Nonparametric Smoothing of Time Series

James A Shine*   James E Gentle†

Abstract
Smoothing of data helps in understanding the data-generating process, but, by its nature, it often obscures interesting observations. In time series data, one kind of interesting observations is a changepoint, that is, a point in the time series at which the data-generating process undergoes a change. The change may be in the general trend of the data, or it may be a change in volatility. These kinds of changes are of particular interest in economic and financial data, so it is important that a smoothing method not obscure the changepoints. The nonparametric alternating trends smoothing (ATS) technique is based on explicit identification of changepoints in the trends of the means in a time series. ATS performs favorably in changepoint detection when compared to other nonparametric smoothing methods such as moving averages. A disadvantage of the ATS approach, however, is that the underlying model assumes that changing trends change in sign.

Key Words: time series, smoothing, changepoints

1. Introduction: Smoothing Time Series

We consider the problem of smoothing a discrete time series, $x_1, x_2, \ldots$. Smoothing of the time series yields a related time series $\tilde{x}_1, \tilde{x}_2, \ldots$ that has less noise; it is smoother.

We smooth time series data for several different reasons. We may do retrospective smoothing of a given set of time series data in order to understand the data-generating process. If a parametric model such as an autoregressive integrated moving average (ARIMA) or a generalized autoregressive conditional heteroscedasticity (GARCH) fits a given set of time series well, the parameters in the fitted model provide information about autocorrelations, about trends in the data, and something about the variance-covariance structure. Nonparametric smoothing generally does not yield a model with parameters that are directly interpretable, but nonparametric retrospective smoothing may identify patterns in trends in the data that indicate important properties of the time series. Retrospective smoothing is done on a given dataset; each observation in the dataset is already known. While retrospective smoothing may provide useful information about the nature of the time series, in applications, we are generally not interested in the past. Smoothing of time series is generally sequential; that is, the smoothing process proceeds forward in time. Two reasons for smoothing time series are to forecast future values and to identify changepoints in the data-generating process.

1.1 Forecasting

Forecasting is an important objective of time series analysis. Forecasting presupposes some regularity in the time series. Given a fitted parametric model such as an ARIMA model, forecasting is trivial. The assumptions underlying the model determine the procedure, which only involves plugging data into the fitted model.

A nonparametric analysis does not yield a simple model that can be used in forecasting. The model underlying a nonparametric approach is a more general statement about the data

*U.S. Army Corps of Engineers, retired
†George Mason University, retired

generating process, possibly only that the data have the same mean over time, or that the data have a linear trend in time.

Given data from a time series, $x_1, \ldots, x_{t-1}$, in forecasting at time t, we are interested in the values x_t, and possibly $x_{t+1}, x_{t+2}, \ldots$. Sequentially smoothing the time series from time t forward, the forecasted values are $\tilde{x}_t, \tilde{x}_{t+1}, \tilde{x}_{t+2}, \ldots$.

In this paper, we occasionally refer to the "best forecast". In this we are not using precise statistical language, but even with a definite model, we are motivated by the statistical concepts of unbiased and minimum variance.

1.2 Changepoints

Changepoints are points in time at which characteristics of the data-generating process changes, or the model itself changes. The exact meaning of a changepoint depends on the model. Some models explicitly include changepoints. The knots in a spline model, for example, may be considered to be changepoints. In an ATS model, for example, the points at which the trends alternate are changepoints. In other cases, changepoints are those points at which the underlying model changes. This may be a fundamental change in the nature of the model, or it may be changes in specific aspects of the same general model (for example, changing the parameter values in an ARIMA model).

2. Nonparametric Smoothing of Time Series

Smoothing presupposes some regularity in the time series. Each point in the time series may be considered to be a random variable, and regularity of the series may be because some aspects of the distributions of the random variables are the same, or at least the same locally. Another possible aspect of regularity within a time series is that the distribution of the random variable at a given point in time may be conditional on the realizations of variables at previous points in time. As we indicated above, we will not consider retrospective smoothing where all the data are available; rather, we will focus on sequential smoothing methods, that is, methods that use data up to the current point in time and predict future values. We will also consider the use of these smoothing methods for identifying points at which the underlying data-generating process undergoes a change.

