

2015-8-19

金融工程 (专题报告)

金融工程

变点理论在择时中的应用：CUSUM 控制图

分析师 范辛亭

☎ (8621)68751859

✉ fanxt@cjsc.com.cn

执业证书编号：S0490510120008

联系人 刘胜利

☎ (8621)68751787

✉ liusl4@cjsc.com.cn

联系人 杨靖凤

☎ (8621) 68751636

✉ yangjf@cjsc.com.cn

相关研究

《【长江金工】平行线法择时》2015-6-9

报告要点

■ 变点理论

变点理论是统计学中的一个经典分支，其基本定义是在一个序列或过程中，当某个统计特性（分布类型、分布参数）在某时间点受系统性因素而非偶然性因素影响发生变化，我们就称该时间点为变点。变点识别即利用统计量或统计方法将该变点位置估计出来。

■ CUSUM 图

CUSUM 图作为工业应用检验变点的三大控制图之一，其利用假设检验和极大似然估计的相关统计原理，构建累积和统计量，不断累积观察值与基线水平的差值，放大观察数据出现的波动，从而更加迅速敏感地探测到微小的异常情况，检验出变点位置。其最大的特点是对系统性变化的敏感性，不需要积累太多的样本，因而能较好的控制风险。

■ 择时原理

本文基于变点 CUSUM 图的基本原理，以股价对数收益率符合局部正态分布为假设前提，构建 CUSUM 的上下统计量，一旦统计量突破阈值即判断出现变点。以股价上升时对数收益率出现上升变点作为买入时机，以股价下降时对数收益率出现下降变点作为卖出时机。统计量中的两个参数，允偏量 k 设置为动态变化自适应的形式，阈值 h 则需要根据直观进行设定。

■ 择时效果

固定参数下，对于不同行情、不同指数都取得了不错的择时效果，夏普比率都得到比较明显的提高，胜率基本在 50% 以上，CUSUM 择时具有较强的稳健性和可操作性；最优参数下效果明显，均能较大程度地跑赢买入持有策略；加入做空机制和高频交易的股指期货交易中，IF1506 和 IF1507 年化收益均达到 180% 以上，胜率在 60% 左右。

目录

一、变点检测	3
二、CUSUM 控制图	4
三、CUSUM 统计量择时	7
四、交易策略及结果	13
五、总结	18

图表目录

图 1: 变点的误检和漏检	3
图 2: 布林带 (休哈特控制图)	4
图 3: CUSUM 统计量的具体应用	7
图 4: k 值的自适应算法	9
图 5: 不同样本距离下均值 μ /标准差 σ 估计值的分布	10
图 6: 对数收益率变点与股价走势	11
图 7: 行情拐点判别	11
图 8: CUSUM 算法流程图	12
图 9: 上证综指趋势行情 CUSUM 择时效果	13
图 10: 上证综指震荡行情 CUSUM 择时效果	14
图 11: 沪深 300 择时效果	15
图 12: 上证 50 择时效果	15
图 13: 中证 500 择时效果	15
图 14: 沪深 300 最优参数择时效果($h=4.8$)	16
图 15: 上证 50 最优参数择时效果($h=2.7$)	16
图 16: 中证 500 最优参数择时效果 ($h=0.7$)	17
图 17: IF1506 择时效果	17
图 18: IF1507 择时效果	18
表 1: CUSUM 总结	7
表 2: 不同考察对象自相关性比较	8
表 3: 不同行情择时效果比较	14
表 4: 股指期货择时效果比较	18

一、变点检测

1. 什么是变点

在一个序列或过程中，当某个统计特性（分布类型、分布参数）在某时间点受系统性因素而非偶然性因素影响发生变化，我们就称该时间点为变点，具体定义如下：

假设存在一个数据集 $\{x_n\}$ ，每个数据观测值相互独立，如果在某一时刻 t ，模型中的某个或某些变量突然发生了变化，即存在 t_k ，使得当 $t < t_k$ 时 X 服从分布函数 F_1 ，当 $t > t_k$ 时， X 服从分布函数 F_2 ($F_2 \neq F_1$)，则 t_k 为该数据集的变点。

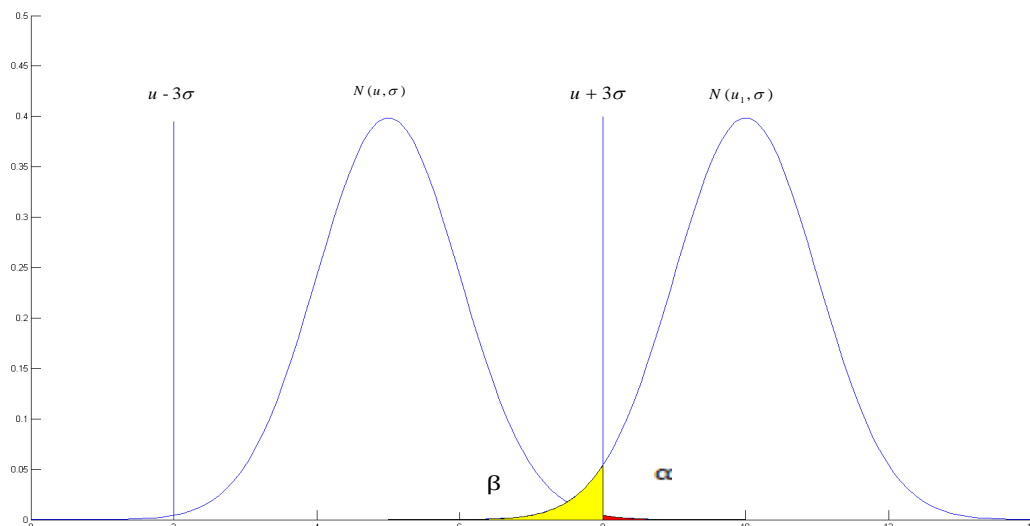
变点识别即利用一定的统计指标或统计方法，对时间序列的状态进行观测，以便准确有效的估计出变点的位置。变点问题作为统计学中的一个重要课题，最早应用于工业质量控制，后延展到金融经济、计算机、气象学、流行病学等多个领域。

2. 金融序列的变点

金融市场上很多时序数据容易受到各种外在因素的影响，产生收益率的突变，因而存在很大的不确定性。突变主要分为两种，系统性变化和偶然性变化。我们希望能够根据金融序列侦测系统性变化，并剔除偶然性变化的影响，但在实际操作中很容易出现误判和漏判的情况。

误判即错误地将偶然性误差当成系统性误差检测出来，将假的弄成真的，即统计中的第一类错误（阿尔法错误）；漏判即错误地将系统性误差当成偶然性误差过滤掉，将真的当成假的，这就是我们所说的第二类错误（贝塔错误）。

图 1：变点的误检和漏检



资料来源：长江证券研究部

准确有效地衡量金融序列的突变点，对于投资工作具有十分重大的意义。首先，作为一种技术指标，处理大规模的金融数据，对市场指标的异动及时作出反应和指示；其次，对于不同变点区间的金融序列数据，进行模型参数的调整，减少建模的不准确或错误建模；最后，应用于交易择时策略，对投资的时点进行一定的判断。

3. 变点研究方法类别

自 20 世纪 70 年代以来，许多统计学家投入到了变点问题研究领域，并取得了部分成果，一些估计和检测变点问题的方法也不断发展完善，如累计平方和方法、迭代累计平方和方法、Bayes 方法、极大似然法、局部比较法、小波分析法等。

从监测目的来看，分为事中变点（连续抽样）和事后变点（固定样本），前者指连续观察某一随机过程，监测到变点时停止检验，不运用到未来数据，主要用于事件预警，后者从已获得的时序数据中检测过去的变点位置，主要用作历史检验。

