

A Look at Sequential Normal Scores and How They Apply to Financial Data Analysis

W. J. Conover¹, Víctor G. Tercero-Gómez², Alvaro E. Cordero-Franco³

¹Texas Tech University, Lubbock, Texas, USA

²Tecnológico de Monterrey, Monterrey, México

³Universidad Autónoma de Nuevo León, Monterrey, México

Email: jay.conover@ttu.edu, victor.tercero@gmail.com, lalo.cordero@gmail.com

How to cite this paper: Conover, W.J., Tercero-Gómez, V.G. and Cordero-Franco, A.E. (2018) A Look at Sequential Normal Scores and How They Apply to Financial Data Analysis. *Journal of Applied Mathematics and Physics*, 6, 787-816.
<https://doi.org/10.4236/jamp.2018.64069>

Received: February 27, 2018

Accepted: April 21, 2018

Published: April 24, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Statistical methods for analyzing economic data need to be timely, accurate and easy to compute. To accomplish this, parametric models are often assumed, but they are at best approximate, and often lack a good fit in the tails of the distribution where much of the interesting data are concentrated. Therefore, nonparametric methods have been extensively examined as alternatives to the constrictive assumptions of parametric models. This paper examines the use of Sequential Normal Scores (SNS) for transforming time series data with unknown distributions into time series data that are approximately standard-normally distributed. Particular attention is directed toward detecting outliers (out-of-control values), and applying subsequent analytic methods such as CUSUMs and Exponentially Weighted Moving Average (EWMA) schemes. Two examples of stock market data are presented for illustration.

Keywords

Sequential Analysis, Sequential Ranks, Transformation, Extreme Values, Outliers

1. Introduction

Many popular statistical methods for detecting outliers and change points in time series data fall into the category of Statistical Process Management (SPM), also known as Statistical Process Control (SPC), and include Shewhart Charts, Cumulative Sums (CUSUMs), and Exponentially Weighted Moving Averages (EWMA). This paper is the first to apply a new method of SPM, called Sequential Normal Scores, to economic time series. The Sequential Normal Scores

transformation transforms any sequence of independent and identically distributed random variables to independent and identically distributed scores that are approximately standard normal random variables. Sequential Normal Scores can easily identify outliers, and allow CUSUM and EWMA methods to be applied to data, without assuming any distributional form, and are thus distribution-free, or nonparametric. Advantages of Sequential Normal Scores over other SPM procedures in analyzing economic time series are demonstrated.

1.1. Sequential Normal Scores

Parametric methods for analyzing time series involve assuming a probability distribution with parameters estimated from the data. These methods suffer from the fact that assumed distributions are approximate at best, with weaknesses in the fit generally in the tails of the distribution where the outliers occur. The most popular nonparametric methods usually involve ranks, and many nonparametric methods are computationally intensive. “Sequential ranks” were introduced in [1] to greatly reduce the computational complexity, because only the most recent observation is ranked (relative to the previous observations)—the previous sequential ranks remain unchanged. Reference [2] and others have developed methods of analysis such as CUSUMs based on sequential ranks. These methods are not simple to use, and thus counteract the speed and convenience intended by using sequential ranks. This complexity has perhaps been the chief reason methods based on sequential ranks are not popular in time series analysis.

Sequential ranks in this study are converted to normal scores, called Sequential Normal Scores [3], to alleviate this problem of complexity in statistical methods. That is, each new observation is converted to a rank relative to the n previous observations in the series, or in a fixed-width moving window of previous observations, but the previous ranks are not changed. This is called a sequential rank.

The sequential rank is converted to a rankit, $(Rank - 0.5)/n$, and then to a Normal Score using the inverse function of a standard normal distribution, as described in more detail in [3]. These are called Sequential Normal Scores (SNS) and are approximately standard normal random variables that are highly correlated with the original data, so parametric methods that assume normality can easily be used for analysis without the need for estimating any unknown parameters.

The ordinary ranks for a sequence of observations have a one-to-one relationship with the sequential ranks for that same sequence of observations. That is, knowledge of only the ordinary ranks for a sequence of observations enables one to construct the sequential ranks for that same sequence, and vice-versa. The same is true for the Normal Scores computed from the ranks, and the Sequential Normal Scores computed from the Sequential Ranks. Thus all the information about the original data that is captured by the ordinary ranks, is also captured by

Normal Scores computed from the original ranks, and by the sequential ranks, and by the Sequential Normal Scores. It is with this principle in mind that we propose using Sequential Normal Scores to perform nonparametric analyses of data when the probability distribution of the original data is unknown.

1.2. Statistical Process Management (SPM) Methods

Shewhart Charts. In this paper, we convert time series observations to Sequential Normal Scores and use standard Shewhart Charts with 3-sigma limits. That is, observations are declared to be outliers if they exceed the 3-sigma limits from the mean, in either direction. Because Sequential Normal Scores are approximately standard normal in distribution, they are declared outliers if they are greater than 3.0 in absolute value.

Most sequential analysis methods in Statistical Process Management assume a Phase I series of observations when the process is known to be “in control”, to establish parameters sufficient for setting control limits. Then the subsequent observations are matched against those control limits to determine when a process is out of control. With economic data there is no “in control” set of observations, so “self-starting” methods are appropriate, to begin the analysis at any time.

Sequential Normal Scores are self-starting, and do not require a Phase I to set the parameters. However, when using 3-sigma control limits, more than 370 observations (trading days) are required before a new observation can have a sequential rank small enough or large enough for the subsequent standard normal score to exceed 3 in absolute value. In economic data with daily observations this converts to 18 months of data, during which time Sequential Normal Scores are collected but all will be within the 3-sigma (equal to plus or minus 3 in our case) limits. We suggest at least two years of data to establish “recent history” against which new observations are matched. Our examples will use two years of data (a moving window of 500 trading days) but the results are similar to the results using longer moving windows.

Although there is nothing preventing the use of the complete available history of observations, comparing new observations with values from distant history may not be as meaningful as comparing observations with only more recent data. Therefore we will look at eight cases of moving windows of past observations. With a moving window, observations are matched against only values in the previous $w = 250k$ days, for $k = 1$ through 8. With stock market data 250 trading days equals approximately one calendar year, so our windows correspond roughly to years of history, from one year to eight years.

If the variance of a sequence increases sharply, the new observations have a greater probability of being declared outliers. Thus an increase in the frequency of outliers may be an indication of an increase in the variability in the sequence, whether it be measured by z -scores or by sequential ranks. Because the observations (daily yields) are assumed to be independent, the sequential ranks and

therefore the Sequential Normal Scores will be independent.

Although Shewhart Chart outliers indicate unusual activity regarding the daily yield relative to historical values, it is also of interest to detect smaller changes in mean yield that are sustained over a period of time. Two popular methods for measuring sustained changes in mean yield are the cumulative changes as measured by a method known as CUSUMs [4] or exponentially weighted moving averages known as EWMA methods [5]. The purpose of this examination is not to evaluate the merits or failings of the CUSUM or EWMA procedures, but to demonstrate that the CUSUM (or EWMA) on the Sequential Normal Scores resembles the CUSUM (or EWMA) on the z -scores of the raw data in many respects, but eliminates many of the alarms that are likely to be false alarms, caused by outliers in the original data. In this sense the CUSUM (or EWMA) on the Sequential Normal Scores can be used to determine analytics on the original data, and may be superior to the analytics obtained by calculating the CUSUM (or EWMA) on the original data.

Cumulative Sums. The CUSUM method examines cumulative sums in the positive direction $S\text{-}pos$ as well as cumulative sums in the negative direction $S\text{-}neg$ for independent standard normal random variables $Z_j, j = 1, 2, \dots$, using the equations

$$S\text{-}pos(Z_1) = 0, S\text{-}pos(Z_j) = \max(0, Z_j + S\text{-}pos(Z_{j-1}) - k), j > 1 \quad (1)$$

$$S\text{-}neg(Z_1) = 0, S\text{-}neg(Z_j) = \min(0, Z_j + S\text{-}neg(Z_{j-1}) + k), j > 1 \quad (2)$$

for some suitably chosen constant k . The rationale behind the CUSUMs is to accumulate z scores that are indicating a possible shift in the mean, either up or down. A penalty of k standard deviations is applied at each step, and the accumulation is not allowed to change signs. We are using $k = 0.5$ and declaring “significance” if $S\text{-}pos > +4.774$ or if $S\text{-}neg < -4.774$ because, according to the R-program `spc`, this will result in a probability of declaring a false positive (Type 1 error) of 0.0027, matching the 3-sigma limits for the Shewhart charts. That is, the average run length between successive CUSUM outliers is about 370 for standard normal data, the same as we are using in the Shewhart Charts.

Exponentially Weighted Moving Averages. The EWMA E_j on the j th observation Z_j is given by

$$E_1 = 0 \text{ and } E_i = \lambda Z_i + (1 - \lambda) E_{i-1} \text{ for } i > 1 \quad (3)$$

for some suitably chosen value λ . According to Table 5.1 on page 188 of [6] the choice of $\lambda = 0.2$ and $\rho = 2.859$ will result in boundaries of ± 0.953 to attain a false positive probability of 0.0027, again matching the 3-sigma limits for the Shewhart charts, if the observations are independent standard normal random variables.