2.1 Time Series Models

There are various models that are useful in expressing the behavior of time series, both the individual observations and their relationship to each other.

2.1.1 Nonparametric Models

A nonparametric model is a general model whose main objective is to smooth the data or to provide predictions.

In nonparametric analysis, we generally do not assume specific distributions, such as normal or Cauchy, but we may make assumptions about specific aspects, such as the mean or median. We may assume, for example, that the means of the random variables constituting the time series are the same, or at least are the same locally; that is, for example, the mean of the random variable at time t, X_t, is the same as the mean of X_{t-1}. Under that assumption, given the observed values x_{t-1} and x_t, the smoothed values of the two, $\tilde{x}_{t-1}$ and $\tilde{x}_t$, should be close together, and without further assumptions about characteristics of the data-generating process, there is no reason for one smoothed value to be larger than the other.

The general form of a nonparametric model for a time series model is

$$x_t \approx f(x_{t-1}, x_{t-2}, \ldots), \tag{1}$$

where the function f just expresses some relationships that characterize the data-generating process. It may be the case that x_t does not depend explicitly on previous values, only that their distributional properties are related. If there are no relationships among the points in the time series, or there are no distributional properties in common, we cannot proceed with a meaningful statistical analysis. The function f in the model above often cannot be written as some simple equation involving $x_{t-1}, x_{t-2}, \ldots$, and we do not attempt to express the functional relationship f completely.

2.1.2 White Noise and Related Time Series

A process $\{w_t\}$ that is a sequence of random variables with mean 0, constant finite variance, and 0 autocorrelations at all lags is called white noise. The 0 autocorrelation condition can be stated as

$$\mathrm{Cor}(w_t, w_{t+h}) = 0 \quad \text{for } h = 1, 2, \ldots \tag{2}$$

A special type of white noise is Gaussian white noise, defined as white noise in which the random variables have a normal (Gaussian) distribution.

The main differences from one white noise process to another is the variance, $\mathrm{V}(w_t) = \sigma_w^2$.

A white noise process can serve as an approximate model for the returns of a time series of prices of financial assets, $(P_t - P_{t-1})/P_{t-1}$.

The variance, or the standard deviation, is important in modeling financial time series (the standard deviation is called the *volatility* or the *risk*.) An important type of changepoint is a point at which the volatility changes significantly, (of course, in practice, the volatility seems to undergo small changes constantly).

Because white noise processes include no relationships among the terms in a time series, they are rarely useful models for observed time series. Another use of white noise is in modeling the error term in time series models, for example reexpressing equation (1) as

$$x_t = f(x_{t-1}, x_{t-2}, \ldots) + w_t. \tag{3}$$

2.1.3 Random Walk

A simple model of a time series is the random walk, in which the model function f above is just x_{t-1},

$$x_t = \phi(x_{t-1}) + w_t, \tag{4}$$

and $\{w_t\}$ is a white noise.

The mean of a random variable following a random walk process starting at x_0 is

$$\mathrm{E}(x_t) = \mathrm{E}(x_0) + \mathrm{E}\left(\sum_{i=1}^{t} w_i\right) = x_0; \tag{5}$$

that is, the unconditional mean of a random walk is constant, depending only on x_0. The variance of a random walk variable, on the other hand, is

$$\mathrm{V}(x_t) \;=\; \mathrm{V}(x_0) + \mathrm{V}\left(\sum_{i=1}^{t} w_i\right)$$

$$\begin{aligned} &= 0 + \sum_{i=1}^{t} \mathrm{V}(w_i) \\ &= t\sigma_w^2, \end{aligned}$$

that is, it depends on t and in fact, grows linearly as t grows.

A random walk is sometimes used as an approximate model for the movement of asset prices. (The *random walk hypothesis* basically states that a random walk is an adequate model of stock prices.)

Given data $\ldots, x_{t-2}, x_{t-1}$, the best forecast under a random walk model for $x_t, x_{t+1}\ldots$ is just $x_{t-1}, x_{t-1}\ldots$.

