TIME SERIES TREND AND MEAN REVERSION SCORING

ARUN VERMA, STEPHEN TAYLOR

1. Goals and Guidelines

Consider a time series of length n with ordered components $\{x_i\}$ for $i = 1, \ldots, n$. Our first aim is to associate a score between -100 and 100 to $\{x_i\}$ based on how strongly the time series is trending. Scores near ± 100 will be associated with strongly positively/negatively trending series, and scores near 0 will be assigned to series that do not exhibit trending characteristics. If a trend score Z for a given series is contained in the interval $[-25, 25]$, (i.e. the series exhibits little to no trending characteristics), then we will compute a mean reversion score for the series. The mean reversion score will take values between 0 and 100; scores near 100 will correspond to highly mean reverting series, and those near 0 will be associated with series which exhibit little to no mean reversion.

We now mention a few guidelines that we incorporate into the trend indicator's construction below. First, we take any time series sampled from a line as our prototypical trending graph which will be given a score of ± 100 depending on the sign of the slope of the regression line. The magnitude of the slope of such a line will not be relevant to the trend score, i.e. we assign a time series sampled from a line with slope 1 and a time series with slope 5 identical scores of 100. We make this choice since if a trader where to identify either of these ideal trending series in the market and held the view that historically trending prices strongly predict future returns, then he/she would expect to make a guaranteed profit in either of these situations. Stated another way, we base the trend score around surety of profit from a buy/hold trading strategy rather than the magnitude of the strategy's expected return. With this in mind, we score time series according to how dispersed they are about their corresponding linear regression (best fit) lines. If a series closely follows its best fit regression line, it will have a high trend score; if it is highly dispersed, then it will have a low trend score.

Finally, we turn to considering the problem of regime change detection for bond spread time series. We first give an example which shows how we can find regime change points in a given time series. In order to identify change points, we first have to construct a regime change score associated with certain subsections of the base time series. We lastly provide details related to the construction of the score.

The document is organized in the following manner: In section 2, we describe bond test data taken from the FIRV screen which we design our trend and mean reversion indicators around. In section 3, we provide the definition (methodology) for our trend indicator and score the twenty bond spread test series. In section 4, we provide the definition (methodology) for our mean reversion indicator and score the bond spread test series whose absolute trend scores are less than 25. In section 5, we outline our regime change detection algorithm by applying it to an example of a bond spread time series. We then provide details of how we calculate the regime change score.

2. Bond Spread Test Data

We first extract twenty distinct end-of-day bond spread time series of varying lengths from the FIRV terminal screen. The range of each graph is given in terms of yield differences between a specified bond and a comparable security which is defined by its Spread type. For instance, G – Sprd gives the spread of the original bond to some comparable treasury, and I – Sprd gives the yield difference of the bond and associated domestic swap curve.

In the table below, we identity different bond spread time series which will test our trend and mean reversion scores on. In particular, we give the twenty base bond names along with the type

Date: July 5, 2011.

of bond spreads that we are considering. We plot the twenty bond spread time series in Figure 1 in blue.

Spread No.	Bond Name	Spread Type
1	IBM 7.5 06/15/13 Corp	Sprd - Bnch
2	IBM 7.5 06/15/13 Corp	I - Sprd
3	IBM 7.5 06/15/13 Corp	Z - Spad
4	IBM 7.5 06/15/13 Corp - ORCL 4.95 04/15/13 Corp	Hist Sprd
5	ORCL 4.95 04/15/13 Corp	Sprd - Bnch
6	F 7.45 07/16/31 Corp	Sprd - Bnch
7	PG 4.7 02/15/19 Corp	Sprd - Bnch
8	MET 2.375 02/06/14 Corp	Sprd - Bnch
9	ORCL 4.95 04/15/13 Corp	Sprd - Bnch
10	C 5 09/15/14 Corp	Sprd - Bnch
11	BAC 4.625 02/18/14 Corp	Sprd - Bnch
12	HBI 8 12/15/16 Corp	Sprd - Bench
13	HLS 7.25 10/01/18 Corp	Sprd - Bench
14	Y 5.625 09/15/20 Corp	I - Sprd
15	TTMTIN 0 07/12/12 Corp	G - Sprd
16	FCX 8.25 04/01/15 Corp	Sprd - Bnch
17	X 7 02/01/18 Corp	Sprd - Bnch
18	TWC 4.125 02/15/21 Corp	Sprd - Bnch
19	TWC 4.125 02/15/21 Corp	I - Sprd
20	NWSA 9.25 02/01/13 Corp	Sprd - Bnch

