Monitoring Active Portfolios: The CUSUM Approach

Thomas K. Philips

Chief Investment Officer

Paradigm Asset Management

Agenda

- Portfolio monitoring: problem formulation & description
- Sequential testing and process control
- The Cusum scheme
  - Economic intuition, simplified theory & implementation
- Issues that arise in practice
  - Optimality and robustness
  - Causality

The Investor's / CIO's Problem

- Invested in / responsible for many active products
  - Fallout of the asset allocation / manager selection / sales / reseach / product development process

- Lots of data coming in from portfolio managers
  - Returns, portfolio holdings, sector allocations, risk profiles, transactions etc.

- Not clear which portfolios merit extra attention
  - Investor / CIO will ideally focus on products that might be in trouble

Observations on The Investment Environment

- First order approximation: market is efficient
  - Performance measurement plays a vital role in evaluation
- Hard to differentiate luck from skill
  - Alpha (signal)= 1%, Tracking Error (noise)=3%
  - t-test for $a > 0$ requires *N*>36 years for *t*>2
- Subtle, and wrong, assumptions
  - Alpha and tracking error are stationary for 36years
  - *t*>2 is necessary to make a decision about the portfolio

Performance Measurement in Practice

- Performance measurement is rooted in classical statistics
- Measure benchmark relative performance over fixed & rolling intervals
  - 3 to 5 years is common
  - Extrapolate trends in rolling benchmark or peer group relative performance
  - Focus attention on the underperforming products
- Cannot identify regime changes or shifts in performance
  - t-stats are too low to be meaningful
  - Not clear if attention is focused on the right products at the right time

Monitoring Performance

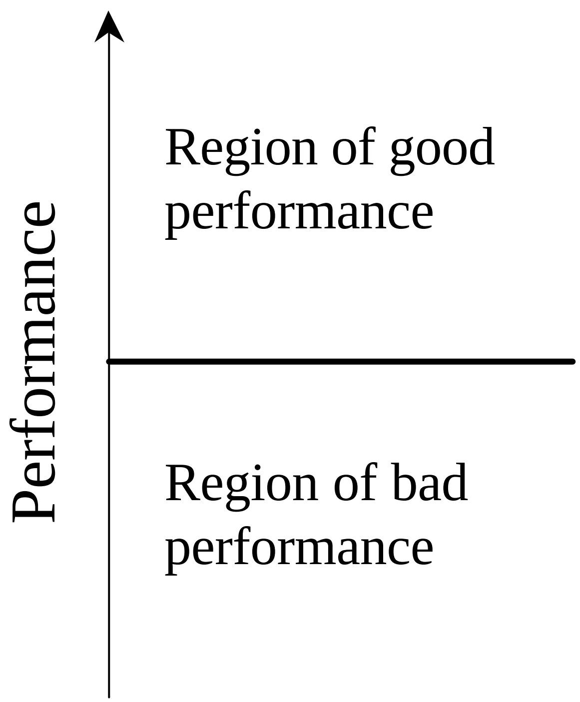
- Performance monitoring is rooted in decision theory & hypothesis testing
  - First step: define what constitutes good & bad performance.
    - Null hypothesis H_0: Performance better than X, manager is good.
    - Alternative hypothesis H_1: Performance worse than Y, manager is bad
  - Next step: Continuously test data against these two hypotheses
    - Raise alarm when sufficient evidence accrues to conclude that manager is bad
  - Key point: *N* is variable – use only as many observations as needed

- Abraham Wald's Sequential Probability Ratio Test
  - Observe a stream of data. Do the following after each observation:
    - Examine the log-likelihood ratio = $\log\left(\dfrac{\Pr[\text{Observations } 1,2,\text{☹},i \mid \text{ Manager is bad}]}{\Pr[\text{Observations } 1,2,\text{☹},i \mid \text{Manager is good}]}\right)$
    - Accept a hypothesis as soon as the likelihood ratio exceeds a threshold

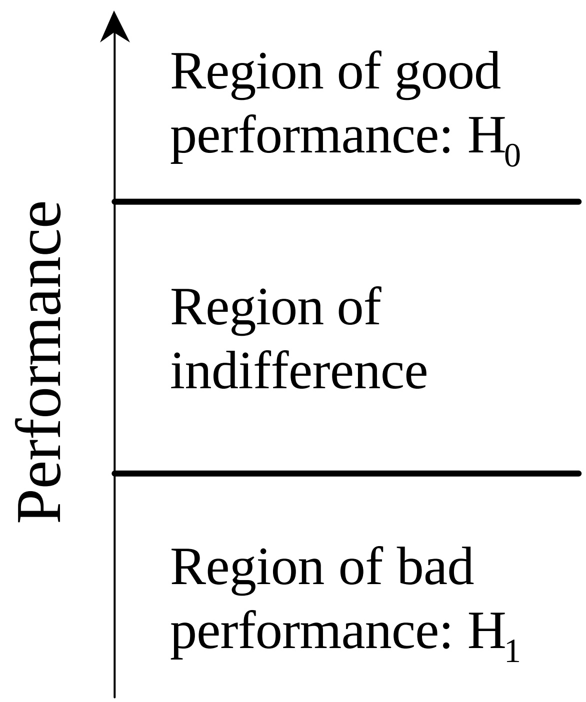
Paradigm Asset Management Co.

Measurement vs. Monitoring: Geometric Interpretation

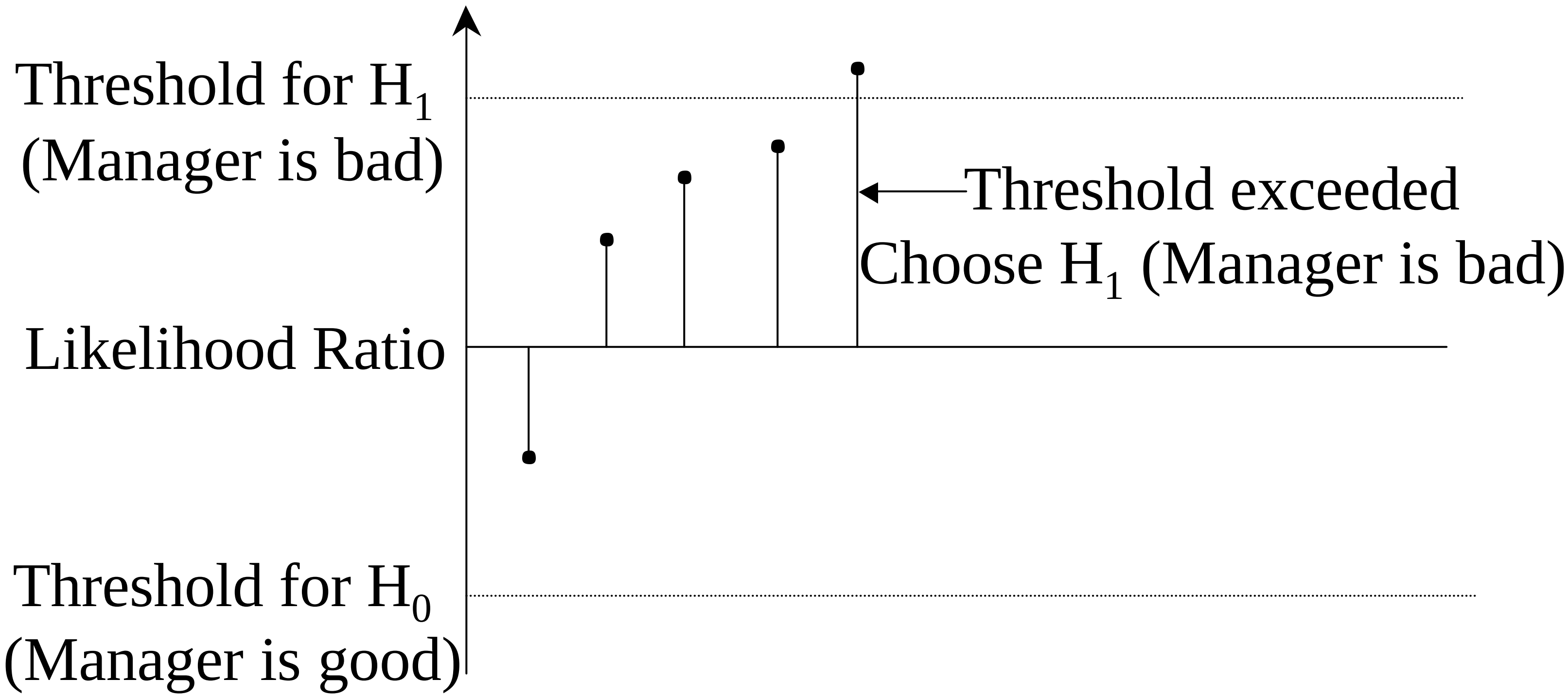
Performance Measurement



Performance Monitoring



Sequential Testing: Visual Explanation



Paradigm Asset Management Co.