A changepoint is a point at which the model changes, so for a random walk, it may be a point at which the variance of the white noise changes, or it may be due to a jump, or, in financial parlance, a "gap", where x_t, instead of being $x_{t-1} + w_t$ is $\delta + x_{t-1} + w_t$.

2.1.4 Random Walk with Drift

A variation is a random walk with drift, given by

$$x_t = \delta + x_{t-1} + w_t, \tag{6}$$

where, again, $\{w_t\}$ is a white noise.

Given a random walk with drift δ, and the data $\ldots, x_{t-2}, x_{t-1}$, the conditional expectation of X_t is The has $\mathrm{E}(x_t) = t\delta + x_0$.

$$\mathrm{E}(X_t|x_{t-1}, x_{t-2}, \ldots) = \mathrm{E}(X_t|x_{t-1}) = \delta + x_{t-1}. \tag{7}$$

(The process is Markovian.)

The mean of the random walk with drift is dependent on the time.

Many financial analysts, without explicitly mentioning the model, use an approximate random walk with drift model. The "trendlines" drawn by technical analysts are line segments with a slope of δ.

Given data $\ldots, x_{t-2}, x_{t-1}$, the best forecast under a random walk model with drift for $x_t, x_{t+1}\ldots$ is just

$$x_{t-1} + \hat{\delta},\ x_{t-1} + 2\hat{\delta}\ \ldots,$$

where $\hat{\delta}$ is a "best" estimate of δ.

A changepoint is a point at which the model changes, so for a random walk with drift, it may be a point at which the variance of the white noise changes, or it may be due to a break in the trend where δ changes in value. A type of changepoint important in financial analysis is one where the sign of δ changes, a "top" or a "bottom".

2.2 Global Smoothing: Fitting Lines and Curves

A simple nonparametric model for a time series is that all observations come from distributions with the same mean or median. In this model the "best" smoothed series is just the constant series equal to the mean or the median of all available observations.

Another simple nonparametric model is that all observations come from distributions with a mean or median that is linear in time. Although this may seem grossly oversimplified, many technical stock analysts use a variation of this model over selected time periods.

In realistic versions of these models, they are limited to restricted time periods, which obviously results in the question of where are the changepoints.

A regression line is an example of global smoothing that fits a trend,

$$\tilde{x}_t = a + bt, \tag{8}$$

where appropriate values of a and b may be determined by least squares regression.

A global model of course could be restricted to a given time regime, and another global model could be used in a subsequent regime.

2.3 Local Smoothing: Moving Averages and Kernel Smoothing in Time Series

Given data from a time series, $x_1, \ldots, x_{t-1}$, we may reasonably assume that the mean of the random variable X_t is close to the mean of the more recent observations. Hence, we may choose the "locally" smoothed value of X_t, $\tilde{x}_t$, as the mean of some previous observations. We let h be a positive integer with $h \leq t-1$, and let $\tilde{x}_t$ be the sample mean

$$\tilde{x}_t = \frac{1}{h}\sum_{i=1}^{h} x_{t-i}. \tag{9}$$

This smoother is a moving average with a window of length h. For a time series in which the unit of time is a day, this is an "h-day moving average". In smoothing daily financial time series such as stock prices, a 50-day moving average or a 200-day moving average is common. The larger is h, the more slowly the moving average changes from one day to the next.

2.3.1 Weighted Moving Averages

At time t, instead of averaging all previous h values equally, it may be better to use a weighted average in which values closer in time are weighted more heavily. We have a smoother of the form

$$\hat{x}_t = \frac{1}{h}\sum_{i=1}^{t-1} x_{t-i} K\left(\frac{t-i}{h}\right), \tag{10}$$

where h is a positive number less than t. The function $K(\cdot)$, which must be nonnegative, is called a kernel function and the smoother is a time series kernel smoother.

Kernel functions and kernel-based smoothing methods are used in many areas of nonparametric data analysis. Retrospective smoothing of time series is similar to other types of nonparametric smoothing, but the more practical sequential smoothing of a time series at any given point can only depend on values before that point in time. Note that the summation includes all observations up to time t.

