6-2015

Dow Theory's Peak-and-Trough Analysis Justified

Chrysostomos Stylios
Technological Education Institute of Epiros, stylios@teiep.gr

Vladik Kreinovich
The University of Texas at El Paso, vladik@utep.edu

Comments:
Technical Report: UTEP-CS-15-54
To appear in: Martine Ceberio and Vladik Kreinovich (eds.), *Constraint Programming and Decision Making: Theory and Applications*, Springer Verlag, Berlin, Heidelberg.

Dow Theory's Peak-and-Trough Analysis Justified

Chrysostomos D. Stylios[1] and Vladik Kreinovich[2]
[1]Laboratory of Knowledge and Intelligent Computing
Department of Computer Engineering
Technological Educational Institute of Epirus
47100 Kostakioi, Arta, Greece
stylios@teiep.gr
[2]Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, Texas 79968, USA
vladik@utep.edu

Abstract

In the analysis of dynamic financial quantities such as stock prices, equity prices, etc., reasonable results are often obtained if we only consider local maxima ("peaks") and local minima ("troughs") and ignore all the other values. The empirical success of this strategy remains a mystery. In this paper, we provide a possible explanation for this success.

1 Formulation of the Problem

Peak-and-trough analysis. In the early 20th century, a theory – known as Dow Theory – was developed for forecasting the behavior of different prices, such as stock prices, equity prices, etc. The main idea behind this theory is that:

- similarly to calculus, where the important first step in the analysis of a function is finding its local minima and maxima,
- the important information about the changes in stock market prices can be obtained if we mark local maxima ("peaks") and local minima ("troughs"); see, e.g., [4, 6].

This analysis is still in use. The resulting peak-and-trough analysis was widely used in the 1920s and early 1930s, until a paper [3] showed the deficiency of the corresponding forecasting techniques.

This paper used then-prevalent expected-return values to analyze the quality of the Dow Theory recommendations. By the 1990s, however, it became clear that when comparing different stock recommendations, it is important to also take into account the corresponding *risks*.

It turns out that if we take risk into account, then the Dow Theory recommendations are not inferior at all, these predictions are actually reasonably good; see, e.g., [7]. As a result, the peak-and-trough analysis has been revived – and it is still used in financial analysis.

Comment. The actual dependence of the stock prices (and other prices) on time t comes with *noise*: random fluctuations caused by many random factors. From the purely mathematical viewpoint, this means that the dependence oscillates all the time, so almost every moment of time has it local minima and local maxima. What the peak-and-trough analysis suggest, of course, is *not* to use all these moments of time, but only to use moments of *true* local minima and maxima, i.e., moments when we can be sure that the local extremum is not cause by the noise itself.

So, to apply this analysis, we need first to be able to distinguish between local extrema which may be due to noise and the real local extrema. There exist efficient algorithms for making this distinction. For example, in situations when all we know about the noise $n(t)$ is that its absolute value $|n(t)|$ is bounded by some value n_0 ($|n(t)| \le n_0$), there is an efficient (linear time) algorithm for detecting real local extrema; see, e.g., [12].

Similar ideas works well in engineering as well. When we only take into account the local extrema, this means that:

- for all the moments of time between a local maximum and the following local minimum, the value $x(t)$ decreases; we do not have any information about how exactly it decreases, we only know that is decreases;
- similarly, for all the moments of time between a local minimum and the following local maximum, the value $x(t)$ increases; we do not have any information about how exactly it increases, we only know that is increases.

In other words, for each moment of time t, we only have one of the following three pieces of information about how the signal $x(t)$ changes in the small vicinity of this moment t:

- we may know that there is a local extremum in this vicinity; in this case, in this vicinity, the value $x(t)$ practically does not change,
- we may know that the value $x(t)$ decreases in this vicinity,
- or, alternatively, we may know that the value $x(t)$ increases in this vicinity.

Interestingly, many efficient methods of signal compression – starting with the so-called *delta-modulation* – are based on recording, for each moment of time, exactly one of these three situations: 0 (no change), $-$ (decrease), or $+$ (increase); see, e.g., [1, 2, 5, 8, 10, 13].

Why is peak-and-trough analysis efficient? What is not clear is why the peak-and-trough analysis, an analysis that ignores all monotonicity segments and only takes into account the local extrema, is efficient.