CUSUM: Good and Bad Levels of Performance

- Good and bad managers defined by their information ratio
  - Allows use in almost any asset class without modification
  - Good manager: Information Ratio > 0.5
  - Bad Manager: Information Ratio < 0
- Corresponding boundaries of regions of good and bad performance
  - H_0: Information ratio = 0.5
  - H_1: Information ratio = 0

Measurement vs. Monitoring: Differences

- Performance Measurement
  - **Good**: Simple math
  - **Good**: Robust to distribution of returns
  - **Bad**: Slow to detect change in performance

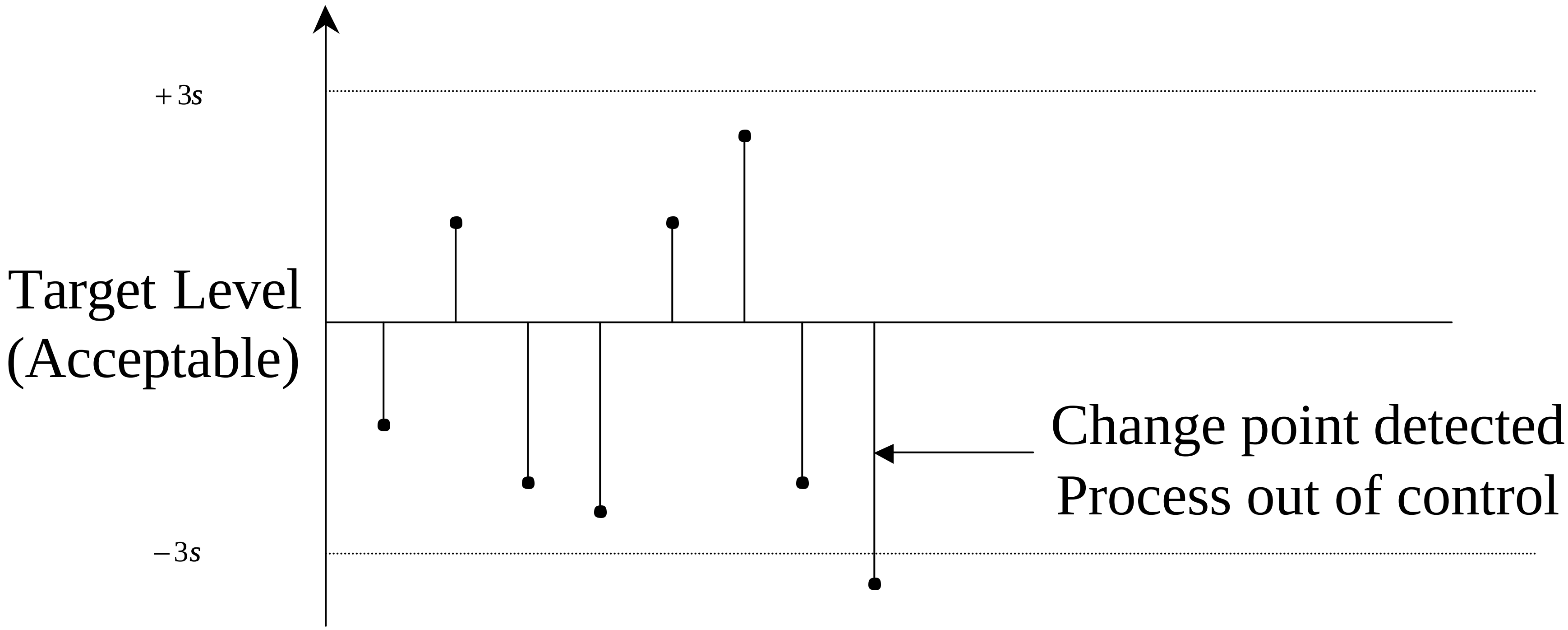
- Performance Monitoring
  - **Bad**: Complex math
  - **Bad**: Sensitive to distribution of returns
  - **Good**: Quick to detect change in performance

- CUSUM: best of both worlds
  - **Good**: Simple math (for users), complex math (for theoreticians)
  - **Good**: Exceptionally robust to distribution of returns
  - **Good**: Exceptionally quick to detect change in performance

Statistical Process Control

- Developed at Bell Labs in the 1930's by Walter Shewart
  - Originally used to monitor Western Electric's telephone production lines
- Traditional process control: focus on process
  - Tweak the machines on the production line
  - If they operate well, products should be good
  - Similar in spirit to performance measurement
- Walter Shewart's great insight: focus on results
  - The product is what counts
    - If it's good, the process is good
    - If it's bad, the process is bad
  - Similar in spirit to performance monitoring

The Shewart Chart



Shewart Chart: Strengths and Limitations

- Strengths
  - Extremely simple
  - Graphical
  - Rapidly detects big process shifts
- Limitations
  - Very slow to detect small process shifts (-10 bp/mo)
  - Sensitive to probability distribution
- Shewart was aware of these limitations
  - Did not succeed in developing a clean solution

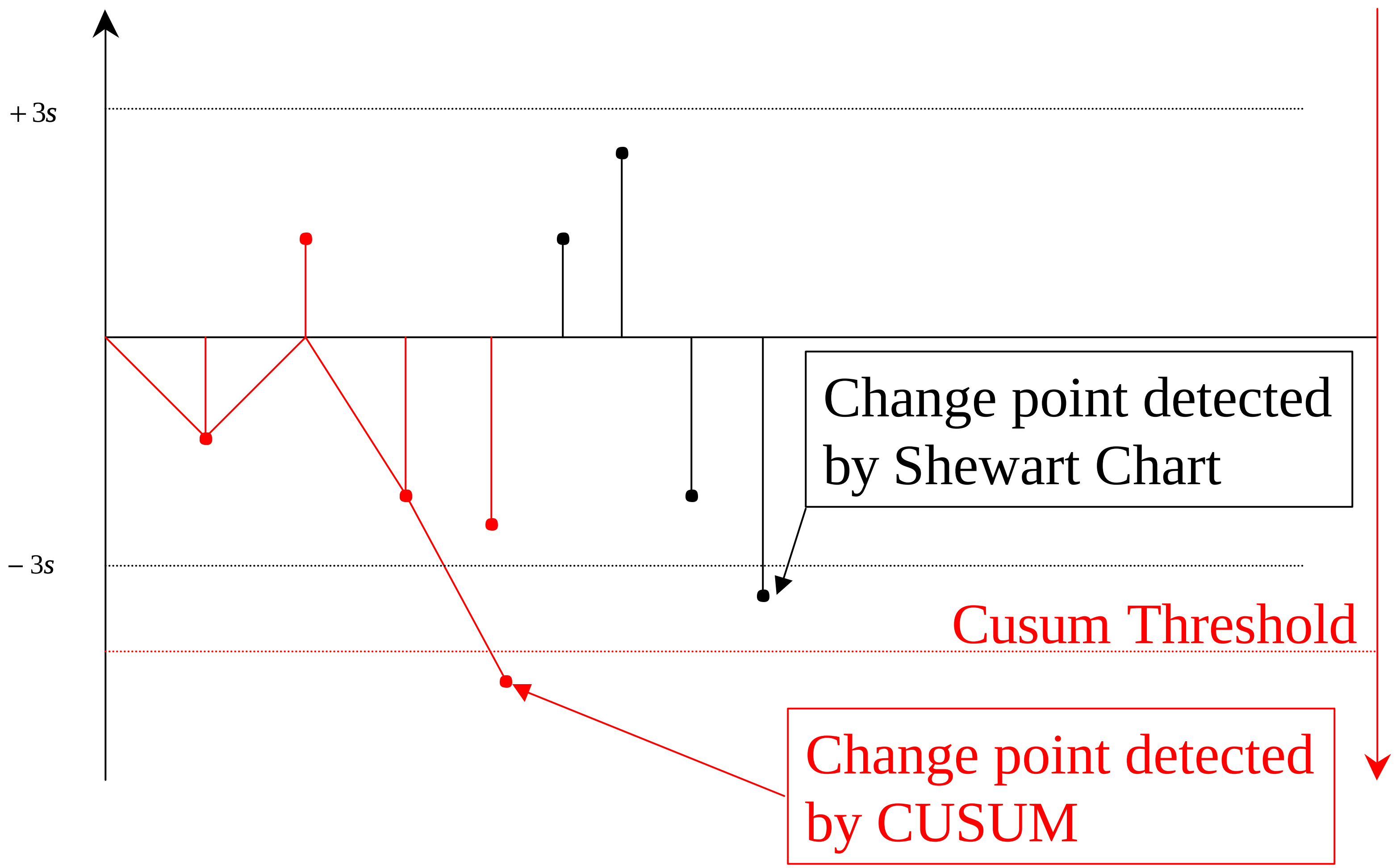
The CUSUM Technique

- Created by E.S. Page in 1954
- Addresses the limitations of the Shewart chart
  - Reliably detects small process shifts
  - Insensitive to probability distribution
- Provably optimal: detects process shifts faster than any other method
  - Proof (Moustakides) is very hard. Uses optional stopping theorem.
  - Extremely robust, good under almost any definition of optimality
    - Much better than exponentially weighted moving average