1.3. Outliers

First we will concentrate on identifying outliers, or out-of-control observations,

defined as observations whose Sequential Normal Scores are greater than 3 in absolute value, corresponding to 3-sigma limits with normal data. There are many ways of classifying observations as outliers and they all involve much subjective reasoning. We feel that the qualities of a good outlier test include the following basic characteristics.

- 1) The results should be consistent with other tests for outliers.
- 2) The results should include “obvious” outliers.
- 3) The results should not include too many false positives.
- 4) The results should not include too few declared outliers (too many false negatives).
- 5) The method should be intuitive to the average person.

Therefore we will examine several methods for identifying outliers and compare them on their merits as defined above. These methods are all self-starting, and therefore do not require “Phase 1” observations to estimate parameters or distributions. The four basic methods are as follows.

1) USING THE RAW DATA: The daily yield will be converted to a z -score by subtracting the mean of the previous observations (as few as 2 are required) and dividing by the standard deviation of the previous observations, and declared an outlier if the result is greater than 3 in absolute value. Only previous observations within a moving window of fixed width are used, to keep the comparison current. If the z -scores agree with the normality assumption the probability of being declared an outlier is close to 0.0027, a very small number.

2) USING THE LOG-TRANSFORMED DATA: The log transformation is popular for financial data of this type for two reasons. First, if the data have outliers resulting from a right-skewed distribution, the log transformation may make the data resemble more closely a normal distribution. Second, as prices tend to trend upwards or downwards over time the percentage change may be a more realistic statistic to use over long time periods rather than a simple arithmetic difference. Therefore the difference between the log price for one day, minus the log price for the previous trading day, which equals the log of the ratio of the two prices, will be the basic observation, and will be treated as above. That is, it will be converted to z -scores and compared with 3 in absolute value. This method may not be as intuitive as the previous method, but it may be superior for other reasons. If the transformed data are approximately normal, the probability of an observation being declared an outlier is close to 0.0027.

3) SEQUENTIAL NORMAL SCORES: The data may be converted to Sequential Normal Scores (SNS), which are approximately standard normal in distribution, and which are highly correlated with the original raw data. An observation is declared an outlier if its SNS is greater than 3 in absolute value. This is a non-parametric method whose exact probability of declaring an outlier can be computed easily as a function of the length of the sequence, or the window size of the previous observations used in the ranking.

4) z -SCORES OF SEQUENTIAL NORMAL SCORES: In the event that SNS

values may stray by chance from their presumed mean of zero or standard deviation of one due to randomness, the z -scores of the SNS values may achieve more intuitive results. That is, the z -score of each SNS is computed by subtracting the mean of the previous SNS values and dividing by the standard deviation of the previous SNS values, to bring the mean and standard deviation closer to zero and one respectively.

All of these methods are self-starting, so they begin on the second or third (in the case of z -scores) observation in the series. Once they reach the desired size of the moving window, be it as few as 250 observations, or as many as 2000 observations, z -scores and Sequential Normal Scores are computed only on the observations in that moving window. With four different methods for declaring outliers, and eight different window sizes, comparisons can be made with the desired qualities stated above for 32 methods that identify outliers.

1.4. Economic Time Series

The value of a stock or a portfolio of investments continually changes over time. This paper examines only investments that are traded openly on the market, with the closing price on each trading day, listed on *yahoo.financial*, as the data being analyzed. Although stock prices are highly auto-correlated, the changes in stock prices from one trading day to the next appear to be independent observations, driven by the tendency for the market to be “efficient”. Actually, changes in stock prices, like many economic time series, have an almost negligible serial correlation, and appear to have no “memory”, like a random series, but their absolute values usually have a non-negligible serial correlation, unlike a random series.

We will show that Sequential Normal Scores convert original data to data that appear to have come from a standard normal distribution, while maintaining a high correlation with the original data, and retaining the non-random characteristics of the original data such as a slight serial correlation, a large serial correlation of absolute values, and a tendency for extreme observations to cluster. The result is a data sequence with a “known” probability distribution, that of the standard normal distribution, so statistical methods designed around standard normal random variables may be used to analyze the original time series. Our analysis is restricted to the change in price from one trading day to the next, and the percentage change from one trading day to the next, called the “yield” or the “relative yield”, which may be positive (a gain, or profit) or negative (a loss).

Although the change in price appears to be a sequence of independent observations, it has been known for many years that almost all financial or economic sequences have subtle dependencies that can be exploited to assist in predicting subsequent values. As noted in [7], “large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes.” This phenomenon has come to be known as “volatility clustering” and has been studied by many researchers. GARCH models such as presented in [8] and [9]

are among the first to be applied to this phenomenon, so sometimes this volatility clustering is called the GARCH effect. Other models such as those studied in [10], [11], and [12] produce values with little or no serial correlations, but whose absolute values or squares have substantial serial correlations and thus result in volatility clustering. Another theory, advanced in [13] and [14], attributes volatility clustering to the “herd” behavior of investors who respond almost simultaneously to other economic indicators. Our interest in this paper is not in producing models that simulate volatility clustering, but rather in detecting the presence of volatility clustering in observed sequences.

When dealing with autocorrelated processes, Statistical Process Management techniques make use of the marginal distribution of the stationary process under monitoring to detect outlying observations. We can see this approach in [15]. This situation is extended when we use SNS on an autocorrelated process as long as this process is stationary.

From [3] we see that SNS can be expressed as a function of sequential empirical quantiles. It was proven that empirical quantiles converge to the true quantiles when dealing with a series of independent observations. This fact was used to prove that SNS converge to the quantiles of a standard normal distribution. If we are dealing with autocorrelated observations, the results in [3] can be extended to stationary time series, using the theoretical results in [16]. From [16] we know that, within some general conditions usually satisfied by ARMA and GARCH processes, the empirical quantiles used in the calculation in SNS are unbiased and consistent estimators of the true quantiles of the marginal distribution of the stationary process. As a corollary, when observations in a stationary time series are transformed into SNS, SNS quantiles still converge to quantiles of the normal distribution, but this normal distribution represents the transformation of the marginal distribution of the stationary process.

1.5. Ordinary Ranks Compared with Sequential Ranks

There are three major differences between ordinary ranks and sequential ranks.

1) **Speed of computation.** In a sequence of observations, each time a new observation appears a new ordinary ranking involves re-computing the ranks of all previous observations and all of the statistics based on those ranks, while a new sequential ranking involves computing the rank of only the new observation – the previous observations retain their previous sequential ranks and the previous statistics computed on those ranks retain their previous values. In a sequence of hundreds of observations the computing time for re-ranking and re-computing is hundreds of times greater for ordinary rank methods than the computing time for sequential rank methods. Also, if an ordinary ranking procedure detects an outlier at some time point, a more extreme observation occurring later in the sequence may be declared an outlier at the expense of the new rank of the previous observation, which may no longer meet the definition of an outlier. Similarly, CUSUM procedures build on the previous observation’s CUSUM, and if

the previous observation's rank has changed, its CUSUM based on ranks has changed also, complicating any analysis based on CUSUMs. The same is true for EWMA. Analyses based on SNS have none of these disadvantages.

2) **Independence.** Sequential ranks, and their derived Sequential Normal Scores, in a series of independent observations are independent of each other, while ranks, and their derived Normal Scores, are not independent. They have a slight negative covariance and thus a slight negative correlation. Analytic methods based on the assumption of independent observations may no longer be accurate when applied to statistics based on ordinary ranks. This is not a problem with statistics based on sequential ranks or sequential normal scores, because they constitute a series of independent statistics if they are derived from independent observations.

3) **Flexible number of outliers.** A sequence of $n = 371$ observations, when converted to Normal Scores through "rankits" $(rank - 0.5)/n$ using ordinary ranks, is guaranteed to have exactly one Normal Score less than -3.0 (the observation with rank 1) somewhere in the sequence and exactly one Normal Score greater than $+3.0$ (the observation with rank 371). Thus it is guaranteed to have two and only two outliers (*i.e.*, Normal Scores greater than 3 in absolute value) until the sequence reaches a length of 1112 or more, at which point it will have exactly 4 outliers. When the sequence reaches 1852 observations in length it will then have exactly 6 outliers using ordinary ranks, and so on. That is not the case with Sequential Normal Scores, which may have 0 or 1 SNS less than -3 , and 0 or 1 SNS greater than $+3$ when it reaches 371 observations in length. A sequence with more than 371 observations, say $370 + k$, ($k > 1$), observations, may have up to k values of SNS less than -3 (because previous ranks of 1 remain ranks of 1) or greater than $+3$ (because previously declared "outliers" do not change their designation). That is, a sequence with 370 or fewer observations will not have any SNS greater than 3 in absolute value, but a sequence with $370 + k$ observations may have up to k SNS greater than 3 in absolute value. Thus SNS have more flexibility in declaring outliers than ordinary ranks.

The z -scores may have a slight dependence built in which we will ignore. In general the occurrence of outliers should be independent Bernoulli random variables with probability close to 0.0027 of occurring. Unlikely clusters of outliers may be the result of a violation of the independence assumption.

Clusters of outliers, statistically significant in their close proximity to each other, may signify an increase in variability in the sequence, or a sharp increase in the mean, or a sharp decrease in the mean. A simple nonparametric test may be used to determine if a cluster of outliers is statistically significant, as follows.