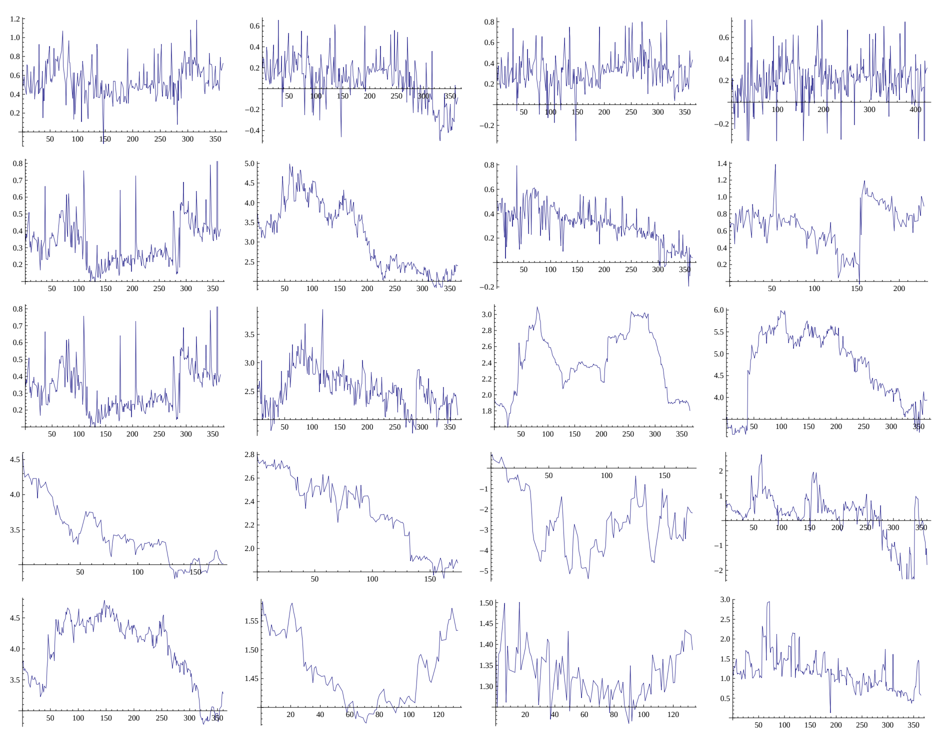

FIGURE 1. Here we plot bond spread time series numbers 1 to 20 ordered from left to right, then top to bottom. Times are in real days (i.e. not 252 trading days) and the range values represent yield spreads.

Note that these time series exhibit a wide variety of characteristics: some series are trending, others have jumps, others appear to be Brownian motions, and still other graphs appear to be generated from mean reverting processes.

3. Trend Score Formula

We examined a wide variety of techniques to score trending time series. The below simple scoring mechanism turned out to produce scores that were most in line with our pictorial/profitable trading strategy notions of trending time series when compared with other alternatives.

Let $\{(t_i, x_i)\}$ be n data points in a time series and let $\bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i$ and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, be the mean time and mean asset price of the time series. Then define the time series correlation by

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(t_i - \bar{t})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(t_i - \bar{t})^2}}. \tag{1}$$

Define the score Z associated to a given time series by

$$Z = \mathrm{Rnd}\left(100 \times \mathrm{sgn}(\rho)|\rho|^{\alpha}\right), \tag{2}$$

where here Rnd $(\cdot)$ is the round function, e.g. $\mathrm{Rnd}(4.34) = 4$ and $\mathrm{Rnd}(15.68) = 16$, $\mathrm{sgn}(\cdot)$ is the sign function, i.e. $\mathrm{sgn}(-14) = 14$ and $\mathrm{sgn}(11) = 11$, and we have fixed $\alpha = 3$. Note that $Z \in [-100, 100]$ by construction; moreover, one only realizes the extremal scores in the event that $\{(t_i, x_i)\}$ are collinear. Scores near -100 correspond to series which have a strong negative trend. Scores near 0 exhibit little to no trend characteristics.

The scores of the twenty example bond spread plots which we previously considered are given in the following table:

0	-16	0	-1
0	-56	-36	0
0	-5	0	-2
-72	-86	-6	-22
-7	-3	-1	-23

We now comment on how one can convert these numerical scores to qualitative scores. We assign qualitative score names to our numerical trend scores based on the following table:

Score Range	Qualitative Score
$Z \in [-100, -51]$	Strongly Negatively Trending
$Z \in [-50, -26]$	Weakly Negatively Trending
$Z \in [-25, 25]$	Not Trending
$Z \in [26, 50]$	Weakly Positively Trending
$Z \in [51, 100]$	Strongly Positively Trending

4. Mean Reversion Score Formula

We now outline how to construct a mean reversion score Z. We only compute this mean reversion score for a time series if the associated trend score is between -25 and 25.