Comment. Similarly to financial applications, from the theoretical viewpoint, the engineering-oriented empirical success of delta-modulation techniques is also largely a mystery.

What we do in this paper. In this paper, we provide a possible explanation for the efficiency of peak-and-trough techniques.

2 Our Explanation

In the first approximation, it is reasonable to only process the most important values. Ideally, we should take into account the values $x(t)$ of the stock price at all moments of time t. The problem is that there is a large amount of these data points, and without a clear understanding of the underlying processes, it is difficult to meaningfully process all this data.

It is therefore reasonable, in the first approximation, to only concentrate on the *most important* stock price values and ignore the less important values.

Which values should we take into account? As the stock price fluctuates, it attains different values x. Some values appear more frequently, some values appear more rarely. It therefore makes sense to concentrate on the prices that appear the largest number of times.

Of course, from the practical viewpoint, very close values x can be viewed as identical. So, when we talk about the time that a value x appears, we mean the time when the value $x(t)$ is within an interval $[x-\delta, x+\delta]$ for some small $\delta > 0$.

How to decide which values are most frequent? The values $x(t)$ are rarely stable, then usually change with time. Thus, the time period during which the value is within a given interval $[x-\delta, x+\delta]$ is small. If we had $x(t_0) = x$ for some moment t_0, this means that the neighboring moments of time t at which $x(t) \in [x-\delta, x+\delta]$ are close to t_0, i.e., have the form $t = t_0 + \Delta t$, where $\Delta t \ll t_0$. For such small values Δt, we can ignore quadratic and higher order terms in the dependence of $x(t)$ on t, and use the linear approximation

$$x(t_0 + \Delta t) \approx x(t_0) + x'(t_0) \cdot \Delta t = x + x'(t_0) \cdot \Delta t. \tag{1}$$

Thus, the length ℓ of the time interval during which

$$x(t) = x(t_0 + \Delta t) \in [x-\delta, x+\delta]$$

is equal to

$$\ell = \frac{2\delta}{|x'(t_0)|}. \tag{2}$$

Resulting explanation. We have decided to only consider the values $x(t_0)$ for which this time interval ℓ is large. According to the formula (2), this means that we should only consider the values $x(t_0)$ at the moments t_0 at which the derivative $x'(t_0)$ is close to 0 – i.e., only the values in the vicinity of points where the derivative is equal to 0. These points are exactly local minima and local maxima – as well as possible non-minimum and non-maximum stationary points.

Thus, we indeed have an explanation of why the peak-and-trough strategy is successful.

3 Additional Theoretical Confirmation of Our Justification

Another situation where extreme points frequently occur. B. S. Tsirelson noticed [11] that in many cases, when we reconstruct the signal from the noisy data, and we assume that the resulting signal belongs to a certain class, the reconstructed signal is often an *extreme* point from this class. For example:

- when we assume that the reconstructed signal is monotonic, the reconstructed function is often (piece-wise) constant;
- if we additional assume that the signal is smooth (one time differentiable, from the class C^1), the result is usually one time differentiable but rarely twice differentiable, etc.

This situation has an explanation. To explain this phenomenon, Tsirelson provided the following *geometric* explanation to this fact: namely, when we reconstruct a signal from a mixture of a signal and a Gaussian noise, then the *maximum likelihood* estimation (a traditional statistical technique; see, e.g., [9]) means that we look for a signal that belongs to the priori class, and that is the closest (in the L^2-metric) to the observed "signal+noise".

In particular, if the signal is determined by finitely many (say, d) parameters, we must look for a signal $\vec{s} = (s_1, \ldots, s_d)$ from the a priori set $A \subseteq R^d$ that is the closest (in the usual Euclidean sense) to the observed values

$$\vec{o} = (o_1, \ldots, o_d) = (s_1 + n_1, \ldots, s_d + n_d),$$

where n_i denotes the (unknown) values of the noise.