2. The Cluster Test for Outliers

For independent events (like out-of-control days) the binomial distribution can be used to find the probability of k or more out-of-control days (one minus the probability of $k - 1$ or fewer out-of-control days) in any given set of n observa-

tions, such as observations following an observed out-of-control day. In other words, the probability that an out-of-control observation is the beginning of a statistically significant cluster of k or more out-of-control observations within n consecutive trading days is given as follows. Let X equal the SNS or the z-score, let t be the trading day when $abs(X) > 3$, and let Y equal the number of additional out-of-control days observed within the next $n - 1$ consecutive trading days. Then:

$$\begin{aligned}
 &P(X \text{ is the first day in a cluster of at least } k \text{ out-of-control days} \\
 &\text{within } n \text{ trading days}) \\
 &= P(Y \geq k - 1 | Y \text{ is binomial with parameters } p = 0.0027 \text{ and } n - 1) \quad (4) \\
 &= 1 - P(Y \leq k - 2 | Y \text{ is binomial with parameters } p = 0.0027 \text{ and } n - 1) \\
 &= 1 - \text{BINOMDIST}(k - 2, n - 1, 0.0027, 1)
 \end{aligned}$$

using Excel’s BINOMDIST function that produces cumulative binomial probabilities. We refer to this as the *cluster test*. A brief table of statistically significant clusters is given in **Table 1**.

A cluster of k outliers (observations >3 or <-3) within n observations for parametric methods (column N(0,1)) and for SNS using various window sizes w , is significant at $alpha = 0.05$ using Equation (4), for different probabilities p . Reject the hypothesis of randomness at the 5% level if there are k or more outliers within the number of consecutive trading days given in **Table 1**.

This points out one advantage of using Sequential Normal Scores: the outlier probability is known exactly, and is distribution-free, for each choice of window size w , while the exact outlier probability is not known for the other methods because their distribution is not known exactly, just assumed to be approximately standard normal.

For example, according to **Table 1** the occurrence of 5 or more outliers within 507 consecutive trading days (approximately 2 years) is statistically significant at the 5% level for independent standard normal random variables. The p -value for an observed cluster of k outliers within n trading days is given by Equation (4). The value $n = 507$ changes in the SNS case, depending on the exact probability of an outlier for the corresponding window size.

Table 1. Length of a cluster of k or more observations to be significant at the 0.05 level.

Window	N (0,1)	$w = 500$	$w = 750$	$w = 1000$	$w = 1250$	$w = 1500$	$w = 1750$	$w = 2000$
$p =$	0.0027	2/500	2/750	2/1000	4/1250	4/1500	4/1750	6/2000
$k = 2$	19	13	20	26	17	20	23	18
$k = 3$	132	90	134	179	112	134	156	119
$k = 4$	304	206	308	410	257	308	359	274
$k = 5$	507	343	514	684	428	514	599	457
$k = 6$	969	494	740	987	617	740	863	658

The occurrence of a significant cluster indicates an increased level of variability in daily yields, or a sharp upward or downward trend in the mean yield, or a combination of both. In short, it indicates a marked instability in the sequence of yields relative to the historical values.

The Poisson distribution can be also used to find the probability of k or more out-of-control days, as an approximation to the binomial distribution if the probability of success is small and the trials are either independent or “weakly dependent” [17]. Let X and Y be defined as in the Binomial test, then:

$$\begin{aligned} &P(X \text{ is the first day in a cluster of at least } k \text{ out-of-control days} \\ &\text{within a time period of } n \text{ days}) \\ &= P(Y \geq k - 1 | Y \text{ is Poisson with } \lambda = 0.0027(n - 1)) \\ &= 1 - P(Y \leq k - 2 | Y \text{ is Poisson with } \lambda = 0.0027(n - 1)) \\ &= 1 - \text{POISSON.DIST}(k - 2, 0.0027(n - 1), \text{TRUE}) \end{aligned}$$

using Excel’s POISSON.DIST function that produces cumulative Poisson probabilities. A comparison of the Poisson probabilities with the binomial probabilities leading to **Table 1** shows no appreciable difference.

3. Example Using Host Hotel and Resorts, Inc. Data

Reference [18] uses real data from the closing stock price of shares of Host Hotel and Resorts, Inc. (HST) found on the website <https://www.hosthotels.com/>. It uses data from January 3, 2000 to March 30, 2007 to detect change-points in the daily returns. We are using data from January 3, 2000 to December 29, 2017 to bring the example closer to current times. The data have a definite quarterly effect, and steady growth. It is a good typical stock to study. The additional years we added to the study cover the real-estate “bust” of 2007, the recession years, and the recovery years until recently.

We find this stock especially interesting because [18] states that the daily returns follow a normal distribution, and uses that model in its analysis. It “proves” this assumption with a Q-Q plot that shows clearly that the bulk of the daily returns is indeed Gaussian, but there is a definite departure in the tails, which it ignores. It is the tails that interest us. The tails prove to be “fatter” than the Gaussian model would suggest and are likely to affect our methods for detecting outliers. This stock is also interesting because of the wide range of ups and downs of the stock over the past 18 years.

This example looks at 18 years of data on the price per share of Host Hotel and Resorts, obtained from the website <https://finance.yahoo.com/>. According to [18], HST is “the largest American lodging and real estate investment trust.” It assumes a normal distribution for the daily change in price. Indeed, the bar chart in **Figure 1** for the first year (250 trading days, representing year 2000) of yields could easily be mistaken for being approximately normal, with the exception of the two obvious outliers.

The entire series of 18 years of closing prices shows a highly autocorrelated series of observations, and is given in **Figure 2**. The time series of closing prices in

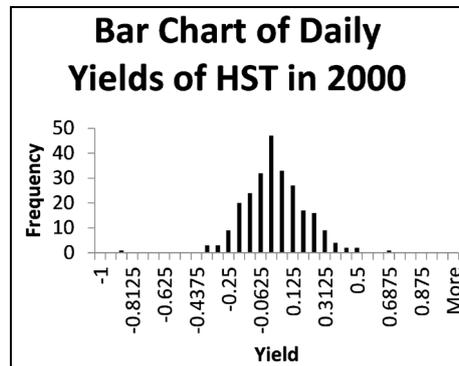


Figure 1. Bar chart of daily change in the first year of HST data examined.



Figure 2. Graph of daily stock prices for 18 years for HST data.

Figure 2 clearly shows the dependence from one market day to the next, with a serial correlation (lag 1) over the 18 years of 0.998. However, the change in price from one market day to the next is a series of essentially independent observations with a serial correlation of only -0.075 . Also the daily change is also almost uncorrelated with the closing price in **Figure 2** (0.031). The z -scores may change the dependence in the sequence. In this data set the z -scores, using a 500-day window, have a serial correlation of -0.015 , much smaller than that for the original sequence of daily changes.

Sequential Normal Scores are independent if they are calculated on independent observations (as are sequential ranks), and since they are calculated on the daily change they may be regarded as essentially independent. The serial correlation of Sequential Normal Scores (using a 500-day window) for this sequence is -0.021 , between the serial correlation of the original data (-0.075) and the z -scores on the daily changes (-0.015). Note that these serial correlations are small, almost negligible, but statistically significantly different from zero because of the length of the series (4528 observations).

It is also interesting to note that the serial correlation of the absolute values of the daily yields is 0.249 , large enough to account for the phenomenon called volatility clustering, and large enough to contain possibly useful information for the prediction of subsequent observations. Using a 500-day window on both series,

the z -scores have a serial correlation of absolute values equal to 0.141 and the serial correlation of the SNS is 0.137, almost the same. Truly independent observations in a sequence this long will have a serial correlation much closer to zero, as will any function of those observations such as absolute values or squared values.

The cross correlation of the Sequential Normal Scores with the z -scores of the raw data, both relative to the previous 500 trading day window, is 0.966, almost perfect correlation. Thus the Sequential Normal Scores are an almost perfect replication of the z -scores of the original data with regard to cross correlation as well as serial correlation of the observations, and serial correlation of the absolute values of the observations, for the 18 years from 2000 to 2017, and have the advantage of a known distribution, the standard normal distribution, so the probability of an outlier can be measured more accurately. Sequential Normal Scores convey basically the same information as the daily profits, but have the distinct advantage of being approximately standard normal, in contrast to the daily profits, which have an unknown distributional form. Thus the Sequential Normal Scores allow probability statements to be made regarding the size of the original observations.

Analysis of the logarithms of the stock prices results in more dependence. The serial correlation of the daily changes in the logarithms for this sequence is -0.127, a substantial correlation, and the serial correlation of the absolute values of the changes in logarithms is 0.420, an even more substantial correlation. For this reason we will not attempt to convert changes in logarithms to Sequential Normal Scores.

A histogram of Sequential Normal Scores for the first 250 trading days is given in **Figure 3**, for comparison with the bar graph of the raw data in **Figure 1**. Note the general shape of a standard normal density, but with randomness due to the data. A histogram of Normal Scores from ordinary ranks of these same observations would follow a rigid bell-shaped curve with no randomness. The correlation coefficient of these 250 Sequential Normal Scores in **Figure 3** with the 250 values of daily profit on which they were based in **Figure 1** is 0.964, a very close correlation, even though the two figures (**Figure 1** and **Figure 3**) might suggest otherwise.