Let $X = \{x_1, \ldots, x_n\}$ be a time series of length n. Then the quadratic variation of this time series is defined by,

$$QV(X) = \sum_{i=2}^{n}(x_i - x_{i-1})^2. \tag{3}$$

The average value of the process is given by

$$\bar{X}(X) = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{4}$$

and the variance of X is defined by

$$\sigma^2(X) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X}(X))^2. \tag{5}$$

Then the mean reversion score $Z(X)$ assigned to a time series X is defined by

$$Z(X) = \mathrm{Rnd}\left(100 \times 2^{-k\sigma^2(X)/QV(X)}\right), \tag{6}$$

where here $\mathrm{Rnd}(\cdot)$ is the round function and $k = 15$ is chosen to give an appropriate normalization for the scores of our test data. The value of k was chosen to be fifteen in order to force our mean reversion and trend scores to be approximately comparable, i.e. a very strongly trending series and a very strongly mean reverting series will get roughly the same scores; they same will also hold for weakly trending/mean reverting series. Mean reversion score values are defined for any positive value of k; for instance in the case $k = 30$, the scores of our test data are given by:

k = 30			
94	91	95	96
93	*	*	77
93	87	2	14
*	*	34	65
13	7	85	83

Here we use a star for in four places in the above table to indicate that we do not compute the mean reversion score for any of the graphs whose absolute trend scores are greater than 25, since these series are already classified as trending series. We finally outline how one can convert these numerical mean reversion scores to qualitative scores by the rules given in the below table.

Score Range	Qualitative Score
$Z \in [0, 50]$	Not Mean Reverting
$Z \in [51, 100]$	Strongly Mean Reverting

5. Regime Change Detection

We now turn to the separate problem of detecting regime changes in a time series. We use the term "regime change" to loosely indicate that a time series qualitatively changes behavior from a given regime (time period) to a subsequent regime. For example, consider the time series in Figure 2. Note that there are several regime changes in this series. In particular, there is a weak down trending regime over the $t \in [0, 20]$ time interval, a strong uptrend for $t \in [21, 55]$, a strong downtrend for $t \in [56, 95]$, a weak mean reversion regime for $t \in [115, 145]$, two strong mean reversion periods for $t \in [151, 172]$ and $t \in [176, 207]$, a strong down trend when $t \in [205, 230]$, and a strong mean reverting regime when we have $t \in [230, 260]$. The aim of regime shift detection is to determine at which points in the time series a change from one regime to another occurs.

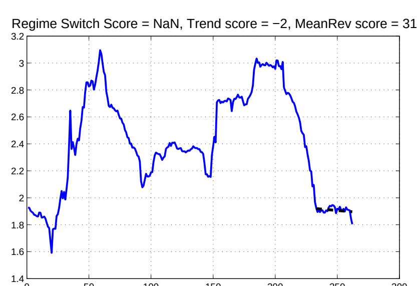

Figure 2. Time series 11 from the above bond spread data which exhibits a variety of different regime changes.

We first outline how our regime shift detection algorithm determines such points using the series in Figure 3 as an example.

In the first subfigure of Figure 3, we plot the first 15 points of our example time series. Note that the mean reversion score for this sub-series is -61, so this sub-series is classified as strongly negatively trending. Next, in the second subfigure, we plot the first 24 days of the series. Note that the trend score is no longer as strong as in the previous case. This is due to the fact that the series moved upward for $t \in [20, 24]$ times, i.e. the previous downward trend did not continue; this sub-series is classified as weakly downtrending which we indicate with a thin red line. Since the time series has changed from a downward to an upward trending series, we should expect that a potential regime shift indication should appear shortly if this upward trend continues. In the third sub-figure, we plot the first 25 values of the time series. Since the upward trend has continued, we now have enough information to signal that a true regime shift has occurred. We then find the optimal beginning and end points of the previous downtrending regime which we indicate with a black and red dot. Stated another way, we signal that a change has occurred from the previous $[0, 20]$ downtrending regime at time $t = 25$.

After the $t = 20$ change point has been detected, we continue to sample the time series and search again for when we deviate from the upward trending regime. Note that in subfigure 4, the upward trend has continued through the 35th point of the time series. We indicate with a dark green line that a strong upward trend is present. Then in the fifth subfigure, we see that from times $t \in [35, 40]$, that the time series has started to enter into a mean reversion regime. The yellow point indicates that it is likely we will signal a regime shift shortly if the mean reversion continues.

In the first subfigure of Figure 4, we note that the mean reversion portion of the time series did not continue since the series reverted back to an upward trending series. We thus did not signal a regime change during this period. Also note here that during times $t = [45, 48]$, the strong upward trend has ceased. The light blue dot gives a weak signal that regime shift has occurred.

