

基于参数调整的协整配对交易策略 理论模型及应用

张河生 闻岳春

(同济大学经济管理学院,上海 200070)

摘要:股指期货与融资融券的启动,开启了中国做空和对冲时代的大门,协整配对的套利交易将会扮演重要角色。在构建交易策略过程中,目前大多数交易者都在交易区间和止损区间的设置上采取经验值,典型的做法是分别取一倍和两倍的标准差。事实上,这种做法不能保证获得最大的套利利润,甚至会出现亏损。本文认为区间的参数选择是一个慎重抉择的过程,应该通过不断地调试以选择最优参数,以使该交易策略体系更好地发挥作用。

关键词:参数调整;协整;配对交易策略

中图分类号:F830.31

文献标识码:B

文章编号:1674-0017-2013(1)-0011-06

一、引言

20世纪八十年代,Morgan Stanley 的团队利用一系列基于统计学与数学构建的量化投资策略获得了巨额利润,其中就包括配对交易策略。此后配对交易作为一种市场中性投资策略,受到了机构投资者和对冲基金的广泛认可。

基于统计套利的配对交易策略是指从市场上找出历史价格走势相近的资产进行配对,当配对的资产价格差偏离历史均值时,则做空价格较高的资产,同时买进价格较低的资产,等待它们回归到长期均衡关系,由此赚取两个资产价格收敛的报酬,其属于一种市场中性策略。配对交易策略要求配对资产的价格序列具有长期稳定的关系,捕捉的是长期均衡中两个资产的短暂价格偏离所带来的套利机会。

在实证研究中,很多学者证明协整配对交易可在我国获得盈利,并提出了改进,如构建 GARCH 模型来刻画价差序列的时变方差性,并表明考虑异方差的配对交易效果更加稳健,亏损的风险大大减少。但是前人的研究对于交易区间、止损区间的设定都是采用历史经验值,事实证明,这种做法可能导致亏损。交易策略需要进行参数调试以获得最优的参数,使其收益最大化。对于配对交易,其交易区间、止损区间的最优参数也需要不断反复地测试才能得到。

近几年来,国内的学者和机构研究人员也开始对协整配对交易进行了专门研究,大多数实证结果表明协整配对套利在我国是可以盈利的。仇中群(2009)、于玮婷(2011)分别对沪深 300 股指期货的高频仿真交易数据及融资融券标的股票价格数据运用协整的统计套利实证分析,表明基于协整的统计套利策略模型是有效的。刘富兵、蒋瑛琨、何苗(2010)与李世伟(2011)考虑了异方差性及 ARCH 效应,建立基于 GARCH 模型的协整套利策略,实证分析表明该模型能取得更好的套利效果。

由此可见,运用协整理论可以在我国股票市场和股指期货市场进行统计套利,并在考虑异方差问题之后能够获得更加稳健的盈利。但是仅仅考虑异方差仍然不能达到最令人满意的结果,这就涉及到区间参数的选择问题,本文将就此问题进行深入研究。

二、理论分析

(一)协整关系

将两个非平稳的时间序列做线性回归: $Y_t = \lambda X_t + \mu_t$,在统计学上可能得到显著性差异,即回归是有效的,而事实上,这个时间序列之间却不存在任何关系,这种现象称为谬误回归,也称为“伪回归”。可见,两个非平稳时间序列经常可能出现谬误回归现象,这会影响到我们的研究。但是如果可以证明回归式中的残差项 μ_t 是平稳的,则可以保证 Y_t 与 X_t 存在一种长期稳定的关系,我们称 X_t 与 Y_t 存在协整关系。

由此可以看出,检验两个时间序列是否存在协整关系的常用方法就是 Engle-Granger 检验法。其思路是

收稿日期:2012-9

作者简介:张河生(1987.8-),男,湖北阳新人,硕士研究生,现就读于同济大学经济与管理学院。

闻岳春(1965.12-),男,浙江余姚人,教授,博士生导师,现供职于同济大学经济与管理学院。

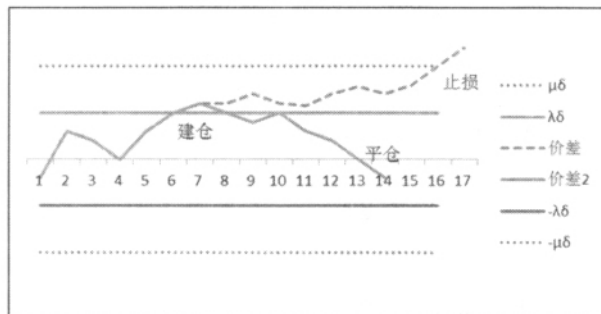
注:本文是国家自然科学基金项目(项目批准号:71273190)、教育部人文社科研究规划项目(编号:10YJA790196)的部分成果。