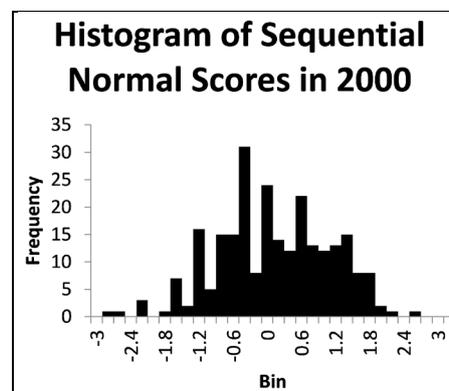


Figure 3. Histogram of the first year's sequential normal scores of the HST data.

A self-starting Shewhart graph may be obtained from the daily profits, starting as early as the third trading day, by subtracting the sample mean of the previous observations and dividing by the sample standard deviation of the previous observations. This results in a series of z -scores with approximate mean of zero, and approximate standard deviation of 1.0. When the z -scores are compared with the Shewhart bounds of ± 3 it is obvious that the tails of the distribution are much fatter than the tails of the standard normal distribution. A moving window of 500 observations is used in this example, so the first 500 observations are not counted in the interest of fairness, even though they could be used. Thus the data represent the years 2002-2017.

An examination of the z -score values shows that 31 of the 4025 z -scores are less than -3 , for 0.77%, much greater than the theoretical 0.135% for the standard normal distribution. The same holds true to a lesser extent for the upper tail, where 16 z -scores exceed $+3$ for 0.40%, again much larger than the target 0.135%. This indicates that even after converting to z -scores, the number of identified “outliers” using z -scores is 47, 1.17% of the total, more than four times the 10.9 (0.27% of the total) that one would expect from a standard normal random variable. It would be misleading to declare all 47 observations to be outliers.

An easy way to determine the statistical significance of the daily variations in profit is obtained by converting the daily profits to Sequential Normal Scores. Now it is easy to spot 3-sigma deviations in a Shewhart graph of Sequential Normal Scores. There are 11 Sequential Normal Scores less than -3 and 8 Sequential Normal Scores greater than $+3$, for a total of 19 out of 4025 days, or 0.47%, much closer to the target 0.27% for the two tails combined. In fact, the actual probability can be computed exactly for a 500-day window of SNS, because of their distribution-free nature, to be $2/500 = 0.40\%$, in good agreement with the observed value of 0.47%. The observed difference in percentages for the z -scores may be attributed to the fact that the original data are not well approximated by a normal distribution in the tails. This discrepancy is corrected almost entirely by using Sequential Normal Scores.

For purposes of discussion, let's look at the significant values of SNS using a 2-year window, listed as “YES” in the “500 trading day window” column in **Table 2**. The outliers (“YES”) in boldface mark a statistically significant cluster of outliers has occurred. The symbol * means that during startup the window was shorter than designated. Starting from day 1 in the series (instead of day 501 as we did above) there are a total of 21 “outliers” detected (values greater than 3 in absolute value), as compared to the expected number of 12.2 using the standard normal probability of 0.0027 for the total of 4525 observations. But here the advantage of being completely distribution-free can be utilized to compute the exact probability of getting a Sequential Normal Score above 3 (in absolute value) for a window of 500 observations. The sequential ranks are equally likely to be any integer from 1 to 500 (barring ties), and only the integers (ranks) 1 and 500 result in Sequential Normal Scores outside the $(-3, +3)$ interval, for an exact probability of

Table 2. Dates of statistically significant outliers in HST daily yields.

Trade Day	Date	DECLARATION OF "OUTLIER" BY WINDOW SIZE: SEQUENTIAL NORMAL SCORES (* = short window)								
Start	1/3/2000	CLOSE	500	750	1000	1250	1500	1750	2000	Direction
426	9/17/2001	\$8.86	YES*	NEGATIVE						
482	12/5/2001	\$9.13	YES*	POSITIVE						
645	7/31/2002	\$11.25	YES	YES*	YES*	YES*	YES*	YES*	YES*	POSITIVE
1258	1/5/2005	\$16.27	YES	NO	NO	NO	NO*	NO*	NO*	NEGATIVE
1475	11/14/2005	\$16.65	YES	YES	NO	NO	NO*	NO*	NO*	NEGATIVE
1621	6/15/2006	\$20.86	YES	YES	NO	NO	NO	NO*	NO*	POSITIVE
1650	7/27/2006	\$21.53	YES	YES	YES	NO	NO	NO*	NO*	NEGATIVE
1703	10/11/2006	\$22.52	YES	YES	YES	YES	YES	YES*	YES*	NEGATIVE
1731	11/20/2006	\$24.94	YES	YES	YES	YES	YES	YES*	YES*	POSITIVE
1783	2/7/2007	\$28.71	NO	NO	NO	YES	YES	YES	YES*	POSITIVE
1787	2/13/2007	\$28.71	NO	NO	NO	NO	YES	YES	YES*	POSITIVE
1885	7/5/2007	\$26.01	YES	YES	YES	YES	YES	YES	YES*	POSITIVE
1911	8/10/2007	\$22.35	NO	NO	NO	YES	NO	NO	YES*	NEGATIVE
1996	12/11/2007	\$18.56	NO	NO	NO	YES	YES	NO	YES*	NEGATIVE
2024	1/23/2008	\$17.41	NO	NO	NO	YES	YES	YES	YES	POSITIVE
2071	4/1/2008	\$17.01	NO	NO	NO	NO	NO	NO	YES	POSITIVE
2189	9/17/2008	\$13.56	NO	NO	NO	YES	YES	YES	YES	NEGATIVE
2190	9/18/2008	\$17.00	YES	YES	YES	YES	YES	YES	YES	POSITIVE
2192	9/22/2008	\$14.48	YES	NEGATIVE						
2197	9/29/2008	\$12.39	YES	NEGATIVE						
2205	10/9/2008	\$7.85	YES	NEGATIVE						
2206	10/10/2008	\$9.28	NO	NO	NO	NO	NO	NO	YES	POSITIVE
2218	10/28/2008	\$8.98	NO	NO	NO	NO	NO	NO	YES	POSITIVE
2241	12/1/2008	\$5.92	NO	NO	NO	YES	YES	YES	YES	NEGATIVE
2917	8/8/2011	\$11.86	YES	NO	NO	YES	YES	YES	YES	NEGATIVE
3666	7/31/2014	\$21.74	YES	NO	NO	NO	NO	NO	NO	NEGATIVE
3805	2/19/2015	\$21.87	YES	NEGATIVE						
3936	8/26/2015	\$17.71	YES	YES	NO	NO	NO	NO	NO	POSITIVE
4148	6/29/2016	\$16.05	YES	YES	YES	NO	NO	NO	NO	POSITIVE
4244	11/14/2016	\$17.38	YES	YES	YES	YES	NO	NO	NO	POSITIVE
4357	4/28/2017	\$17.95	YES	NO	NO	YES	NO	NO	YES	NEGATIVE
Total			21	17	14	20	18	17	23	

$2/500 = 0.004$, slightly above the target of 0.0027. So the actual expected number of outliers in SNS is 18.1, in good agreement with the observed number of 21.

Continuing on in **Table 2** down the column for the 500 day window, in the first six and a half years (January 2000 through September 2006) there are exactly seven 3-sigma shifts; the first is a drop caused by the terrorist attacks on September 11 2001, and six others. Over a period of about 1700 trading days (the first six and a half years) an event expected to occur once every 370 days (or 250 days, using the exact probability) occurred only seven times. That is reasonable, in fact expected. None of these are clustered close enough to be statistically significant.

Then the next significant outlier occurred on October 11, 2006, trading day number 1703 in our series. This is only 82 trading days from the outlier on June 15, 2006, and there was another outlier between the two, on July 27, 2006. That means three outliers occurred in 83 trading days, which is statistically significant at the 5% level according to the cluster test and **Table 1**, to mark the first significant cluster of outliers in almost seven years. If that was not enough to ring an alarm bell, 28 trading days later another outlier occurred, on November 20, 2006. That means four outliers occurred within 111 trading days, which has a p-value of 0.01 according to Equation (4). Two of the four were positive outliers and two were negative, suggesting a change in variation rather than a shift in the mean. Maybe it is a coincidence that the real estate bubble burst three months later in February 2007, sending stock prices into a steep slide that lasted several years, and maybe not. We have to leave that conclusion up to the experts in the field.

Five more outliers (two positive and three negative) occurred in the next two years (2007 and 2008) maintaining a “code red” alert, culminating with four outliers in the month from September 18, 2008 to October 9, 2008. This period of two unstable years coincides with the decline in stock price from about \$28 to about \$3. Then there are no more outliers for almost three years, while the stock price turned around and started a two year climb, peaking again in February 2011 at about \$20. During this climb, the “red alert” sign came down on trading day 2397, which is July 16, 2009, because of a lack of outliers. Then there are no significant clusters for the remainder of our series. Note that for longer windows, 5 years to 8 years, another significant cluster occurred on December 1, 2008.