Since the noise is Gaussian, we can usually apply the *Central Limit Theorem* [9] and conclude that the average value of $(n_i)^2$ is close to σ^2, where σ is the standard deviation of the noise. In other words, we can conclude that

$$(n_1)^2 + \ldots + (n_d)^2 \approx d \cdot \sigma^2.$$

In geometric terms, this means that the distance

$$\sqrt{\sum_{i=1}^{d}(o_i - s_i)^2} = \sqrt{\sum_{i=1}^{d} n_i^2}$$

between $\vec{s}$ and $\vec{o}$ is $\approx \sigma \cdot \sqrt{d}$. Let us denote this distance $\sigma \cdot \sqrt{d}$ by ε.

Let us first, for simplicity, consider the case when $d = 2$, and when A is a convex polygon. Then, we can divide all points p from the exterior of A that are ε-close to A into several zones depending on what part of A is the closest to p:

- one of the *sides*, or
- one of the *edges*.

Geometrically, the set of all points for which the closest point $a \in A$ belongs to the *side* e is bounded by the straight lines orthogonal (perpendicular) to e. The total length of this set is therefore equal to the length of this particular side; hence, the total length of all the points that are the closest to all the sides is equal to the *perimeter* of the polygon. This total length thus does not depend on ε at all.

On the other hand, the set of all the points at the distance ε from A grows with the increase in ε; its length grows approximately as the length of a circle, i.e., as const$\cdot\varepsilon$.

When ε increases, the (constant) perimeter is a vanishing part of the total length. Hence, for large ε:

- the fraction of the points that are the closest to one of the sides tends to 0, while
- the fraction of the points p for which the *closest* is one of the *edges* tends to 1.

Similar arguments can be repeated for any dimension. For the same noise level σ, when d increases, the distance $\varepsilon = \sigma \cdot \sqrt{d}$ also increases, and therefore, for large d, for "almost all" observed points $\vec{o}$, the reconstructed signal is one of the extreme points of the a priori set A.

A similar explanation can be applied to our case as well. In our case, as we showed in the previous section, extreme values are also much more frequently observed than others. Thus, our argument can be viewed as a particular case of the general geometric explanation proposed by Tsirelson.

Acknowledgments

This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and

DUE-0926721. This work was performed when C. Stylios was a Visiting Researcher at the University of Texas at El Paso.

The authors are thankful to Djuro Zrilic for valuable discussions.

References

[1] G. I. Bourdopoulos and A. Pnevmatikakis, *Delta-Sigma Modulators: Modeling, Design and Applications*, Imperial College Press, London, UK, 2003

[2] L. Breems and J. Huijsing, *Continuous-Time Sigma-Delta Modulation for A/D Conversion in Radio Receivers*, Springer Verlag, New York, 2013.

[3] A. Cowles, "Can stock market forecasters forecast?", *Econometrica*, 1933, Vol. 1, No. 3, pp. 309–324.

[4] W. P. Hamilton, *The Stock Market Barometer: A Study of its Forecast Value Based on Charles H. Dow's Theory of the Price Movement*, Barrons, New York, 1922.

[5] E. Janssen and A. van Roermund, *Look-Ahead Based Sigma-Delta Modulation*, Springer, Dordrecht, Heidelberg, London, New York, 2011.

[6] R. Rhea, *The Dow Theory*, Barron's, New York, 1932.

[7] J. Schannep, *Dow Theory for the 21st Century: Technical Indicators for Improving Your Investment Results*, Wiley, New York, 2008.

[8] R. Schreier and G. C. Temes, *Understanding Delta-Sigma Data Converters*, IEEE Press, Piscataway, New Jersey, and Wiley, Hoboken, New Jersey, 2004.

[9] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall/CRC, Boca Raton, Florida, 2011.

[10] H. Sira-Ramírez, *Sliding Mode Control: The Delta-Sigma Modulation Approach*, Birkhäuser, Basel, Switzerland, 2015.

[11] B. S. Tsirel'son, "A geometrical approach to maximum likelihood estimation for infinite-dimensional Gaussian location. I", *Theory Probab. Appl.*, 1982, Vol. 27, pp. 411–418.

[12] K. Villaverde and V. Kreinovich, "A linear-time algorithm that locates local extrema of a function of one variable from interval measurement results," *Interval Computations*, 1993, No. 4, pp. 176–194.

[13] D. G. Zrilic, *Circuits and Systems Based on Delta Modulation: Linear, Nonlinear and Mixed Mode Processing*, Springer, Berlin, Heidelberg, 2010.