We continue the series in the next subfigure where we note that there has been a change from the upward trending $t \in [20, 58]$ regime to a negatively trending series during times $t \in [58, 66]$. At time $t = 66$, we signal that a change has indeed occurred and again use black and red dots to indicate the beginning and end of the previous trend. We continue onward in the third subfigure where we denote the beginning and end of the previous regime with a red and black dot. In the fourth subfigure, we have changed from the previous trending portion of the time series to a mean reverting region. Note that we have a high mean reversion score for the subseries which starts at time $t = 150$ and ends near $t = 170$. We depict this strong mean reversion score graphically with several black triangles.

Finally, in the last two subfigures, we summarize all change points which we detected. Specifically, in subfigure 5, we plot red dots on each point of the time series where a regime change was detected. In the sixth subfigure, we show the running mean reversion score for the time series in the top plot, the twenty day running trend score in the middle graph, and the twenty day running mean reversion score in the lower graph. Note that discontinuities in these graphs correspond to points where a change was detected.

6. Adaptive Trend Construction and Regime switch Indication Methodology

We now provide details for the full algorithm which was described informally in the previous section. The current version of MATLAB code for this algorithm is given in U:/AVERMA/matlab/mean reversion vs trend/ regime_breakout_latest.m.

Initially set previous_break_out to 1 (which corresponds to the first point of the time-series). Denote the full time series by $x(1 : N)$ where here N is the total number of data points in the time series. Let M be the minimum length of a segment (we use M=15) on which we with to calculate the trend classification score on.

Repeat the following steps until done:

(1) Identify the sub-series from $k = \mathit{previous\_break\_out}$ to the current point N, lets denote this by $x(k : N)$
(2) Calculate the current adaptive trend classification of the sub-series by calculating the trend and mean reversion score of every segment $x(j : N)$ where $j = k, k+1, \ldots, N-M+1$. This will identify an **optimal starting point** for the current trend as start_trend. This step is described in detail in the sub-section on **Detailed Construction of the Adaptive Trend classification**.

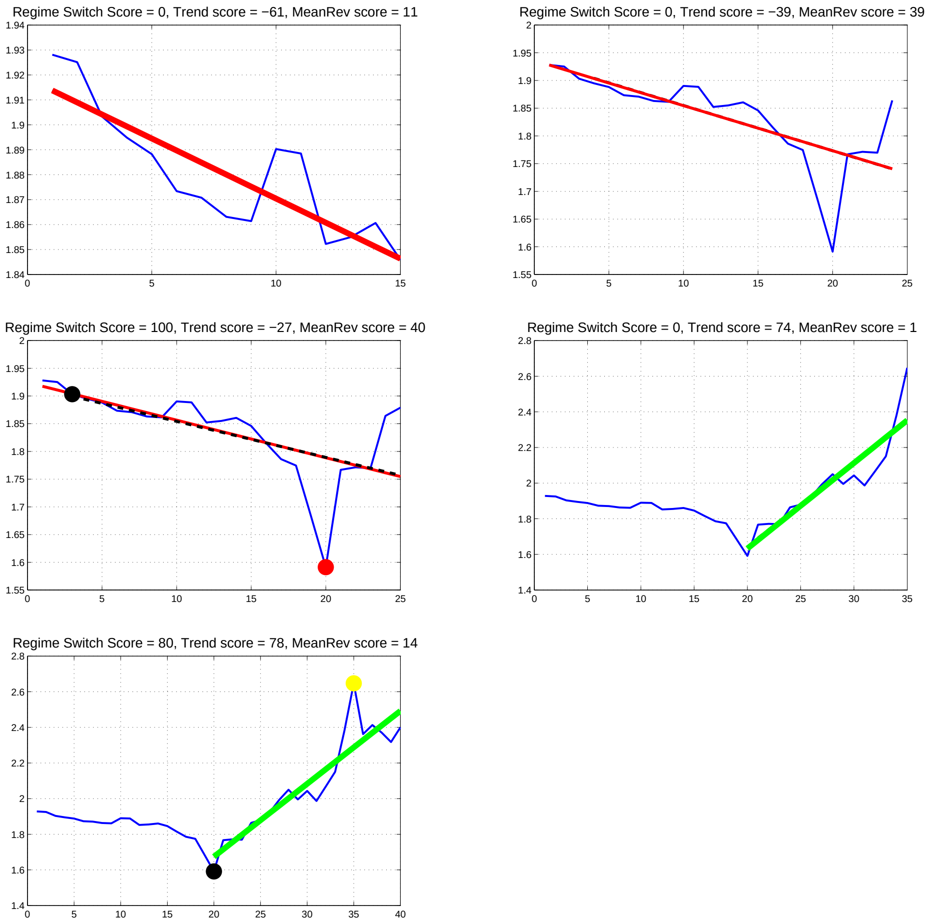

FIGURE 3. Time series 11 from the above bond spread data which exhibits a variety of different regimes.