Six outliers over a span of almost three years, as occurs near the end of the series for the 500-day window, would be statistically significant for standard normal random variables according to **Table 1**. This may not be of interest from a practical standpoint because of the long time span, even though it is statistically significant. We are arbitrarily assuming in this paper that clusters less than one year (250 trading days) in duration from beginning to end are of more interest to financial analysts than clusters that span longer periods of time, so we are marking in boldface in **Table 2** only outliers that mark a significant cluster of total length less than 250 trading days (approximately one year) in length. Then it becomes apparent that for all window sizes of length 2, 3, or 4 years the outliers mark statistically significant clusters on November 20, 2006. Even though the windows of

longer length (5, 6, 7, and 8 years) had outliers on November 20, 2006, they did not mark statistically significant clusters until their next outlier which occurred on February 7, 2007, shortly before the collapse of the real estate market that month. All seven windows signaled statistically significant clusters of outliers shortly before the two year collapse in the market for this stock. This agreement in identifying outliers across the choice of window size is apparent in **Table 2**, as well as the agreement in declaring significant clusters of outliers. For this reason the SNS approach to identifying outliers is recommended.

Table 3 emphasizes the year-by-year comparison of significant SNS and *z*-scores (using a 500-day window), showing that years with many significant outliers, especially negative outliers, coincide with steep declines in the price of HST stock values. A day-by-day comparison of declared outliers in **Table 4** shows that of the 21 days identified as outliers using SNS, 20 are also identified as outliers by the *z*-scores, for excellent agreement.

Table 3. Year by year comparisons of numbers of high and low SNS outliers with high and low *z*-scores, for HST data.

YEAR	LOW SNS	HIGH SNS	LOW <i>z</i> -score	HIGH <i>z</i> -score	Starting Price	Change in year
2000	0	0	1	1	8.4375	4.375
2001	1	1	6	1	12.8125	-3.6125
2002	0	1	1	1	9.2	-0.3
2003	0	0	0	0	8.9	3.35
2004	0	0	0	1	12.25	4.75
2005	2	0	3	4	17	2.18
2006	2	2	3	2	19.18	5.38
2007	0	1	5	3	24.56	-7.28
2008	3	1	4	3	17.28	-9.55
2009	0	0	0	0	7.73	4.09
2010	0	0	0	0	11.82	6.53
2011	1	0	4	0	18.35	-3.48
2012	0	0	0	0	14.87	1.27
2013	0	0	0	0	16.14	3.12
2014	1	0	1	0	19.26	4.55
2015	1	1	6	0	23.81	-8.65
2016	0	2	3	2	15.16	3.97
2017	1	0	1	0	19.13	0.72
TOTAL	12	9	38	18	19.85	(12/29/2017)

The point here is that by using Sequential Normal Scores there is no need to continually estimate the mean and the variance, as usual in Shewhart charts, because the mean is zero and the standard deviation is one in a stable period. An unstable period with unusual variability is easy to detect, and the likelihood of false positives is greatly reduced.

A comparison of the SNS method with the z -score methods for all eight choices of window sizes is interesting as shown in **Table 5**. For a fair comparison, the first 2000 trading days are not considered, and the comparison begins only after all the various window sizes are complete.

The first difference is that the Shewhart Chart on the original raw data identifies more than twice as many outliers as the SNS method, about four to seven times as many as expected if the probability of an outlier is truly 0.27%. This suggests that the true probability of a z -score being declared an outlier is much greater than 0.27%, but unknown. The second difference is that with this many declared outliers the cluster test on z -scores is practically useless. Almost every declared outlier raises the alarm of a cluster occurring. This suggests that the Shewhart Chart on the raw data is declaring many false positives. Using the log transformation results in even more outliers declared which exacerbates the problem of false positives. These results are consistent for all eight choices of window length for z -scores on the raw data. We do not list all the z -score outliers in this paper because there are so many.

Table 4. A frequency table of the day by day comparisons of “Outliers” as declared by SNS with “Outliers” as declared by z -scores, for HST data 2000-2017.

	High z	Low z	Neither	TOTAL
High SNS	8	0	1	9
Low SNS	0	12	0	12
Neither	10	26	4468	4504
TOTAL	18	38	4469	4525

Table 5. The number of HST declared outliers from trading day 2001 to the end of the series.

Window size	250	500	750	1000	1250	1500	1750	2000
Raw data	20	24	26	28	40	43	46	46
(Expected count)	6.82	6.82	6.82	6.82	6.82	6.82	6.82	6.82
log(data)	38	58	76	87	89	91	82	78
(Expected count)	6.82	6.82	6.82	6.82	6.82	6.82	6.82	6.82
Seq. Normal Scores	0	11	8	7	11	9	9	13
(Expected count, exact)	0.00	10.11	6.74	5.05	8.09	6.74	5.78	7.58
z-scores on SNS	11	10	12	10	11	10	6	1
(Expected count)	6.82	6.82	6.82	6.82	6.82	6.82	6.82	6.82

Analysis of the raw data shows not only an excessive numbers of outliers, very much in excess of the expected number for a 3-sigma chart, but also that there is no consistency in results over the different window sizes, almost doubling the number of outliers as the window size goes from 500 trading days to 2000. The same is true for the log-transformed data. In contrast, the Sequential Normal Scores are consistent in their more modest results over all seven window sizes (not including the 250-day window because it cannot declare SNS outliers). Converting SNS to z -scores brings in a different level of inconsistency, where the number of outliers diminishes dramatically as the window size increases. The conclusion seems to be that for a reasonable number of false positives, and a consistent declaration of outliers, the Sequential Normal Scores method appears to be the best of the various methods examined.

4. CUSUM and EWMA with Sequential Normal Scores

Next we compare the CUSUM method on Sequential Normal Scores for detecting changes in the mean yield, with the CUSUM method on the z -scores of the raw data. We use the 500-day window for this comparison, but the results are very similar for all the other window sizes examined in this paper.

The preceding analysis looked at variability, and showed that Sequential Normal Scores are good for identifying outliers, without producing as many false positives as the original observations produce. Outliers may indicate unusual volatility, or they may indicate a large shift (positive or negative) in the mean. The CUSUM method and the EWMA method were invented to detect smaller, but consistent, changes in the mean yield. Sequential Normal Scores also allow for CUSUM and EWMA analyses, unlike ordinary rank methods where the previous ranks, and therefore the previous CUSUMs and EWMA, change with each new observation.

We performed a CUSUM analysis on the z -scores of the raw data (500-day moving window) to see if the CUSUMs detect a shift in the mean daily profit. For comparison we performed the same analysis on the Sequential Normal Scores (500-day moving window) to see if we can detect a shift in the mean daily profit. As stated earlier, we used $k = 0.5$ as the adjustment in Equations (1) and (2), and $h = \pm 4.774$ as the boundaries, which results in the probability of a Type I error close to the 0.0027 that results from 3-sigma limits in a Shewhart Chart for normally distributed data.

The correlation coefficient of the positive CUSUMs for the z -scores on the raw data with the positive CUSUMs of the Sequential Normal Scores is a very high 0.961, and the corresponding negative CUSUMs have a correlation of 0.945. This indicates that the CUSUMs on the Sequential Normal Scores convey much of the same information that the CUSUMs on the z -scores of the raw data conveys. The advantage of the nonparametric approach is that the boundary value is distribution-free, while the probability of exceeding the z -score boundaries is dependent on the underlying distribution.

Tables 6-8 present a comparison of the CUSUMs and EWMA for HST data in the years 2002-2017 (4027 days). Both series used 500-day windows. **Table 6** gives the yearly totals, and shows that for the years 2002-2017 the raw data *z*-scores resulted in 110 days with significant CUSUMs (25 positive, 85 negative) out of 4027 days total, for 2.7%, ten times the 0.27% expected for normal distributions. The SNS resulted in 73 days with significant CUSUM values (20 positive, 53 negative), for 1.8%, also much larger than, but much closer to, the target.

Table 6. Yearly comparisons of numbers of significant values of CUSUMs and EWMA.

YEAR	CUSUMs				EXP. WT.MOVING AVE.			
	SNS		z-SCORE		SNS		z-SCORE	
	HIGH	LOW	HIGH	LOW	HIGH	LOW	HIGH	LOW
2002	2	0	1	1	3	2	1	0
2003	1	0	0	0	1	0	0	0
2004	0	0	0	0	0	0	0	0
2005	0	1	0	2	0	0	0	1
2006	0	1	2	4	1	1	3	1
2007	4	21	9	25	4	6	5	8
2008	0	16	1	17	0	3	1	2
2009	0	0	0	0	0	0	0	0
2010	0	0	0	0	0	0	0	0
2011	0	11	0	18	0	3	0	6
2012	0	0	0	0	0	0	0	0
2013	0	0	0	0	0	0	0	0
2014	9	0	9	0	1	0	1	2
2015	3	3	2	17	2	2	1	7
2016	1	0	1	1	1	0	1	2
2017	0	0	0	0	0	0	0	0
TOTAL	20	53	25	85	13	17	13	29
EXPECTED	5.44	5.44	5.44	5.44	5.44	5.44	5.44	5.44

Table 7. Daily agreements and disagreements with CUSUMs for SNS and *z*-scores.

	Low z	High z	Neither	TOTAL
Low SNS	51	0	2	53
High SNS	0	17	3	20
Neither	34	8	3913	3955
TOTAL	85	25	3918	4028

Table 8. Daily agreements and disagreements with EWMA for SNS and z-scores.