Most kernels used in smoothing are positive functions symmetric about 0 and they integrate to 1. Some common kernels are rectangular, triangular, and Gaussian (normal).

The symmetric rectangular kernel, for example, is

$$K_{\mathrm{T}}(t) = \begin{cases} 1 & \text{if } |t| \leq 1/2 \\ 0 & \text{otherwise;} \end{cases}$$

the triangular kernel is

$$K_{\mathrm{T}}(t) = \begin{cases} 2-|t| & \text{if } |t| \leq 1/2 \\ 0 & \text{otherwise;} \end{cases}$$

and the Gaussian kernel is

$$K_{\mathrm{G}}(t) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}.$$

The kernel used in sequential time series smoothing, however, must be one-sided. A kernel for use in time series can be formed from a symmetric kernel by folding the function about 0. To preserve the area under the curve, either the width or the height must be changed.

A moving average smoother corresponds to a kernel smoother in which the kernel is a folded rectangular kernel; that is,

$$K_{\mathrm{FR}}(t) = \begin{cases} 1 & \text{if } -1 \le t \le 0 \\ 0 & \text{otherwise.} \end{cases}$$

A folded version of any symmetric kernel can be used. A folded triangular kernel, for example, is

$$K_{\mathrm{FT}}(t) = \begin{cases} 1+t & \text{if } -1 \le t \le 0 \\ 0 & \text{otherwise.} \end{cases}$$

A folded kernel with no lower bound allows all observations in a time series prior to time t to be included in the fit of x_t.

2.3.2 *Exponentially Weighted Moving Average*

A variation of a moving average is called an exponentially weighted moving average (EWMA). An EWMA is a weighted average of the current price (in financial applications, the previous day's price) and the previous EWMA. If at time t (with $t \ge 2$), the most recent price is p_{t-1}, and the proportion assigned to the current price is α, then

$$\mathrm{EWMA}_t = \alpha_t p_{t-1} + (1-\alpha_t)\mathrm{EWMA}_{t-1}, \quad \text{for } t \ge 2, \tag{11}$$

with the weighting factor α_t often taken as $2/t$, and

$$\mathrm{EWMA}_1 = p_1.$$

2.3.3 *Forecasting Using Moving Averages*

A forecast based on a moving average is just the moving average using data up to that point, either the actual observed data, or for farther beyond the current time, observed data plus other forecasts.

2.3.4 *Changepoints Identified by Moving Averages*

When a changepoint occurs, a moving average may or may not reflect the change.

Change in Mean

If the changepoint is because of a change in mean of the time series, a moving average with a smaller window width will reflect the change better than one with a larger window width. A moving average that places more weight on recent values will also better reflect the change. Of course, in any case, the issue is whether the change in mean is significant with respect to the overall variability in the time series.

Change in Variance

If the changepoint is because of a change in variance of the time series, a moving average also becomes more variable, but it does not directly indicate the change. If the time series is differenced, and a moving average is formed on the time series of the absolute values of the differences, the change in variance will be manifested by a change in the mean of the absolute values of the differences. As mentioned above, a moving average with a smaller window width will reflect the change better than one with a larger window width, and a moving average that places more weight on recent values will also better reflect the change. Also, as mentioned above, in any case, the issue is whether the change in the mean of the absolute differences is significant.

Change in Linear Trend

Trend lines are the basic elements of many patterns in the technical analysis of stock prices, and so changepoints that involve a change in trend are of particular interest. The differenced series represents the trend from one point to the next. A trend in the time series would yield a sequence of differences that are similar. As up trend would yield a differenced time series of positive values, and a down trend would yield a differenced time series of negative values.

If the changepoint is because of a change in the linear trend of the time series, then a moving average of the differenced time series will reflect that change. A moving average with a smaller window width will reflect the change better than one with a larger window width, and a moving average that places more weight on recent values will also better reflect the change.

Another way that moving averages can be used to identify changes in trends is by comparing two moving averages with different windows.

