Pattern Recognition and Statistical Learning in Financial Time Series

James E. Gentle
George Mason University

Email: jgentle@gmu.edu
URL: mason.gmu.edu/~jgentle
Outline

• Defining and identifying patterns in general time series
  – Changepoints
  – Smoothing
  – Location and scale registration
  – Scale registration
  – Trends in the derivative

• Statistical learning: clustering and classification of time series
  – Similarity measures
  – Clustering methods
  – Classification methods
Patterns in Time Series Data

There are many types of patterns in a time series.

Consider this simple time series.
There is a simple pattern. The values increase, then decrease, then increase to a higher point, then decrease, then increase to a lower point, then decrease.

In a time series of financial asset prices, this pattern is called a “head and shoulders”.
Patterns in Time Series Data

This pattern is characterized by “changepoints”.

What about time series that have the same pattern, but that look different?

Although the pattern is the same, the time axis can be distorted.

The slopes (rates of change) are different.
Patterns in Time Series Data

Even if the changepoints occur at the same times, the relative values can be different.

The slopes (rates of change) are different.
Patterns in Time Series Data

Both the changepoints and the relative values can be different.

The slopes (rates of change) are different.
Patterns in Time Series Data

The pattern may be inverted.
Patterns in Time Series Data

The same pattern can be distorted by noise or intermediate values.
Patterns in Time Series Data

The changepoints that define a pattern may be of interest for other time series that may be related.
Patterns in Time Series Data

Patterns in time series generally are based on changepoints, as we have seen in the preceding examples.

Between two changepoints, we identify some characteristic of the series, such as an upward or downward trend.

Other things can characterize the pattern, such as the average value of the series between changepoints.

In some cases, the exact point of change may be somewhat arbitrary.
Patterns Determined by Changes in Average Value
Patterns in Time Series Data

Another thing that can characterize a pattern in a time series is the volatility of the series between changepoints.
Patterns Determined by Changes in Volatility
Patterns in Time Series Data

Another thing that can characterize a pattern in a time series is not just the direction of a trend (up or down) but the general value of the slope of the series between changepoints.
Patterns Determined by Changes in Slopes of Trends
Similarity/Dissimilarity Measures in Time Series Data

A primary objective in mining time series data is to identify time series that are similar to each other.

There are various ways of doing this. One of the most common is to measure dissimilarity by a metric induced by an $L_p$ norm, and of course the most common of these is the $L_2$ or Euclidean norm.

A measure of this sort uses the distances between values of the two time series at each point.

This, of course, assumes that the two series are registered with respect both to location and to scale.
Similar Time Series
The series $x_1$ and $x_2$ are “similar”; $x_3$ is “similar” only after a location registration.
Smoothing and Dimension Reduction in Time Series Data

The question is how do we measure the difference (or similarity) between two time series.

It is difficult to apply a metric to each point in each time series.

A better approach is to smooth each time series first.

One of the simplest methods of smoothing in to replace the time series with a piecewise constant series ("piecewise aggregate approximation, or PAA").
Piecewise Constant Approximation
Other Smoothing Approximations

There are many other types of approximations, both local and global.

Additional approximations may be built on a preliminary approximation such as piecewise constant fits.

For example symbolic aggregate approximation, or SAX, proposed by Lin et al. (2003), further discretizes the individual constants into a fixed (small) number of normal quantiles, represented by symbols.

The first step in using SAX is to standardize the data.

SAX transforms a given standardized time series

$$x_1, \ldots, x_n$$

into

$$A_1, \ldots, A_w, \text{ where } w \ll n,$$

$$A_i \in \{"S_1", \ldots, "S_k"\}, \text{ and } k \text{ is of order } 1.$$
Alternating Trends Smoothing (ATS)

Alternating trends smoothing, or ATS, is a new method of piecewise linear smoothing of a time series.

Each line segment covers a trend regime, and the slope of the line corresponds to the strength of the trend.

The line segments correspond to linear splines with knots at the reversal points.

ATS is more useful than PAA in smoothing financial time series.
Example of ATS

Zillow Prices 2011-07-20 through 2012-07-19
Example of ATS

ATS of Zillow Prices 2011-07-20 through 2012-07-19
Modifications of ATS

Notice that the ATS tends to overshoot the peaks and valleys. This is because the identification of a reversal point is delayed until the reversal is confirmed by a true trend.

There are, of course, several possible modifications of ATS.

The most obvious is to splines with knots at the changepoints identified by ATS.

The splines could be linear, which is equivalent to simple linear regressions constrained to join at the changepoints. Higher degree splines, such as cubics, might produce a better fit.

Approximations fit by splines will not yield invariant metrics,

Another useful modification is to fit a simple linear regression line in each trend regime without the requirement that the line segments join at the reversal points.

The simple linear regression lines will not be continuous, but they will yield invariant metrics.
Symbolic Trend Patterns (STP)

Smoothing is the first step in pattern recognition. The next step is to reduce the dimension of the problem even further.

Replacing continuous numeric representations by a symbolic representation is often an effective way of doing this.

The same idea of SAX, which uses PAA, can be applied to ATS.
The set of symbols I currently use correspond to the syllables formed by selection of a consonant

J, K, L, M, N

that represents duration of an upward trend, or of a consonant

P, Q, R, S, T

that represents duration of an downward trend, and selection of a vowel

A, E, I, O, U

that represents magnitude of a trend.

Thus, the prices in the Zillow example would be transformed into

RU, LU, QA, KA, RU, LI, QE, JI, RU, NU, QE, MO, RE, KO, PE, JA, SI, MO

Of course, because the trends alternate, if a single direction is given, then there would be no need for different symbols to be used to designate up and down moves.
Clustering Time Series Data Using Patterns

Similarity/dissimilarity measures allow clustering of time series.