(3) Calculate the $\widehat{RSS}$ value on $x(start\_trend : N)$ as described in the sub-section on **Automatic detection of Regime switch**. Indicate regime switch has occurred if discovered and set new break out to the end of "previous regime" in the $\widehat{RSS}$ calculations.
(4) If no direct regime switch was identified, check for manual switch over-ride. Indicate regime switch if discovered and set new break out to the end of "previous regime" found in the new $\widehat{RSS}$ calculations. For detailed steps on this please see sub-section **Manual Regime Switch Indicator**.
(5) Go to the next day and back to step 1.

Now we describe each of the above steps in more detail.

6.1. **Detailed Construction of the Adaptive Trend classification.** Given a series $x(k : N)$ calculate the current trend classification of the sub-series by calculating the trend and mean reversion score of every segment $x(j : N)$ where $j = k, k+1, \ldots, N-M+1$. This is done exactly as described in earlier sections on trend and mean reversion scoring.

For each j, first calculate the trend and mean reversion scores as described earlier in the document on the series $x(j : N)$. Assign $score(j) = |trend\_score(j)|$ if $|trend\_score(j)| >= 0.25$ else assign $score(j) = mean\_reversion\_score(j)$. We then apply a penalty term to the score by using : $score(j) = score(j) * \frac{2}{\pi} \arctan(\frac{N-j+1}{2})$. This favors the long-lasting trends more than short-lasting trends. This gives a trend of length 2 a relative weight of 50% compared to the weight of 87% for length 5, 93.5% for length 10 and 95% for length 20.

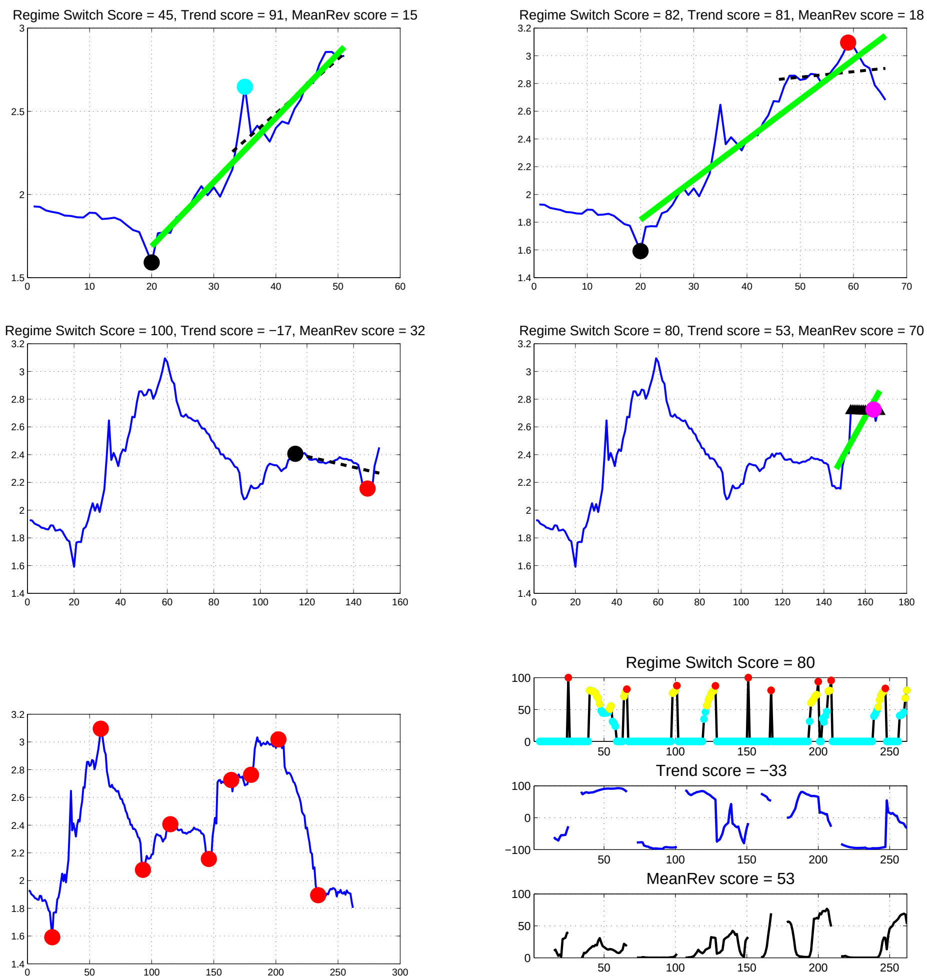

FIGURE 4. Time series 11 regime shift detection continued and a summary of regime change detection times for the full time series.

Find the maximum score from this vector of scores, $score(j), j = k, k+1, \ldots, N-M+1$, and assign the location for the maximum as the start_trend index to be used in the automatic regime detection calculations in the next section.