首先将两个时间序列做回归,然后针对残差项做平稳性检验,如果是平稳的就认为两者具有协整关系,否则就不存在协整关系。

(二) 配对交易策略

如果能够得到两个资产的价格具有协整关系,我们就可以利用这个协整关系做统计套利。首先要构建交易区间和止损区间。其原理是价差序列(也即回归式中的残差序列是平稳的:均值不变、方差为常数),具有均值回复的趋势。将其去中心化后(也即将残差减去其均值后所得到的残差序列),均值就是 0,价差就有回复到 0 的趋势。

图 1 交易决策



由图 1,我们可以构建 $(-\lambda\delta, +\lambda\delta)$ 的交易区间以及 $(-\mu\delta, +\mu\delta)$ 的止损区间(δ 为残差 μ_t 的标准差, λ, μ 为选取的系数)。交易区间就是价差突破区间上下边界时即开始建仓的波动区间,止损区间就是价差突破上下边界时平仓止损的波动区间,以防止更大的亏损。因为价差有均值回复的趋势,如果偏离过大,就属于不正常的范围,会带来潜在的亏损。例如,当价差序列在第 6 个时刻点突破上边界 $+\lambda\delta$ 时,开始构建一个组合,做多一种资产同时做空另一种资产。我们预期这个价差在未来某一点会恢复到 0 值附近,即价差将会收敛,当价差在 13 时刻时恢复到 0 就平仓,完成一个交易回合;如果价差没有收敛而是继续上扬,在第 16 时刻突破了我们设立的止损区间 $(-\mu\delta, +\mu\delta)$ 的上边界,也要进行平仓止损。

如果价差序列开始突破到下边界 $-\lambda\delta$,就建立一个反方向的头寸,同样是收敛到 0 值时平仓;如果未收敛而继续下行突破 $-\mu\delta$ 时进行平仓止损。

(三) 考虑异方差性

由于在做线性回归时都是在五大经典假设的前提下,其中之一就是假设残差的方差不变性。但事实证明,金融资产价格的时间序列并不满足这一假设,会存在所谓的异方差性,这就意味着我们设立的交易区间与止损区间并不是两条平行的直线,而是由实时的价格数据决定的一个波动的区间。

GARCH 模型是一个专门针对金融数据所量身订做的回归模型,在普通回归模型基础上,GARCH 模型对误差的方差进行了进一步的建模。一般的 GARCH 模型可以用公式(1)和(2)表示,其中 X_t 为 t 时刻价差的值, β_i, β_j 为滞后项的系数, e_t 为价差序列自回归的残差项, σ_t^2 为价差序列 t 时刻的方差, ε 为服从标准正态分布的随机项,K、C 值为常数项。

$$\text{条件均值方程: } X_t = e_t + \beta_i \sum_{i=1}^p e_{t-i} + \beta_j \sum_{j=1}^q X_{t-j} + k \quad (1)$$

$$\text{条件方差方程: } \sigma_t^2 = \beta_i \sum_{i=1}^m \sigma_{t-i}^2 + \beta_j \sum_{j=1}^n \varepsilon_{t-j} + C \quad (2)$$

(四) 参数选择的重要性

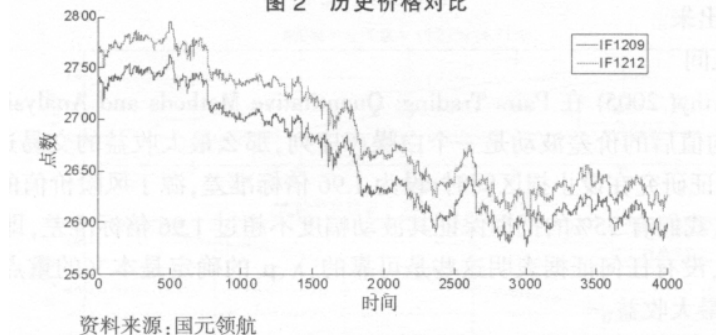
即使考虑了异方差,交易区间与止损区间变成了波动的通道,固定参数下,通道的宽度也不会有特别明显的变化。比如说正负 1 倍的标准差作为交易区间上下界,考虑异方差后的交易通道的宽度一般不会超过 6 倍的标准差。但价差序列在保持平稳的前提下仍然可能会有很大的波动,并在 0 均值线一边徘徊一段时间,这样就会造成频繁交易,考虑成本后就可能大幅降低盈利,甚至出现亏损。而且如果是在止损区间边界徘徊的话,考虑成本后会造成巨大的亏损。因此有必要调整区间的参数使通道宽度合适,让价差序列尽量少接触止损通道的边界,接触交易区间边界次数要在频繁交易带来的高额成本以及多次交易获得盈利之间获得平衡。参数的选择是交易策略有效的一个重要因素,不能轻率地凭经验值确定,而应该通过对历史数据不断地试验以获得最优的参数值。

