Mining for Patterns in Financial Time Series

James E. Gentle*

Abstract
There is a widespread belief that certain patterns of stock prices over time portend specific future types of movement of those prices. We consider various criteria for identifying these types of patterns, and briefly look at some historical price data. We then look at various specific types of breakout movements, and then attempt to determine bags of patterns that tend to precede these breakouts. In the absence of concomitant data, patterns within a single time series of stock prices seem to have very little predictive power. We consider types of patterns over multivariate time series, methods of identifying such patterns, and whether multivariate patterns have more predictive power.

Key Words: clustering of time series subsequences, streaming data, rule discovery

1. Introduction

Patterns in time series data are of interest for a variety of reasons. Some types of time series are related to underlying physical processes and discovery of patterns in such time series enable better understanding of the physical process. Physical processes often have a natural periodicity, but even in such processes, the observed data either is contaminated by some superimposed noise or else results from interacting processes that cause the patterns to be difficult to detect. In climate dynamics, for example, there are many patterns, such as the classic ENSO, that result from interacting physical processes that have yet to be understood.

In economics, the time series of interest (probably!) do not result from underlying physical processes, and any patterns in the time series are unlikely to result from any natural periodicity or other deterministic drivers. The patterns, while often cyclic, are not periodic. Nevertheless, there is a persistent belief that patterns have predictive value. Whether or not patterns can be used for prediction, identification of patterns in economic or financial time series can be useful in understanding relationships among various economic time series.

Interest in identification of patterns in time series is often motivated by comparison of two time series or, equivalently, comparison of two subsequences within a given time series. General reviews of transformations on time series and their use in classification of time series are provided by Fu (2011) and Gama (2010). In the rest of this section, I will briefly mention some of the issues, and then in Section 2 turn to the relevant methods for financial time series.

1.1 Transformations in Time Series

The time series we will consider are ordered sequences of real- or vector-valued observable random variables, $X_t, X_{t+1}, X_{t+2}, \ldots$ where t is some fixed integer. We often focus on a finite subsequence, and we usually consider realizations of the random variables: $x_t, x_{t+1}, \ldots, x_{t+n}$. We often represent a time series by a single symbol, x, for example.

We will use various transformations of the data. Many important classes of transformations are one-to-one, for example, the affine transformations of the form $g(x) = ax + b$,

*George Mason University, Fairfax, VA 2030, jgentle@gmu.edu

where a is a non-zero real number. We are generally interested in transformations T that are equivariant with respect to some group of mappings $\mathcal{G}$ on the elements of a time series. This means if g is a mapping of elements of the time series, then there is a mapping g^* on the transformed time series such that for a time series x

$$T(x) = g^*(T(g(x))). \tag{1}$$

If there is a group $\mathcal{G}$ of mappings of a time series x and an induced group $\mathcal{G}^*$ of mappings of the transformed time series $T(x)$ such that equation (1) holds for any $g \in \mathcal{G}$ with a particular $g^* \in \mathcal{G}^*$, then we say the transformation is equivariant with respect to $\mathcal{G}$.

Other useful types of transformation of the time series results in a simpler representation of the time series. A transformation that yields a "simpler representation" is usually not one-to-one. It often is a transformation into a lower dimensional representation or it may yield a discretized representation. Given a time series, $x_t, x_{t+1}, \ldots, x_{t+n}$, we will denote a transformation as $T_{t,n,d}(x)$ as some arbitrary real d-vector where $d \leq n + 1$. We often use the simpler notation $T(x)$. Such a transformation may be used to associate some interesting pattern with the underlying time series.

We seek transformations that reduce the amount of data while at the same time retaining as much information as possible. We also seek transformations that are invariant to, or nearly invariant to summary measures of the data. If we have some measure ρ for comparing two time series x and y for example, we would want a transformation T to which the measure to is invariant: $\rho(x, y) = \rho(T(x), T(y))$. Invariance may be too strong a requirement. Instead we may seek a transformation under which ρ is monotonic in the sense that for three time series x, y, and z, if $\rho(x, z) < \rho(y, z)$ then $\rho(x, z) = \rho(T(y), T(z))$.