从监测内容来看，分为单变点研究和多变点研究。目前大多数研究集中于单变点区域，即假设研究的时序数据中至多只有一个变点。但实际情况一般都存在多个变点，由此衍生了多变点的研究问题，但在研究阶段，大部分学者都是先确定变点数目，再探究变点位置，由此研究方法大打折扣。

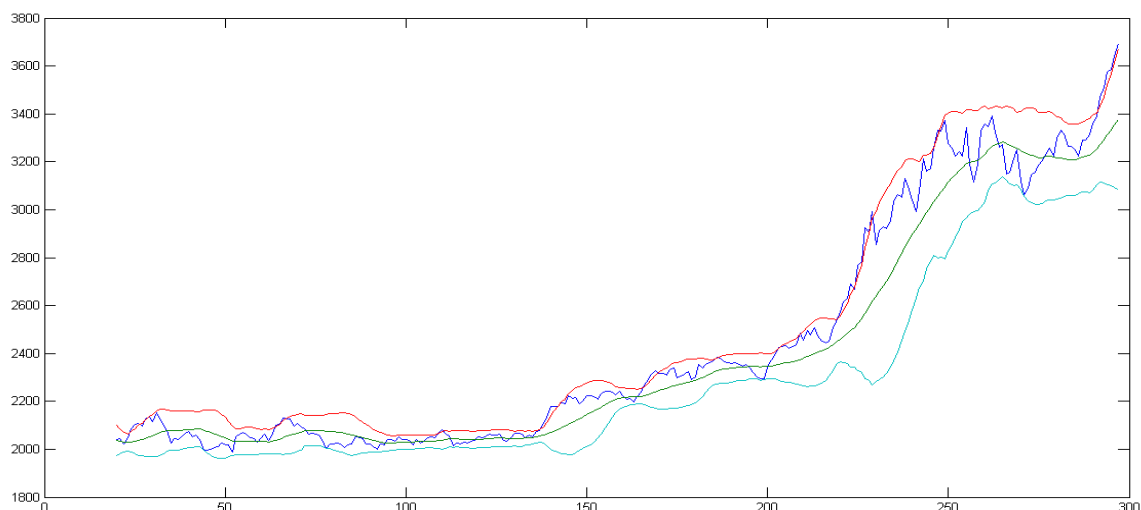
二、CUSUM 控制图

1. CUSUM 简介

传统的变点检测大多基于统计原理，包括最小二乘法、极大似然法、贝叶斯方法等。随着统计控制过程（SPC）的兴起，控制图技术迅速发展开来，并广泛应用于工业生产之中。控制图即运用典型的数理统计方法，判断产品是否偏离典型分布，甄别是否存在异动。

目前在统计过程控制中最成熟的是三大控制图：休哈特控制图（Shewhart）、累积和控制图（CUSUM）以及指数加权滑动平均控制图（EWMA）。其中最传统的休哈特控制图（ $\mu-3\sigma$, $\mu+3\sigma$ ）在金融投资中的应用是布林带，EWMA 则同指数加权均线比较接近。

图 2：布林带（休哈特控制图）



资料来源：长江证券研究部

累积和图不使用原始数据、均值、极差或标准差记录数据，取而代之的是有代表性的近似数，通过不断累积偏差，并将微小异动不断放大，从而对系统偏离稳态的情况进行预警。

基本上，使用累积和图的原因可以归结为敏感性，能够更加灵敏地侦测均值的微小偏移，这在样本数偏小时优势尤为明显。敏感性高带来的另外一个好处是可以降低风险。

2. CUSUM 原理

CUSUM 控制图的设计思想是对信息加以累积，将过程的小偏移累加起来，达到放大的结果，从而提高检验小偏移的灵敏度。CUSUM 作为一个统计量，其由来具有严格的数学推理，总的来说，是一个变点假设检验通过极大似然法推导得到的统计量。

令 x_1, x_2, \dots, x_t 为独立的 $N(0,1)$ 同分布， $x_{t+1}, x_{t+2}, x_{t+3}, \dots$ 为独立的 $N(\delta, 1)$ 同分布，其中 t 为未知变点，对于给定的观察序列 $x_1, x_2, x_3, \dots, x_n$ ，假设 $t = v (v < n)$ 。如此构成一个假设检验问题：

$$H_1: X_i \sim N(0,1), i = 1, 2, \dots, n$$

$$H_1: \exists v < \infty, s.t. X_i \sim N(0,1), i = 1, 2, \dots, v; X_i \sim N(\delta, 1), i = v, v+1, \dots, n$$

则似然比统计量为(以 $\phi(\cdot)$ 表示标准正态分布 $N(0,1)$ 的分布密度函数)：

$$L_{n,v} = \frac{\prod_{i=1}^v \phi(x_i) \prod_{i=v+1}^n \phi(x_i - \delta)}{\prod_{i=1}^n \phi(x_i)} = \frac{\prod_{i=v+1}^n \phi(x_i - \delta)}{\prod_{i=v+1}^n \phi(x_i)} = \exp \left\{ \delta \sum_{i=v+1}^n \left(x_i - \frac{\delta}{2} \right) \right\}$$

对数化为

$$A_{n,v} = \ln L_{n,v} = \delta \sum_{i=v+1}^n \left(x_i - \frac{\delta}{2} \right)$$

假设变量 x_1, x_2, \dots, x_t 与 $x_{t+1}, x_{t+2}, x_{t+3}, \dots$ 有偏移，则其对数似然统计量为

$$\Lambda_n = \max_{1 \leq v < n} \Lambda_{n,v} = \max \left\{ \delta \sum_{i=v+1}^n \left(x_i - \frac{\delta}{2} \right) \right\}$$

若我们检验的为向上偏移，即 $\delta > 0$ ，上述的对数似然统计量等价于下面统计量

$$Z_n = \max_{1 \leq v < n} \sum_{i=v+1}^n \left(x_i - \frac{\delta}{2} \right)$$

设 $n-1$ 个观测值没有均值偏移，即

$$Z_i \leq h, i = 1, 2, \dots, n-1$$

h 为门限。如果在时刻 n ，满足

$$\begin{cases} x_n - \delta/2 > h \\ x_n + x_{n-1} - \delta > h \\ x_n + x_{n-1} + x_{n-2} - 3\delta/2 > h \\ \vdots \\ x_n + x_{n-1} + \dots + x_1 - n\delta/2 > h \end{cases}$$

则这个过程发生了均值偏移。

记 $\tilde{x}_i = x_i - \frac{\delta}{2}$ ，且 $\tilde{x}_0 = 0$ ，并记 $\tilde{S}_k = \sum_{i=0}^k \tilde{x}_i$ ，于是可得：

$$\begin{aligned} Z_n - Z_{n-1} &= \tilde{x}_n - \min\{0, \tilde{S}_n - \min_{0 \leq v \leq n-1} \tilde{S}_v\} \\ &= \max\{\tilde{x}_n, \tilde{x}_n - \tilde{S}_n + \min_{0 \leq v \leq n-1} \tilde{S}_v\} \\ &= \max\{\tilde{x}_n, \min_{0 \leq v \leq n-1} \tilde{S}_v - \tilde{S}_n\} \\ &= \max\{\tilde{x}_n, -Z_{n-1}\} \end{aligned}$$

用不定参数 k 代替 $\frac{\delta}{2}$ ，就得到了 Z_n 的递推公式：

$$Z_n = \max\{0, Z_{n-1} + x_n - k\}, n = 1, 2, \dots$$

若设定报警门限为 $h > 0$ ，如果在第 n 个观察点满足

$$Z_n > h (Z_i \leq h, i = 1, 2, \dots, n-1)$$

则报警，确定在 n 以前的统计量发生了均值向上偏移，判断有系统性因素而非偶然性因素存在。

向下偏移也可通过类似的推导得到。

3. CUSUM 统计量的具体应用

CUSUM 方法在金融市场中也有初步应用，首先是针对股票市场连续收益率的均值漂移问题、结构性变点问题作了初步的研究，其实是部分学者根据 CUSUM 的警示作用创造了简单的技术交易策略。