三、实证分析

本文基于 Matlab 软件运用协整理论对股指期货配对交易做实证分析,选取的是 IF1209 与 IF1212(分别表示 2012 年 9 月到期和 2012 年 12 月到期)两个品种 2012 年 5 月 4 日 10:06 到 2012 年 5 月 24 日 15:15 的分钟数据,各有 4000 个数据。将 IF1209 的数据放在 series 变量的第一列,IF1212 的数据放在 series 变量的第二列,创建一个 4000*2 的矩阵变量 series(当然也可以选取 IF1206 和 IF1207,合约对的选择可以随意)。

两个合约的历史收盘价格数据如图 2 所示,初步发现两者的走势有一些类似之处,可能存在协整关系。

图 2 历史价格对比



资料来源:国元领航

(一)协整关系检验

分别对两个合约产品的时间序列做 ADF 检验,在 Matlab 里需要用到 adfresst 函数,返回的 h 值为 0,证明两个产品的历史价格数据都是非平稳的。然后对两个时间序列数据做一阶差分后再做平稳性检验,证明两个序列是一阶单整的,可能存在协整关系。

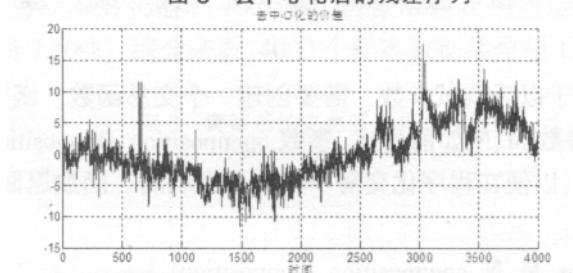
Matlab 里用 egcitest 函数对两个序列的协整关系进行检验。其步骤是:首先将 series 的第一列数据对第二列做回归,即让 IF1209 的数据对 IF1212 的数据做回归,得到公式(3),然后检验回归式中残差值是否具有平稳性,返回的 h 值为 1,证明两者具有协整关系。

$$IF1209 = 0.9585 * IF1212 + 80.7276 + \varepsilon_t \quad (3)$$

(二)考虑异方差

将公式(3)中得到的残差序列中心化后(图 3)做 engle test 检验,我们用 archtest 函数,返回的 h 值为 1,表示残差具有异方差性。

图 3 去中心化后的残差序列



考虑到异方差后,我们需要对残差序列建模用来刻画残差的异方差性,这意味着我们后面所构建的交易区间和止损区间不再是两条平行的直线,而是波动的平行通道。首先对 indicate 这个变量用 ARMA 模型的 AIC 定阶法则定阶,得到 R=1、M=1 时,AIC 值最小,因此可以判断 indicate 序列属于 ARMA(1,1)模型;其次用 GARCH(1,1)模型来刻画 indicate 序列的时变方差,这是基于通常的做法(一般认为 GARCH(1,1)模型能够很好地刻画时变方差)。于是本次案例中运用 ARMA(1,1)-GARCH(1,1)来刻画残差序列,并且运用于后面的交易区间和止损区间的构建,见公式(4)、公式(5)。

$$X_t = e_t - 0.7932 * e_{t-1} + 0.9748 * X_{t-1} - 0.0005 \quad (4)$$

$$\sigma^2 = 0.0302 * \varepsilon_{t-1}^2 + 0.9504 * \sigma_{t-1}^2 + 0.0055 \quad (5)$$