6.2. **Automatic detection of regime switch.** We now give a detailed explanation of the calculations which are involved in the construction of our mean reversion, trend, and regime shift scores depicted in subfigure 5 of Figure 4. The lower plot in subfigure 5 is just a plot of a adaptive mean reversion score. Similarly, the middle plot is a graph of adaptive trend score. Finally, the top graph corresponds to our regime shift score, whose construction we now outline.

Let $\{(t_i, x_i)\}$ for $i = 1, \ldots, n, \ldots, N$ be a given time series, where here x_1 and x_n denote the initial and final points of the previous regime. Here x_1 is identified to be the start_trend location as described in the previous section.**NOTE: This is $x(start\_trend : N)$ segment and we reset the indexing 1 to actually start_trend**.

We keep track of the current score and current classification every day. The trend score and mean reversion score using the optimal start_trend point are stored and a classification is made based on these two values as described by the appropriate range based classification on trend score followed by a classification on mean reversion score if necessary.

6.2.1. *Initial Computations On Previous Regime.* We describe three calculations on the candidate previous regime given a choice of n.

Previous Regime Classification

First, calculate the trend score and the mean reversion score on the previous regime. In particular, calculate the trend and mean-reversion scores on $x(1:n)$ segment - we will call this previous regime classification.

Trend line calculation

Denote the best fit regression line corresponding to the previous regime values $\{(t_i, x_i)\}$ by $\mu(t) = a + bt$, for $i = 1, \ldots, n$.

Here a and b are given by

$$b = \frac{\sum_{i=1}^n x_i t_i - \frac{1}{n}\sum_{i=1}^n x_i \sum_{i=1}^n t_i}{\sum_{i=1}^n t_i^2 - \frac{1}{n}(\sum_{i=1}^n t_i)^2}, \quad a = \frac{1}{n}\sum_{i=1}^n x_i - \frac{b}{n}\sum_{i=1}^n t_i. \tag{7}$$

NOTE: you will simply use $t_i = i$ as this is linear time (in days) but we use the notation t_i for more clarity.

Define μ_i for $i = 1, \ldots, n$ as sampled values of $\mu(t) = a + bt$ at times t_i, i.e. the time series $\{(t_i, \mu_i)\}$ $i = 1, \ldots, n$, is just a sampling of μ at times t_i, i.e. $\mu_i = \mu(t_i)$.

Now, shift the regression line to pass exactly through the final point by adding $(x_n - \mu(n))$ to this regression line. We then bend this regression line as follows. If the series is strongly upward trending or downward trending (based on Trend Score) then we reduce the slope b by factor of two. This ensures that if you were in a upward trending regime before, then the regime shift would be measured against a slightly lower upward trend as a reference. As a result, if the series continues to trend with same or even a slightly reduced slope (and indeed higher slope) then it would not be signaled as a regime shift.

If the previous regime was random or mean-reverting, then no such slope shift is applied. In particular,

$\mu(t) = a + bt + (x_n - (a + bt_n))$ for times $t \le t_n$, and,

$\mu(t) = a + bt_n + \frac{b}{2}(t - t_n) + (x_n - (a + bt_n))$ for times $t > t_n$ if previous regime is upward or downward trending; otherwise we extrapolate using the first function $\mu(t) = a + bt + (x_n - (a + bt_n))$ over the full time range $t_1, \ldots, t_N$.

Volatility Calculation

Finally, define the daily returns of the price process by $r_i = x_i - x_{i-1}$, and let σ be the standard deviation of $\{r_2, \ldots, r_n\}$.

6.2.2. *Computations After End of Previous Regime.* We now wish to assign a regime shift score to different sub-series of $\{x_{n+1}, \ldots, x_N\}$, which satisfy the following three constraints. First, the length of the sub-series must be greater than fifteen. Second, the final change point (detection point) must be within five and twenty days of the current time t_N, and finally, the initial point x_{n+1} must have a greater time value than the end of the previous series.

Alternatively, we can state these constraints as follows: Let x_1 be the initial point of the new regime that we wish to detect, let x_n for $1 \le n \le N$ be the final point, and let x_N be the current day. Then we must have

(i) $n - 1 \ge 15$,

(ii) $N - n \in [5, 20]$,

Given these constraints, we want to compute a trend score for each admissible subseries of $\{x_{n+1}, \ldots, x_N\}$. We first compute the values of Z_j associated to every j in the detection region $(n+1) : N$ which are defined by

