为实数, M_t 表示两支股票价格差中的均值回归部分, R_t 表示两支股票价格差中的随机游走部分,一阶自回归系数 $\rho \in (-1, 1)$, $\epsilon_{M,t}, \epsilon_{R,t}$

分别为期望为 0, 方差为 σ_M^2, σ_R^2 的独立高斯白噪声过程.

由上式可以看出时间序列 X_1 和 X_2 服从残差为 $W = (W_t)_{t \in T}$ 的偏自回归关系. 均值回归的方差在总残差方差中的占比是偏自回归模型的关键统计量,记为

$$R_{M,R}^2 = \frac{\text{Var}[(1-B)M_t]}{\text{Var}[(1-B)W_t]} = \frac{2\sigma_M^2}{2\sigma_M^2 + (1+\rho)\sigma_R^2},$$

$$R_{M,R}^2 \in [0, 1],$$

其中, B 是滞后算子. 当 $R_{M,R}^2 = 0$ 时,表示自回归部分占比为 0,全部都是随机游走部分,当 $R_{M,R}^2 = 1$ 时,意味着随机游走部分占比为 0,全部都是自回归部分.

由于 M_t 和 R_t 是不可直接观察的,根据卡尔曼滤波极大似然函数法估计参数 $\beta, \rho, \sigma_X, \sigma_M, \sigma_R$,通过卡尔曼滤波等式计算得到 M_t 和 R_t 的值.

记标准化后的均值回归部分值为

$$Z\text{-score}_t = \frac{M_t}{\sigma_M} \quad (2)$$

1.4 配对交易模型设置说明

说明 1 训练集和测试集

该部分参考文献[17]的设置,选择 2009-01-01~2018-12-31 共十年的指数成分股日复权收盘价作为总数据样本,其中,将 2009-01-01~2013-12-31 共五年的数据作为训练集,用来进行协整检验和部分协整检验,计算和保存 β 和 Z-score 值,通过遗传算法计算带止损条件的最优阈值及无止损的最优阈值,在 2014-01-01~2018-12-31 共五年时间的测试集上进行回测和验证.

说明 2 仓位、收益的计算方式

该部分参考文献[1,30,35]的计算方式. 假设配对股票为 P 和 Q ,其股票价格关系为 $X_{2,t} = \alpha + \beta \cdot X_{1,t} + W_t$,其中 $X_{1,t}$ 表示 t 时刻股票 P 的价格, $X_{2,t}$ 表示 t 时刻股票 Q 的价格. 这意味着,买多一手 Q ,同时要卖空 β 手 P . 买多 1 元的 Q ,代表买了 $1/X_{2,t}$ 手的 Q ,此时,应卖空 $\beta X_{1,t}/X_{2,t}$ 的 P ,代表买了 $\beta/X_{2,t}$ 手的 P . 很显然, $\beta X_{1,t}/X_{2,t}$ 与开始时的 1 元不等价. 为了保持市场中性,应设置:

①若 $\beta X_{1,t}/X_{2,t} \leq 1$,则投资 V 元在 Q 上,买 $V/X_{2,t}$ 手 Q ,同时投资 $V \cdot \beta X_{1,t}/X_{2,t}$ 在 P 上,买 $\beta \cdot V/X_{2,t}$ 手 P .

②若 $\beta X_{1,t}/X_{2,t} > 1$,则投资 V 元在 P 上,买 $V/X_{1,t}$ 手 P ,同时投资 $V X_{2,t}/\beta X_{1,t}$ 在 Q 上,买 $V/\beta X_{1,t}$ 手 Q .

本文设置初始资金 V_0 为人民币 1000 万元,交易手续费为单笔万分之五,每个交易日实时更新当前资金情况,通过资金曲线计算收益. 每次开仓由初始资金计算仓位,止损按照初始资金的比例进行判断,其余资金不做任何操作,排除连续复利导致的收益虚高的情况,使实证结果更贴合实际.

说明 3 开平仓、止损规则

设遗传算法计算得到的最优开平仓阈值为 $L_{10}, L_{1c}, L_{20}, L_{2c}$,分别表示上、下两个方向的开平

仓交易阈值,同一配对的交易阈值为确定值,不同配对的交易阈值有所区别。止损阈值为 L_{loss} (注:本文三个模型分别设置三种止损方式,最优止损阈值为遗传算法计算得出的 $0\sim 10\%$ 范围内的某一确定值,是与最优开平仓阈值一同计算得出的,不同配对的止损阈值有所区别。固定止损阈值是指止损阈值设置恒为 10% ,对所有配对都是固定不变的。不止损是指无止损设置,即所有配对的止损阈值 L_{loss} 均为 100%), Z -score 以 Z_t 表示。