	Low z	High z	Neither	TOTAL
Low SNS	13	0	4	17
High SNS	0	8	5	13
Neither	16	5	3977	3998
TOTAL	29	13	3986	4028

Table 7 is a frequency table of the number of days where there is agreement between the CUSUMs on the z-scores and the CUSUMs on the SNS in declaring outliers for the HST data over the 18 years examined. It shows that of those 73 days with significant CUSUMs on the SNS, 68 were the same days identified by using the z-scores on the raw data (17 of the 20 “high” and 51 of the 53 “low”), for a pretty good agreement. Most of the excess in significant CUSUMs for the z-scores occurs immediately following an unusually large (in absolute value) z-score, a “carryover effect” not shown by the SNS analysis.

Table 8 shows that the day-by-day agreement for EWMA is also very good. Of the 30 outliers identified by EWMA on SNS, 21 are the same days identified by EWMA on the raw data z-scores. The correlation of the EWMA on the z-scores and the EWMA on the SNS is a near-perfect 0.980.

A CUSUM or EWMA chart built specifically for detecting scale changes might also be found useful in practice, and the results obtained in this research can be used analogously.

5. Example Using the S & P 500 Index

The previous example examined the behavior and analysis of a single company’s stock, known as HST. This example looks at a portfolio of 500 stocks, the S & P 500, that is one basis for representing the entire market with a single number. Shares in the S & P 500 cannot be bought or sold, but many companies have mutual funds trying to mimic the makeup of the S & P 500 which makes this an important example. The S & P 500 is a weighted average of the price per share of 500 different companies that represent in some sense the entire economy of publicly traded companies. The S & P 500 seems to be a logical place to start, to see if Sequential Normal Scores can be useful in analyzing a portfolio of stocks.

To demonstrate the versatility of Sequential Normal Scores we will analyze the daily percent change of the closing price of the S & P 500, rather than the actual daily change in price which we analyzed with the HST data in the previous example. Otherwise our approach is the same as described in the HST example. The historical closing price of the S & P 500 is obtained from the website <https://finance.yahoo.com/> for 21 years from January 2, 1997 to December 29, 2017, amounting to 5285 trading days. The closing price is highly dependent on the previous day’s closing price, but the change in closing price (the yield) appears to behave as a series of independent random variables. Rather than analyze

the actual daily change we will divide the daily change by the previous day's closing price to get a percentage change each day. This will allow a fairer comparison as prices go up and down.

The first order serial correlation of the daily yields for S & P 500 data over the 21 years is a low -0.058 , but the serial correlation of the absolute values of these same data is a more substantial 0.172 , illustrating the principle of “volatility clustering” described earlier. On the other hand the percentage change of the closing prices has a slightly greater (in absolute value) serial correlation, -0.070 , as do the absolute values of the percentage change, 0.237 , compared with the raw data. The z -scores on the percentage change in the raw data, using a 500-day window, have a serial correlation of -0.062 , and the serial correlation of the absolute values is 0.189 , both close to, but between, the corresponding numbers for the actual daily changes and the percentage changes.

The Sequential Normal Scores calculated on the percentage change (500-day window) are highly correlated (0.971) with the z -scores on the percentage change in raw data (500-day window), indicating that conclusions based on the analysis of the SNS may be carried over to the raw data itself. The sequence of Sequential Normal Scores using a 500-day window has a serial correlation of -0.052 and the absolute values of the SNS have a serial correlation of 0.149 , both in keeping with the percentage changes in the original data (-0.070 and 0.237), and the z -scores using a 500-day window (-0.062 and 0.189), but slightly smaller in absolute value.

The histogram of the percentage yield of the first 250 trading days (representing the year 1997) is given in **Figure 4**. Besides a slight skewness, there are two obvious extreme outliers that challenge the assumption of normality. Nevertheless we will proceed with the analysis as we did in the previous example, comparing the results of assuming normality in the measurements with the nonparametric approach that uses Sequential Normal Scores.

The closing value of the S & P 500 over the 21 years from January 3, 1997 to December 29, 2017 is the basis for **Figure 5**. **Figure 5** shows the range of values from a low of about 750 to a high of over 2600, suggesting that looking at percentage change might result in a more reasonable analysis than looking at absolute change.

One quality of interest is identifying the percentage daily yield as an “outlier” or not, as in the previous example. We will declare a day's percentage return as an outlier if its z -score or SNS falls outside the 3-sigma limits for a standard normal random variable. As before, this will correspond to converting percentage yields to z -scores, by subtracting an estimate of the mean and dividing by an estimate of the standard deviation (Method 1), by converting the logarithm of the closing price to percentage daily log yields, and then converting to z -scores (Method 2), by converting the daily percentage yield to Sequential Normal Scores (Method 3) and finally by converting the SNS to z -scores (Method 4). Eight different windows of history will be used for each, ranging from one year

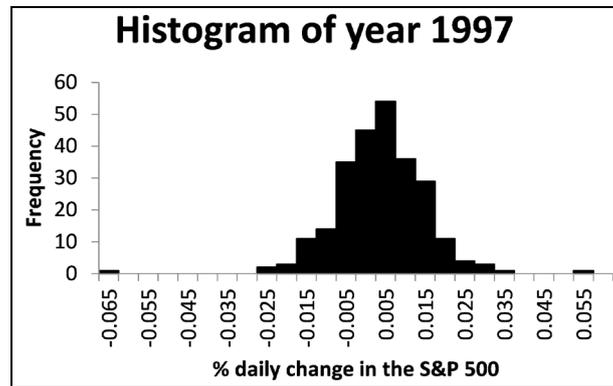


Figure 4. Histogram of the daily percent changes of S & P 500 representing the year 1997.



Figure 5. The closing value of the S & P 500, 1997-2017.

to eight years, as with the previous example, for 32 different methods of identifying outliers.

A histogram of the Sequential Normal Scores for the first 250 percentage daily yields representing year 1997 of the S & P 500 is given in **Figure 6**. It is seen to be approximately the shape of a standard normal density function, with variations due to randomness of the data. The outliers obvious in **Figure 4** become part of the curve because of the nature of ranks pulling in the outliers.

To examine the longer term distributions, the z -scores on the percentage yields, using a 500-day window, are compared with the Shewhart bounds for the years 1999 to 2017. For the 4784 trading days examined after the first 500 observations, representing years 1999-2017, the expected number of observations exceeding +3 is 6.46, while the observed number of z -scores exceeding +3 is 35. The same is true for the lower tail, where the expected count is again 6.46 but the observed count is 46. Thus the tails are much “fatter” than expected.

By comparison the exact exceedance probability for SNS using a 500-day window is $1/500 = 0.002$ in each tail, compared with the standard normal value of 0.00135. The actual number of exceedances for the SNS with a 500-day window is 14 above +3, and 10 below -3, in closer agreement with the exact

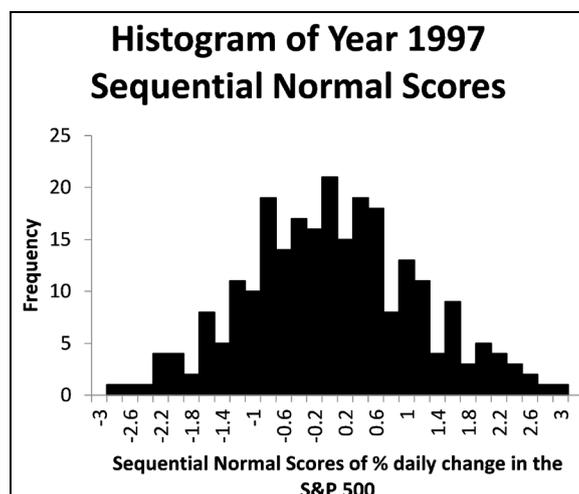


Figure 6. Histogram of the SNS of the S & P 500, representing Year 1997.

expected counts of 9.56 in each tail. As for agreement with the z -scores, 23 of the 24 significant Sequential Normal Scores are also significant z -scores on the raw data (all of the 14 above +3, and 9 of the 10 below -3) showing good agreement of the two methods.

The total number of outliers resulting from each of the four methods, under the eight different window sizes, over 3284 trading days from trading day 2001 (December 15, 2004) to trading day 5284 (December 29, 2017) for 13 years of S & P index daily percent changes, which includes all but the first 2000 trading days (8 years), for a fair comparison, is given in **Table 9**. It is seen again, as in the HST example, that the number of outliers identified is quite large for the raw data and for the logs of the data, at times seven times the expected value of 8.9 days, suggesting the occurrence of many false positives. The Sequential Normal Scores have the lowest number of outliers identified, but still slightly more than the expected number. The z -scores on the SNS result in slightly larger numbers of observed values, but not as consistent over the eight different windows as we would like.

Large numbers of outliers suggest underestimation of the probability of an outlier. The number of outliers identified by the SNS method is consistent over all window sizes, and closer to the target values, so this method appears to be the most satisfactory method, of the ones examined, for identifying outliers.

Table 10 shows the dates of the outliers identified by the Sequential Normal Scores method for S & P-500 percentage yields over the various window sizes, like **Table 2** showed for HST data. These dates are determined by 3-sigma limits on a Shewhart Chart of Sequential Normal Scores (5284 trading days from January 2, 1997 to December 29, 2017). The outliers (“**YES**”) in boldface indicate a statistically significant cluster of outliers has occurred. The asterisk (*) indicates outliers that occurred during startup while the window was shorter than designated. **Table 10** shows good agreement among the seven (ignoring the 250 day window, which cannot produce an outlier) different windows for determining

Table 9. A comparison of the number of outliers identified by the 32 different methods.