Page's Great Insight

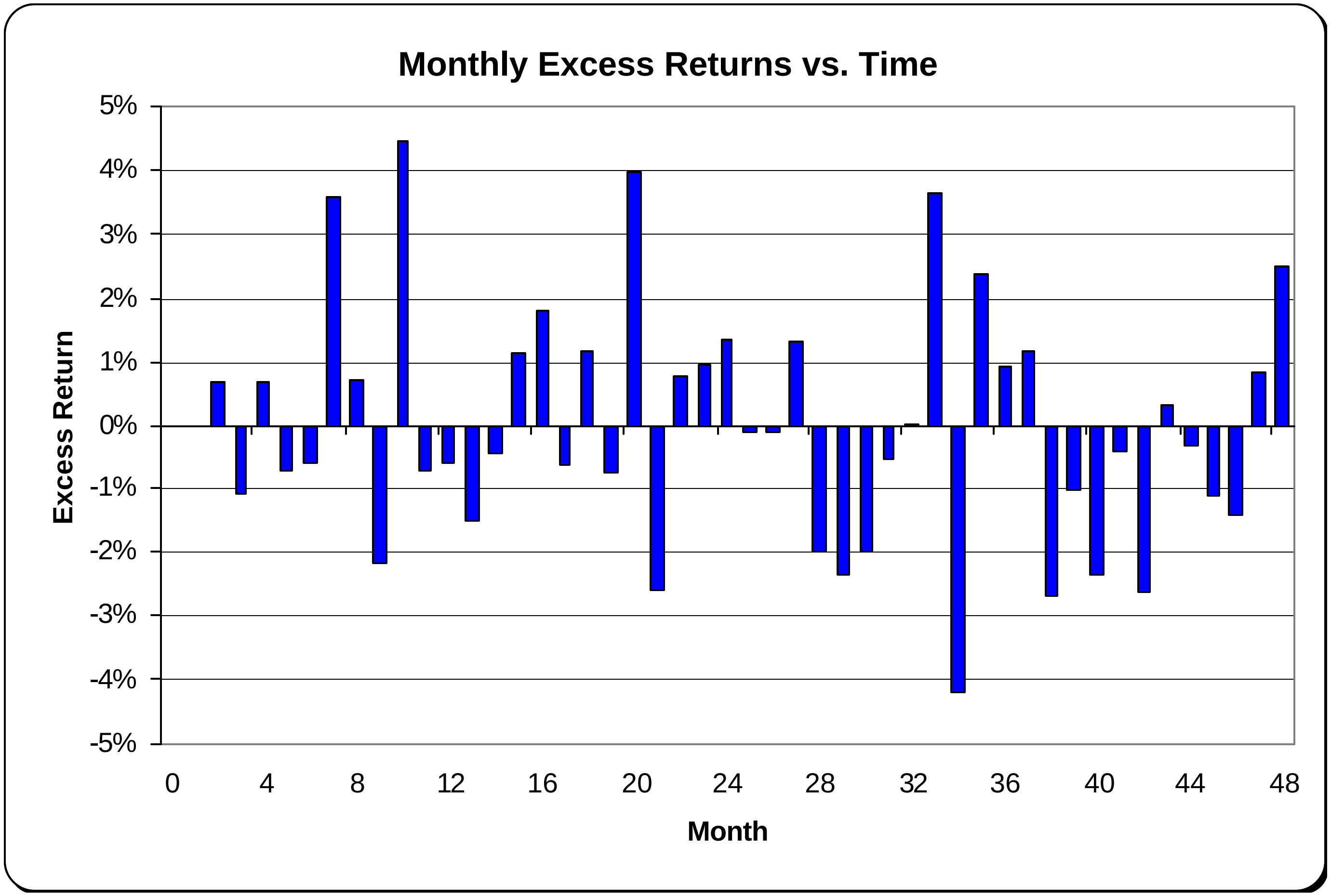
- Plot the cumulative *arithmetic* sum of residuals (e.g. excess returns)
  - Cumulating filters noise, strengthens signal
    - Positive process mean $\Leftrightarrow$ Positive slope
    - Negative process mean $\Leftrightarrow$ Negative slope
    - 0 process mean $\Leftrightarrow$ 0 slope (flat)
- Changes in slope are easily detected, both visually and mathematically
  - Cusum is a very clever variant of the Sequential Probability Ratio Test
  - Raise an alarm if the cumulative sum becomes large and negative
- Works about as well as the Shewart chart for large process shifts
  - Works much faster for small process shifts
  - Particularly well suited to money management

The Cusum Plot



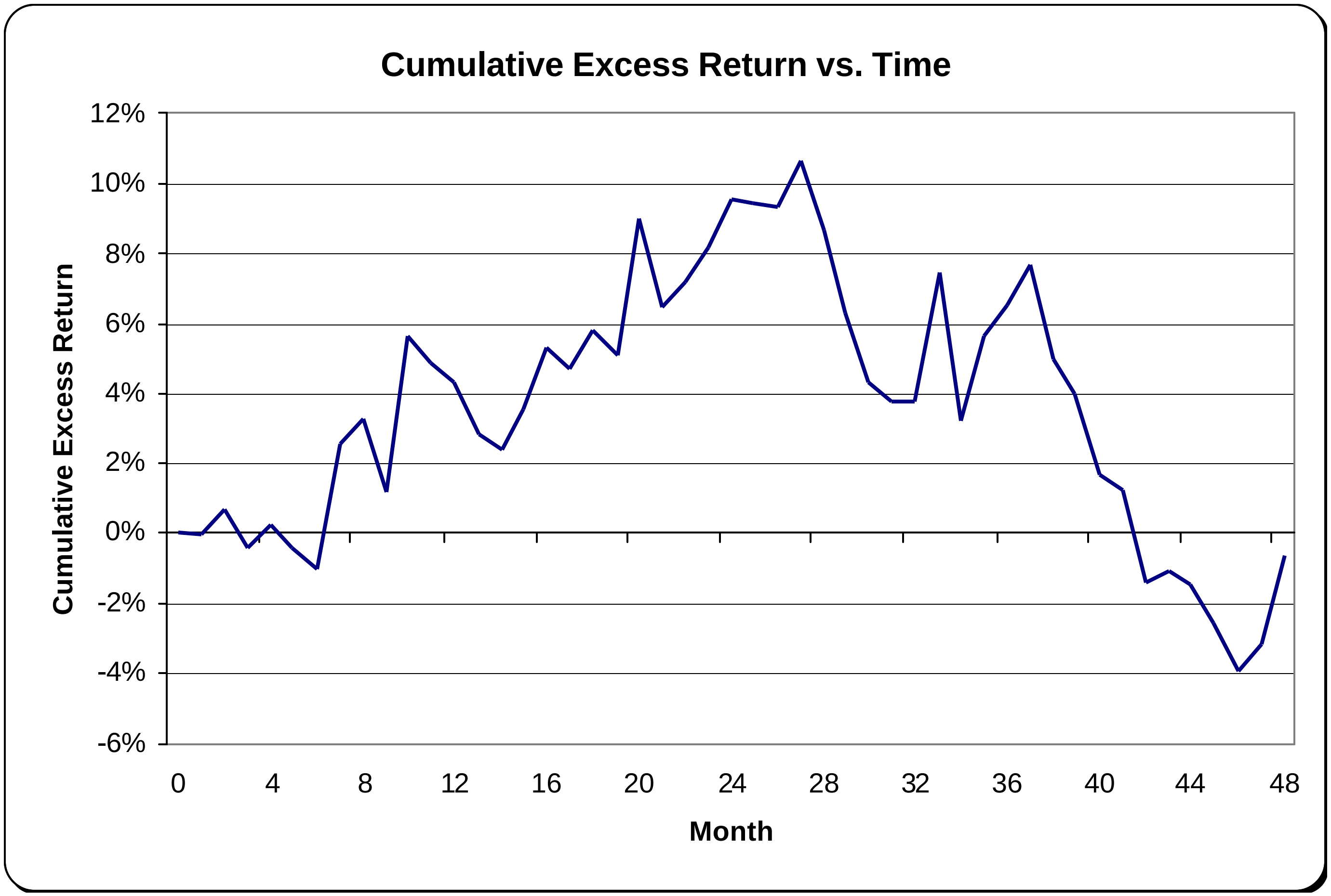
Paradigm Asset Management Co.

CUSUM: Visual Example – I



Paradigm Asset Management Co.

CUSUM: Visual Example – II



CUSUM – Intuitive Explanation

- Compute *current* performance
  - Discard old returns that are unrelated to current performance
  - Raise an alarm when current performance is reliably negative

- CUSUM is a backward looking SPRT. At time N
  - Compute likelihood ratio based on the k most recent observations, k=1,2,…,N

- Find k^*, the value of k which *maximizes* the likelihood ratio
  - Compare the maximum of these likelihood ratios to a threshold
  - Raise an alarm if it is sufficiently high

- CUSUM is optimal because it maximizes the likelihood ratio!
  - Also simplifies math and makes it insensitive to distribution of returns

CUSUM – Simplified Math

- Define $\hat{IR}_N$ to be a maximum likelihood estimate of the information ratio based on a single observation at time N

- Excess Return$_N$ = $\log\left(\dfrac{1+r_N^{Portfolio}}{1+r_N^{Benchmark}}\right)$

- Tracking error is estimated recursively (variant of Von Neumann's estimator)

$$\hat{s}_0^2 = \hat{s}_1^2 = s_0^2$$

$$\hat{s}_N^2 = 0.9\times\hat{s}_{N-1}^2 + 0.1\times\frac{12\times\left(e_N - e_{N-1}\right)^2}{2}, \qquad N = 2,3,\text{😐}$$

$$\hat{s}_N = \sqrt{\hat{s}_N^2},\ \ N = 0,1,2,3,\text{😐}$$

- Information Ratio$_N$ = Excess Return$_N$ /Tracking Error$_{N\text{-}1}$

CUSUM – Simplified Math

- At time N, Cusum computes

$$L_N(k^*) = \max_{1 \le k \le N} \log \left[\frac{f\left(\hat{IR}_{N-k+1}, \hat{IR}_{N-k+2}, \text{☹}, \hat{IR}_N \mid \text{Information Ratio} = 0 \right)}{f\left(\hat{IR}_{N-k+1}, \hat{IR}_{N-k+2}, \text{☹}, \hat{IR}_N \mid \text{Information Ratio} = 0.5 \right)} \right]$$

- When excess returns are normal, it reduces to a simple recursion!

$$\begin{aligned} L_0 &= 0 \\ L_N &= \max\left[0, L_{N-1} - \hat{IR}_N + 0.25 \right], \quad N = 1,2,\ldots \end{aligned}$$

- Compare L_N to a threshold – if it exceeds it, raise an alarm

Paradigm Asset Management Co.