①开仓条件: $L_{10} < Z_t < L_{1c}$ (或 $L_{2c} < Z_t < L_{20}$) 且此时无仓位时开仓,这里的开仓是同时开多仓的 Q (或 P) 和空仓的 P (或 Q)。

②平仓条件: $Z_t \geq L_{1c}$ (或 $Z_t \leq L_{2c}$) 且此时有该方向的仓位时平仓。

③止损条件:每个交易日实时监控当前总资金,若当前总资金 V_{t2} 与开仓前的总资金 V_{t1} 相比处于亏损状态,且亏损比例 $(V_{t2} - V_{t1})/V_0 \geq L_{loss}$,则进行止损平仓,等待下个交易机会。

④强制平仓条件:若回测时间到达最后一个有数据的交易日,而此时还有未平仓仓位时,则进行强制平仓。

图3为招商证券(600999.SH)和光大银行(601818.SH)协整 Z -score 值的开平仓线说明,以辅助读者理解交易逻辑。

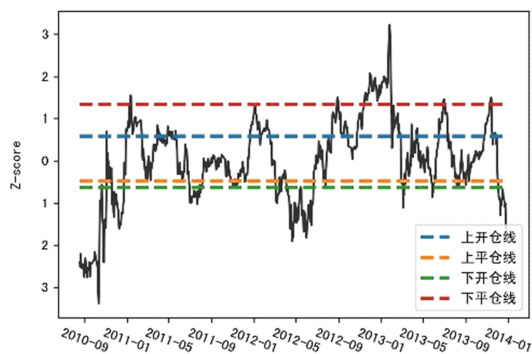


图3 开平仓说明图

Fig. 3 Opening and closing diagram

说明4 回测指标计算说明

①夏普比率:计算公式为 $\text{Sharpe Ratio} = (E(R_p) - R_f) \cdot \sqrt{252} / \sigma_p$, $E(R_p)$ 表示投资组合预期回报率,这里是指以日为单位计算得到的收益率; R_f 是以日为单位无风险利率; σ_p 表示投资组合标准差,通过计算每日收益率的标准差来表示。单日夏普比率 $(E(R_p) - R_f) / \sigma_p$ 乘 $\sqrt{252}$ 得到年化夏普比率。

②历史最大回撤:通过收益曲线动态历史最高点和最低点的差值计算得出,是动态累积记录的过程。图4展示了招商证券(600999.SH)和光大银行(601818.SH)交易期间的动态收益图。其中,历史最大回撤区间发生在2012年10月至2013年2月间,共损失23%的总利润。

③单日最大回撤:单个交易日的动态收益的最大回撤。

④交易次数:回测期间总交易次数,一次开仓

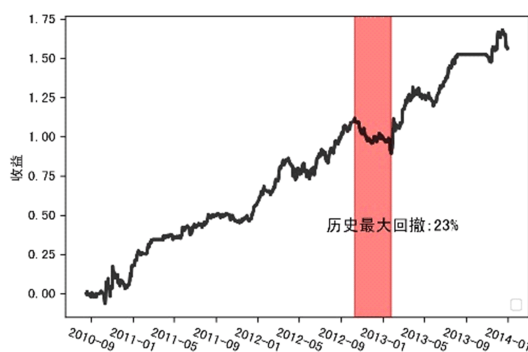


图4 历史最大回撤说明图

Fig. 4 Historical maximum recall

和一次平仓综合成为一次交易。

⑤交易频率:总交易日与交易次数的比值。

2 实证检验

本文以指数分行业成分股2009-01-01~2013-12-31共五年的日加权收盘价数据作为训练集,2014-01-01~2018-12-31共五年的日加权收盘价数据为测试集。行业分类标准参照MSCI发布的全球行业分类标准,共分为材料、房地产、工业、公共事业、金融、可选消费、能源、日常消费品、信息技术、医疗保健和电信业务11个行业。各行业符合协整性和部分协整性的股票对数量如表1所示。可见,工业的配对股票数量较多,金融、材料、可选消费、信息技术、医疗保健行业次之,房地产、能源等行业的配对数量相对较少,而电信业务行业无符合条件的配对。下面将分协整配对和部分协整配对两个部分检验最优带止损阈值的表现。

表1 各行业协整配对和部分协整配对数量

Tab. 1 Number of cointegration pairs and partial cointegration pairs

行业名称	沪深300指数		中证500指数	
	协整配对数量	部分协整配对数量	协整配对数量	部分协整配对数量
材料	104	29	236	148
电信业务	0	0	0	0
房地产	20	11	30	69
工业	413	140	322	678
公共事业	20	13	8	17
金融	325	99	2	6
可选消费	116	37	228	154
能源	4	3	8	4
日常消费品	6	4	16	63
信息技术	102	33	256	241
医疗保健	21	8	36	130