Window size	250	500	750	1000	1250	1500	1750	2000
Raw z-scores	56	70	74	72	68	60	50	51
Target %	0.0027	0.0027	0.0027	0.0027	0.0027	0.0027	0.0027	0.0027
Actual %	0.0170	0.0213	0.0225	0.0219	0.0207	0.0183	0.0152	0.0155
log z-scores	57	72	73	74	67	57	52	53
Target %	0.0027	0.0027	0.0027	0.0027	0.0027	0.0027	0.0027	0.0027
Actual %	0.0174	0.0219	0.0222	0.0225	0.0204	0.0174	0.0158	0.0161
Seq. Nor. Scores	0	22	16	13	14	11	10	12
Target %	0.0000	0.0040	0.0027	0.0020	0.0032	0.0027	0.0023	0.0030
Actual %	0.0000	0.0067	0.0049	0.0040	0.0043	0.0033	0.0030	0.0037
z-scores on SNS	18	26	33	40	44	30	21	17
Target %	0.0027	0.0027	0.0027	0.0027	0.0027	0.0027	0.0027	0.0027
Actual %	0.0055	0.0079	0.0100	0.0122	0.0134	0.0091	0.0064	0.0052

outliers using Sequential Normal Scores on the 21 years of S & P 500 data. It also shows that the summary of detected outliers reduces to only 32 different trading days out of the 5284 days in the sequence. For example, using a window of size 500 results in 24 significant outliers being detected. The total of the other window sizes includes only 8 additional days that are considered outliers.

Many of these outliers trigger alarms using the cluster test. Those outliers that trigger cluster alarms are given in boldface type. For example, the first outlier in the table occurred on trading day 1010 (January 3, 2001), but it is not in boldface in **Table 10** because it does not trigger the occurrence of a significant cluster of outliers. The next two outliers occur 18 months later in July 2002, within five days of each other, triggering a cluster alarm for the longer windows (5 - 8 years) using **Table 1**. The next significant outlier cluster is signaled on June 29, 2006, for the 500 day window, and the S & P index began a 20% climb in the next 15 months. Then two more outliers in August 2007 announce a significant outlier cluster in the shorter windows (2 - 4 years). Four months later the S & P 500 dove from over 1500 to about 700 in 16 months. During this period there are many significant outlier clusters identified in all seven windows. In March 2009 the significant clustering ended, and the S & P 500 turned around and started a six year recovery during which no new outlier clusters are identified. Maybe these are coincidences, but they are interesting to observe.

A recent paper [19] used a non-parametric test for variance to study the S & P 500 based on monthly averages from July 2004 to June 2009. It found (using a retrospective analysis) a significant change in variance occurred in February 2006, which is close to the finding in our study of a significant cluster of outliers in June 2006, without using a retrospective analysis.

Table 10. Dates of statistically significant outliers (“YES”) in S & P 500 daily percent changes.

Trade		Window		Size							Direction
Day	Date	Close	500	750	1000	1250	1500	1750	2000		
1010	1/3/01	1347.56	YES	NO	NO	NO*	NO*	NO*	NO*	POSITIVE	
1397	7/24/02	843.43	YES	YES	YES	YES	YES*	YES*	YES*	POSITIVE	
1400	7/29/02	898.96	NO	NO	NO	YES	YES*	YES*	YES*	POSITIVE	
2277	1/20/06	1261.49	YES	NO	NO	NO	NO	NO	NO	NEGATIVE	
2378	6/15/06	1256.16	YES	NO	NO	NO	NO	NO	NO	POSITIVE	
2388	6/29/06	1272.87	YES	NO	NO	NO	NO	NO	NO	POSITIVE	
2553	2/27/07	1399.04	YES	YES	NO	NO	NO	NO	NO	NEGATIVE	
2664	8/6/07	1467.67	YES	YES	YES	NO	NO	NO	NO	POSITIVE	
2673	8/17/07	1445.94	YES	YES	YES	NO	NO	NO	NO	POSITIVE	
2694	9/18/07	1519.78	YES	YES	YES	NO	NO	NO	NO	POSITIVE	
2814	3/11/08	1320.65	YES	YES	YES	YES	NO	NO	NO	POSITIVE	
2819	3/18/08	1330.74	YES	YES	YES	YES	NO	NO	NO	POSITIVE	
2940	9/9/08	1224.51	NO	NO	NO	YES	NO	NO	NO	NEGATIVE	
2944	9/15/08	1192.70	YES	YES	YES	YES	YES	YES	YES	NEGATIVE	
2946	9/17/08	1156.39	YES	NEGATIVE							
2947	9/18/08	1206.51	YES	YES	YES	YES	YES	NO	NO	POSITIVE	
2954	9/29/08	1106.42	YES	NEGATIVE							
2955	9/30/08	1166.36	YES	POSITIVE							
2960	10/7/08	996.23	NO	NO	NO	YES	YES	YES	YES	NEGATIVE	
2962	10/9/08	909.92	NO	NO	NO	YES	YES	YES	YES	NEGATIVE	
2964	10/13/08	1003.35	YES	POSITIVE							
2966	10/15/08	907.84	YES	NEGATIVE							
2975	10/28/08	940.51	NO	NO	NO	YES	YES	YES	YES	POSITIVE	
2987	11/13/08	911.29	NO	NO	NO	NO	NO	NO	YES	POSITIVE	
2998	12/1/08	816.21	NO	NO	NO	YES	YES	YES	YES	NEGATIVE	
3074	3/23/09	822.92	NO	NO	NO	NO	NO	NO	YES	POSITIVE	
3672	8/4/11	1200.07	YES	NO	NO	NO	NO	NO	NO	NEGATIVE	
3674	8/8/11	1119.46	YES	NO	NO	NO	NO	NO	NO	NEGATIVE	
3675	8/9/11	1172.53	YES	NO	NO	NO	NO	NO	NO	POSITIVE	
4690	8/21/15	1970.89	YES	YES	NO	NO	NO	NO	NO	NEGATIVE	
4691	8/24/15	1893.21	YES	YES	YES	NO	NO	NO	NO	NEGATIVE	
4693	8/26/15	1940.51	YES	YES	NO	NO	NO	NO	NO	POSITIVE	
TOTAL			24	17	14	16	13	12	14		
Expected			21.1	14.1	10.6	16.9	14.1	12.1	15.9		

Table 11 shows the agreement between the SNS outliers and the z-score outliers on a year by year basis, where both are using a 500 day window. It is interesting to see the agreement between the two methods, and the fact that the most outliers occur during the worst-performing S & P 500 years.

Table 12 details the exact days of the outliers, comparing the SNS with the z-scores again, both with 500-day windows, for 5284 trading days of the S & P 500 from January 2, 1997 to December 29, 2017. Note that 23 of the 24 SNS “outliers” are also declared “outliers” by the z-scores, for excellent agreement. The results for the other window lengths are consistent with these results.

Table 11. Year-by-year comparisons of high and low SNS outliers with high and low z-scores, and current value of the S & P-500 index, with the yearly change.

Year	Low SNS	High SNS	Low z-score	High z-score	Starting Value	Change in year
1997	0	0	2	2	737.01	238.03
1998	0	0	3	3	975.04	253.06
1999	0	0	0	0	1228.1	227.12
2000	0	0	2	2	1455.22	-171.95
2001	0	1	2	2	1283.27	-128.6
2002	0	1	0	3	1154.67	-245.64
2003	0	0	0	0	909.03	199.45
2004	0	0	0	0	1108.48	93.6
2005	0	0	0	0	1202.08	66.72
2006	1	2	0	2	1268.8	147.8
2007	1	3	11	5	1416.6	30.56
2008	4	5	18	12	1447.16	-515.36
2009	0	0	0	2	931.8	201.19
2010	0	0	0	0	1132.99	138.88
2011	2	1	4	3	1271.87	5.19
2012	0	0	0	0	1277.06	185.36
2013	0	0	0	0	1462.42	369.56
2014	0	0	2	1	1831.98	226.22
2015	2	1	6	3	2058.2	-45.54
2016	0	0	1	0	2012.66	245.17
2017	0	0	0	0	2257.83	415.78
TOTAL	10	14	51	40	2673.61	(12/29/2017)

Table 12. Daily comparisons of “Outliers” as detected by SNS and z-scores.

	High z	Low z	Neither	TOTAL
High SNS	14	0	0	14
Low SNS	0	9	1	10
Neither	26	42	5192	5260
TOTAL	40	51	5193	5284

6. CUSUM and EWMA Analyses on the S & P 500 Data

Because the original daily percentage changes are considered to be independent observations, the Sequential Normal Scores are also considered to be independent observations, and both sets of observations lend themselves to analysis using CUSUMs and EWMA. **Tables 13-15** compare the CUSUM and EWMA computations on the *z*-scores using a 500-day window, with the SNS using a 500-day window, starting with trading day 501 (December 28, 1998) and extending through December 29, 2017. Similar results occur with the other window sizes but are not reported here.

Table 13. Yearly totals of significant CUSUMs and EWMA.