Some widely-used transformations include the discrete fourier transform (DFT), the discrete wavelet transform (DWT), piecewise linear, and piecewise constant models (PAA), and singular value decomposition (SVD). The usefulness or "goodness" of a transformation depends on the purpose for which it is being applied. The nature of the information that is lost through the transformation is one criterion, and another is the ability of the transformation to preserve and indeed to emphasize the particular aspects of the data. In many cases a categorical transformation, such as a symbolic representation, is useful because then classification of time series is facilitated. Lin et al. (2007) summarize a variety of time series representations, using a classification tree.

1.2 Patterns in Time Series

It is often of interest to identify patterns in time series. A "pattern" is a collection of transformed time series, usually characterized by some general characteristic, such as "increasing" or "head-and-shoulders". Identification of patterns in time series is facilitated by transformations, and indeed detection of patterns is one of the main reasons for performing transformations.

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 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.

There are many types of approximations, both local and global. Additional approximations may be built on a preliminary approximation, such as piecewise constant fits. For example **s**ymbolic **a**ggregate appro**x**imation, 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 then transforms a given standardized time series

$$x_1, \ldots, x_n$$

into

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

and

$$A_i \in \{\text{“}S_1\text{”}, \ldots, \text{“}S_k\text{”}\},$$

where k is of order 1.

If the series have been smoothed by fitting to a minimum norm, then it is obvious that the the distance between the two smoothed series measured by the same norm provides a lower bound on the distance between the original series (measured by the same norm).

For example, in the case of piecewise constant approximations by means within windows, we have the trivial Pythagorean relationship within each window:

$$\|x - y\|^2 = \|c_x - c_y\|^2 + \|x - c_x - (y - c_y)\|^2.$$

The same metric applied to any further ordinal discretization of an initial smoothing of two time series would likewise provide a lower bound on the metric applied to the original raw series. (For the L_2 metric, this follows from the wellknown decomposition SSA+SSW=SST.)

Lower bounding yields an invariance (up to within-variation) of dissimilarity metrics. (“Invariance” here means “transitive”.)

Similarity/dissimilarity measures allow clustering of time series. The objective is to put separate time series into groups of series that have similar patterns. SAX, for example, has been found to be very effective for this purpose (see Lin et al., 2007, e.g.).

For any numeric time series, SAX yields a sequence of symbols, for example, from the PAA

$$c_1, c_2, c_3, c_4, c_5, c_6$$

we might get

$$a,\ b,\ a,\ c,\ c,\ a.$$

Two such segments could be compared just as SNPs in DNA sequences.

1.3 Motifs in Patterns

A pattern must be understood to capture the gestalt of the data; not its details. That is the primary challenge in identifying patterns; the smoothing must capture the salient overall characteristics, while ignoring noise and irrelevant details. (Deciding relevancy, of course, is the central issue.)

One of the main purposes in identifying patterns is to be able to compare different time series and to decide the extent to which two time series are similar. Our objective would be to categorize the patterns into a relatively small number of classes. As in other areas of data mining, we may define a bag-of-patterns; that is, a set of specific shapes, structures, and motifs (see, for example, Lin and Li, 2009).

If the time series are assumed to have a common beginning and end, and equal sampling frequencies, the comparison is straightforward; otherwise, there is a preliminary problem of registration.

This, of course, assumes that the two series are registered with respect both to location and to scale.