(三)交易策略

1、滚动取样交易

交易是基于前面 M 个时间段的数据,每隔 N 个时间段来进行的。如果前面 M 个时间段内两个时间序列

存在协整关系,就可以按照上文阐述的策略进行交易。假设 $M=420, N=60$,也就是,从现在开始进行协整配对交易,我们参考的是前面 420 分钟(也即前 7 个小时)的历史数据来判断有无协整关系,60 分钟(也即 1 个小时)以后,我们重新换一批数据,用 1 个小时以后的之前 7 个小时数据进行协整配对交易,也就是用现在的前 6 个小时和未来 1 个小时的数据进行协整来判断交易机会。之所以用滚动方式不断调整用于判断协整关系的历史数据,是因为协整关系本身的不稳定性。两个时间序列的协整关系可能随着时间的推移突然消失,一般是由于某个变量发生了巨大冲击事件,使得时间序列数据彻底改变了以往的路径;同时,协整关系也可以在之前没有协整关系的两个时间序列间突然存在,这是因为随着历史数据的增加,两个具有某种联系的变量的协整关系才逐渐显示出来。

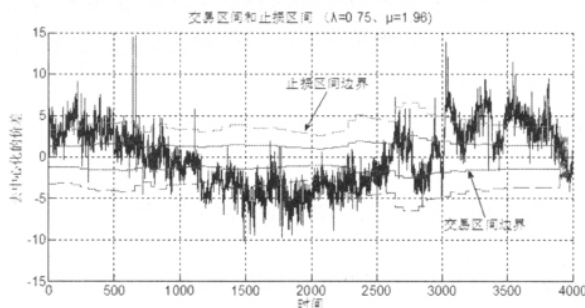
2、交易区间、止损区间

Ganapathy Vidyamurthy(2005) 在 Pairs Trading: Quantitative Methods and Analysis 一书中通过随机模拟得到如下规律:假设去均值后的价差波动是一个白噪声序列,那么最大收益的交易边界条件是正负 0.75 倍的标准差,即 $\lambda=0.75$ 。实证研究在设止损区间时,设为 1.96 倍标准差,源于风险价值的思想,假设去中心化的对数价差服从正态分布,我们有 95% 的把握保证其波动幅度不超过 1.96 倍标准差,即 $\mu=1.96$ 。

但这只是经验数据,没有任何证据表明这些是可靠的, λ, μ 的确定是本文的重点,我们将在后文着重分析如何调试参数以获得最大收益。

考虑异方差之后的交易区间、止损区间如图 4 所示。根据我们的交易策略,是每隔 60 分钟观察前 420 分钟的数据进行协整配对交易,所以交易的初始数据是样本数据的第 421 个。前 420 个数据的区间通道是由程序自动画出,但没有用处,即使价差不断触碰区间上下限也不会有任何交易指令发出。同理,在 3960 个数据之后的区间也是没有用处的,即使图 4 显示此时的价差不断触碰区间的上下限。

图 4 交易区间与止损区间



3、交易函数

在构建交易策略时。为了便于以后调试参数,需要创建一个交易函数。函数 pairs 具有五个输入变量,series 是去中心化的价差序列,参数 M, N 如前所述,参数 openposition、stopposition 就是 λ, μ ,这几个参数需要通过样本内数据不断运行调试,以便在程序化交易中获得最大利润。函数返回一个交易信号,并做出损益图。

```
function varargout =pairs(series, M, N, openposition, stopposition)
```

4、成本与损益的计算

期货公司进行程序化交易时需要考虑的成本有两种:滑点与手续费。滑点是指下单的点位与最后成交的点位之间的差距。滑点的产生有可能是由于网络延迟或当时行情波动剧烈,也有可能是因为不正规的交易商故意做手脚。任何一个平台都会存在这个问题,这取决于平台软件的质量以及服务器的稳定性等。期货公司的滑点一般为 0.4 个点。手续费也是机构投资者必不可少的一项成本,沪深 300 股指期货合约对交易所来说,交割手续费标准为交割金额的万分之一;期货公司一般会高于交易所的标准,为万分之一点五。