CUSUM: Algorithmic Description

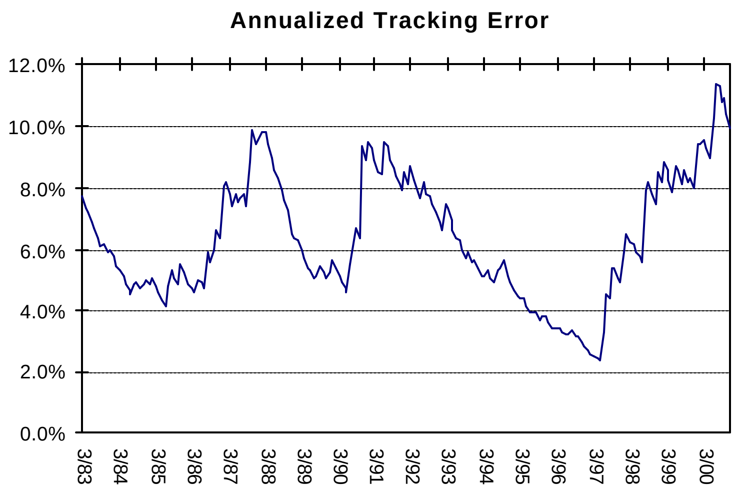
- Step 0: Initialize Tracking Error, set likelihood ratio to 0
- Each time a new return is recorded, perform the following 3 steps
  - Step 1: Compute excess return, tracking error and information ratio
  - Step 2: Update the likelihood ratio using simple recursion
  - Step 3: Compare the likelihood ratio to a threshold
    - If it does not exceed the threshold, do nothing, wait for the next return
    - If it exceeds the threshold, raise an alarm, launch an investigation
- If investigation suggests that this is a false alarm
  - Reset likelihood ratio to 0, restart CUSUM
- If evidence suggests that a problem exists, take corrective action

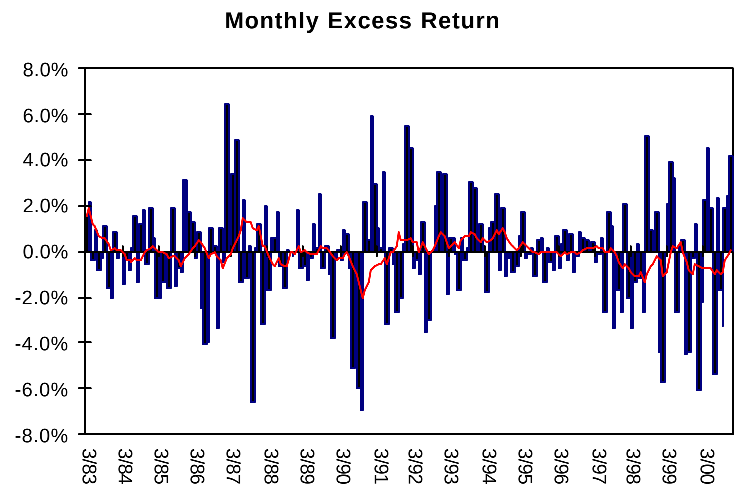
CUSUM: Setting The Threshold For An Alarm

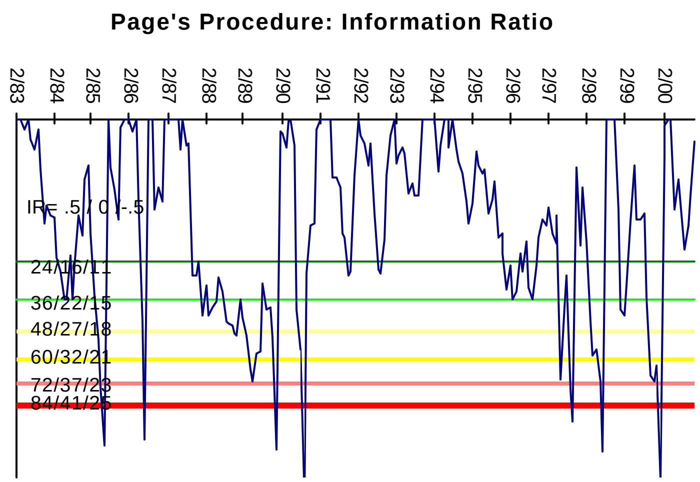
- Must make a trade-off between detection speed and rate of false alarms

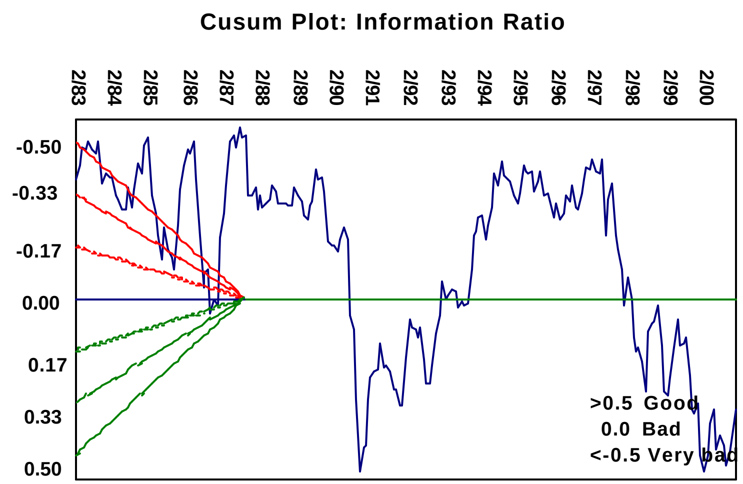
- Our choices:
  - Average time to detect a bad manager: 41 months (10x faster than t-test)
  - Average time between false alarms for a good manager: 84 months

CUSUM: Large Value Manager vs. Russell 1000 Value





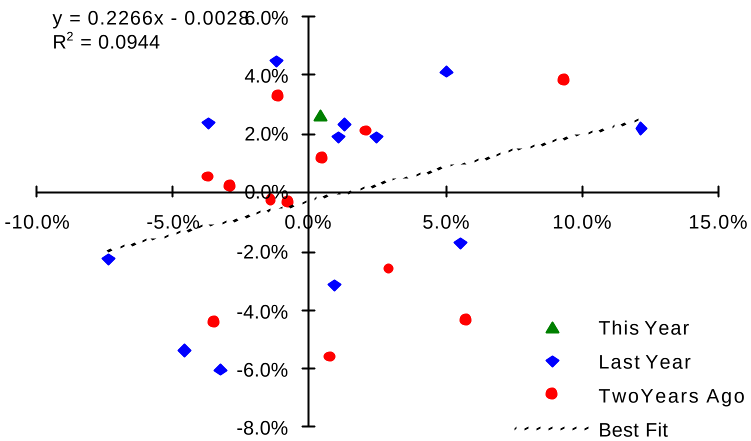




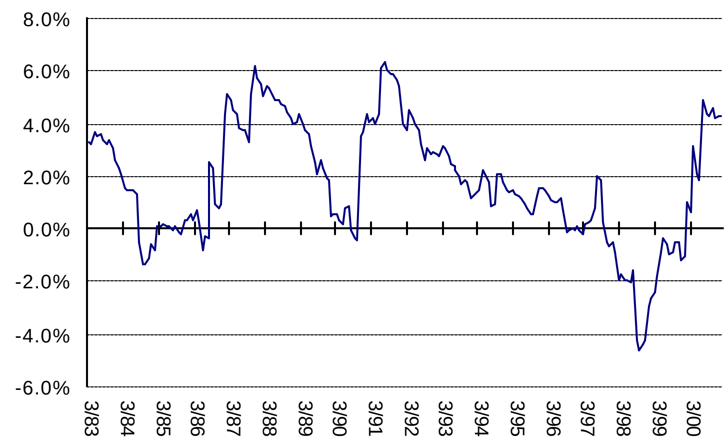
Paradigm Asset Management Co.