2.1 协整配对模型

本文使用遗传算法对沪深300指数和中证500指数分行业成分股的训练集里符合协整性的配对股票分别求其带止损条件的最优阈值和无止损条件的最优阈值,再将求得的阈值结果分别放入测试集进行回测,测试带止损的最优阈值、无止损限制的最优阈值以及设置10%固定止损的最优阈值三个配对交易模型的表现情况。

其中,沪深 300 指数分行业成分股中所有配对股票训练、测试期总体表现如表 2 所示。可见,带止损的最优阈值配对交易模型在取得较高收益的基础上,有效降低了回撤。在训练期间,带止损的最优阈值配对交易模型(以下简称最优止损模型)有效控制历史最大回撤为 27%,较无止损最优阈值模型、固定止损最优阈值模型(以下简称无止损模型、固定止损模型)分别降低 13%和 7%,单日最大回撤为 10%,较固定止损模型和无止损模型分别降

低 1%和 3%。在降低模型整体交易风险的基础上,最优止损模型的年化收益率达到了 56%,超过了固定止损模型 2%,与无止损模型基本相当。从夏普比率的角度入手,最优止损模型的年化夏普比率为 1.74,较固定止损模型和无止损模型约提高了 0.08,排三者之首。可见,无论是从风险、收益还是夏普比率的角度上看,最优止损模型在沪深 300 指数协整配对交易的训练期内均表现优异。

表 2 沪深 300 协整配对交易三个模型汇总表现

Tab. 2 Results of cointegration pairs trading in CSI 300

	模型名称	总收益	年化收益率	夏普比率	单日最大回撤	历史最大回撤	交易次数	交易天数	交易频率
训练期	无止损	1.99	0.60	1.66	(0.13)	(0.40)	11.15	899.79	157.32
	固定止损	1.75	0.54	1.65	(0.11)	(0.34)	13.81	899.79	116.10
	最优止损	1.83	0.56	1.74	(0.10)	(0.27)	10.48	899.79	172.32
测试期	无止损	1.53	0.37	0.59	(0.27)	(1.01)	4.45	1077.94	441.04
	固定止损	1.06	0.25	0.49	(0.21)	(0.76)	9.09	1077.94	181.48
	最优止损	1.10	0.27	0.52	(0.20)	(0.71)	10.04	1077.94	181.56

[注]总收益、年化收益率、回撤的单位为 100%,夏普比率的单位为 1,交易次数的单位为次,交易天数的单位为天,交易频率的单位为(天/次),下表同。

测试期间模型的表现与训练期间基本一致。最优止损模型单日和历史最大回撤仍为三者最低,其中,单日最大回撤较固定止损和无止损分别降低了 1%和 7%,历史最大回撤较二者分别降低了 5%和 30%。从收益和夏普比率角度入手,与固定止损模型相比,最优止损模型将年化收益提高了 2%,夏普比率提高了 0.03。从测试期间的模型整体效果可见,止损的设置有效控制了交易风险,极大程度地降低了交易回撤。且最优止损的设置较固定止损而言,在降低交易风险的同时,还有效提高了模型的整体收益。

表 3 展示了中证 500 指数协整配对交易模型的

表 3 中证 500 协整配对交易三个模型汇总表现

Tab. 3 Results of cointegration pairs trading in CSI 500

	模型名称	总收益	年化收益率	夏普比率	单日最大回撤	历史最大回撤	交易次数	交易天数	交易频率
训练期	无止损	2.60	0.78	1.79	(0.14)	(0.45)	16.09	897.29	96.79
	固定止损	2.32	0.70	1.78	(0.12)	(0.37)	20.33	897.29	70.17
	最优止损	2.36	0.72	1.92	(0.11)	(0.30)	16.49	897.29	95.85
测试期	无止损	1.87	0.47	0.68	(0.30)	(1.04)	5.20	1044.83	343.07
	固定止损	1.29	0.32	0.58	(0.22)	(0.79)	11.32	1044.83	132.92
	最优止损	1.19	0.30	0.58	(0.21)	(0.74)	11.47	1044.83	146.52

综合以上两个指数回测的整体表现可知,首先,止损的设置是必要的。无论是在无止损最优阈值的基础上设置固定止损,还是使用带止损的最优阈值,止损的设定都可以有效降低回撤,控制风险,规避潜在危机。其次,本文提出的最优止损的设定较固定止损而言更为合理,可以在进一步降低回撤的基础上,整体提高策略收益。

由训练期和测试期的回测结果对比可见,在测试期间,模型的整体回撤较训练期有所增加,收益和交易频率有所降低。本文选取回测中典型的配对股票[招商证券(600999.SH),光大银行(601818.SH)]来解释出现该现象的原因。训练期的 Z-score 值如图 3 所示,收益如图 4 所示,训练期得到的最