由 CUSUM 的推导分析，我们得到 CUSUM 统计量，即

$$S_t = \sum_{i=1}^t (X_i - k) = S_{t-1} + (X_t - k)$$

其中 $S_0 = 0$ ， X_t 为金融序列。

令 $y_i = X_i - k$ ，其中 k 为允偏量。上下预警指标分别为

$$\begin{cases} C_i = \max(C_{i-1} + y_i, 0) \\ C'_i = \min(C'_{i-1} + y_i, 0) \end{cases}$$

若预警指标分别达到上下阈值 h ，则对应地产生上变点和下变点。

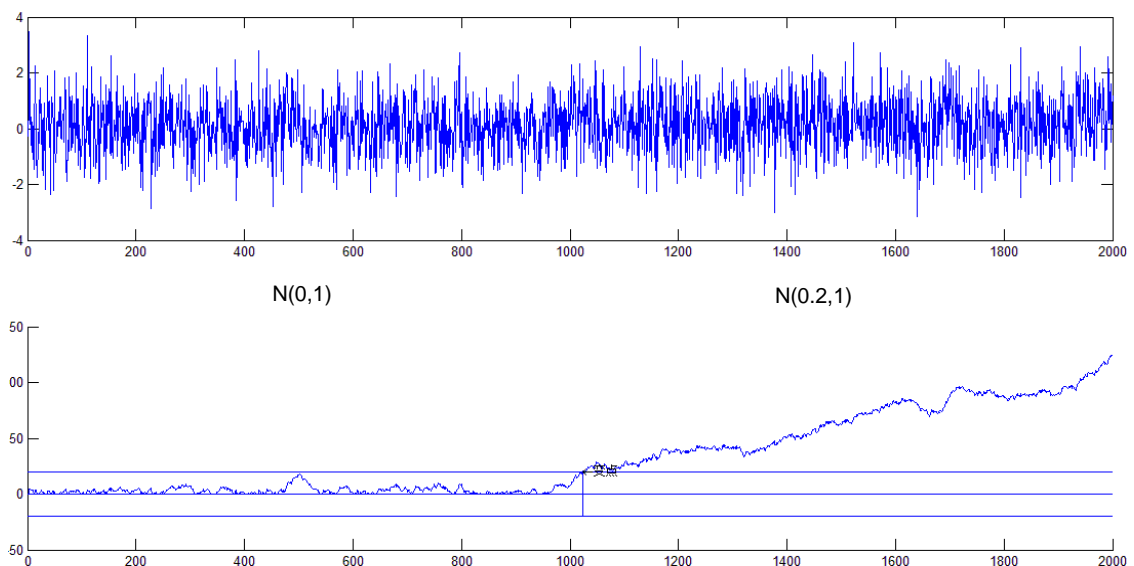
$$C_i \geq h, C_i' \leq -h$$

具体我们通过下面的图进行说明，以上 CUSUM 统计量和上变点为例。

选取 2000 个随机数，其中前 1000 个服从标准正态分布，后 1000 个服从均值偏离为 0.2 的，方差为 0 的正态分布。在序列分布已知的情况下，我们需要将分布开始变化的位置即变点估计出来。

在此我们设定允偏量 $k = 0.1$ ，阈值 $h = 20$ ，根据上述 CUSUM 统计量的计算得到得到该序列的上 CUSUM 统计量，并观察该累计值是否超过阈值。在不断累积过程中，CUSUM 控制图在不到 1100 的位置将这微小的差距识别了出来，效果相对还比较明显，一般统计方法很难甄别如此微小的变化。

图 3: CUSUM 统计量的具体应用



资料来源：长江证券研究部

三、CUSUM 统计量择时

1. CUSUM 法总结

经过上文对 CUSUM 统计量的推导和分析，总结 CUSUM 法如下：

表 1: CUSUM 总结

前提假设	序列之间无自相关性；数据符合正态分布
推导依据	假设检验、极大似然估计
应用方式	参数 CUSUM 法、非参数 CUSUM 法
变点触发条件	$y_i = X_i - k,$ $C_i = \max(C_{i-1} + y_i, 0)$ $C_i = \min(C_{i-1} + y_i, 0)$ $C_i \geq h, C_i \leq -h$

资料来源：长江证券研究部

CUSUM 算法具有严格的推导过程，其应用领域也愈趋广泛。若应用得当，应该能够对金融序列中的异动作出预警。

CUSUM 法通过设置允偏量 k 和阈值 h 来对序列的变点进行检测。 k 越大，在 Z_n 中出现正值的可能性就越小，因而累加起一个较大的正值的可能性就越低。门限 h 越大，误报率越低，但监测时间就越长。参数 k 、 h 的设置都涉及到误检率和漏检率的选择问题，在后文中，将针对参数选择进行具体讨论。

2. 数据处理

首先分析一下我们所选取的研究对象。

CUSUM 方法的一个前提是数据之间无自相关性，对于股价、股价变动额、收益率、对数收益率，我们随机选取上证综指某段时间的样本，对其自相关性进行了检测：

表 2：不同考察对象自相关性比较

滞后阶	股价	增长额	收益率	对数收益率
1	0.9986	0.0098	0.0086	0.0072
2	0.9972	-0.0347	-0.0207	-0.0192
3	0.9958	0.0583	0.0318	0.0328
4	0.9943	0.0599	0.0470	0.0473
5	0.9927	-0.0099	-0.0044	-0.0039

资料来源：长江证券研究部

股价的自相关性较高，而增长额、收益率、对数收益率的自相关性相对较低。为接近假设，首先排除股价指标，其他三个指标对应股价分别有不同的含义：

增长额稳定时 $x_t - x_{t-1} = p \Rightarrow x_t = pt + x_0$ 反映在股价上为斜线

增长率稳定时 $(x_t - x_{t-1}) / x_{t-1} = p \Rightarrow x_t = (p+1)^t x_0$ 反映在股价上为幂函数线

对数收益率稳定时 $\log(x_t / x_{t-1}) = p \Rightarrow x_t = e^{pt} x_0$ 反映在股价上为幂函数线

CUSUM 算法的基本思路是尽量使得考察对象贴近我们的假设和推导，因而一般会选择对数收益率作为考察对象。数据方面对数收益率已满足独立性这个假设前提，同时股票的对数收益率符合标准正态分布的假设也被广泛运用于金融学的经典理论之中，如期权定价 B-S 公式、资产组合理论、资产资本定价理论 CAPM、套利定价理论 APT 等。但就实际情况而言，这一假设并不是很成立。

本报告中将这个假设条件进行放松，假设股票的对数收益率在局部范围之内是比较稳定的，符合 $N(u, \delta)$ 分布，检验的变点区域则符合 $N(u_1, \delta)$ 分布，这在一定程度上是合理的。对于此类问题（变点的连续问题），令

$$Y_i = \frac{X_i - \mu}{\sigma}, \quad \delta = \frac{\mu_1 - \mu}{\sigma}$$

将金融序列转化为标准正态分布，则问题转化为

$$H_1: Y_i \sim N(0, 1), i = 1, 2, \dots, n$$

$$H_1: \exists v < \infty, s.t. Y_i \sim N(0, 1), i = 1, 2, \dots, v; Y_i \sim N(\delta, 1), i = v, v+1, \dots, n$$