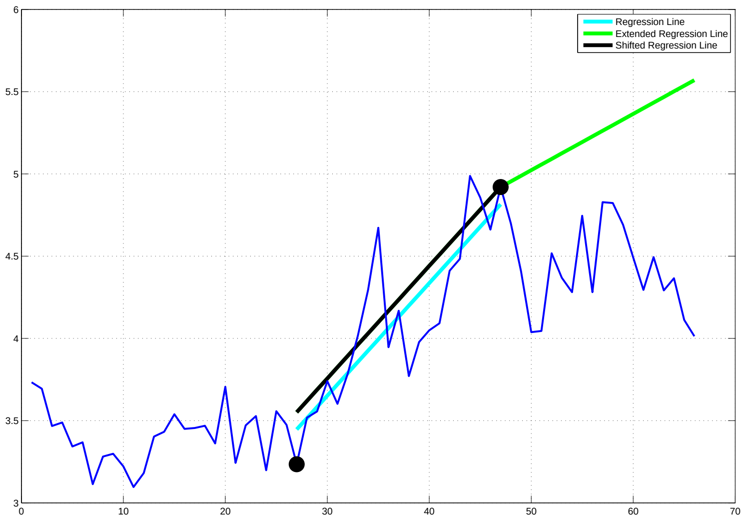

FIGURE 5. Here we plot an example of the best fit regression line for the previous regime $t \in [27, 46]$ in light blue, and the shifted regression line above it in black. We then plot the extended/slope adjusted regression line whose construction is outlined above in green.

$$Z_j = \frac{x_j - \mu_j}{\sigma f(j)}, \quad f(j) = \frac{\arctan(j-n)}{\arctan(1)}, \tag{8}$$

where σ and μ_j are taken from the computations in the previous subsection.

Secondly, for an admissible subseries starting at start_trend (x_1) and terminating at x_n, we define the regime shift score $RSS(n)$ according to

$$RSS(n) = \frac{\sum_j w_j Z_j}{\sum_j w_j}, \quad w_j = (\frac{1}{2})^{\frac{N-j}{2}}, \quad j = n+1, \ldots, N \tag{9}$$

We also form the final $RSS(n)$ value as

- If previous regime classification is mean-reverting then, $RSS(n) = |RSS(n)|$
- if previous regime classification is random then reset $RSS(n) = 0$
- If previous regime classification is trending up then, $RSS(n) = -RSS(n)$. If any of the Z values is positive then $RSS(n) = 0$.
- if adaptive classification is trending down and any of the Z values is negative then $RSS(n) = 0$.

6.3. **Optimal break out detection.** From the matrix of RSS scores of all admissible subseries, we pick up the highest absolute score which identifies the ending point j of the previous regime. The starting point i is fixed to start_trend point identified in the first step of the algorithm.

Finally, make a transformation to convert the above score to a new score in a range from 0 to 100. We do this via mapping the cumulative normal score of the RSS value to between 0 to 100 and then use linear interpolation. We map $RSS = 0$ to a score of 0, $RSS = 1.25$ to a score of 50, $RSS = 2.5$ to a score of 80, and $RSS = \infty$ is mapped to 100. Given a RSS value, we simply linearly interpolate using $normcdf(RSS)$ and above mappings to produce the transformed score. We denote this transformed score by $\widehat{RSS}$. In particular, $normcdf(0) = 0.5$ is mapped to 0, $normcdf(1.25) = 0.8944$ is mapped to 50, $normcdf(2.5) = 0.9938$ is mapped to 80 and finally *normcdf*$(\infty) = 1$ is mapped to 100. For interpolation between these values you would need an implementation of *normcdf*, i.e. cumulative normal probability function.

If $\widehat{RSS} \in [0, 0.5]$, then we plot a black dot on $x_1 = start\_trend$ in the time series graph and a blue dot on x_j. If $\widehat{RSS} \in [0.50, 0.80]$, we plot a yellow dot on x_j instead of a blue dot. Similarly, if $\widehat{RSS} \in [0.8, 1.0]$, we plot a red dot on x_j and signal that a regime shift has occurred. We then repeat the above process again and try to detect a signal change for the next regime.

6.4. **Manual Regime Switch Indicator.** We indicate a manual regime switch if the adaptive classification as of current date has changed significantly from yesterday's classification. In particular, one of the following scenarios needs to happen for us to indicate a manual regime switch(note we store the history of **adaptive trend and mean reversion scores** to be able to do this analysis):

- From up or down trend (weak or strong) to mean-reverting.
- From up (or down) trend to opposite down (or up) trend. Both trends could be either weak or strong.
- From random to **strong** up or down trend. A switch to weak trend doesn't qualify as regime change.
- From mean-reverting to up or down trend (weak or strong). We also require the trend score to have changed by at least 10% (absolute value change in trend score should be greater than 0.1) because of boundary cases.

Finally for all cases above, we require that the new trend classification should not be identical to the last available manual classification. In particular, when we get into a manual mode we calculate the new trend classification we are going into. When the next manual detection comes along (without any automatic detection in between) and shows the same new trend then we disqualify that. Additionally, the last available classification is nullified whenever we have an automatic regime switch occur. After a manual switch detection is complete, the last available classification is reset to the new classification as of the current day.