回测结果。在训练期内,最优止损模型与无止损模型相比,历史最大回撤降低了 15%,单日最大回撤降低了 3%;与固定止损模型相比,在历史最大回撤降低 7%,单日最大回撤降低 1%的基础上,年化收益率还提高了 2%。最优止损模型在训练期内的年化夏普比率为 1.92,较固定止损模型和无止损模型分别提高了 0.14 和 0.13,为三者最高。在测试期内,最优止损模型较无止损模型历史最大回撤降低 30%、单日最大回撤降低了 9%;较固定止损模型历史最大回撤降低了 5%,且保持年化收益率和夏普比率基本相当。可见,模型能够在有效降低回撤的基础上,保持较好的收益能力和夏普比率。

优止损阈值为 3.47%。其训练期的 Z-score 及收益都较为稳定,无较大回撤波动。给出测试期的 Z-score 图像及训练期得到的最优止损阈值如图 5 所示,测试期三个模型的整体回测结果如图 6 所示。

由图 5、图 6 红色区域可见,在 2014 年年底至 2015 年上半年间股价出现了较大幅度的不一致波动情况,这是由于这段时间中国股市出现了较为极端的牛熊市,招商证券股份有限公司随之大幅波动,而光大银行的股票波动较小导致。期间,模型出现了较大的回撤,降低了策略的整体收益。在样本外数据中,由于市场的波动性和不确定性,无法确保股票对始终符合协整性要求,止损的设定可以使交易风险可控,从而降低未来不确定性的影响。可见,设

置最优止损能使模型在收益和损失二者之间找到平衡点,即在保持一定收益的基础上,能够较好地控制持仓风险。

2.2 部分协整配对模型

与协整配对阈值研究部分的步骤类似,本文测试了沪深 300 指数和中证 500 指数分行业成分股中符合部分协整性的配对股票在带止损的最优阈值、无止损限制的最优阈值以及设置 10% 固定止损的最优阈值三个配对交易模型中的表现情况。

沪深 300 指数的整体表现如表 4 所示。在训练期内,最优止损模型的历史最大回撤为 28%, 单日最大回撤为 9%, 为三者中的最低值。在控制交易回撤风险最低的同时,最优止损模型的年化收益率达到 55%, 夏普比率为 1.75, 均为三者中的最高值。其中年化收益率较固定止损模型提高 6%, 较无止损模型提高 3%, 效果明显。测试期的整体表现与训练期基本一致。尽管与训练期相比测试期的整体回测表现稍显逊色,但最优止损模型在测试期更能展现其优越性。在测试期间,较无止损模型而言,最优止损模型的单日最大回撤降低了 3%, 历史最大回撤降低了 45%, 同时还将在年化收益率提高了 4%。固定止损模型虽然也达到了类似的控制风险

的效果,但在风险控制和盈利能力上均逊色于最优止损模型。由此可见,最优止损模型在沪深 300 指数的符合部分协整性的股票对中表现较其他两个模型相比,无论是在训练期还是在测试期,都有更高的总收益、夏普比率、交易频率以及较低的回撤,表现优异。

中证 500 指数的整体表现如表 5 所示。由回测结果可见,与沪深 300 指数回测结果表现类似,最优止损模型较固定止损模型和无止损模型而言,总收益、夏普比率和交易频率为三者最高,同时保持着最低的回撤。在训练期中,最优止损模型的历史最大回撤为 39%, 年化收益率为 65%, 年化夏普比率为 1.65, 分别较固定止损模型和无止损模型回撤降低了 4% 和 2%, 收益提高了 6% 和 4%, 夏普比率提高了 0.12 和 0.08。除此之外,交易次数和交易频率也为三者的最高值。在测试期中,最优止损模型在降低回撤和提升收益的表现更为突出。较无止损模型而言,有效的风险控制将历史最大回撤降低了 33%, 较为合理的交易机制将总收益提高了 13%。固定止损模型也能达到有效地控制风险的效果,但在风险控制能力和盈利能力上均逊色于最优止损模型。