如此就可以进行后续的操作和计算。本文在拟合参数的估计中，采取统计里常见的样本估计方式，均值取样本均值替代，标准差以样本标准差表示，则有：

$$u = \sum_{i=1}^m X_i$$

$$\sigma = \sqrt{\frac{1}{m-1} \sum_{i=1}^m (X_i - u)^2}$$

3. 参数设定

实际操作中，我们所处的环境是不断变化的，过去的最优参数不一定适用于现在和未来，同时在处理不同的对象中，参数也应该会有所差别，也就是说固定参数进行择时的方法无论从空间维度还是从时间维度讲都不太实用。本报告将重点研究参数自适应的方法。

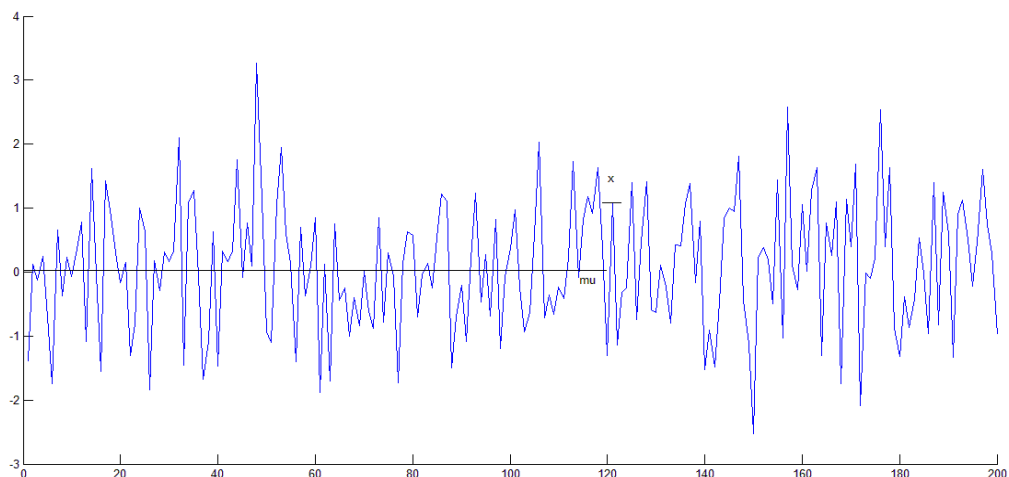
(1) 允偏量 k 值

我们讨论经过正态分布标准化后的假设检验，即变点前的数据服从 $N(0,1)$ 分布，变点后的数据服从 $N(\delta, 1)$ 分布。累积和控制图的算法设计是基于既定的观测值偏移量，即控制图参数 k，理论上 $k=\delta/2$ 时控制图的效果最好，其中， δ 即为观测值 X 的偏移量。但是实际情形下， δ 是未知量，也是我们需要检测的值，同时 δ 会随采样时间 t 变化，其大小决定了累积和控制图参数 k 的取值，并通过 k 的变化影响控制图的统计性能，所以有必要对 δ 进行动态预测和更新。

这里我们基于已检测的历史数据和待检测的点作一个动态设定。运用已知历史序列拟合参数 μ 和 σ ，再将历史数据和待检测点进行标准正态化，则历史数据的均值为 0，而待测点的值为 x。在某种程度上，可将标准化后的历史数据看作 $N(0,1)$ 分布，而将待检测点看作 $N(x, 1)$ 分布，按照 CUSUM 的推导，设置

$$k = \frac{x}{2}$$

图 4: k 值的自适应算法



资料来源：长江证券研究部

(2) 阈值 h

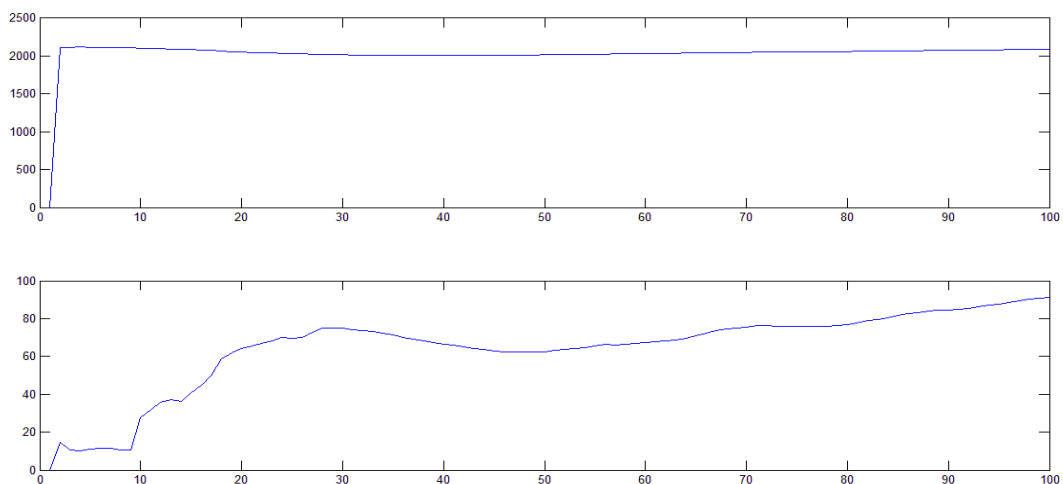
在参数检验中，为使得观测数据尽量复合假设，需要将观测到的数据正态标准化，所以阈值 h 的设置在一定程度上相当于标准正态分布参数的选定，参考布林带的做法，我们将 h 定为 2（布林带标准差的倍数一般为 2）。

此外，考虑到估计的均值和方差在样本量较小时，我们通过样本估计值得到的正态分布参数容易出现较大波动，有必要调整 h 的设置。在样本数少于 10 个时，我们将阈值的设置放宽

$$h_0 = 3h$$

其中 h 为样本数大于 10 时的我们设定的阈值。

图 5：不同样本距离下均值 μ /标准差 σ 估计值的分布



资料来源：长江证券研究部

4. 变点判断

CUSUM 图在某种程度上需要定义起点，以便计算样本的均值和方差。从一个起点开始，从前往后迭代，不断更新样本参数，进行标准正态化，同时计算 CUSUM 的上下统计量，判断是否达到上下限阈值，并确定变点的位置。

当然也可采取类似于布林带的做法，固定一个窗口期间，将起点定义在这个窗口区间起点，在窗口滚动过程中，不断更新样本和参数。但是这样很容易受样本的影响，可操作性较低。本文采取的是另外一张方法。

CUSUM 法不同于其他控制图的一个关键是 CUSUM 需要累积。通过观察 CUSUM 图（图 3）可以发现，随着行情往后迭代，只要序列的统计特性不变，起点的位置与变点的识别关系不大。即只要在前面分布的 $N(0,1)$ 分布中定义一个起点，变点总是能够识别出来，并且基本相差不大。

我们在判断当前点是否是变点时，选取离当前点足够远的任一个起点，从前往后迭代，不断更新正态分布参数，判断当前点是否恰好为变点（从起点到该点之前无其他变点）。若不是，定义初始起点的下一个点作为新的起点，重新往后迭代，判断当前点是

否为变点，直到遍历所有的起点为止。只要有一个起点对于当前点为变点，则当前点为变点的结论成立；若所有起点对应当前点均不为变点，则当前点不为变点。

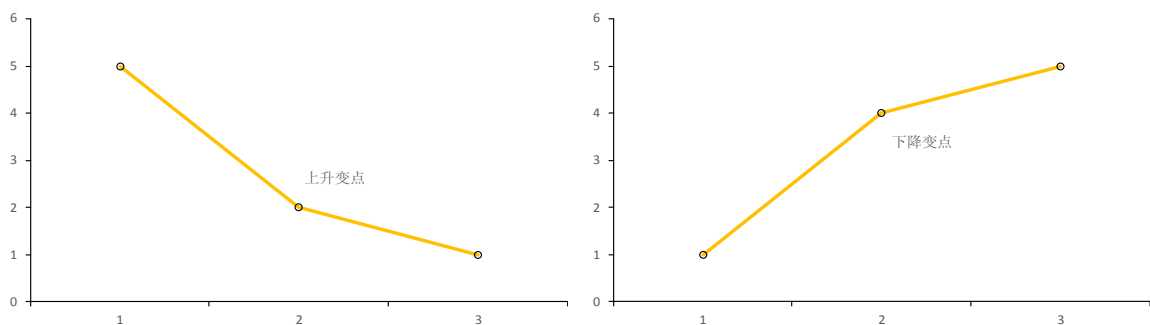
实际上，我们可以讲算法进一步简化，即我们在判断上一个拐点时，其必定有对应的最后一个起点，我们在判断该拐点之后的行情点时，都可以从该起点之后开始迭代，由此识别拐点过程就得到很大的简化。