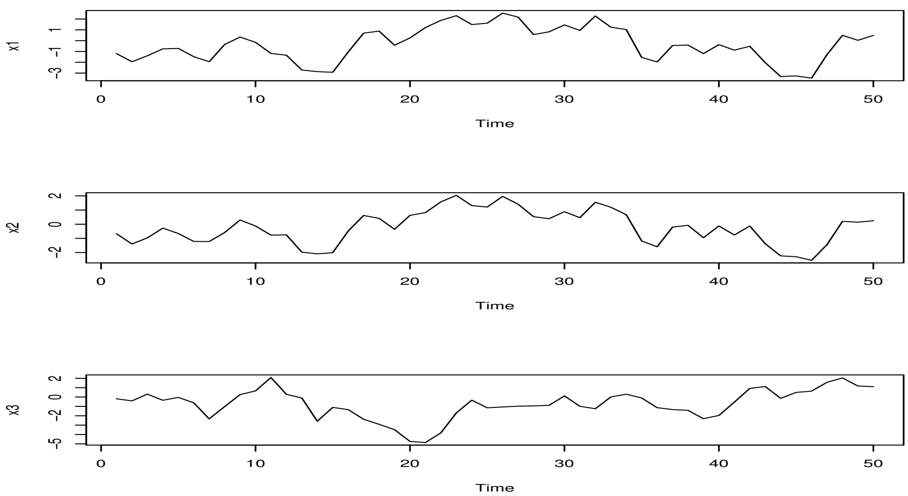

Figure 1: Time Series with Time-Shifts

1.4 Scaling and Shifting the Time Axis

In most cases, our interest in whether two time series exhibit similar patterns does not require that the patterns be synchronous. This implies that we may need to shift (or "register") the time axis of one series before its pattern seems to be the same as the pattern of the other.

In Figure 1.4, the series x_1 and x_2 are "similar"; x_3 is "similar" only after a location registration.

Another, more difficult type of registration involves scaling of time. In simple patterns where a "frequency" can be identified, this scaling is merely registration or scaling of the frequencies to be the same.

There has been much research directed toward the registration problem. One rather old method that continues to be moderately successful is dynamic time warping (DWT) (see Gama, 2010, for example).

2. Finding Patterns in Financial Time Series

There are various types of financial time series, and the interrelationships present interesting problems in data mining, Gentle (2009).

One of the important types of financial time series is a series of asset prices (or an index of asset prices) and the rate of change of the prices. Gentle and Härdle (2012) discuss several ways of modeling these prices. Tsay (2010) applies traditional time series modeling approaches to financial time series. ("Traditional" here includes newer methods that allow heterschedasticity.)

In this paper, I am interested in prices of a given asset or of an index of asset prices. We often focus on the rate of change of prices of an asset or of an index over time. Figure 2.1 shows the rate of change of the S&P500 index over a period of five years. (The rate of change shown is the log-rate.)

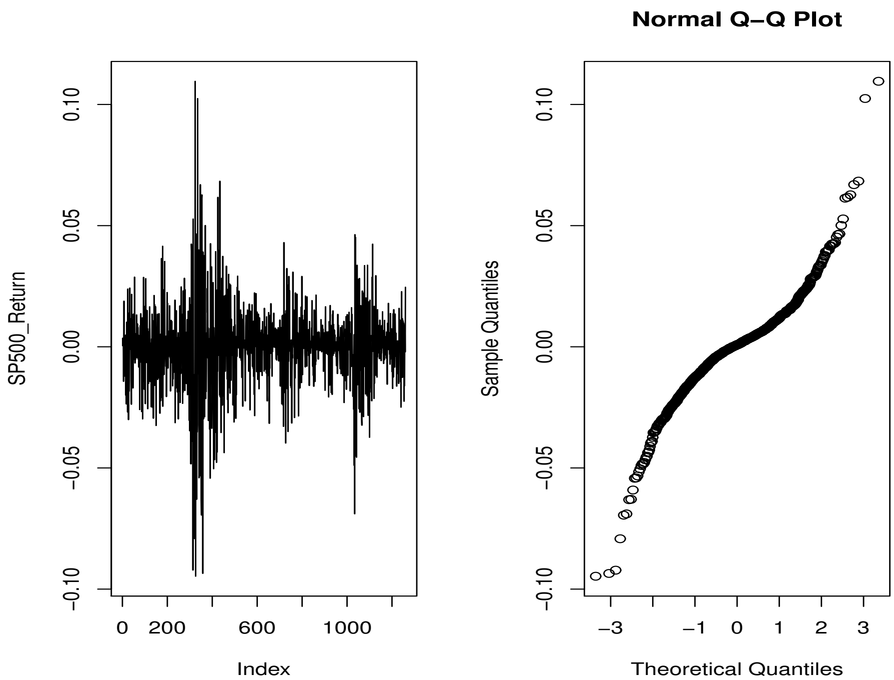

Figure 2: Five Years of S&P500 Returns through 2012-06-30

2.1 The Nature of Asset Price Changes

There are some stylized properties of empirical financial data that are widely recognized.