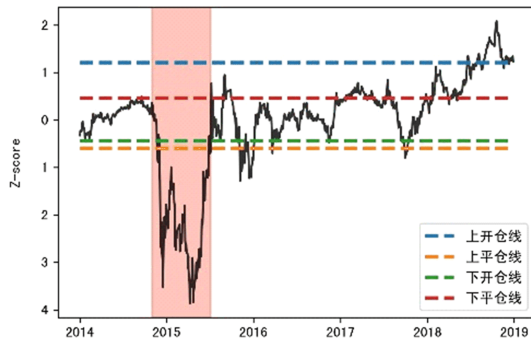


图 5 测试期 Z-score 及最优止损阈值

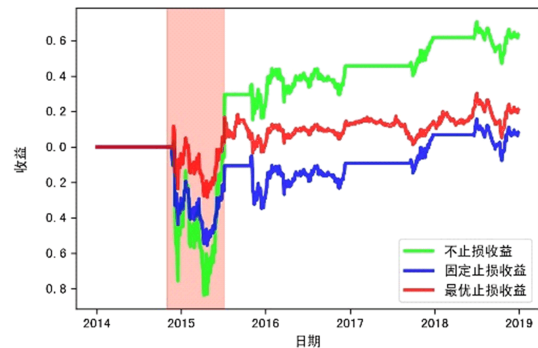


图 6 测试期三个模型回测结果

Fig. 5 Z-score and optimal stop loss threshold in test period

Fig. 6 Back test results in test period

表 4 沪深 300 指数部分协整配对交易三个模型汇总表现

Tab. 4 Results of partial cointegration pairs trading in CSI 300

模型名称	总收益	年化收益率	夏普比率	单日最大回撤	历史最大回撤	交易次数	交易天数	交易频率
训练期 无止损	1.78	0.52	1.69	(0.10)	(0.29)	19.78	918.05	74.94
训练期 固定止损	1.68	0.49	1.65	(0.10)	(0.30)	22.63	918.05	62.35
训练期 最优止损	1.87	0.55	1.75	(0.09)	(0.28)	26.11	918.05	56.71
测试期 无止损	0.02	0.01	0.09	(0.23)	(1.25)	9.44	1089.02	268.55
测试期 固定止损	0.10	0.02	0.06	(0.16)	(0.84)	18.51	1089.02	87.20
测试期 最优止损	0.19	0.05	0.12	(0.16)	(0.80)	26.51	1089.02	58.88

表 5 中证 500 指数部分协整配对交易三个模型汇总表现

Tab. 5 Results of partial cointegration pairs trading in CSI 500

模型名称	总收益	年化收益率	夏普比率	单日最大回撤	历史最大回撤	交易次数	交易天数	交易频率
训练期 无止损	2.40	0.61	1.57	(0.13)	(0.41)	23.64	1028.97	84.88
训练期 固定止损	2.32	0.59	1.53	(0.12)	(0.43)	28.38	1028.97	66.34
训练期 最优止损	2.56	0.65	1.65	(0.12)	(0.39)	32.44	1028.97	63.70
测试期 无止损	0.30	0.07	0.20	(0.25)	(1.22)	11.41	1058.94	241.50
测试期 固定止损	0.34	0.08	0.16	(0.18)	(0.91)	21.79	1058.94	81.40
测试期 最优止损	0.43	0.10	0.20	(0.18)	(0.89)	29.92	1058.94	58.90

综合上述分析可见,与协整配对股票的表现相比,部分协整配对股票的交易频率更高,这与文献[30]的结论一致.但在测试期的收益、回撤表现中,部分协整配对股票的整体表现不尽人意,与文献[30]的结论相违背.为此,本文选择一对符合部分协整性的股票对[交通银行(601328.SH),工商银行(601398.SH)]来具体解释产生该现象的原因,并提出相应的建议.两个股票训练和测试期的股价图像如图 7、8 所示.部分协整配对交易模型回溯期间表现不佳主要有以下三方面的原因:

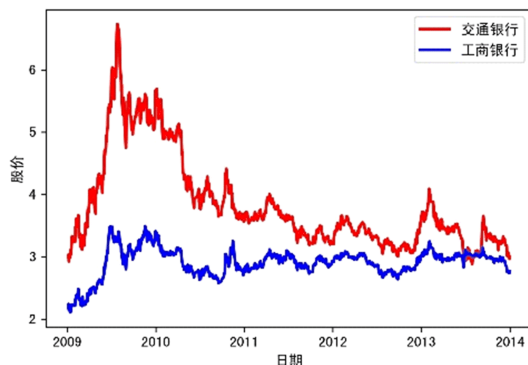


图 7 训练期股价图像
Fig. 7 Share price in training period

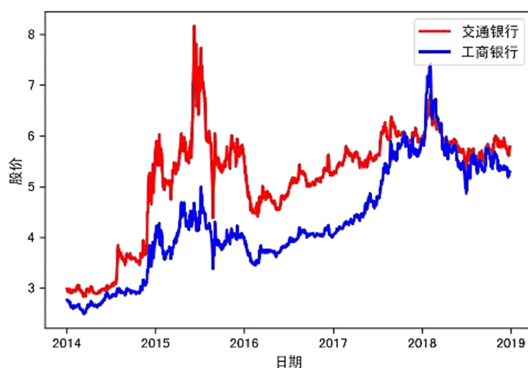


图 8 测试期股价图像
Fig. 8 Share price in test period

首先是回溯期较长导致的符合部分协整性的配对减少和参数失灵.由于部分协整配对交易更适合短期交易,长期符合部分协整性的配对股票数量较少,与文献[30]设置的三年训练期和半年测试期相比,各五年的训练、测试期导致配对股票的不确定性增加,由图 9 与图 10 的对比可见,测试期的 Z-score 值与训练期的差别较大,较为不稳定,这会导致计算得到的最优开平仓和止损阈值无法发挥在训练期的作用,造成损失.