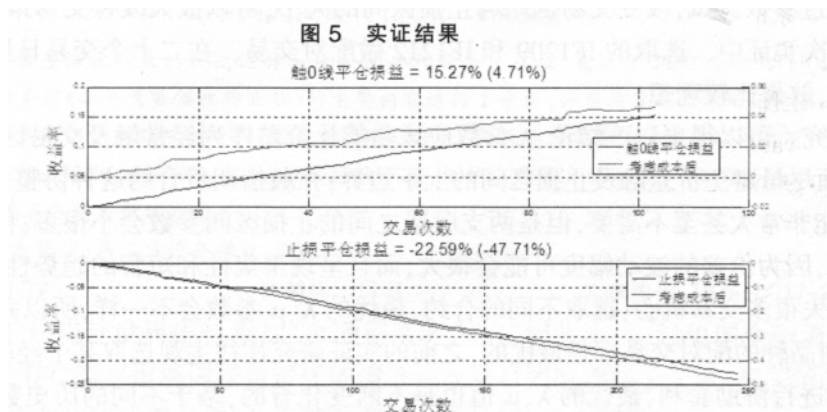
由于做的是套利交易,每一个交易回合涉及到两种合约:IF1209 和 IF1212 (下文中用 1 代表 IF1209,2 代表 IF1212),而每一回合中有两次交易,比如开仓时是做多 1 与做空 2,平仓时就是做空 1 与做多 2。同时,每次交易反方向合约数量的固定比例为 β ,即交易一手 1 就会按相反方向交易 β 手 2 (此例中 β 值为 0.9578,即公式(3)的贝塔系数)。

考虑滑点与手续费作为成本,成本就是交易的四种合约点数之和加上滑点后乘以手续费率,即[(空 1+ β * 空 2+多 1+ β * 多 2)+0.4*4]*0.00015;收益的计算是用交易的空单点数减去两次交易多单点数,即(空 1+ β * 空

2)-(多 $1+\beta^*$ 多 2)。由于期货交易是保证金交易,收益率的计算是用所得的损益除以所交的保证金,计算公式为: $(\text{空 } 1+\beta^* \text{ 空 } 2 + \text{多 } 1+\beta^* \text{ 多 } 2) \times 0.15$ 。

(四) 初步结果

初步结果如图 5 所示。根据上述交易策略,本次实证模拟中,去中心化的价差接触 0 线平仓次数为 101 次,累计盈利 15.27%,但是考虑交易成本后仅盈利 4.71%;价差触碰止损通道平仓的次数为 246 次,累计亏损 22.59%,考虑交易成本后亏损 47.71%。综合来看,4000 个样本数据共亏损 7.32%,但是考虑交易成本后,亏损 43%。

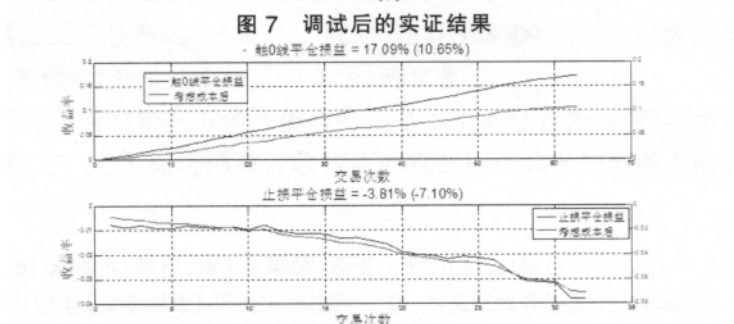
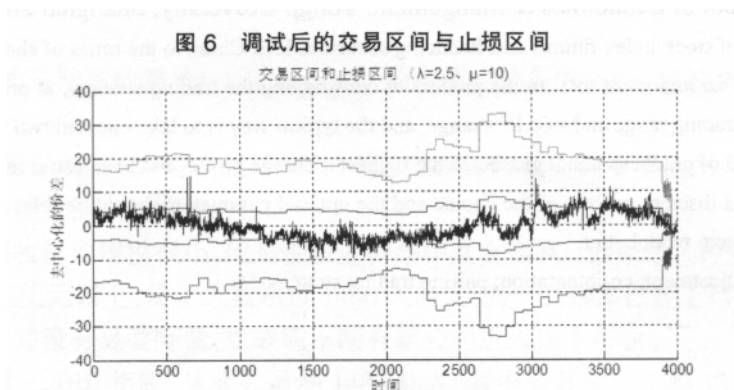


(五) 参数的调试及结果

由上所述,选择 $\lambda=0.75$ 、 $\mu=1.96$ 作为区间的参数,导致了亏损。从图 5 分析得到,因为交易次数太频繁导致成本高于盈利的点数;即使价差触 0 线的交易有一定盈利,但止损平仓导致的亏损因为交易次数的频繁而非常巨大,最终导致了整体的亏损。从图 4 可以看出,是由于价差不断地触碰交易区间和止损区间导致交易次数增加,于是考虑把 λ 、 μ 参数设的大些,从而使交易次数减少,特别是止损的次数要减少。但是如果把参数设得太大,区间过于宽广,价差无法触及区间边界,交易次数也就少,无法实现盈利。