Changepoints in trends can also be identified by crossovers of moving averages over varying windows, although the detection of the changepoint occurs beyond the actual changepoint.

2.4 Local Smoothing: Trends in Time Series

In many applications of time series, the objective is to identify and model trends. Although interesting trends may be curvilinear ("parabolic" in the parlance of financial analysts), here, for simplicity, we will focus on linear trends. A change in the slope of a trendline is a changepoint, and hence, smoothing based trends is linked with the identification of changepoints.

Retrospectively, trends can often be identified by visual inspection of a plotted time series, and following retrospective smoothing, trends may be even more obvious visually.

Our interest is in the current and future trends.

2.4.1 Differencing to Detect Trends

Analysis of the first-order integrated series can identify linear trends because if the differences of first order of lag 1 are constant, then the time series has a linear trend.

In a time series that has linear trends over various subregions, the integrated series is constant over each subregion. Changepoints can be identified as the points at which the constant value of the integrated series changes.

2.4.2 Alternating Linear Trends and Changepoints

A method called alternating trends smoothing, or ATS, identifies changepoints in a time series using alternating (up and down) linear trends. ATS uses a smoothing parameter, h, , which specifies the step size within which to look for changepoints.

ATS provides changepoints and their corresponding trendlines. A trendline can be used to forecast values within a given regime.

ATS has been used in the past to detect specific patterns in financial data such as a Head and Shoulders pattern, post-pattern trends, and volatility measures before and after a pattern (Shine, Gentle and Perry, 2015,2017).

The algorithm for ATS is given below. It is reprinted by permission from Gentle and Wilson (2018).

Alternating Trends Smoothing (h)

1. Set $d = 1$ (changepoint counter)
2. While (more data in first time step)
   - (a) for $i = 1, 2, \ldots, m$, where $m = h$ if h additional data available or else m is last data item:
input x_i;
   - (b) set $b_d = 1$; $c_d = x_1$
   - (c) determine $j_+, j_-, x_{j_+}, x_{j_-}$ such that
$x_{j_+} = \max x_1, \ldots, x_h$ and $x_{j_-} = \min x_1, \ldots, x_h$
   - (d) set $s = (x_k - x_i)/(k - i)$ and $r = \text{sign}(s)$
   - (e) while $r = 0$, continue inputting more data; stop with error at end of data
3. Set $j = i$ (index of last datum in previous step); and set $d = d + 1$
4. While (more data)
   - (a) for $i = j + 1, j + 2, \ldots, j + m$, where $m = h$ if h additional data available or else $j + m$ is last data item:
input x_i;
     - i. while $(\text{sign}(s) = r)$
       - A. set $k = \min(i + h, n)$
       - B. if $(k = i)$ break
       - C. set $s = (x_k - x_j)/(k - j)$
       - D. set $j = k$
     - ii. determine j_+ such that rx_{j_+} is the maximum of $rx_{j+1}, \ldots, rx_{j+m}$
     - iii. set $b_d = j_+$; and set $c_d = x_{j_+}$
     - iv. set $d = d + 1$; set $j = j_+$; and set $r = -r$
   - (b) set $b_d = j_+$; and set $c_d = x_{j_+}$

∎

Within a given regime, a regression trendline, such as in equation (8), could be fit to the data between the changepoints defining the regime.

The trendline

$$x_t = a + bt$$

may be fitted by least squares, but least absolute values yield a more robust fit. The least absolute values or L_1 determines a and b so as to minimize

$$\sum_{i=1}^{n} |x_i - a + bi|.$$

3. Examples

3.1 Comparing Different Smoothing Approaches

Figure 1 shows the daily closing values of Intel stock for first three quarters of 2017, along with smoothing estimates from moving averages, an EMWA, and the ATS algorithm.

The values themselves are in black. The light green curve is a 20-day moving average, and the dark blue curve is a 40-day moving average, the purple curve is an exponentially weighted moving average, and the red alternating lines are the results from the ATS algorithm, using a value of $h = n/10$ where n is the total number of points.