- **Heavy tails**. The frequency distributions of rates of return decrease more slowly than $\exp(-x^2)$. This is seen in the normal Q-Q plot in Figure 2.1.
  - Can an outlier-generating distribution model fit well?
- The **volatility** is **not constant**.
  - Also, the volatility clusters. This is seen in the left-hand time series plot in Figure 2.1.
- **Intraday** data is **different** from **interday** data.
  - The trading market has a microstructure.

There are many interesting questions to pursue, beginning with a search for workable models. 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.

•extension of the piecewise aggregate approximation to include min and max (see Lkhagva et al., 2006)

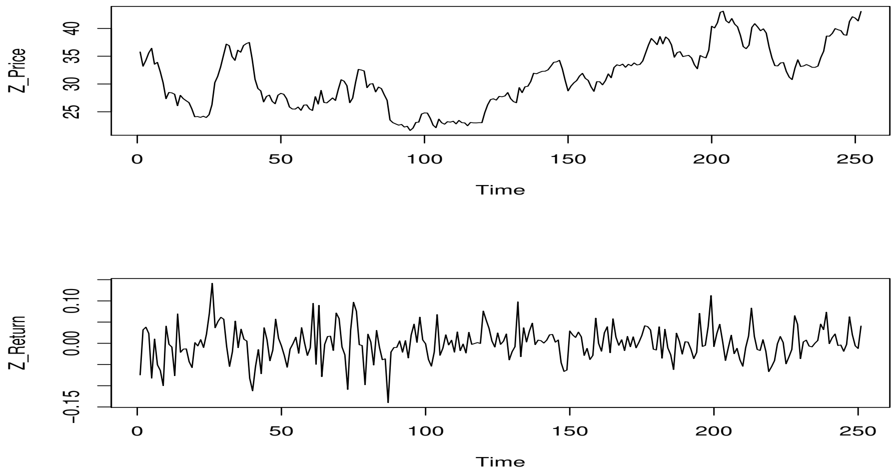

Figure 3: Prices and Returns of Zillow, Inc. (Z) 2011-07-20 (IPO) through 2012-07-19

•use adaptive break points; helps with departures from normality (see Mörchen and Ultsch, 2005)

The fundamental problem, however, is nonstationarity, which affects both prices and returns, which is apparent in Figure 2.1.

2.2 Interesting Patterns in Financial Time Series

What characterizes interesting patterns in financial time series is often somewhat different for what is interesting in other types of time series, In the analysis of financial time series, just as with almost any other type of time series, one of the most important objectives is to find meaningful trends.

In financial data, the primary interest is often in identifying important points (turning points) instead of just the usual types of patterns (see Fu et al., 2008, and Phetking and Selamat, 2008).

We are interested in how long trend last and when a trend in a different direction. We are also interested in measuring the strength of the trends, hence, any transformation of the time series data must involve the derivative, or at least preserve information in the derivative.

2.2.1 Trend Charts

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. See Dorsey (2007) for a complete description of point and figure charts.

A point and figure chart transforms a given time series

$$x_1, \ldots, x_n$$

into

$$d_1, \ldots, d_w,$$

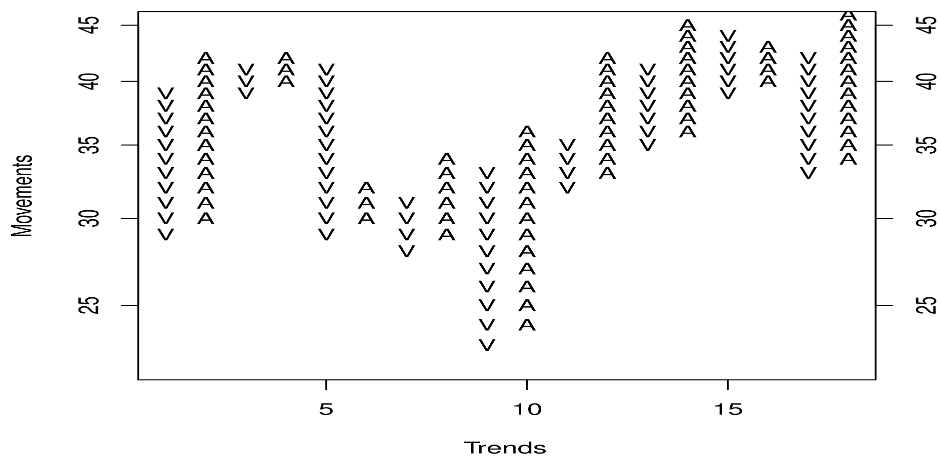

Figure 4: Trend Chart of Z 2011-07-20 (IPO) through 2012-07-19

where d_i represents the number of unit changes from x_{j_i} to $x_{j_i+t_i}$ with $d_{i+1} = -d_i$ and $w \ll n$.