5. 拐点判断

我们通常所说的行情拐点一般是指行情到了一个高点开始下行或者到了一个低点开始反弹，即行情的稳态发生变化，我们就称之为拐点。而本文所说的变点判断则是对行情的对数收益率走势进行分析，一旦对数收益率的稳态变化，就产生变点。二者分析的对象不同，因而在拐点判断上存在区别。

股价的拐点一般情况下都是对数收益率的变点（股价拐点对应的对数收益率变化相当大），而对数收益率的变点不一定是行情的拐点，易出现股价下降趋势减缓被检测为上升变点、上升趋势减缓被检测为下降变点的情况。总的来说，二者的考察对象不同，稳态偏离规则不同。我们需要通过一定的方式，将二者结合或等同起来。

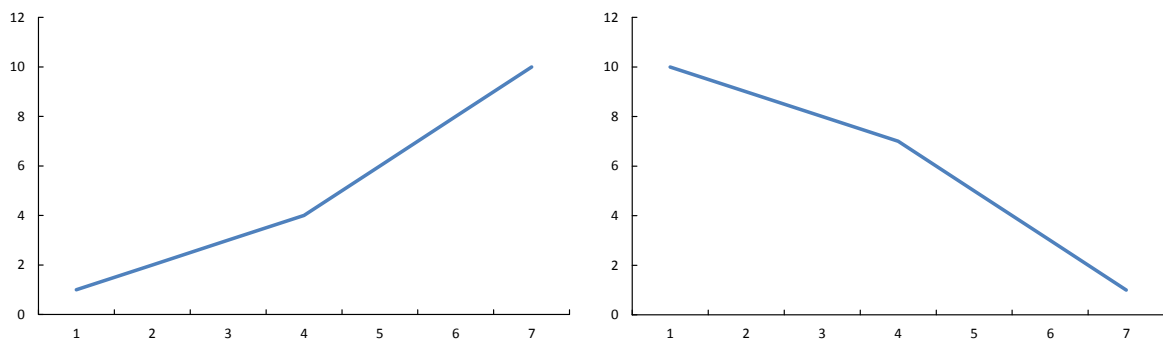
图 6：对数收益率变点与股价走势



资料来源：长江证券研究部

我们希望能够在漏检率和误检率之间获得平衡，但实际上任何一种技术择时方法都很难做到这一点。本文的思路是尽量降低误检率，将那些无法明确判断行情方向的变点剔除，只需要甄别在行情上升趋势中对数收益率的上变点（行情加速上升）、行情下降趋势中对数收益率的下变点（行情加速下降），从而将上下拐点都识别出来。

图 7：行情拐点判别

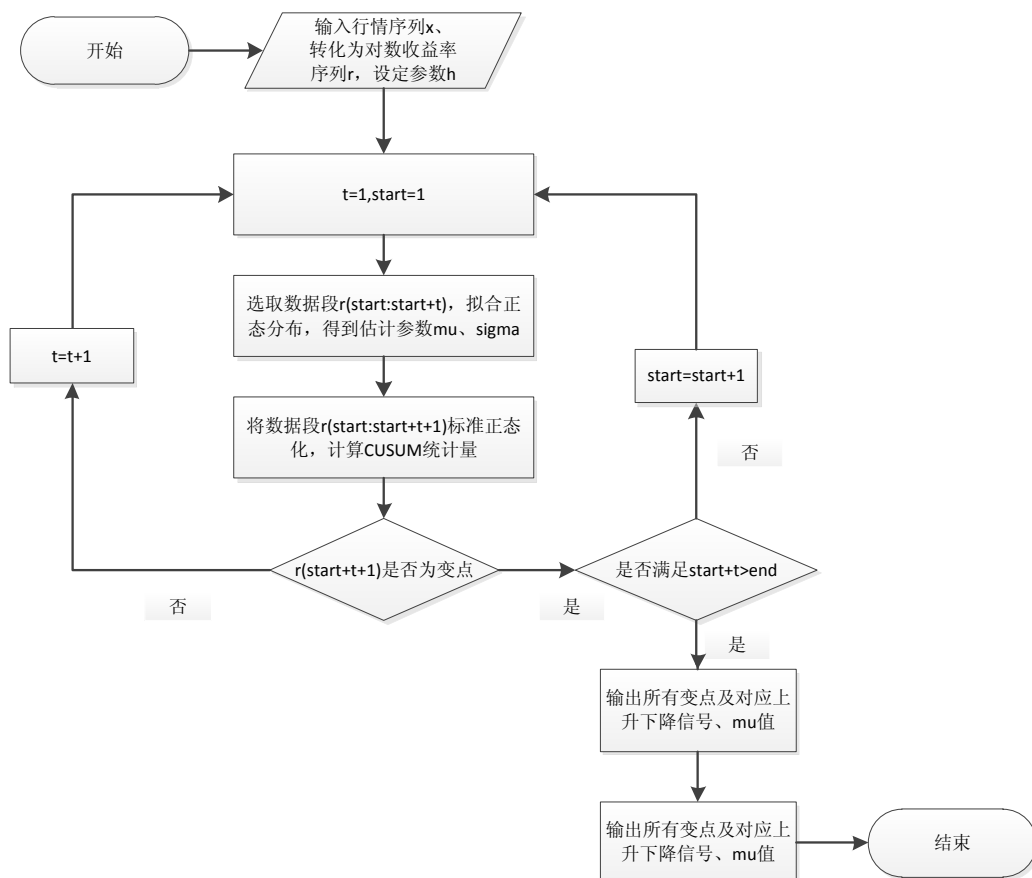


资料来源：长江证券研究部

6. 算法步骤

- 选取数据段 x_0 ，计算对数收益率 r ；
- 从数据段 x_0 的第一个数据开始，以第一第二个数据为初始数据段 $startdata$ ，进行正态分布拟合，得到均值 μ 和标准差 σ ，选择数据段外的第一个数据（整体样本的第三个数据），合成实验数据段 $data$ ，并利用 μ 和 σ 值进行归一化，得到标准正态分布序列（近似） x ；
- 允偏量 k 采取动态变化，为归一化后序列 x 的最后一个数值的一半，计算上下 CUSUM 统计量，判断 x 序列最后一个数是否超出 CUSUM 阈值，若超出，对 x 序列最后一个点进行变点标记；若不超出，往后迭代，直至出现变点标记为止；
- 从 x_0 第二个点开始，重现选择数据段，依照 a)、b) 步骤标记变点，直至标记完所有的起点为止；
- 若该点存在标记数，则判断该点为变点；
- 在变点中选择处于股价上升路径的上升变点以及股价处于下降路径中的下降变点，以此确定拐点和拐点方向。