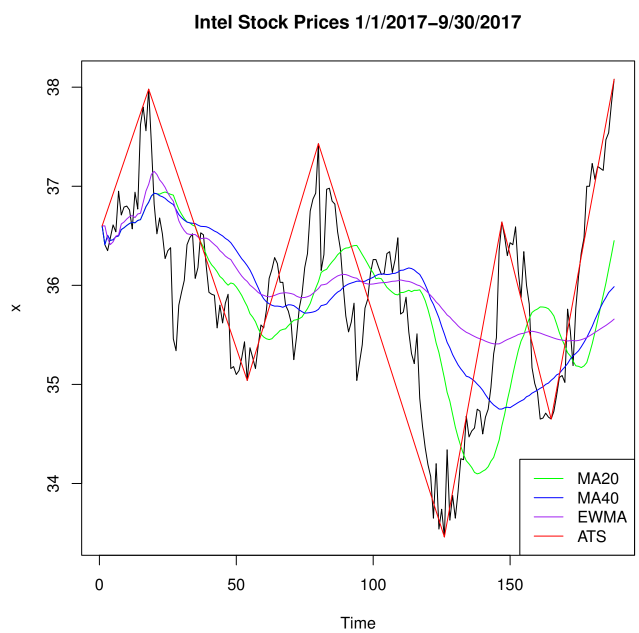

Figure 1: Comparison of different smoothing approaches

This example shows the precision of the ATS algorithm in identifying changepoints. We have shown the optional graphic produced by ATS, but the algorithm also lists the actual changepoints x_i, y_i in its output, as below. The moving average and EWMA curves are not precise in identifying changepoints and require manual interpretation.

```
ATS(x,addtoplot=TRUE,col="red")
      [,1]   [,2]
[1,]     1 36.60
```

```
[2,]   18 37.98
[3,]   54 35.04
[4,]   80 37.43
[5,]  126 33.46
[6,]  147 36.64
[7,]  165 34.65
[8,]  188 38.08
```

3.2 ATS with different step sizes

The h (step size) parameter in ATS can be adjusted to look for short-term or long-term trends. Figure 2 shows the ATS algorithm used on randomly generated data with a $N(0, 25)$ distribution. The red lines show the algorithm with a step size of 5, and the blue lines show the algorithm with a step size of 25. With a smaller step size, more changepoints are found.

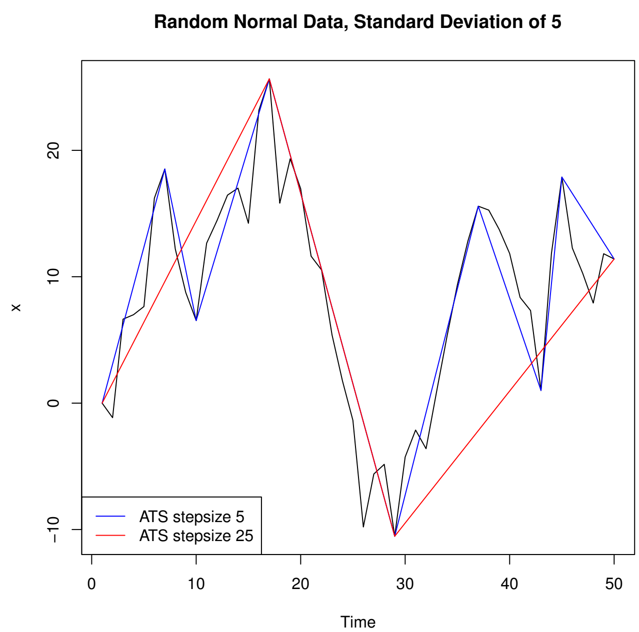

Figure 2: ATS with different step sizes

3.3 Two Random Walks with Two Different Drifts

Figure 3 shows ATS applied to 4 different random walk models. In these models, the first 50 points follow a random walk with drift δ_1 and the next 100 points follow a random walk with drift δ_2. For a small δ ATS does not pick up the changepoint, as in the upper left graph, but for larger δs ATS does identify the changepoint, as in the other three graphs.