模型在测试期表现不佳的第二个原因是随机游走部分带来的影响无法忽视.由图 11 和图 12 对比可知,在训练期的随机游走部分在 $[0.9, 1.8]$ 区间波动,而测试期内随机游走部分在 $[1.5, 4.2]$ 之间,波动区间明显增大.部分协整配对交易模型本身允许随机游走部分的存在,随机游走部分比例较大、较剧烈或持仓时间较长的配对股票会在交易过程中受到较多随机游走部分带来的影响,从而导致整体收益的下滑.

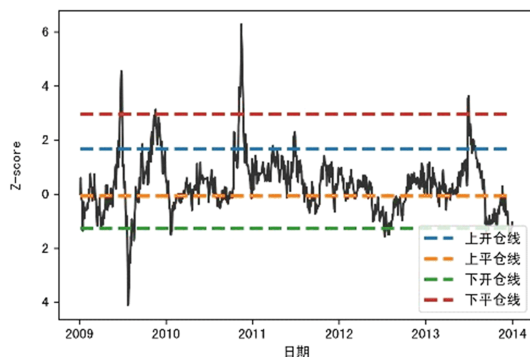


图 9 训练期 Z-score 及最优止损阈值
Fig. 9 Z-score and optimal stop loss threshold in training period

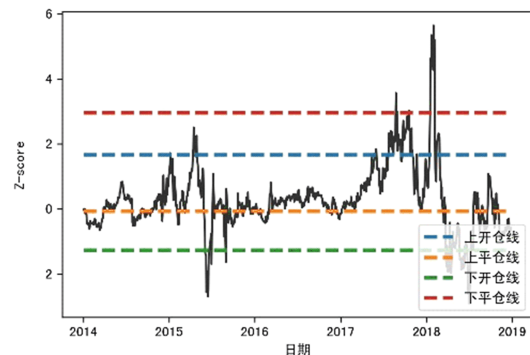


图 10 测试期 Z-score 及最优止损阈值
Fig. 10 Z-score and optimal stop loss threshold in test period