CUSUM: Large Value Manager vs. Russell 1000 Value

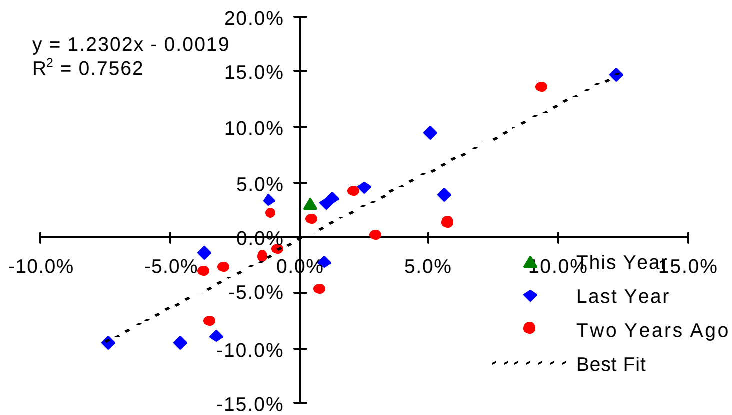
Excess Return vs. Benchmark



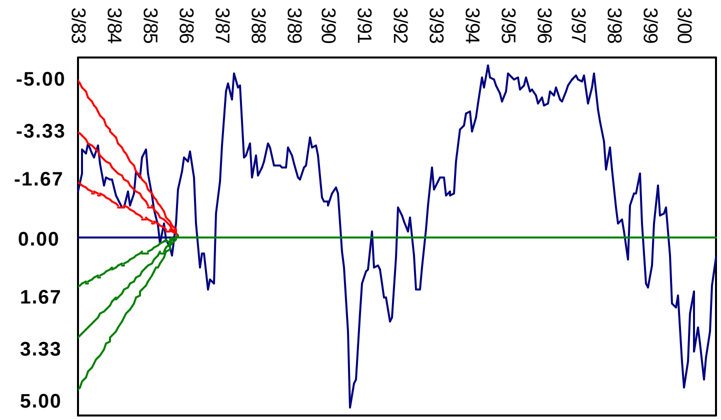
Excess Volatility Relative To Benchmark



Total Returns: Manager vs. Benchmark

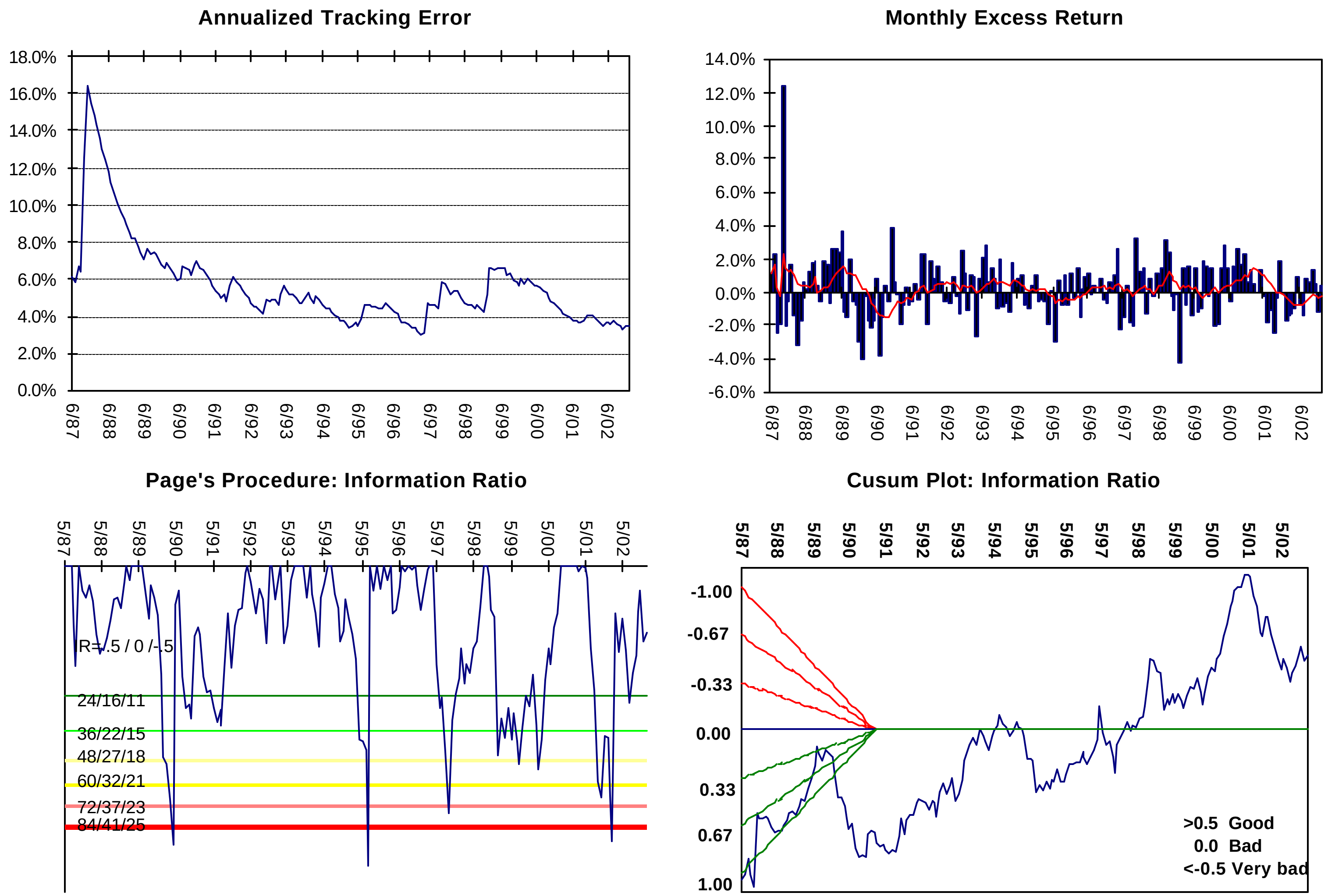


Cusum Plot: Annualized Excess Return



Paradigm Asset Management Co.

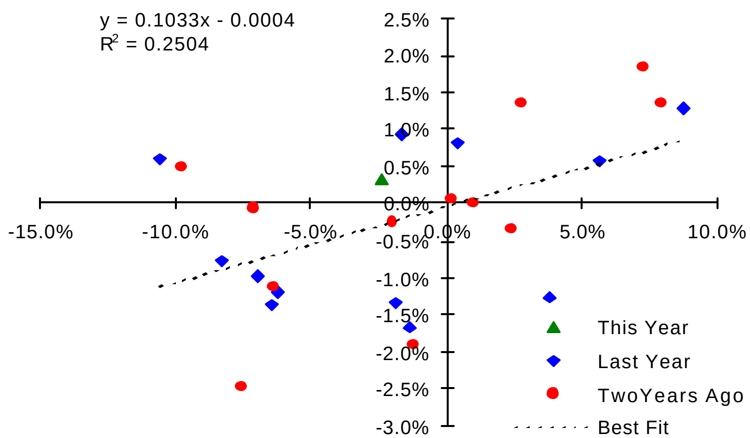
CUSUM: Large Growth Manager vs. Custom Index



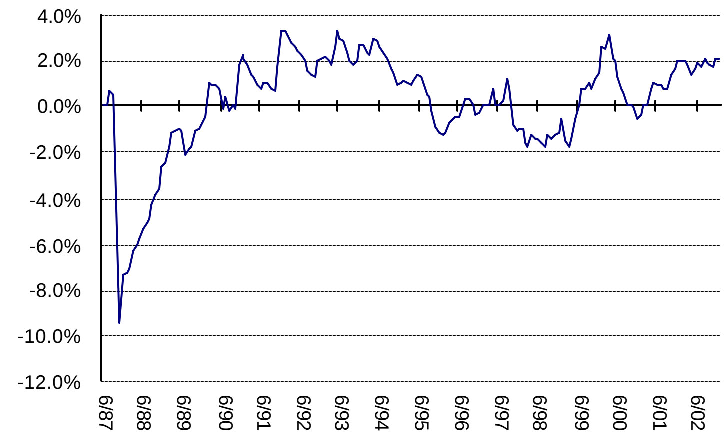
Paradigm Asset Management Co.

CUSUM: Large Growth Manager vs. Custom Index

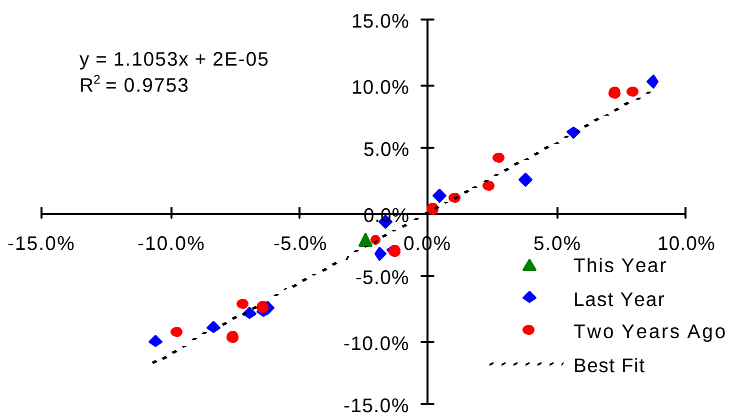
Excess Return vs. Benchmark



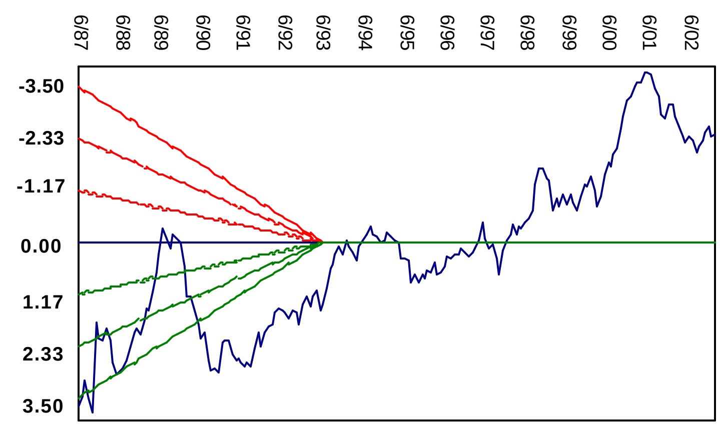
Excess Volatility Relative To Benchmark



Total Returns: Manager vs. Benchmark



Cusum Plot: Annualized Excess Return



Paradigm Asset Management Co.

CUSUM: Strengths

- Detects underperformance exceptionally fast
  - Provably optimal, though proof is very hard
- Robust to distributions of returns
  - Likelihood ratio is weakly dependent on return distribution
- Adapts to changes in tracking error
  - Can use it in any asset class without modification
- Very easy to implement
  - Can be done in Excel or in specialized SPC packages

CUSUM: Limitations

- Thoughtless use can lead users astray
  - Wrong benchmark is the most common error
- Does not provide a causal explanation for a change in performance
  - Use it to launch investigations, not as a hire/fire tool
- Somewhat sensitive to correlation
  - If correlation coefficient < 0.5, just raise the threshold
  - For higher correlation coefficients, must rework the recursion
  - Best solution: use the right benchmark

CUSUM in Practice

- Cusum is extraordinarily powerful, but can be abused
  - Extreme robustness can lead to abuse
- Do not run it on autopilot as a hire / fire tool
  - It is a *monitoring* and *investigative* tool
- Run additional tests when an alarm is raised
  - Determine *why* the manager underperformed
- Ensure that the benchmark is good
  - Excess returns should be uncorrelated
- Thresholds are chosen to work well in practice
  - Don't second guess Cusum before an alarm is raised

Summary

- Cusum detects underperformance rapidly
  – Over 10 times faster than standard techniques
- Very powerful and reliable technique
  – Extremely robust- works across styles & asset classes
  – Very few false alarms in practice
  – Focuses attention on managers who require it
- In daily use at a number of large institutions
  – Plan sponsors, asset managers and consultants
  – Used to monitor over $500 billion in actively managed assets

Using Statistical Process Control To Monitor Active Managers

Thomas K. Philips
Chief Investment Officer
Paradigm Asset Management
650 Fifth Avenue
New York, NY 10019

Phone: (212) 771-6115
e-mail: tkp@paradigmasset.com

Emmanuel Yashchin
Research Staff Member
IBM T.J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10547

Phone: (914) 945-1828
e-mail: yashchi@us.ibm.com

David M. Stein
Managing Director
Parametric Portfolio Associates
7310 Columbia Center
701 5th Avenue
Seattle, WA 98104

Phone: (206) 386-5594
e-mail: dstein@paraport.com

This version: 2/24/2003

Preprint: Do not quote without permission

To appear in the Journal of Portfolio Management

ABSTRACT

Investors and CIOs who are invested in (or bear responsibility for) many active portfolios face a resource allocation problem: To which products should they direct their attention and scrutiny? Ideally they will focus their attention on portfolios that appear to be in trouble, but these are not easily identified using classical methods of performance evaluation.

In fact, it is often claimed that it takes forty years to determine whether an active portfolio outperforms its benchmark. The claim is fallacious. In this article, we show how a statistical process control scheme known as the CUSUM can be used to reliably detect flat-to-the-benchmark performance in forty months, and underperformance faster still. By rapidly detecting underperformance, the CUSUM allows investors and CIOs to focus their attention on potential problems before they have a serious impact on the performance of the overall portfolio.