4. Conclusions and Future Directions

Based on preliminary experiments, the ATS algorithm demonstrates promise as a nonparametric smoothing approach for identifying changepoints in time series data.

In our ongoing work, we are generating data sets and evaluating our method on these data sets. This will allow us to better determine how well this model works and what improvements might be necessary.

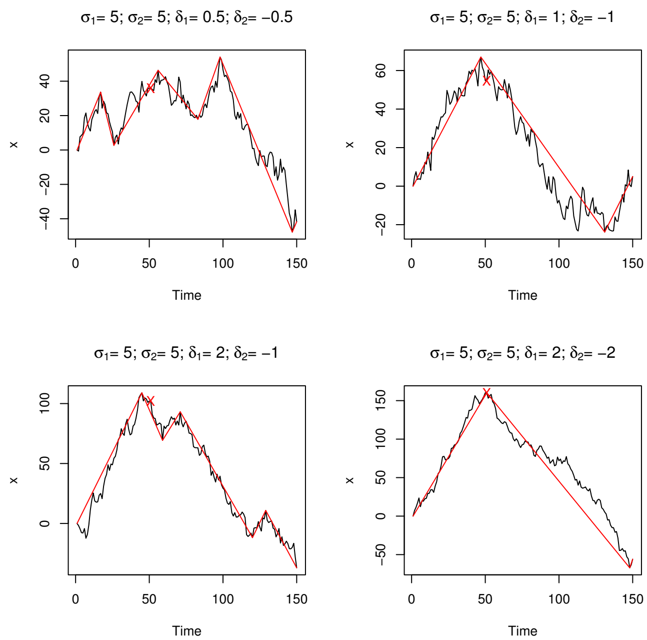

Figure 3: Changepoints in Two Simulated Random Walks with Drifts

A key feature we are working on is the ability to identify a changepoint with ATS without a change in the slope. Currently the lines are alternating in slope, but we want to allow a changepoint to be identified with the two slopes (before and after) retaining the same sign. We also are developing a metric to determine the ”strength” of an ATS changepoint, to determine which changepoints are ”strongest” or ”best”. We also have done preliminary work on ”bounding lines” around the data for each segment identified by ATS, and hope to develop that work further.

REFERENCES

Gentle, G., and Wilson, S. (2018), ”Finding Patterns in Time Series”, in *Handbook of Big Data Analytics*, eds. W.K. Hardle, H.H. Lu and X. Shen, Switzerland: Springer, 125-150.

Hardle, W. (1991), *Smoothing Techniques with Implementation in S*, New York, USA: Springer-Verlag.

Hirukawa, M. (2018), *Asymmetric Kernel Smoothing: Theory and Applications in Economics and Finance*, Singapore: Springer.

Loader, C., (1999), *Local Regression and Likelihood*, New York: Springer.

Schimek, M.E. ed (2000), *Smoothing and Regression: Approaches, Computation, and Application*, New York: Wiley.

Shine, J.A., Gentle, J.E., and Perry, C. (2015), Discovering and Evaluating Trend Patterns in Financial Time Series, in *Proceedings of 2015 Joint Statistical Meetings*, Seattle, WA.

Shine, J.A., Gentle, J.E., and Perry, C. (2017), Trends and volatility of stock prices following occurrence of specific technical patterns, in *Proceedings of 2017 Joint Statistical Meetings*, Baltimore, MD.

Siegmund, D. (1985), *Sequential Analysis: Tests and Confidence Intervals*, New York: Springer Verlag.

Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis*, Boca Raton, FL: Chapman and Hall.

Simonoff, J.S. (1996), *Smoothing Methods in Statistics*, New York: Springer-Verlag.

Tartakovsky,A., Nikiforov, I., and Basseville, M. (2015), *Sequential Analysis: Hypothesis Testing and Change-point Detection*, Boca Raton, FL: CRC Press.

Wald, A. (1947), *Sequential Analysis*, New York: Dover Publications.

Wand M.P., and Jones, M.C. (1995), *Kernel Smoothing*, New York: Chapman and Hall.