The x_{j_i} are **reversal points**.

The index j_i of a reversal point, corresponding to the i^{th} transformed point, depends on two things: the amount of change that corresponds to a "box" (an "X", that is, "up", or an "O", that is, "down"), and the number of boxes in a different direction to confirm a trend reversal.

A simple modification of a point and figure chart, which I call a "trend chart", measures change by use of a fixed rate, rather than fixed amounts within certain ranges. For a given base rate, the actual rate for an up move is larger than the rate for a move down, so that two opposite moves cancel each other out. Another change I have made is to ignore intraday highs and lows. A trend chart for Zillow from its IPO is shown in Figure 2.2.1.

2.2.2 Patterns in Trend Charts

One reason point and figure charts are so popular it that they vividly display patterns of stock price changes.

There are various "trend lines" that stand out. Some of these lines have particular meaning for technical analysts, and go by such names as "resistance lines" (either bullish or bearish) and "support lines" (again, either bullish or bearish).

Patterns of levels of the reversal points, whether the next trend is up or down, also have meaning for technical analysts. These patterns go by such names as "double top", "head and shoulders", "catapult" (either bullish or bearish), and so on.

Technical analysts use both trend lines and patterns to forecast the direction of the movement of prices (see, for example, Dorsey, 2007).

2.2.3 Properties of Trend Charts

Trend charts have a number of useful properties:

- data reduction.

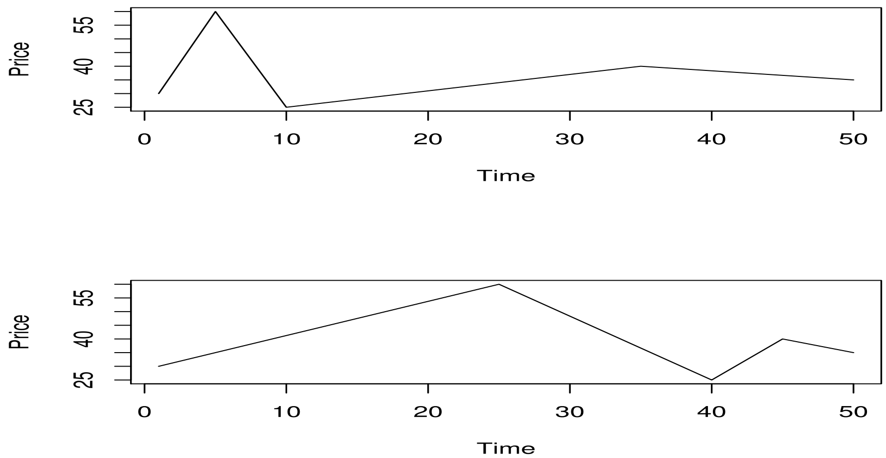

Figure 5: Two Simple Series that Would Have Identical Trend Charts

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

There is also a major limitation to trend charts:

- **all** time information is lost (see Figure 2.2.3).

The reversal-trend transform applied to these two time series would yield the same trend series.

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).

2.2.4 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. To do this, we transform a given time series