图 8: CUSUM 算法流程图



资料来源：长江证券研究部

四、交易策略及结果

1. 交易策略

针对 CUSUM 择时，我们构建的择时交易策略如下：

- 判断基准：日开盘价

为了减少滞后，选取开盘价作为时点判断和选择的基准。

- 买入信号：第一个上升拐点

若前一拐点为上升拐点，不对其进行操作；前一拐点为下降拐点，买入。

- 卖出信号：第一个下降拐点

若前一拐点为下降拐点，不对其进行操作；前一拐点为上升拐点，卖出。

- 交易价格：均价

最高价与最低价之和的二分之一。本文策略的净值计算也以均价为基准。

- 交易成本：千分之二

指数：单边 0.2% 的交易成本，双边 0.4%。股指期货：单边万分之五。

- 涨跌限制：8%

若涨幅或跌幅超过 8%，一般难以买进卖出或资金成本过高，则下次交易。

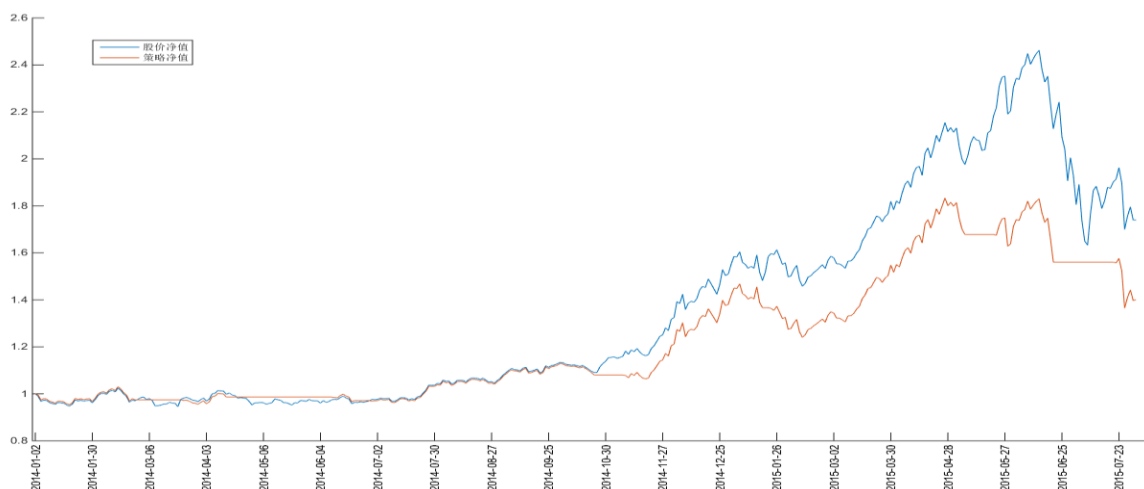
2. 固定参数策略效果

按照开始设定的参数观察择时效果，允偏量 k 为动态变化的参数，阈值 h 设定为 2。

(1) 不同行情

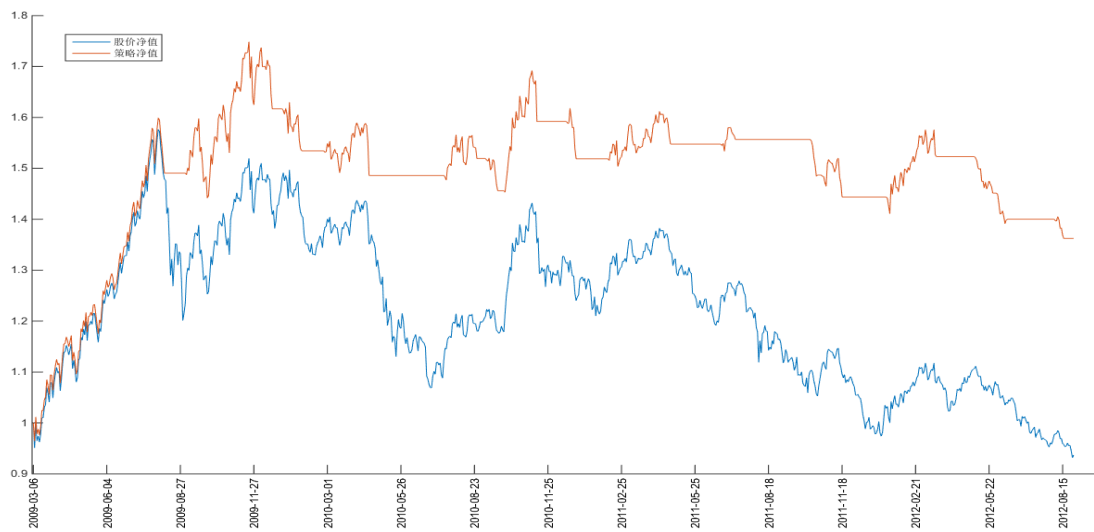
选取上证综指趋势行情和震荡行情下进行 CUSUM 择时，并观察择时效果。

图 9：上证综指趋势行情 CUSUM 择时效果（2014.01.01-2015.07.31）



资料来源：长江证券研究部

图 10: 上证综指震荡行情 CUSUM 法择时效果 (2009.03.06—2012.08.29)



资料来源: 长江证券研究部

无论是趋势行情还是震荡行情, CUSUM 法择时都取得了不错的效果。趋势行情下, 虽然没能跑赢大盘, 但仍能够抓住大部分收益。震荡行情下, 不仅跑赢了买入并持有策略, 夏普比率也得到了较大的提升。胜率表现均较为突出。

表 3: 不同行情择时效果比较

比较项	趋势	震荡	比较项	趋势	震荡
交易区间	386	841	买入准确率	37.50%	37.50%
交易次数	16	32	卖出准确率	100.00%	56.25%
总收益_无成本	44.71%	45.24%	最大套牢	17.18%	8.67%
总收益_有成本	40.15%	36.22%	最大踏空	-13.58%	-7.28%
日均收益_无成本	0.12%	0.05%	买入平均滞后阶数	3.80	3.81
日均收益_有成本	0.10%	0.04%	卖出平均滞后阶数	3.25	3.97
超额收益_无成本	-29.24%	51.53%	买入平均收益率	5.64%	2.83%
超额收益_有成本	-33.81%	42.52%	卖出平均收益率	2.65%	-2.32%
最大回撤	23.50%	21.40%	年化总收益_有成本	26.00%	10.77%
买入持有最大回撤	33.66%	40.90%	年化超额收益_有成本	-21.90%	12.64%
买入胜率	57.14%	66.67%	年化夏普比率	99.94%	41.79%
卖出胜率	71.43%	46.67%	买入持有年化夏普比率	118.31%	-15.22%

资料来源: 长江证券研究部

(2) 不同指数

接下来观察其他几个指数 (沪深 300、上证 50、中证 500) 的择时表现。

图 11: 沪深 300 择时效果 (2005.01.05—2015.07.31)



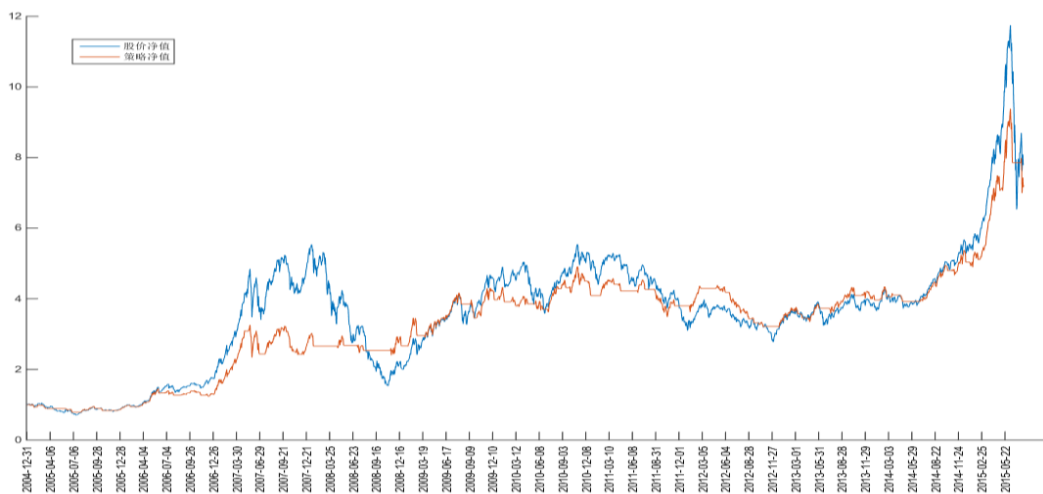
资料来源: 长江证券研究部

图 12: 上证 50 择时效果 (2004.01.02—2015.07.31)