The objective is to put separate time series into groups of series that have similar patterns.

For example another series, which after the same transformation as before yields

RO, LU, QA, KA, SU, LI, QE, KI, RU, NU, QE, MO, RE, KO, PE, JA, SI, MO

would be put in the same cluster as the Zillow series from the previous slide:

RU, LU, QA, KA, RU, LI, QE, JI, RU, NU, QE, MO, RE, KO, PE, JA, SI, MO

We need a measure of similarity to say how close or how similar the two series are.

The measure could be a simple metric, such as sum of squared differences.
Clustering Time Series Data Using Patterns

Two such segments of a time series represented in this way could be compared just as SNPs in DNA sequences.

Sometimes it is necessary to register the sequence.

We may have

\[
\text{SI, MO, RU, LU, QA, KA, RU, LI, QE, JI, RU, NU, QE, MO, RE, KO, PE, JA}
\]

which does not look like

\[
\text{RO, LU, QA, KA, SU, LI, QE, KI, RU, NU, QE, MO, RE, KO, PE, JA, SI, MO}
\]

until we shift it:

\[
\text{RU, LU, QA, KA, RU, LI, QE, JI, RU, NU, QE, MO, RE, KO, PE, JA}
\]

We then apply the metric and conclude (in this case) that the two series are quite similar.
Classification of Time Series Data Using Patterns

Clustering is sometime called “unsupervised learning”.

Supervised learning is classification.

It is useful when we have some time series with a given characteristic, or that belongs to a given group.

For example, one group may be time series of stock prices that are followed by an increase in price, and another group that are followed by a decrease in price.

We inspect each group and determine differences between the time series in each group.

This is where our measure of similarity/dissimilarity comes into play.
Classification of Time Series Data Using Patterns

There are many methods that can be used for classification, including classification trees along with variations such as random forests and boosting, support vector machines, linear or quadratic discriminant analysis, and so on.

Any of these methods can be used for classification after the time series has been transformed in the manner I have described.
Applications in Financial Data

Financial data present many interesting questions to pursue.

Can patterns be used to forecast future trends?

Our present concerns are trends and patterns. (There are many people, called “technical analysts”, who believe that useful **forecasts** can be made using the observed patterns.)

Can simple techniques such as SAX yield useful information?

First of all, standardization and the use of normal quantiles are not going to work well with financial data because financial data is characterized by heavy tails.

- extension of the piecewise aggregate approximation to include min and max (see Lkhagva et al., 2006)
- use adaptive break points; helps with departures from normality (see Mörchen and Ultsch, 2005)
Finding the Important Trends

In financial data, the primary interest is often in identifying changepoints (see Fu et al., 2008, and Phetking and Selamat, 2008).

We are interested in trends.

A useful transformation must involve the derivative, or at least preserve information in the derivative.

An extreme type of transformation is used in “point and figure charts”. The basic ideas go back to the “Figuring” methods of Charles Dow in the late 1800s.

By the 1940s a simplified form similar to the present one had evolved.

It is eminently suited for hand construction.
Properties of Point and Figure Charts and Trend Charts

Good:

- data reduction
- effective smoothing; small changes in direction are ignored
- patterns of changes stand out

Bad:

- all time information is lost
Loss of Time Information in Point and Figure Charts and Trend Charts

The reversal-trend transform applied to these two time series would yield the same trend series.
Similarity/Dissimilarity Measures in Trend Charts

Metrics are used on time series to rank and/or group different series.

A desirable property of a transformation on a time series is that the rankings and groupings that result from a metric that is applied to two original series will be the same as the rankings and groupings that result from use of the same metric on the transformed series.

There is no metric that is invariant to a point-and-figure transformation (even up to within-variation).
Accounting for Time in Trend Charts

A major failing of point and figure charts is the absence of information about time; hence, another objective is to add information to point and figure charts.

We want to incorporate derivative information between reversal points.

Transform a given time series

\[ x_1, \ldots, x_n \]

into

\[
\left( \begin{array}{c}
  d_1, \ldots, d_w \\
  t_1, \ldots, t_w \\
\end{array} \right),
\]

where \( d_i \) represents the number of unit changes from \( x_{j_i} \) to \( x_{j_i+t_i} \) and, as before, \( d_{i+1} = -d_i \) and \( w \ll n \).

The transformed time series is

\[ d_1/t_1, \ldots, d_w/t_w, \]

which is the slope of the trend-line within a given trend regime.
Metrics on the Transformed Series

Next, we seek a metric that is both useful for identifying patterns and is “invariant” (except for within-variation).

The metric $\rho$ for transformed series of the form

$$d_1/t_1, \ldots, d_w/t_w,$$

is approximately invariant.

A better transformation, which I call the “reversal-trend” transformation, uses the methods outlined above to identify the reversal points, and then fits a trendline to the data between each pair of successive reversal points, using the metric $\rho$.

That metric is then invariant (except for within-variation) on the transformed time series

$$b_1, \ldots, b_w,$$

where the $b_i$ are the slopes.
Summary

The reversal-trend transformation

• is effective in identifying important change points

• is useful for identifying patterns in identifying patterns

• leads to ATS (alternating trends smoothing)

• leads to STP (symbolic trend patterns)
Additional Issues

Things I am currently exploring with one of my graduate students:

- adjusting the reversal points post hoc so that the ATS does not overshoot the peaks and valleys
- investigate alternative symbolic transformations for STP
- incorporation of more derivative information
- identification of jumps
- multi-scale techniques (to study volatility clustering, etc.)
Thank You!

Questions?