Once a manual switch is detected, we use a similar methodology to find the optimal break-out point as done in the automatic switch detection including the previous regime and computations after the end of the previous regime.

We change the bounds in the constraints to a minimum of 10 days for the trend (from 15) and minimum 3 days for the breakout period (from 5). This will help manual detection work in a more timely fashion, as typically manual breakouts are triggered a few days after they have occurred and we do not want to add longer constraints to further delay the indicator. Also, note that current optimal point for trend is not valid as manual trigger occurs when regime has already changed. We go back in time and find the best starting point and finish point of the trend that we have recently switched from.

In particular, we need to search over a matrix of RSS scores for the best one. Let x_m be the initial point let x_n for $m \leq n \leq N$ be the final point, and let x_N be the current day. Then we must have

(i) $n - m \geq 10$,
(ii) $N - n \in [3, 20]$,
(iii) $m \geq 1$

We build RSS for each (m, n) pair that qualifies. Also, note that previous regime classification for the purposes of RSS calculations is done over the $x(m : n)$ segment.

The optimal breakout is identified by searching over the 2-D array for the maximum RSS value. Note that we do not further look at the $\widehat{RSS}$ value and indicate the optimal point of breakout irrespective of the magnitude of the regime switch score. We manually set the regime switch score to 0.8 (lowest value for the "red" breakout).

7. Speedup improvements

The trend and mean-reversion calculation for everyday between previous break out and now has common calculations that can be shared. we are looking at calculating scores for $x(t : N)$..

7.1. **Mean Reversion Calculation.**

(1) For mean reversion we can build a running QV function which is simply adding daily squared difference each day - so the QV for starting point t, is simply equal to QV for starting point $(t-1)$ minus $(x(t)-x(t-1))^2$ instead of calculating from scratch. We recommend doing this loop in reverse order, i.e. Use $QV(t-1) = QV(t) + (x(t)-x(t-1))^2$

(2) The mean value $\bar{X}(t)$ can also be updated as follows, $\bar{X}(t-1) = \frac{\bar{X}(t-1)*(N-t+1)+x(t-1)}{N-t-2}$

(3) the volatility can be updated as $\sigma^2(t-1) = \frac{\sigma^2(t)*(N-t)+x(t-1)^2+\bar{X}(t)^2*(N-t+1)-\bar{X}(t-1)^2*(N-t+2)}{N-t+1}$

7.2. **Trend Calculation.**

(1) For trend calculation we can get speedup using by noting that the terms in denominator, call them $SSx(t) = \sum_{i=t}^{N}(x_i-\bar{x})^2, SSt(t) = \sum_{i=1}^{N}(t_i-\bar{t})^2$ for sum of squares(note the slight change of notation from trend formula to make trend and mean reversion speedup equations consistent).

(2) We can calculate these iteratively using update for the means $\bar{X}(t-1) = \frac{\bar{X}(t)*(N-t+1)+x(t-1)}{N-t+2}$, $\bar{t}(t-1) = \frac{\bar{t}(t)*(N-t+1)+t(t-1)}{N-t+2}$ (here $t(t-1)$ is simply the time equal to $t-1$)).

(3) The sum of squares can be updated as $SSx(t-1) = SSx(t)*(N-t+1)+x(t-1)^2+\bar{X}(t)^2*(N-t+1)-\bar{X}(t-1)^2*(N-t+2)$, $SSt(t-1) = SSt(t)*(N-t+1)+t(t-1)^2+\bar{t}(t-1)^2*(N-t+1)-\bar{t}(t-1)^2*(N-t+2)$.

(4) finally update for the numerator $N(t-1) = N(t)+x(t-1)*t(t-1)+(N-t+1)*\bar{X}(t)*\bar{t}(t)-(N-t+2)*\bar{X}(t-1)*\bar{t}(t-1)$

7.3. **RSS Calculation.**

(1) The RSS calculation uses mean reversion and trend calculation which can be speeded up using the methods described above first.

(2) The calculation for a is similar to calculation of trend, the denominator just using SSt instead of $\sqrt{SSx*SSt}$ and hence can be speeded up the same way.

(3) μ calculation can't really be speeded up.

(4) σ calculations can be speeded up just like volatility calculation in mean reversion.

(5) I would keep Z calculations same as before without any speedup.

7.4. **Manual classification.**

(1) The manual classification also uses RSS calculation which can be speeded up using the methods described above.

8. Filling Gaps

Finally there would be days when there is no classification available (due to lack of data since last breakout). On such days, instead of outputting N/A for the classification, we should go back to previous minimum number of days and output the trend/meanrv classification for that segment. This shouldn't over-ride the main logic of adaptive trend in the algorithm but just purely on the output front tom make sure we fill the gaps.