资料来源: 长江证券研究部

图 13: 中证 500 择时效果 (2004.12.31—2015.07.31)



资料来源: 长江证券研究部

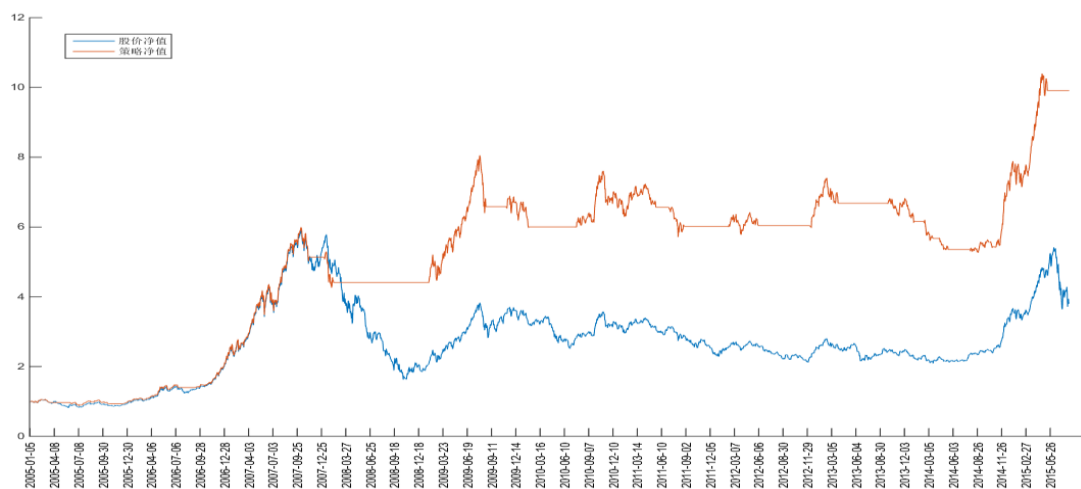
择时的参数选择涉及到两个维度，时间维度和空间维度。时间维度上，同一对象不同时间点的参数设置不一样；空间维度上，同一时间点不同对象的参数设置也不尽相同。传统择时需要同时顾及这两方面的因素影响，但往往不容易控制。

本部分即通过上证综指不同行情下查看时间维度上的参数设置效果，通过不同类别指数（三个股指期货标的）查看空间维度上的参数设置效果。总的来说，就几个结果表现来看，CUSUM 法参数的设置具有较强的稳健性和可控性。

3. 最优参数策略效果

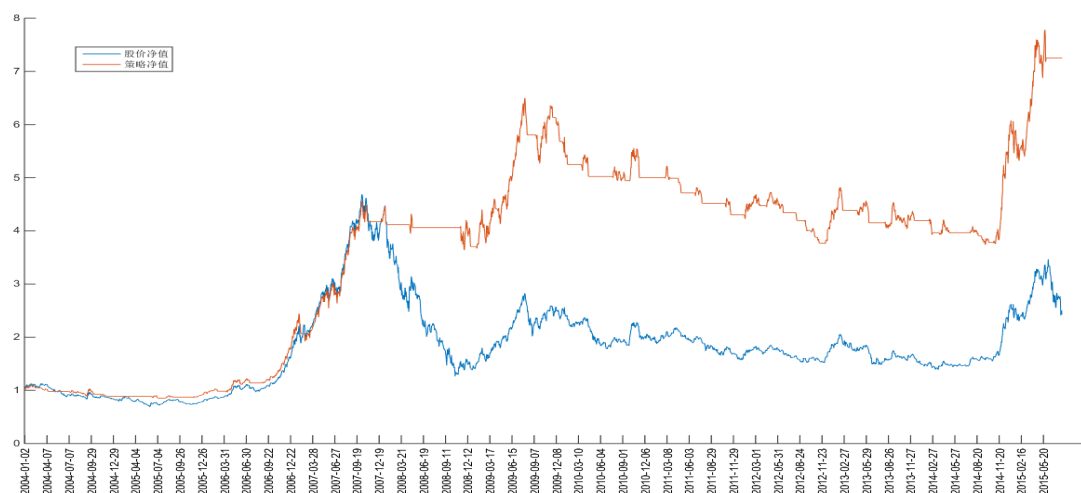
指数不同时，其股票成分不同，指数本身的波动特性也会有所差别，因而实际上不同指数进行选择时，其参数设置应该有所不同。本文选取沪深 300、上证 50 以及中证 500 日线数据为参考对象，检测其最优参数择时效果。其中最优择时效果以年化夏普比率最高为标准。

图 14: 沪深 300 最优参数择时效果 (2005.01.05—2015.07.31, $h=4.8$)

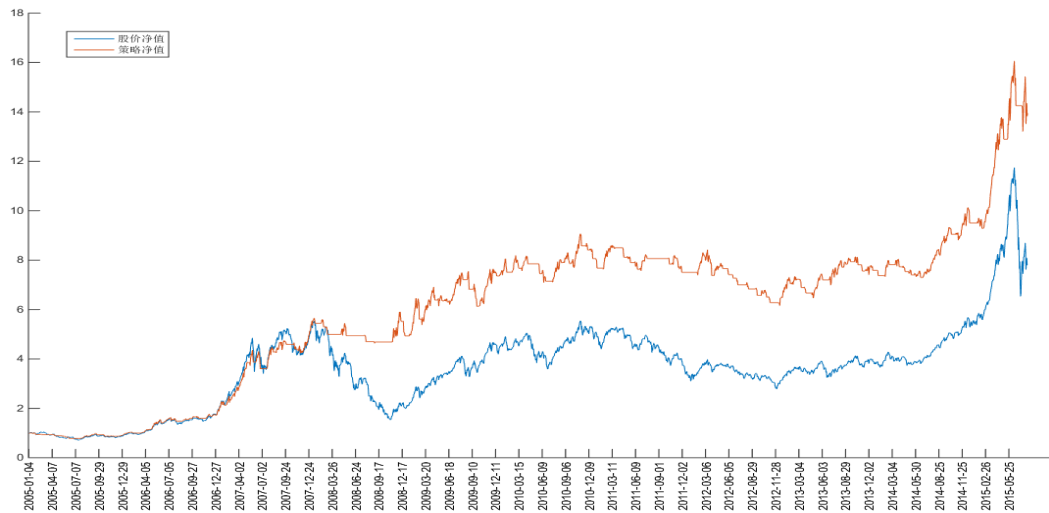


资料来源：长江证券研究部

图 15: 上证 50 最优参数择时效果 (2004.01.02—2015.07.31, $h=2.7$)



资料来源：长江证券研究部

图 16: 中证 500 最优参数择时效果 (2004.12.31—2015.07.31, $h=0.7$)


资料来源: 长江证券研究部

最优参数下的择时结果只是作为一个参考, 我们处于事中判断时, 很难确定我们当前的最优参数是什么。同时在行情不同时, 参数的敏感性也会存在很大差别。实际上, CUSUM 同传统的几种择时方法相比, 在最优参数下, 效果都比较接近。

4. 加入做空机制和高频交易的股指期货

在用于实盘操作时, 择时比较常见的应用领域在于高频交易。考虑到股指期货的特殊性, 既可以实现高频交易, 又可反向做空, 该部分我们选取股指期货作为参考对象进行择时研究。