The CUSUM procedure is provably optimal: for any given rate of false alarms, no other procedure can detect underperformance faster. It is robust to the distribution of excess returns, allowing its use in almost any asset class without modification, and is currently being used to monitor over $500 billion in actively managed assets.

Introduction and Overview

Most traditional performance measurement algorithms (see, for example, Bodie, Kane and Marcus [1999], and the references cited therein) measure the performance of a portfolio over a fixed horizon, typically the three to five most recent years, and suffer consequently from a serious limitation. Good performance in some years can mask poor performance in others, making it difficult to estimate the portfolio's current performance, and harder still to identify transitions from good performance from bad.

Indeed, it is often claimed that it takes 40 years to determine whether an active manager has outperformed his benchmark by a statistically significant margin. While this is true if the manager's mean excess return is stationary and if one requires 95% confidence that it is positive, it is of little use to investors, who should not (and do not) wait 40 years to determine if their portfolio's mean excess return is both stationary and positive.

In fact, time variation in the mean excess return is the norm in an efficient market. As inefficiencies appear and are then arbitraged away, all investment processes will at some times flourish, and at other times stagnate or stumble, necessitating a fundamentally different approach to performance measurement: investors ought to continually estimate the *current* performance of their portfolios, and to rigorously re-evaluate each manager's investment strategy as soon as they can determine that it no longer adds value.

Dynamic performance measurement and change point detection are closely related, and have long been addressed in the fields of Sequential Analysis and Statistical Process

Control (SPC). In his seminal book, Abraham Wald [1947] showed how Sequential Analysis could greatly speed up the detection of change in a wide variety of systems, and his insights led to the development of many algorithms for change point detection.

In this article, we describe one such procedure, the CUSUM, and show that with an appropriate choice of parameters, it can detect a transition from outperformance to flat-to-the-benchmark performance in 40 months, and to underperformance faster still. In fact, Moustakides [1986] proves that for any fixed rate of false alarms, the CUSUM is the fastest way there is to detect a change in performance.

Furthermore, if the portfolio's performance is stationary, i.e. if its expected excess return is constant, the power of the CUSUM test is similar to that of Wald's [1947] Sequential Probability Ratio Test (SPRT), which, being asymptotically optimal, is more efficient than the widely used t-test for a difference in means. The CUSUM offers the user the best of both worlds: when performance is stationary, it discriminates good performance from bad about as well as the SPRT, and better than the t-test. However, when performance is time-varying, it detects changes much faster than both.

Though the CUSUM procedure is easily described, its mathematics are complex. This article is written for practitioners, and is organized as follows. We first specify the measure of performance we think it best to monitor. Following this, we touch briefly on the theoretical underpinnings of the CUSUM. We then describe the CUSUM procedure. algorithmically, along with a brief description of the mathematics underlying each step. Finally, we discuss it implementation and use, along with its strengths and limitations.

It is our objective to describe the CUSUM in layman's language, and to provide sufficient information to allow its implementation. We refer the mathematically sophisticated reader to Wald [1947], Moustakides [1986], Basseville and Nikiforov [1993], Yashchin, Philips and Stein [1997] and the references cited therein.

The Information Ratio

A useful measure of performance must incorporate both risk and return. While there are many such measures of performance, we have found one to be particularly useful. It is the information ratio, which is defined to be the ratio of a portfolio's excess return relative to an appropriate benchmark to its tracking error, or the standard deviation of its excess return relative to the same benchmark. A lucid discussion of the properties of the information ratio can be found in Grinold and Kahn [1999]. Even though investors typically specify a portfolio's performance in terms of its excess return, there are many advantages to monitoring its information ratio instead.

First, a portfolio's excess return is positive if and only if its information ratio is positive. Second, its excess return is the product of its information ratio and its tracking error, readily allowing a translation between one and the other. Third, while the achievable level of excess return varies significantly across asset classes, information ratios are far more stable, and monitoring the information ratio instead of the excess return allows the CUSUM procedure to be used across asset classes without modification. Finally, and most important of all, the information ratio is directly related to a manager's skill.

With an appropriate benchmark, the sequence of excess returns is uncorrelated and approximately normally distributed, and the probability that it will outperform its benchmark over any specified horizon is simply related to its information ratio. Table 1 illustrates the relationship, and contains two surprises. First, even a modest information ratio (0.25) leads to a reasonable probability of outperformance over a five-year horizon. Second, an information ratio of 1 virtually guarantees outperformance.

While the CUSUM is readily adapted to monitor other measures of performance such as the Sharpe Ratio, the Treynor Ratio or the intercept and slope (a and b) in a single factor model, the variance of these parameters tends to be significantly higher than that of the information ratio, slowing the detection of changes in performance.

The CUSUM Procedure: Theory

The CUSUM procedure was first proposed by Page [1954] as a method to rapidly detect changes in the mean of a noisy random process. While the mathematics of Page's solution are complex, we can appreciate his insight every time we compare a chart of monthly and cumulative excess returns for an active manager. The sequence of monthly excess returns is noisy, and it extremely difficult to determine the magnitude, or even the sign, of the mean excess return. It is harder still to detect changes in the mean.

The cumulative excess return chart however, tells a more convincing story, as the slope of the cumulative excess return plot equals the portfolio's average excess return. If the portfolio outperforms its benchmark, the cumulative excess return plot trends upward.

Correspondingly, if the portfolio underperforms its benchmark, it trends downward. Finally, if the portfolio's performance is flat to its benchmark, the cumulative excess return plot is trendless. Changes in slope are easily detected visually, and a quick look at the cumulative excess return plot elicits a great deal of information about a portfolio's performance that is not easily divined from its monthly excess returns.

The CUSUM procedure is a mathematical implementation of this insight. Formally, the CUSUM procedure is a *Backward Looking Sequential Probability Ratio Test*. Each time a new return is recorded (typically every month), it re-computes the optimum interval over which to measure the portfolio's performance relative to its benchmark. It then computes the relative probability of the observed sequence of excess returns over this interval being generated by good and bad investment processes, and raises an alarm when this ratio exceeds a user-defined threshold.

Figures 1 through 4 show the monthly excess return, the exponentially weighted tracking error, the cumulative excess return and the cumulative estimated information ratio of a large value manager relative to the Russell 1000 Value Index. The protractors of annualized excess return (in percent per annum) and annualized information ratio displayed on the left of the cumulative sum plots allow one to quickly estimate the performance over any period of interest by comparing the slope over the period to the slopes displayed in the protractors. The visual power of the cumulative sum plots is striking. At a single glance, one can identify periods of underperformance and outperformance, and the times at which transitions from one regime to the next occurred.

The CUSUM Procedure: Description

Following an initial phase in which the CUSUM's current estimate of the information ratio is set to 0 and its estimate of the tracking error is set to its expected value, a three-step procedure is executed every time a new return is recorded. A detailed explanation of each step follows this simple algorithmic description.

1. Estimate the portfolio's current information ratio using the newly recorded return.
2. Compute both the optimum performance measurement interval and the log-likelihood ratio over this interval. The log-likelihood ratio is the natural logarithm of the ratio of the probability that the observed sequence of returns was generated by a bad manager to the probability that it was generated by a good manager.
3. If the log-likelihood ratio does not exceed a user-defined threshold, do nothing. If it does, raise an alarm: there is sufficient statistical evidence to suggest that the manager's performance has changed from good to bad. Stop monitoring and conduct an investigation to determine the root cause of the underperformance.

Step 0: Initialize the CUSUM

Set the initial estimate of the manager's information ratio to 0, and the initialize the tracking error to its expected value. The expected tracking error can be derived from a time series of the portfolio's historical excess returns relative to its benchmark and an examination of the investment process.

Step 1: Estimate The Current Information Ratio

Define r_i and b_i to be the return of the portfolio and the benchmark respectively in month *i*. The logarithmic excess return in month i, e_i, is defined by

$$e_i = \ln\left(\frac{1+r_i}{1+b_i}\right). \tag{1}$$

The use of logarithmic excess returns ensures that the managers' excess returns are correctly compounded. Furthermore, if e_i is normally distributed, it is a maximum likelihood estimate of the portfolio's current compound excess return.

Next, define s_i and IR_i to be the annualized tracking error and information ratio respectively of the portfolio in month *i*. We estimate the tracking error using an exponentially weighted variant of Von Neumann's [1941] estimator of variance. This estimator exploits that fact that for two uncorrelated random variables with the same mean, $12 \times E[(e_i - e_{i-1})^2] = s_i^2 + s_{i-1}^2$, where $E[.]$ denotes expectation. It adapts to a slowly changing mean and variance, and is given by

$$\hat{s}_0^2 = \hat{s}_1^2 = s_0^2 \tag{2}$$

$$\hat{s}_i^2 = g\hat{s}_{i-1}^2 + (1-g)\frac{12 \times (e_i - e_{i-1})^2}{2}, \qquad 0 < g < 1, \qquad i = 2,3,\mathrm{K} \tag{3}$$

$$\hat{s}_i = \sqrt{\hat{s}_i^2},\ i = 0,1,2,3,\mathrm{K} \tag{4}$$

Setting $g = 1$ is equivalent to assuming that the tracking error equals s_0 at all times. If, on the other hand, $g < 1$, the estimator adapts to changes in tracking error. Clearly, there is a trade-off to be made: lowering g decreases the time taken to respond to changes in tracking error, but at the expense of increasing the noise in the estimate. In practice, setting $g = 0.9$ provides a good compromise between detection speed and noise.

This estimator has two significant advantages over the standard "sum of squared deviations from the sample mean" estimator. First, subtracting adjacent returns effects a cancellation between their means, making the estimator insensitive to slow changes in the process mean. Secondly, an abrupt change in the mean distorts only two point estimates, and its effect is rapidly extinguished through exponential smoothing.

The current information ratio is estimated as

$$\hat{IR}_i = \frac{12e_i}{\hat{s}_{i-1}}. \tag{5}$$

Using $\hat{s}_{i-1}$ instead of $\hat{s}_i$ prevents e_i from appearing simultaneously in both the numerator and the denominator, and ensures that our estimates of the information ratio remain approximately unbiased and uncorrelated.