YEAR	CUSUMs				EWMA			
	SNS		z-SCORES		SNS		z-SCORES	
	HIGH	LOW	HIGH	LOW	HIGH	LOW	HIGH	LOW
1999	0	0	0	0	0	0	0	0
2000	0	0	3	1	0	2	0	2
2001	0	2	1	2	0	2	0	4
2002	5	5	11	3	0	3	1	3
2003	0	0	0	0	0	0	0	0
2004	0	0	0	0	0	0	0	0
2005	0	0	0	0	0	0	0	0
2006	0	0	0	0	0	0	0	0
2007	0	0	1	13	0	1	0	9
2008	5	17	15	32	0	7	0	15
2009	0	0	1	0	0	0	0	0
2010	0	0	0	0	0	0	0	0
2011	0	4	1	10	0	2	0	4
2012	0	0	0	0	0	0	0	0
2013	0	0	0	0	0	0	0	0
2014	1	0	4	4	0	0	0	1
2015	1	5	3	7	0	3	0	3
2016	0	2	0	9	0	1	0	3
2017	0	0	0	0	0	0	0	0
TOTAL	12	35	40	81	0	21	1	44
EXPECTED	6.46							

Table 14. Daily agreements and disagreements with CUSUMs for SNS and z-scores.

	Low z	High z	Neither	TOTAL
Low SNS	33	2	0	35
High SNS	0	11	1	12
Neither	48	27	4658	4733
TOTAL	81	40	4659	4780

Table 15. Daily agreements and disagreements with EWMA for SNS and z -scores.

	Low z	High z	Neither	TOTAL
Low SNS	21	0	0	21
High SNS	0	0	0	0
Neither	23	1	4685	4759
TOTAL	44	1	4735	4780

Table 13 gives the yearly totals, where it is evident that years with high numbers of outliers using the SNS correspond to years with high numbers of outliers using z -scores, for both the CUSUMs and the EWMA. As shown in **Table 13** there are 121 statistically significant CUSUMs on the z -scores of the raw data, 40 positive, 81 negative. This is more than double the number (47) of statistically significant CUSUMs on the Sequential Normal Scores, which had 12 positive and 35 negative. Note that the years of clusters of significant CUSUMs agree with the periods of clusters of significant outliers, but are more sensitive to changes in means, rather than changes in variance.

Table 14 gives the daily agreement figures for CUSUMs. It is interesting to note that out of the 47 significant CUSUMs on the SNS, 44 of them were included as significant in the raw data, 11 of the 12 “high” and 33 of the 35 “low”. The excess of significant CUSUMs on the raw data can be traced back in almost every case to the occurrence of an unusually large CUSUM that carried over to a number of successive trading days that were not counted by the SNS, which minimizes the effect of an unusually large value by virtue of using ranks.

The correlation between the positive CUSUMs (z -scores vs. SNS) is 0.944, and the correlation between the negative CUSUMs is 0.939, indicating an almost perfect proxy using the SNS. These CUSUM values computed from Sequential Normal Scores can be used for further analysis, such as determining “change dates” and other interesting phenomena just as CUSUMs on the raw data are often used, because they mimic the CUSUMs on the raw data without being overly sensitive to unusually large observations spilling over to subsequent dates.

Table 15 is a frequency table that shows the day-by-day agreement and disagreement between the EWMA computed on the SNS versus the EWMA computed on the z -scores, using 500-day windows for both. All 21 significant EWMA values computed on the SNS are also significant EWMA values for the z -scores.

It is curious to note that there are no EWMA values above the threshold 0.953 for the SNS values, and only 1 for the z -scores, but this is consistent with a similar analysis on the actual daily differences instead of the percentage daily differences. The total of 21 significant EWMA values is closer to the expected count of 12.9 expected from a sequence of independent normal random variables, than is the 45 significant EWMA values observed on the z -scores. The correlation between the EWMA on the z -scores and the EWMA on the SNS is a near-perfect 0.972.

7. Conclusion

Sequential Normal Scores provide a fast and simple way of converting any time series of independent and identically distributed random variables into a stream of independent apparently standard normal random variables. Even for financial data such as yields of individual stocks or changes in the S & P-500 index where the data are not exactly independent, the Sequential Normal Scores are highly correlated with the original data, even mimicking the serial correlation of lag one and the serial correlation of absolute values. The identification of outliers using SNS agrees across seven different windows, and is in agreement with methods based on z -scores of the original data, but without as many false positives. Also the CUSUM and EWMA computations on SNS agree with similar computations on the z -scores of the raw data, without as many false positives. The Sequential Normal Scores provide a distribution-free method of analysis with known probabilities, while the original data have an unknown distribution, similar to a normal distribution, but with tail probabilities much different than normal probabilities. Therefore statistical methods that require exotic methods of analysis perhaps can be replaced by simpler methods based on the analysis of Sequential Normal Scores.

References

- [1] Parent Jr., E.A. (1965) Sequential Ranking Procedures. Doctoral Dissertation, Stanford University, Dept. of Statistics, Palo Alto.
- [2] McDonald, D. (1990) A Cusum Procedure Based on Sequential Ranks. *Naval Research Logistics*, **37**, 627-646.
[https://doi.org/10.1002/1520-6750\(199010\)37:5<627::AID-NAV3220370504>3.0.CO;2-F](https://doi.org/10.1002/1520-6750(199010)37:5<627::AID-NAV3220370504>3.0.CO;2-F)
- [3] Conover, W.J., Tercero-Gómez, V.G. and Cordero-Franco, A.E. (2017) The Sequential Normal Scores Transformation. *Sequential Analysis*, **36**, 397-414.
<https://doi.org/10.1080/07474946.2017.1360091>
- [4] Page, E.S. (1961) Cumulative Sum Control Charts. *Technometrics*, **3**, 1-9.
<https://doi.org/10.1080/00401706.1961.10489922>
- [5] Roberts, S.V. (1959) Control Chart Tests Based on Geometric Moving Averages. *Technometrics*, **1**, 239-250. <https://doi.org/10.1080/00401706.1959.10489860>
- [6] Qiu, P. (2014) Introduction to Statistical Process Control. CRC Press, London.
- [7] Mandelbrot, B.B. (1963) The Variation of Certain Speculative Prices. *Journal of Business*, **XXXVI**, 392-417. <https://doi.org/10.1086/294632>
- [8] Bollerslev, T., Chou, R. and Kroner, K. (1992) ARCH Modeling in Finance. *Journal of Econometrics*, **52**, 5-59.
[https://doi.org/10.1016/0304-4076\(92\)90064-X](https://doi.org/10.1016/0304-4076(92)90064-X)
- [9] Engle, R. (1995) ARCH Models. Oxford University Press, Oxford.
- [10] Ding, Z., Granger, C. and Engle, R. (1983) A Long Memory Property of Stock Market Returns and a New Model. *Journal of Empirical Finance*, **1**, 83-106.
[https://doi.org/10.1016/0927-5398\(93\)90006-D](https://doi.org/10.1016/0927-5398(93)90006-D)
- [11] Ding Z. and Granger, C.W.J. (1996) Modeling Volatility Persistence of Speculative Returns: A New Approach. *Journal of Econometrics*, **73**, 185-215.

- [https://doi.org/10.1016/0304-4076\(95\)01737-2](https://doi.org/10.1016/0304-4076(95)01737-2)
- [12] Cont, R., Bouchaud, J.-P. and Potters, M. (1997) Scaling in Stock Market Data: Stable Laws and Beyond. In: Dubrulle, B., Graner, F. and Sornette, D., Eds., *Scale Invariance and Beyond*, Springer, Berlin, 75-85.
https://doi.org/10.1007/978-3-662-09799-1_5
- [13] Lux, T. and Marchesi, M. (2000) Volatility Clustering in Financial Markets: A Microsimulation of Interacting Agents. *International Journal of Theoretical and Applied Finance*, **3**, 675-702. <https://doi.org/10.1142/S0219024900000826>
- [14] Kirman, A. and Teyssiere, G. (2002) Microeconomic Models for Long-Memory in the Volatility of Financial Time Series. *Studies in Nonlinear Dynamics and Econometrics*, **5**, 281-302. <https://doi.org/10.1162/10811820160130260>
- [15] Lu, C.W. and Reynolds, M.R. (1999) Control Charts for Monitoring the Mean and Variance of Autocorrelated Processes. *Journal of Quality Technology*, **31**, 259-274. <https://doi.org/10.1080/00224065.1999.11979925>
- [16] Dominicy, Y., Hörmann, S., Ogata, H. and Veredas, D. (2013) On Sample Marginal Quantiles for Stationary Processes. *Statistics & Probability Letters*, **83**, 28-36. <https://doi.org/10.1016/j.spl.2012.07.016>
- [17] Ross, S.M. (2014) *Introduction to Probability Models*. Academic Press, New York.
- [18] Pepelyshev, A. and Polunchenko, A. (2017) Real-Time Financial Surveillance via Quickest Change-Point Detection Methods. *Statistics and its Interface*, **10**, 93-106. <https://doi.org/10.4310/SII.2017.v10.n1.a9>
- [19] Villanueva-Guerra, E.C., Tercero-Gómez, V.G., Cordero Franco, A.E. and Conover, W.J. (2017) A Control Chart for Variance Based on Squared Ranks. *Journal of Statistical Computation and Simulation*, **87**, 3537-3562. <https://doi.org/10.1080/00949655.2017.1376327>