具体采取高频投资中比较常见的思路, 以沪深 300 的 15 分钟高频线为例。考虑到做多做空机制, 我们现在并不考虑保证金及止盈止损, 设置交易成本为万分之五, 以 IF1506 和 IF1507 为例。

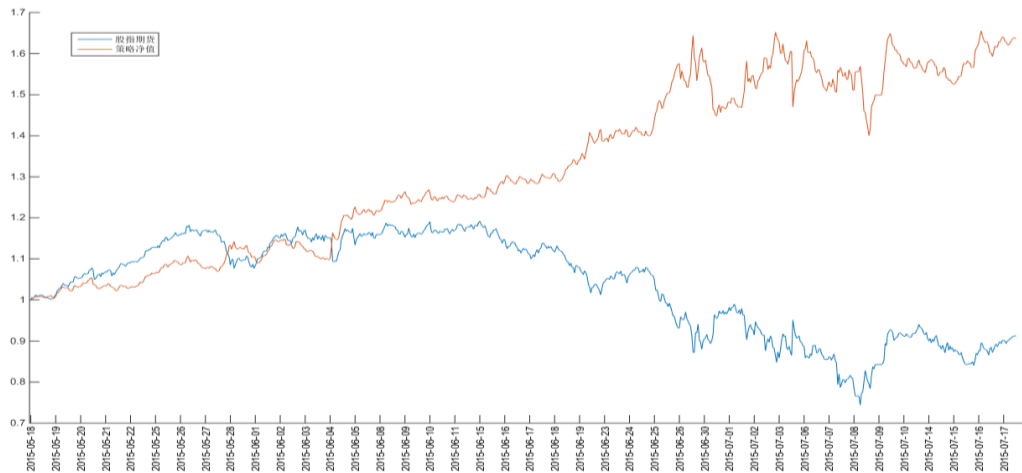
高频数据下, 将阈值 h 的设定降低, 定为 $h=0.2$ 。

图 17: IF1506 择时效果 (2015.04.16-2015.06.19)



资料来源: 长江证券研究部

图 18: IF1507 择时效果 (2015.05.18-2015.07.17)



资料来源：长江证券研究部

表 4: 股指期货择时效果比较

比较项	趋势	震荡	比较项	趋势	震荡
交易频率 (次/天)	2.18	1.81	年化总收益_有成本	187.87%	370.42%
最大回撤	9.19%	15.23%	年化超额收益_有成本	210.41%	421.37%
买入持有最大回撤	15.58%	37.59%	年化夏普比率	520.94%	545.45%
胜率	61.86%	57.14%	买入持有年化夏普比率	-49.71%	-42.58%

资料来源：长江证券研究部

阈值设定并没有进行优化, 但就两个合约的表现而言, 其效果比较明显。年化绝对收益分别达到了 188% 和 370%, 年化超额收益分别高达 210% 和 421%, 夏普比率更是得到了极大的提升。胜率均超过了 50%, 在技术择时中相对来说还比较高。不足的是最大回撤并未做太多的改善, 存在较大的风险。

五、总结

变点理论是统计学中的一个经典分支, 其基本定义是在一个序列或过程中, 当某个统计特性 (分布类型、分布参数) 在某时间点受系统性因素而非偶然性因素影响发生变化, 我们就称该时间点为变点。变点识别即利用统计量或统计方法将该变点位置估计出来。

CUSUM 图作为工业应用检验变点的三大控制图之一, 其利用假设检验和极大似然估计的相关统计原理, 构建累积和统计量, 不断累积观察值与基线水平的差值, 将微小偏差累积, 放大观察数据出现的波动, 从而更加迅速敏感地探测到微小的异常情况, 检验出变点位置。其最大的特点是对系统性变化的敏感性, 不需要积累太多的样本, 因而能较好的控制风险。

本文基于变点 CUSUM 图的基本原理, 以股价对数收益率符合局部正态分布为基本原理, 构建 CUSUM 的上下统计量, 一旦统计量突破阈值即判断出现变点。以股价上升时对数收益率出现上升变点作为买入时机, 以股价下降时对数收益率出现下降变点作为

卖出时机。统计量中的两个参数，允偏量 k 设置为动态变化自适应的形式，阈值 h 则需要根据直观进行设定。

固定参数下，对于不同行情、不同指数都取得了不错的择时效果，夏普比率都能得到比较明显的提高，胜率基本在 50% 以上，CUSUM 择时具有较强的稳健性和可操作性；最优参数下效果明显，均能较大幅度地跑赢买入持有策略；加入做空机制和高频交易的股指期货交易中，IF1506 和 IF1507 年化收益均达到 180% 以上，胜率在 60% 左右。

对本报告的评价请反馈至长江证券机构客户部

姓名	分工	电话	E-mail
周志德	主管	(8621) 68751807	zhouzd1@cjsc.com.cn
甘 露	副主管	(8621) 68751916	ganlu@cjsc.com.cn
杨 忠	华东区总经理	(8621) 68751003	yangzhong@cjsc.com.cn
鞠 雷	华南区总经理	(86755) 82792756	julei@cjsc.com.cn
李敏捷	华北区总经理	(8610) 66290412	limj@cjsc.com.cn

投资评级说明

行业评级	报告发布日后的 12 个月内行业股票指数的涨跌幅度相对同期沪深 300 指数的涨跌幅为基准，投资建议的评级标准为：
看 好：	相对表现优于市场
中 性：	相对表现与市场持平
看 淡：	相对表现弱于市场
公司评级	报告发布日后的 12 个月内公司的涨跌幅度相对同期沪深 300 指数的涨跌幅为基准，投资建议的评级标准为：
买 入：	相对大盘涨幅大于 10%
增 持：	相对大盘涨幅在 5%~10%之间
中 性：	相对大盘涨幅在-5%~5%之间
减 持：	相对大盘涨幅小于-5%
无投资评级：	由于我们无法获取必要的资料，或者公司面临无法预见结果的重大不确定性事件，或者其他原因，致使我们无法给出明确的投资评级。



研究部/机构客户部

上海

浦东新区世纪大道 1589 号长泰国际金融大厦 21 楼
(200122)

电话: 021-68751100

传真: 021-68751151

武汉

武汉市新华路特 8 号长江证券大厦 9 楼
(430015)

传真: 027-65799501

北京

西城区金融大街 17 号中国人寿中心 606 室
(100032)

传真: 021-68751791

深圳

深圳市福田区福华一路 6 号免税商务大厦 18 楼
(518000)

传真: 0755-82750808

0755-82724740

重要声明

长江证券股份有限公司具有证券投资咨询业务资格，经营证券业务许可证编号：Z24935000。

本报告的作者是基于独立、客观、公正和审慎的原则制作本研究报告。本报告的信息均来源于公开资料，本公司对这些信息的准确性和完整性不作任何保证，也不保证所包含信息和建议不发生任何变更。本公司已力求报告内容的客观、公正，但文中的观点、结论和建议仅供参考，不包含作者对证券价格涨跌或市场走势的确定性判断。报告中的信息或意见并不构成所述证券的买卖出价或征价，投资者据此做出的任何投资决策与本公司和作者无关。

本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可升可跌，过往表现不应作为日后的表现依据；在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告；本公司不保证本报告所含信息保持在最新状态。同时，本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

本公司及作者在自身所知情形范围内，与本报告中所评价或推荐的证券不存在法律法规要求披露或采取限制、静默措施的利益冲突。

本报告版权仅仅为本公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布。如引用须注明出处为长江证券研究部，且不得对本报告进行有悖原意的引用、删节和修改。刊载或者转发本证券研究报告或者摘要的，应当注明本报告的发布人和发布日期，提示使用证券研究报告的风险。未经授权刊载或者转发本报告的，本公司将保留向其追究法律责任的权利。