$$x_1, \ldots, x_n$$

into

$$\begin{pmatrix} d_1, \ldots, d_w \\ t_1, \ldots, t_w \end{pmatrix},$$

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

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

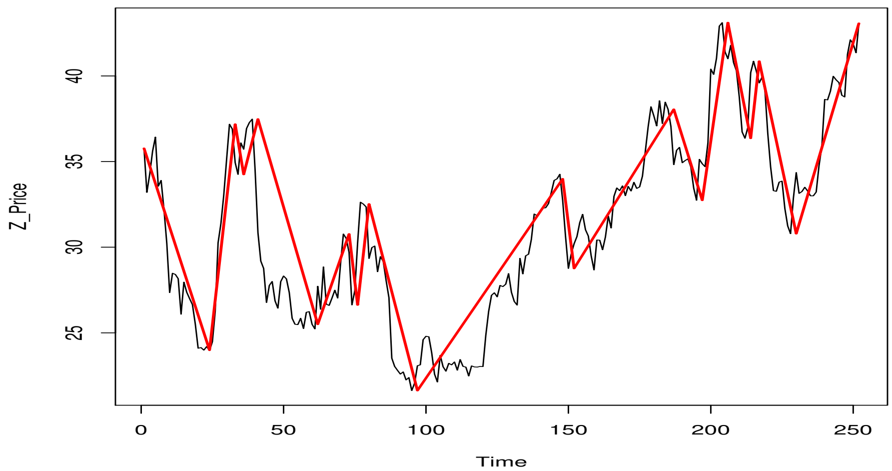

Figure 6: ATS of Z 2011-07-20 through 2012-07-19

which is the slope of the trend-line within a given trend regime.

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

The metric ρ 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 ρ.

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.

2.3 Alternating Trends Smoothing (ATS)

Alternating trends smoothing, or ATS, is a new method of piecewise linear smoothing of a time series. In ATS, 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 because it shows the rate of change.

Figure 2.3 shows ATS applied to Zillow since the date of its IPO.

2.3.1 Other Types of Smoothing Alternating Trends

Another method of determining the line segments is by regression on the points within each trend regime, constrained so that the lines join at the reversal points. Various norms can be used for either the smoothing spline or the regression fits.

The line segments of ATS are not the same as the trend lines of a point and figure chart, because ATS incorporates time information.

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 use higher degree splines, such as a cubic, for example. 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.

2.4 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.

2.5 Properties of ATS

The reversal-trend transformation has a number of useful properties; in particular, it

- 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).

2.6 Modifications of ATS

As mentioned above, there are a number of possible modifications to ATS, such as the basic method of smoothing, whether linear or high order splines, for example.

In addition to those simple modifications, there are various other possibilities that I am currently exploring:

- 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.).

REFERENCES

Thomas J. Dorsey (2007), *Point and Figure Charting: The Essential Application for Forecasting and Tracking Market Prices*, third edition, John Wiley & Sons, Inc., Hoboken.

Fu, Tak-Chung (2011), A review on time series data mining, *Engineering Applications of Artificial Intelligence* **24**, 164–181.

Fu, Tak-Chung, Korris Fu-Lai Chung, Robert Wing Pong Luk, and Chak-man Ng (2008), Representing financial time series based on data point importance, *Engineering Applications of Artificial Intelligence* **21**, 277–300.

Gama, João (2010), *Knowledge Discovery from Data Streams*, CRC Press, Boca Raton.

Gentle, James E., and Wolfgang Karl Härdle (2012), Modeling asset prices, in *Handbook of Computational Finance*, Springer, Heidelberg, 15–33.

Gentle, James E. (2009), Challenges in mining financial data, in *Next Generation of Data Mining*, edoted by Hillol Kargupta, Jiawei Han, Philip S. Yu, Rajeev Motwani, and Vipin Kumar, CRC Press, Boca Raton, 527–547.

Lin, Jessica, Eamonn Keogh, Li Wei, and Stefano Lonardi (2007), Experiencing SAX: A novel symbolic representation of time series, *Data Mining and Knowledge Discovery* **15**, 107–144.

Lin, Jessica, and Yuan Li (2009), Finding structural similarity in time series data using bag-of-patterns representation, in *Statistical and Scientific Database Management*, 461–477.

Lkhagva, Battuguldur, Yu Suzuki, and Kyoji Kawagoe (2006), New time series data representation ESAX for financial applications, *Proceedings of the 22nd International Conference on Data Engineering Workshops (ICDEW'06)*, 17–21.

Mörchen, Fabian and Alfred Ultsch (2005), Optimizing time series discretization for knowledge discovery, *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Datamining*, 660–665.

Phetking, Chaliaw, Mohd Noor Md. Sap, and Ali Selamat (2008), A multiresolution important point retrieval method for financial time series representation, *International Conference on Computer and Communication Engineering - ICCCE*, 510–515.

Tsay, Ruey S. (2010), *Analysis of Financial Time Series*, third edition, John Wiley & Sons, Inc., Hoboken.