Step 2: Determine the optimum estimation interval and compute the log-likelihood ratio

Each time a new return is recorded, compute both the optimum performance measurement interval and the log-likelihood ratio over this interval. The log-likelihood ratio is the logarithm of the ratio of the probability that the observed sequence of information ratios was generated by a bad manager to the probability that the observed sequence of information ratios was generated by a good manager. The optimum performance measurement interval is the interval that maximizes the log-likelihood ratio.

Based on our collective investment experience, we believe that *an information ratio of 0.5 or better constitutes good performance*. As an active portfolio that cannot outperform its benchmark is easily replaced by an index fund, *an information ratio of 0 or worse constitutes bad performance*. Correspondingly, we define a good manager to be one whose information ratio is 0.5 or better, and a bad manager to be one whose information ratio is 0 or worse. If we have N observations since the portfolio's inception, the log-likelihood ratio based on the k most recent observations is given by

$$L_N(k) = \log\left[\frac{\text{Probability}\,[k \text{ most recent observations} \mid \text{Information Ratio} = 0]}{\text{Probability}\,[k \text{ most recent observations} \mid \text{Information Ratio} = 0.5]}\right], \qquad (6)$$

After computing $L_N(1),\ L_N(2),\ \Lambda\ , L_N(N),$ the CUSUM determines the optimum performance measurement interval by identifying k^*, the value of k for which $L_N(k)$ is maximized, thus maximizing its ability to discriminate between good and bad active managers. In practice, it proves convenient to work with just the non-negative part of the log-likelihood ratio, which we define by $L_N = \max[0, L_N(k^*)]$.

If $L_N(k^*) < 0$, the information ratio is more likely to be positive than negative, and no information is discarded by setting L_N to 0. On the other hand, if $L_N(k^*) \geq 0$, $L_N = L_N(k^*)$, and L_N parsimoniously encapsulates all the evidence available at time N to reject the hypothesis that the portfolio's performance is satisfactory. More importantly, if the excess returns of the portfolio are independent and identically distributed normal random variables, L_N can be computed using a simple recursion:

$$
\begin{aligned}
L_0 &= 0 \\
L_N &= \max\left[0, L_{N-1} + L_N(1)\right], \qquad N = 1,2,\ldots. \\
&= \max\left[0, L_{N-1} - \hat{IR}_N + 0.25\right].
\end{aligned} \tag{7}
$$

This conscious maximization of the log-likelihood ratio by determining the optimal estimation interval differentiates the CUSUM from all other performance measurement methods, and is the root source of its ability to rapidly detect a change in performance. It can be shown that the distribution of L_N is only weakly dependent on the actual distribution of excess returns, allowing the use of the CUSUM to monitor active portfolios in almost any publicly traded asset class without modification.

Step 3: Compare the log-likelihood ratio to a threshold and raise an alarm if necessary

If L_N is positive, we ask – *How likely is it that the portfolio's underperformance is caused by a decline in the ability of the investment process to add value*?

Clearly, if the log-likelihood ratio is small, the underperformance is not statistically significant: all active managers are afflicted by an occasional bout of underperformance. If, on the other hand, it is large, it is likely that the underperformance is caused by a decline in the ability of the investment process to add value. The ability to discriminate between a fundamentally flawed investment process and random noise in a viable investment process is an increasing function of the log-likelihood ratio.

The CUSUM procedure explicitly recognizes this trade-off between detection speed and the ability to discriminate between good and bad managers. It requires us to specify a threshold for the log-likelihood ratio below which no decision is made about the state of the investment process. When this threshold is crossed, sufficient evidence has accrued to conclude that its ability to add value has declined.

The threshold determines both the average time taken to detect underperformance by a bad manager and the rate of false alarms (incorrect identifications of a good manager as bad). If the threshold is set low, underperformance is rapidly detected, but we experience a correspondingly high rate of false alarms. If, on the other hand, we set the threshold high, it takes longer to detect underperformance, but the rate of false alarms drops.

Every statistical decision procedure must make this trade-off between detection speed and the rate of false alarms. Moustakides [1986] proves that under some relatively general conditions, the trade-off made by the CUSUM procedure is optimal – for *any* distribution of returns that is drawn from the exponential family[1], and for *any* fixed rate of false alarms, no other procedure can detect underperformance faster. His article represents the culmination of a twenty-five year effort to prove this remarkable result.

In practice, a threshold setting that detects flat-to-the-benchmark performance in three and a half years and allows one false alarm in seven years proves satisfactory. This setting was determined by examining the actual investment results of a large number of domestic and international equity and fixed income managers. It affords a ten-fold improvement in detection speed over the traditional t-test for a difference in means while still maintaining a reasonably low rate of false alarms.

Table 2 shows the thresholds required for various expected times between false alarms. These thresholds were computed using the approximate computational approach described in Yashchin, Philips and Stein [1997] and were verified using a simulator. Appropriate thresholds can be computed for any desired level of good and bad performance – we refer the interested reader to Yashchin, Philips and Stein [1997], Vance [1986] and Woodall [1983] for a detailed description of the numerical procedure.

---

[1] Formally, the exponential family of distributions consists of all distributions whose density function can be written in the form $f(x,q) = h(x)g(q)\exp\left(\sum w_i(q)t_i(x)\right)$, where $h(x)$ and $t_i(x)$ are functions only of *x*, while $g(q)$ and $w_i(q)$ are functions only of q. It is a large family, and includes most distributions that are used to model portfolio returns, including the normal, log-normal, student-t and gamma.

The Cusum Procedure: Implementation and use

The CUSUM procedure is implemented using the following sequence of steps:

1. Initialize the tracking error s_0 to its expected value and set $L_0 = 0$.
2. Each time a new return is recorded,
   a. Compute the logarithmic excess return using equation (1)
   b. Update the estimate of tracking error using equations (2), (3) and (4).
   c. Estimate the current information ratio using equation (5)
   d. Update the likelihood ratio L_N using equation (7)
3. Compare the updated likelihood ratio to the thresholds in Table 2
   a. If it does not exceed the user chosen threshold, do nothing
   b. If it does exceed the user chosen threshold, raise an alarm and investigate the manager's process to determine the cause of underperformance.
      i. If the investment process is found to be satisfactory, consider this to have been a false alarm. Go back to Step 1, reset the CUSUM, and continue monitoring the portfolio.
      ii. If the investment process is found deficient, stop the CUSUM and decide on a course of action.

Figure 5 displays the log-likelihood ratio plot (also known as Page's plot in honor of E.S. Page) for the manager whose performance is shown in Figures 1 to 4. When the portfolio underperforms its benchmark, the log-likelihood ratio increases (i.e. moves downward). When the portfolio outperforms its benchmark, it decreases (i.e. moves upward). We have inverted the direction of the y-axis so that it corresponds to our established notion that good performance points upward while bad performance points down.

The plot has six labeled horizontal levels. Each corresponds to a row in Table 2, and is labeled with three numbers. The first number is the expected time (in months) to cross this level if the manager is good, i.e., if his information ratio is 0.5. The second number is the expected time to cross this level if the manager is bad, or if his information ratio is 0. The third number is the expected time to cross this level if the manager is exceptionally bad, or if his information ratio is –0.5. On average, a good manager's log-likelihood ratio will cross the first level once in 24 months, while a bad manager's log-likelihood ratio will cross it once in 16 months, and an exceptionally bad manager's log-likelihood ratio will cross it once in 11 months.

Notice that the average time to cross a level increases monotonically as we go down levels, and decreases monotonically from left to right at any given level. This is to be expected. The more confident we wish to be of our ability to discriminate between good and bad managers, the longer we must wait. The worse a portfolio's information ratio, the sooner its log-likelihood ratio will cross any given level.

When the log-likelihood ratio crosses the lowest level, an alarm is raised. At this point, sufficient evidence has accrued to warrant an investigation of the manager's process. It takes only 41 months to detect flat to the benchmark performance. However, once in 84 months, a good manager's log-likelihood ratio will cross this level by chance and a false alarm will be raised. Following the alarm, an investigation must be launched. If, after the investigation, it is concluded that the manager's investment process is satisfactory and that the alarm is false, the CUSUM is reset and the monitoring process is restarted.

In our example, two alarms were raised: one in March 1993 and the other in December 2000. The first alarm was, with the benefit of hindsight, a false alarm – the manager's performance recovered after the alarm was raised. The second one was valid – the manager's performance has become far more volatile, and in spite of some sharp reversals, its performance continues to deteriorate.

This raises a very important question – *what ought one to do when an alarm is raised*? The CUSUM procedure detects underperformance, but offers no causal explanation for it. **It should not, therefore, be used as a tool to engage and terminate asset managers based solely upon their recent performance**. It is incumbent upon, and indeed imperative for, the user to launch a thorough investigation into the manager's investment process and to decide if it is likely to result in satisfactory performance going forward.

If the user concludes that the investment process is likely to generate satisfactory returns in the future, the alarm is likely to have been false, the manager ought to be retained and the CUSUM re-initialized. If, on the other hand, she feels that the process is unlikely to perform well in the future, a number of courses of action, up to and including termination, are available, and an appropriate decision must be made. In the example presented, an examination of the portfolio and a discussion with the portfolio manager resulted in an understanding of why its risk has increased so sharply in recent years.

Like any performance measurement technique, the CUSUM has both strengths and limitations, and users must be aware of both to maximize its utility. Its strengths include:

1. A transition from outperformance to underperformance is detected far more rapidly than with traditional methods.
2. The CUSUM equations are simple, and the algorithm it is easily implemented in a spreadsheet, a statistical package such as S-Plus, SAS or Matlab, or a specialized process control package such as SPC/PI+ or SPC-PC.
3. The CUSUM is robust to the distribution of excess returns. It can therefore be used to monitor portfolios in all publicly traded asset classes without changing the threshold at which an alarm is raised.
4. It adapts to changes in tracking error, regardless of their cause, and
5. Investigations launched after an alarm was raised have almost always resulted in the identification of the underlying cause of the performance problem.