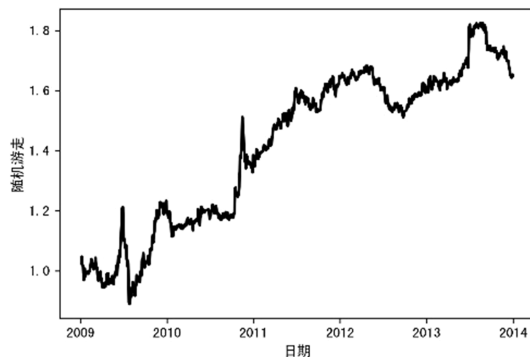


图 11 训练期随机游走值
Fig. 11 Random walk in training period

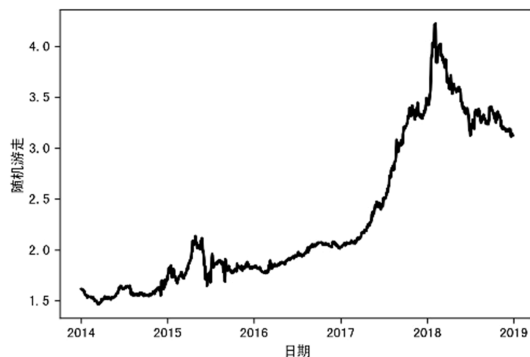


图 12 测试期随机游走值
Fig. 12 Random walk in test period

最后一个原因是模型选股的设置不同.文献

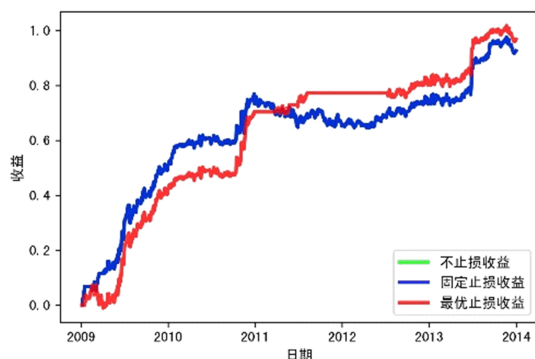


图 13 中固定止损收益曲线与不止损收益曲线基本重合, 是回测中未触发固定止损导致。

图 13 训练期三个模型收益图

Fig. 13 Benefits of three models in training period

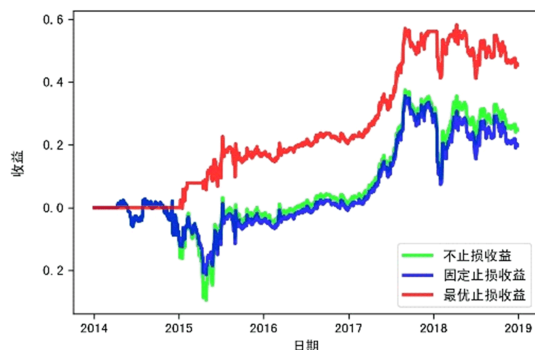


图 14 测试期三个模型收益图

Fig. 14 Benefits of three models in test period

[30]选取训练期夏普比率排名前 10 的配对股票作为股票池, 而本文选择市场所有符合部分协整的股票对作为股票池, 缺少了对训练期股票表现的筛选, 给测试期带来了更多的不确定性。

本文针对模型回测过程中发现的问题给出进行部分协整配对交易的建议: 首先, 在筛选配对股票阶段可通过控制随机游走部分的比例、夏普比率排序等手段筛选配对股票, 从而降低未来不可控因素带来的影响; 其次, 尽量在高频数据下进行股票对的筛选和回测, 控制回测时间。未来可考虑分钟级别、Tick 级别的回测, 通过滑动窗口等方式控制时间间隔, 减少参数失灵的现象。相信以上改进能够进一步发挥模型的特色, 提升模型的整体表现。

另外, 由图 13 和 14 可见, 尽管在测试期内受到若干影响, 本文提出的最优止损模型仍然维持了较高的收益和较低的风险, 能够在维持较高收益, 甚至提高整体收益的基础上, 有效控制和降低交易风险。

3 结论

本文通过遗传算法求解配对交易带止损的最优阈值, 并与固定止损和不止损条件下的最优阈值在符合协整和部分协整配对条件下的沪深 300 指数和中证 500 指数分行业成分股中进行实证检验。

实证检验结果表明, 本文提出的带止损条件的最优交易阈值设置使配对交易在保持较高水平的收益的基础上, 能够有效控制风险和损失。

本文的创新点主要有以下三点。首先, 提出基于遗传算法带止损条件的配对交易最优阈值, 并比较最优止损、固定止损和不止损三者在配对交易中的表现, 发现带止损条件的交易阈值设置使配对交易在保持较高水平的收益的基础上, 有效控制风险和损失。其次, 本文除了根据协整性筛选股票对, 还引入了近年提出的部分协整性检验, 在中国股市中进行回测验证, 针对回测的具体情况提出建议。最后, 与目前配对交易往往选择训练期表现前几名、市值前几名、或只选择一组或几组配对股票进行验证不同, 本文的实证检验涉及两个指数分行业成分股的所有符合配对标准的股票对, 其中符合协整性的股票对共 2273 组, 符合部分协整性的股票对共 1887 组, 使回测结果更具有说服力。

综上本文引入遗传算法计算带止损条件的最优交易阈值, 并在中国股市中进行了广泛而全面的检验。实证结果表明, 智能寻优算法得到的最优止损的设置可以在具有波动性和不确定性的市场中合理平衡收益和止损的关系, 能够做到在保持稳定收益的基础上有效控制风险。除针对常见的协整股票对进行检验外, 本文还将部分协整检验方法在中国市场予以推广, 并根据检验结果提出改进建议。未来还可将模型引入期货市场、海外市场等其他交易环境, 并研究模型在高频配对交易中的表现。

参考文献 (References)

- [1] VIDYAMURTHY G. Pairs Trading: Quantitative Methods and Analysis[M]. Hoboken, NJ: Wiley, 2004.
- [2] JUREK J W, YANG H. Dynamic portfolio selection in arbitrage[DB/OL]. [2020-02-01]. <http://ssrn.com/abstract=882536>.
- [3] SONG Q, ZHANG Q. An optimal pairs-trading rule[J]. Automatica, 2013, 49(10): 3007-3014.
- [4] EKSTR E, LINDBERG C, TYSK J. Optimal liquidation of a pairs trade[C]// Advanced Mathematical Methods for Finance. Berlin: Springer, 2011: 247-255.
- [5] MUDCHANATONGSUK S, PRIMBS J A, WONG W. Optimal pairs trading: A stochastic control approach[C]// 2008 American Control Conference. IEEE, 2008: 1035-1039.
- [6] KUO K, LUU P, NGUYEN D, et al. Pairs trading: An optimal selling rule[J]. Mathematical Control and Related Fields, 2005, 5(3): 489-499.
- [7] TIE J, ZHANG H, ZHANG Q. An optimal strategy for pairs trading under geometric Brownian motions [J]. Journal of Optimization Theory and Applications, 2018, 179(2): 654-675.
- [8] NGO M M, PHAM H. Optimal switching for pairs trading rule: A viscosity solutions approach[J]. Journal of Mathematical Analysis and Applications, 2016, 441(1): 403-425.
- [9] LARSSON S, LINDBERG C, WARFHEIMER M. Optimal closing of a pair trade with a model containing jumps[J]. Applications of Mathematics, 2013, 58(3): 249-268.
- [10] CARL L. Pairs trading with opportunity cost[J]. Journal of Applied Probability, 2014, 51(1): 282-286.
- [11] 杨艳军, 陈思岑. 基于高频数据的我国国债期货市场套