由上述分析得出需着重调试的参数为 λ 、 μ ,于是假定 M 、 N 为默认值 420 和 60,假设调试参数 λ 的范围为 0.5 到 4,步长为 0.25, μ 的范围为 2 到 10,步长为 1,共有 $15 \times 9 = 135$ 组值,依次代入 pairs 函数,将能够使考虑成本后的收益最大的那组值作为最优参数。

如图 6 和图 7 所示,得到最优的 λ 、 μ 值分别为 2.5 和 10,去中心化的价差接触 0 线平仓次数为 60 次,累计盈利 17.09%,考虑交易成本后仍盈利 10.65%;价差触碰止损通道平仓的次数为 27 次,累计亏损 3.81%,考虑交易成本后亏损 7.10%。综合来看,4000 个样本数据共盈利 13.28%,即使考虑交易成本后仍然盈利 3.55%。



由此可以看出 λ 、 μ 参数调试的效果非常明显,其实这种变化也是很好理解的。交易区间与止损区间的变化使得交易触发的条件发生了极大的改变,调试后价差触 0 线平仓的次数由 101 次变为 60 次,盈利由 15.27% 上升为 17.09%;考虑交易成本后,这种上升更为明显,调试前盈利为 4.71%,调试后盈利上升为 10.65%。在止损平仓这一块,调试前后交易次数由 246 次变为 27 次,大大减少了这一部分的亏损,由 22.59% 降到 3.81%;考虑交易成本后,这一部分的效果更为明显,前后亏损由 47.71% 降到 7.10%。

四、结论

实证表明通过参数调试,改变交易区间与止损区间的大小,可以极大改善交易策略的运行结果,并最终实现盈利。在此次实证中,选取的 IF1209 和 IF1212 做配对交易,在二十个交易日里扣除成本后实现盈利 1.98 万元人民币,效果比较理想。

经过实证研究,可以得出以下结论: λ 参数应该是能让价差序列经常触及交易区间边界的值, μ 参数应尽量大一些,从而尽量避免价差触及止损区间的上下边界;在股指期货合约这种协整关系明显且稳定的配对中,止损区间可能非常大甚至不需要,但是两支股票之间的止损区间参数会小很多;协整配对交易一定要考虑异方差的情况,因为价差的波动幅度可能会很大,而且呈现聚集性和短暂的趋势性,固定的区间会导致巨大的亏损或者丧失很多交易机会;选取不同的合约,最优的 λ 、 μ 参数会不一样,所以并不存在一组这样的 λ 、 μ 值,它对于任何品种的配对交易都是最优的,之前的实证研究往往主观选取某个经验值,这是不合理的;在相同的产品对中进行协助套利,最优的 λ 、 μ 值也是不断变化着的,基于不同的历史数据得出的最优 λ 、 μ 值也是稍微有些差异的,这说明需要根据实时数据动态调整 λ 、 μ 值。

参考文献

- [1]仇中群.基于协整的股指期货套利研究[D].安徽:中国科技大学,硕士学位论文,2009。
- [2]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,(2):198-202。
- [3]刘富兵,蒋瑛琨,何苗.股指期货跨期套利研究——股指期货系列报告之二十一[R].国泰君安证券研究所,2010。
- [4]魏奋.运用 Matlab 开发协整和配对交易策略.http://www.mathworks.cn,2011-10-9。
- [5]于玮婷.基于协整方法的统计套利策略的实证分析[J].科学决策,2011,(3):70-85。

The Co-integration Pairing Trading Strategy Based on the Parameter Adjustment: the Theory Model and Application

ZHANG Hesheng Wen Yuechun

(School of Economics & Management, Tongji University, Shanghai 200070)

Abstract: The start-up of stock index futures and financing securities lead China to the times of short and hedge, and co-integration pairing trading will play an important role. In the process of constructing the trading strategy, at present, most traders use experience value when setting the trading range and stop loss range, and the typical way is to take one and two times of the standard deviation respectively. In fact, this kind of practice cannot guarantee the biggest arbitrage profits and even cause losses. The paper thinks that the range parameter selection is a discreet process of the choice and the optimal parameters should be selected constantly by debugging to make the trading strategy system work better.

Keywords: parameter adjustment; co-integration; pairing trading strategy

责任编辑、校对:焦少飞