Its primary limitation is its sensitivity to correlation in the sequence of excess returns, as equation (7) is valid only when returns are independent. In practice, however, serial correlation does not pose a major problem for the following reasons.

First, the choice of an appropriate benchmark will ensure that the sequence of excess returns is essentially uncorrelated. This is easily accomplished in practice: large value managers should be benchmarked against large value indices and fixed income managers should be benchmarked against bond indices of similar duration and credit quality

Second, Yashchin [1993] shows that if the correlation in the series of estimated information ratios is reasonably small (i.e. if $|r| \leq 0.5$), simply modifying the threshold at which an alarm is raised is almost as effective as an exact computation of the log-likelihood ratio. Positive correlation requires the threshold to be raised, while negative correlation requires it to be lowered.

However, a decade of experience with a wide range of domestic and international equity and fixed income managers has shown that if a portfolio is correctly benchmarked, its excess returns and estimated information ratios are almost always nearly uncorrelated. With this elementary precaution (which should be an integral part of *any* manager evaluation), even unsophisticated users can successfully use the CUSUM, for prior to an alarm being raised, there is no need for intervention by a skilled person.

Summary

The CUSUM procedure is an exceptionally powerful tool for monitoring the performance of actively managed portfolios. It detects flat-to-the-benchmark performance in forty months, is robust to the distribution of excess returns, and works well with equity, fixed income and currency portfolios, both domestic and international. It is currently being used by plan sponsors, consultants and money managers on three continents to monitor over $500 billion in actively managed assets, and their experience with it has been uniformly positive.

Like all statistical methods that use only returns, the CUSUM does not provide a causal explanation for a manager's performance. It is, therefore, imperative to view an alarm as a call for an investigation into the root cause of a manager's underperformance, and not as an automatic signal to terminate the manager.

Acknowledgement

The author would like to thank Richard Michaud of New Frontier Advisors for his many thoughtful comments, which led to substantial improvements in the presentation of this paper.

References

Basseville, Michele, and Igor Nikiforov, 1993. *Detection of Abrupt Changes -Theory and Application*, Englewood Cliffs, NJ: Prentice Hall.

Bodie, Zvi, Alex Kane, and Alan Marcus. 1999. *Investments, 4th Edition*, Boston, MA: Irwin/McGraw-Hill

Grinold, Richard and Ronald Kahn. 1999. *Active Portfolio Management, 2nd Ed.*, New York, NY: McGraw-Hill.

Moustakides, George. 1986. "Optimal Stopping Times for Detecting Changes in Distributions." *Annals of Statistics*, v. 14, no. 2:1379-1387.

Page. E.S. 1954."Continuous Inspection Schemes." *Biometrika*, v. 41, no. 1:100-115.

von Neumann, John, R.H. Kent, H.R. Bellinson and B.I. Hart. 1941. "The Mean Square Successive Difference." *Ann. Math. Statistics*, v. 12, no. 2 (June):153-162.

Vance, Lonnie C., 1986. "Average Run Lengths of Cumulative Sum Control Charts for Controlling Normal Means", *Journal of Quality Technology*, v. 18, no. 3:189-193.

Wald, Abraham. 1947. *Sequential Analysis*, New York: Wiley.

Woodall, William. 1983. "The Distribution of the Run Length of One-Sided CUSUM Procedures for Continuous Random Variables", *Technometrics*, v. 25, no. 3:295-301.

Yashchin, Emmanuel. 1993. "Performance of CUSUM Control Schemes for Serially Correlated Observations." *Technometrics*, v. 35, no. 1 (February):37-51.

Yashchin, Emmanuel, Thomas K. Philips, and David M. Stein. 1997. “Monitoring Active Portfolios Using Statistical Process Control.” In *Computational Approaches to Economic Problems*, Hans Amman, Berc Rustem and Andrew Whinston (eds), Dordrecht, Netherlands: Kluwer Academic Publishers.

Information Ratio	***Probability Of Outperforming The Benchmark Over A Given Horizon***		
	1 Year	**3 Years**	**5 Years**
0.00	0.50	0.50	0.50
0.25	0.60	0.67	0.71
0.50	0.69	0.81	0.87
0.75	0.77	0.90	0.95
1.00	0.84	0.96	0.99

Table 1. Probability of Outperformance vs. Information Ratio

Threshold for L_N	*Expected Time Taken To Cross Threshold (Months)*		
	IR=0.5: T_0	IR=0: T_1	IR=-0.5: T_2
11.81	24	16	11
15.00	36	22	15
17.60	48	27	18
19.81	60	32	21
21.79	72	37	23
23.59	84	41	25

Table 2. Detection speed vs. Threshold

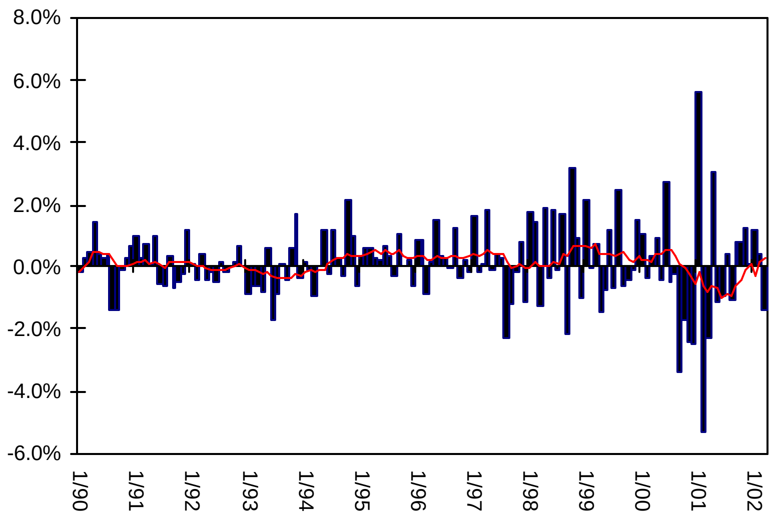


Figure 1: Monthly Excess Returns: Large Value Mutual Fund vs. Russell 1000 Value Index

Figure 2: Exponentially Weighted Tracking Error: Large Value Mutual Fund vs. Russell 1000 Value Index

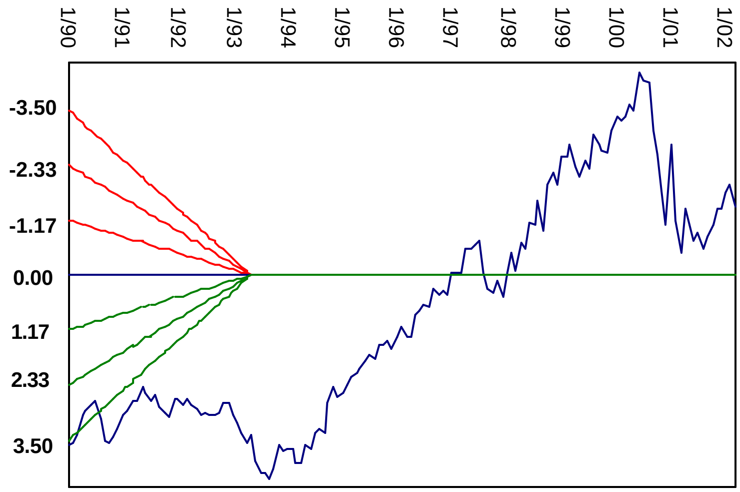


Figure 3: Cumulative Excess Return: Large Value Mutual Fund vs. Russell 1000 Value Index

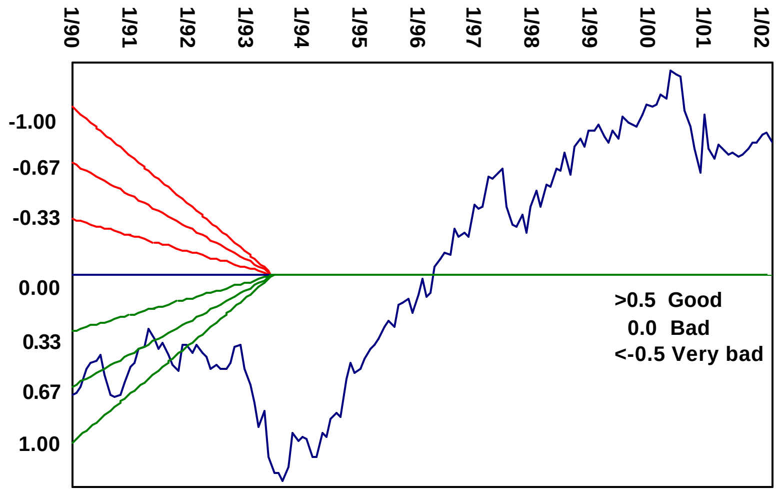


Figure 4: Cumulative Estimated Information Ratio: Large Value Mutual Fund vs. Russell 1000 Value Index

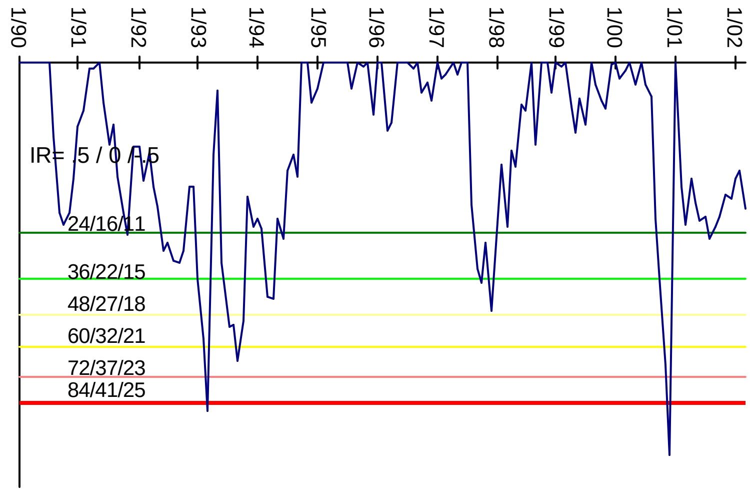


Figure 5: Page's Plot of Log-Likelihood Ratios: Large Value Mutual Fund vs. Russell 1000 Value Index