- 利研究[J]. 财务与金融, 2018, 172(2): 5-10.
- [12] CHEN C W S, CHEN M, CHEN S Y. Pairs trading via three-regime threshold autoregressive GARCH models [C]// *Modeling Dependence in Econometrics*. Berlin: Springer, 2014: 127-140.
- [13] HUCK N. Pairs selection and outranking: An application to the S&P 100 index[J]. *European Journal of Operational Research*, 2009, 196(2): 819-825.
- [14] HUCK N. Pairs trading and outranking: The multi-step-ahead forecasting case[J]. *European Journal of Operational Research*, 2010, 207(3): 1702-1716.
- [15] 龙奥明, 毕秀春, 张曙光. 基于 LSTM 神经网络的黑色金属期货套利策略模型[J]. *中国科学技术大学学报*, 2018, 48(2): 125-132.
- [16] 胡文伟, 胡建强, 李湛, 等. 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(2): 148-160.
- [17] 毕秀春, 刘博, 袁吕宁, 等. 带止损条件的配对交易最优阈值[J]. *系统科学与数学*, 2019, 39(7): 1117-1141.
- [18] ZHANG Q. Stock trading: An optimal selling rule[J]. *Mathematical Control and Related Fields*, 2015, 5(3): 489-499.
- [19] LINDBERG C. Pairs trading with opportunity cost[J]. *Journal of Applied Probability*, 2014, 51(1): 282-286.
- [20] 倪禾. 基于启发式遗传算法的指数追踪组合构建策略[J]. *系统工程理论与实践*, 2013, 33(10): 2645-2653.
- [21] 张鸿彦, 林辉, 姜彩楼. 用混合小波网络和遗传算法对期权定价的研究[J]. *系统工程学报*, 2010, 25(1): 43-49.
- [22] HUANG C F, LI H C. An evolutionary method for financial forecasting in microscopic high-speed trading environment [J]. *Computational Intelligence and Neuroscience*, 2017, 2017: 1-18.
- [23] HUANG C F, LIN C H, CHEN P C, et al. An improved genetic-based forecasting model for high-speed trading [C]// *2017 International Conference on Applied System Innovation*. IEEE, 2017: 1904-1907.
- [24] CLAVERIA O, MONTE E, TORRA S. Evolutionary computation for macroeconomic forecasting [J]. *Computational Economics*, 2017, 53(2): 833-849.
- [25] CLAVERIA O, MONTE E, TORRA S. Tracking economic growth by evolving expectations via genetic programming: A two-step approach [R]. Barcelona, Spain: The Research Institute of Applied Economics (IREA), 2018.
- [26] SAKS P, MARINGER D. Genetic programming in statistical arbitrage [C]// *Applications of Evolutionary Computing*. Berlin: Springer, 2008: 73-82.
- [27] HUANG C F, HSU C J, CHEN C C, et al. An intelligent model for pairs trading using genetic algorithms [C]// *Computational Intelligence and Neuroscience*. London: Hindawi Publishing Corporation, 2015: Article ID 939606.
- [28] CALDEIRA J F, MOURA G V. Selection of a portfolio of pairs based on cointegration: A statistical arbitrage strategy[J]. *Brazilian Rev Finance*, 2013, 11: 49-80.
- [29] JACOBS H, WEBER M. On the determinants of pairs trading profitability[J]. *Journal of Financial Markets*, 2015, 23: 75-97.
- [30] CLEGG M, KRAUSS C. Pairs trading with partial cointegration[J]. *Quantitative Finance*, 2018, 18(1): 121-138.
- [31] HOLLAND J H. Adaptation in natural and artificial systems[J]. *Ann Arbor*, 1992, 6(2): 126-137.
- [32] ENGLE R F, GRANGER C W J. Co-integration and error correction: Representation, estimation, and testing[J]. *Econometrica: Journal of the Econometric Society*, 1987, 55(2): 251-276.
- [33] JOHANSEN S. Statistical analysis of cointegration vectors [J]. *Journal of Economic Dynamics and Control*, 1988, 12(2-3): 231-254.
- [34] HUCK N, AFAWUBO K. Pairs trading and selection methods: Is cointegration superior? [J]. *Appl Econ*, 2015, 47: 599-613.